Fish Age Reading and Otolith

Analysis Using Deep Learning

Arjay Cayetano

Dissertation

2024

Fish Age Reading and Otolith Analysis Using Deep Learning

Universität Bremen

Arjay Cayetano

Dissertation zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

Dr. rer. nat.

Universität Bremen

Fachbereich 2

August 2024

Summary

Fish age reading is a crucial step in the proper management of fisheries. To determine the fish age, several methods have been developed making use of fish structures that give clues on fish growth throughout the different seasons. Supported by years of extensive research and validation efforts, the use of otoliths (ear stones) has become the standard approach.

Within otoliths, there are growth rings (annuli) that form patterns through uneven calcium carbonate deposition influenced mainly by seasonal factors affecting the fish growth. Hence, the traditional age reading methodology works by visual inspection and manual counting of these rings to derive the fish age. However, certain cases make the process problematic and error-prone. As these errors have big impact on fisheries management, it is important to explore ways on how these can be prevented.

The field of computer vision provides a means to make the process of age reading less reliant on subjective interpretations. Using otolith images, many studies applied classical image processing techniques that take advantage of the changes on image intensity when traversing the otolith from the center to the outer edge.

With the progress in artificial intelligence (AI), computer vision methods have taken a new level of sophistication. Using machine learning algorithms, AI models are trained to learn the patterns of growth on the otolith. Early approaches in this direction utilized classical algorithms such as Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) and some also employed feature engineering in order to create meaningful feature sets (e.g., intensity signals) to be used by the algorithms.

Recently, deep learning (DL) algorithms such as Convolutional Neural Networks (CNNs) have gain substantial popularity as they outperform those classical machine learning methods. Early works on DL-based fish age reading have shown excellent accuracy on estimating the fish age based on otolith images. However, the main drawback is that they are formulated as classification or regression making them incompatible with traditional ring counting protocols.

In this thesis project, a different perspective for using deep learning on the task of fish age reading was explored. The methods applied were designed to specifically detect and annotate the annuli which are then counted to derive the fish age. Two different object detection and segmentation algorithms were used namely, Mask R-CNN and U-Net. In this thesis, the effectiveness of the two methods was demonstrated along with the tools developed to make the approaches widely accepted by the community. In addition, the study elucidated advanced techniques to improve the accuracy further and also highlighted additional related tasks for general otolith analysis that both algorithms managed to perform effectively.

Zusammenfassung

Alterslesungen von Fischen sind ein entscheidender Schritt für die effektive Bewirtschaftung der Fischerei. Zur Bestimmung des Fischalters wurden mehrere Methoden entwickelt, die sich Fischstrukturen zunutze machen, die Hinweise auf das Wachstum der Fische im Laufe der verschiedenen Jahreszeiten geben. Unterstützt durch jahrelange, umfangreiche Forschungs- und Validierungsbemühungen hat sich die Verwendung von Otolithen (Gehörsteinen) zum Standardansatz entwickelt.

In den Otolithen befinden sich Wachstumsringe (Annuli), die durch ungleichmäßige Kalziumkarbonatablagerungen Muster bilden, die hauptsächlich durch saisonale Faktoren beeinflusst werden, die das Fischwachstum beeinflussen. Die traditionelle Methode zur Altersbestimmung beruht daher auf der visuellen Inspektion und dem manuellen Zählen dieser Ringe. In bestimmten Fällen ist dieses Verfahren jedoch problematisch und fehleranfällig. Da diese Fehler große Auswirkungen auf die Fischereiverwaltung haben, ist es wichtig, Wege zu finden, wie solche Fehler verhindert werden können.

Der Bereich der Computervision bietet eine Möglichkeit, den Prozess der Altersbestimmung weniger abhängig von subjektiven Interpretationen zu machen. Bei der Verwendung von Otolithenbildern wurden in vielen Studien klassische Bildverarbeitungstechniken angewandt, die sich die Änderungen der Bildintensität beim Durchlaufen des Otolithen von der Mitte zum äußeren Rand zunutze machen.

Mit den Fortschritten in der künstlichen Intelligenz (KI) haben die Methoden der Computer-Vision eine neue Stufe der Raffinesse erreicht. Mithilfe von Algorithmen des maschinellen Lernens werden KI-Modelle trainiert, um die Wachstumsmuster auf dem Otolithen zu lernen. Frühe Ansätze in dieser Richtung nutzten klassische Algorithmen wie Support Vector Machines (SVMs) und Artificial Neural Networks (ANNs), und einige setzten auch Feature Engineering ein, um aussagekräftige Feature-Sets (z. B. Intensitätssignale) zu erstellen, die von den Algorithmen verwendet werden.

In letzter Zeit haben Algorithmen des Deep Learning (DL), wie z. B. Convolutional Neural Networks (CNNs), erheblich an Popularität gewonnen, da sie diese klassischen Methoden des maschinellen Lernens übertreffen. Frühe Arbeiten zur DL-basierten Altersbestimmung von Fischen haben eine ausgezeichnete Genauigkeit bei der Schätzung des Fischalters auf der Grundlage von Otolithenbildern gezeigt. Der größte Nachteil besteht jedoch darin, dass sie als Klassifizierungs- oder Regressionsverfahren formuliert sind, was sie mit traditionellen Ringzählungsprotokollen inkompatibel macht.

In dieser Arbeit wurde eine andere Perspektive für den Einsatz von Deep Learning bei der Altersbestimmung von Fischen erforscht. Die angewandten Methoden wurden so konzipiert, dass sie Wachstumsringe erkennen und beschriften, die dann gezählt werden, um das Fischalter abzuleiten. Es wurden zwei verschiedene Algorithmen zur Objekterkennung und -segmentierung verwendet, nämlich Mask R-CNN und U-Net. In dieser Arbeit wurde die Effektivität der beiden Methoden demonstriert, und es wurden Werkzeuge

entwickelt, um die Ansätze in der Community zu verbreiten. Darüber hinaus wurden in der Studie fortgeschrittene Techniken zur weiteren Verbesserung der Genauigkeit erläutert und zusätzliche verwandte Aufgaben für die allgemeine Otolithenanalyse hervorgehoben, die beide Algorithmen effektiv durchführen konnten.

Contents

EXAMINATION COMMITTEE

Examiner/Chairperson: **Prof. Dr. Matthew Nielsen**

University of Bremen

Fachbereich 2 Biologie/Chemie

Bremen, Germany

Examiner/Reviewer 1: **Prof. Dr. Thomas Brey (supervisor)**

University of Bremen

Fachbereich 2 Biologie/Chemie

Bremen, Germany

Examiner/Reviewer 2: **Prof. Dr. Andreas Birk**

Constructor University

School of Science and Engineering

Bremen, Germany

Examiner/Reviewer 3: **Dr. Christoph Stransky**

Thünen Institute of Sea Fisheries

Deputy Director

Bremerhaven, Germany

Date of colloquium: 29 August 2024

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 Stock Assessments and Fish Age

Determining the age of fish is one of the most basic yet important steps in the accurate assessment of fish stocks, which plays a vital role in fisheries management and monitoring as well as in the analysis of fish population dynamics. The age variable along with the corresponding year-class determination is required in the calculations regarding growth and mortality rates (Campana and Thorrold, 2001) as well as in the multitude of mathematical equations/statistical methods (e.g., catch curves and age-at-length keys) involved in the stock assessment as basis of sustainable fisheries (Martin and Cook, 1990; Thorson and Prager, 2011; Campana, 2001).

Hence, both over- and under-estimation of this biological variable can lead to detrimental effects, with the latter error appearing to be more common (Campana, 2001). One notable example is the ageing error (van den Broek, 1983; Smith et al., 1995) involving orange roughy (*Hoplostethus atlanticus*) in New Zealand where the age was underestimated by a wide margin (i.e. initially thought to attain an age limit of 20-30 years, but was found later to reach around 100 years). This has caused their population to be projected erroneously which led to mismanagement. Other reported ageing errors include those involving *Sebastes* spp in eastern and western Canada (Chilton and Beamish, 1982), and walleye pollock in Central Bering Sea (Beamish and McFarlane, 1995), which also have caused problems related to overfishing resulting to their declining biomasses in a relatively short periods of time. Erroneous estimates of the growth rate, age structures, maturity and lifespan of various fish species and stocks can cause too optimistic projections when it comes to their population productivity, causing mismanagement by setting too high fishing quotas (Campana, 2001).

Obtaining the age of a fish is not a simple process. The most straightforward yet highly inaccurate method is to get the correlation between the length of the fish and age. For example, the so-called Petersen method (Lux and Service, 1971) has been widely applied as a simple way of estimation which works quite well for shorter-lived fish. The idea is to get the frequency of fish in a particular sample having a specific length range to obtain a frequency plot. Figure 1.1 shows a typical plot of fish age versus length that highlights

1

the limitation of this estimation method. As it can be immediately seen, the main problem occurs at higher age values where there is a significant overlap between their lengths, depicted in the plot as the plateauing of length values after reaching certain age.
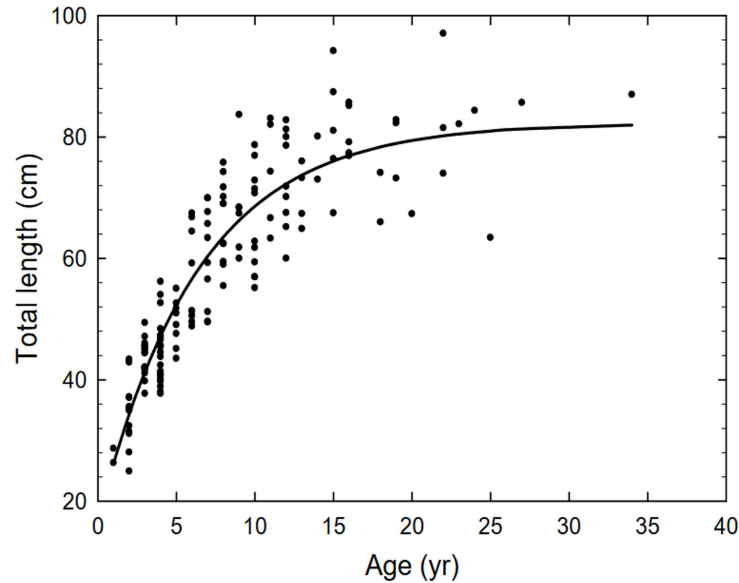
Figure 1.1: A plot of fish age versus length along with a curve-fitting procedure (extracted from Nichols and DeMartini (2008)).

It is therefore important to come up with a more scientifically sound and accurate approach to obtain the fish age. Throughout the decades, many effective approaches have already been employed for fish ageing and it still continuously being improved year after year (Campana, 2001). Moreover, there has been a lot of refinements/improvements in the age reading protocols used which have been disseminated through various events, seminars, workshops and exchanges (Vitale et al., 2019). Also, quality control procedures have been implemented such as having expert-trainee pairs, in a form of apprenticeship where a new reader can be sufficiently guided by the experienced reader. Lastly, it is also sometimes encouraged, whenever possible, to have multiple readers per fish species to help maintain a good quality checking scenario/environment. In the age of artificial intelligence, it would be very interesting to see how it can still be further improved especially with the modern tools that AI provides.

1.2   Otoliths and Traditional Age Reading

Fundamentally, the process of fish age reading has its inspiration from the field of dendrochronology (VanderKooy et al., 2020), where growth patterns among trees are being studied. The cross-section of tree trunks turned out to contain periodic patterns that indicate the number of years of a particular tree. Similarly, researchers tried to find if there is also a fish structure that can also provide reliable patterns that is indicative of the

changing seasons.

The notion that there could be body structures in fish that can indicate periodic/seasonal patterns has been conceptualized even during the early centuries (Jackson, 2007). Anton van Leewenhook through the use of early microscopy have studied different parts of fish that would reveal the fish age. Scales and fins are prime candidates for these season-induced growth pattern formations. First, they have the advantage of easy extraction as they do not require dissecting the fish. However, this also comes with some problems such as the potential to be destroyed in the wild environment of the fish. In this case, these structures will be reabsorbed and regrow causing errors in the age estimate (Campana and Thorrold, 2001).

Due to these problems with external structures, the importance of alternative fish structures which are more stable and do not experience resorption (Campana and Thorrold, 2001) has been greatly highlighted. Focus has then shifted towards internal structures, particularly otoliths, and through the various validation studies (Campana, 2001), it became eventually clear that it gives the most accurate estimate among all the alternatives.

Hence, in recent years, the use of otoliths has then become the standard practice. While it has the primary disadvantage of having to kill and dissect the fish, it makes up for its many useful characteristics. These include continual growth and relative isolation from external conditions as well as the absence of any possibility of resorption (Campana and Thorrold, 2001) .

As part of the inner ear, otoliths are responsible for the balance, movement, and the hearing sense of a fish (Schulz-Mirbach et al., 2019; Campana, 1999). Inside the inner ear, there are three pairs of otoliths situated namely the sagittae, lapilli and the asterisci. As the largest of the three pairs, the sagittae are the commonly used otoliths for age reading. The other two can also be used albeit to a lesser degree. Figure 1.2 shows a simple illustration showing the general structure of the fish inner ear and where the three pairs of otoliths are located inside it.
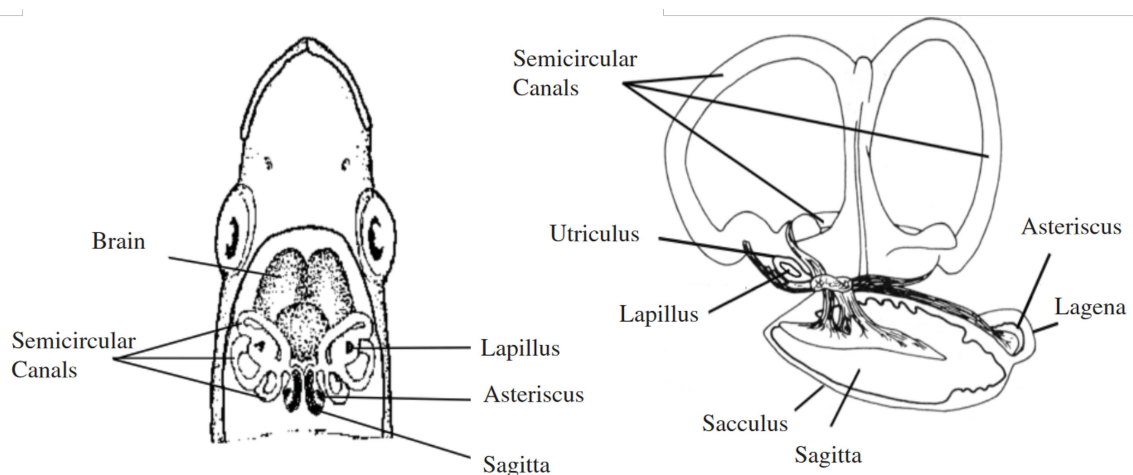
Figure 1.2: An illustration of the the inner ear structure of fish along with the relative locations of the otoliths (extracted from VanderKooy et al. (2020)).

Otolith Preparation

As mentioned earlier, one of the main drawbacks of using otoliths for fish age reading is the fact that one needs to kill and dissect the fish in order to extract the otoliths. Also, the whole sequence of steps is not straightforward and certain specific steps and specialized tools are required. The next subsections enumerate the major steps for the otolith image preparation and some extra considerations taken for proper handling (Bernreuther and Wilhelms, 2013).

- **Dissection of the Fish**

  Here the fish is cut along the ventral side starting from the pelvic fin up until the head region. Using a scalpel or pair of scissors, the head area is split into two and then the obstructing structures are removed (e.g., gills and other tissues). After this step, the otoliths will then become visible.

    Inside the concavity containing the fish brain, otoliths can be found located around the area underneath or near the eye socket. In most fishes, the largest otolith, the sagitta, can be located easily along the vicinity of the skull. Using forceps, the sagitta can then be easily taken out without much resistance.

Figure 1.3: Preparation and dissection of fish.

- **Extraction of Otoliths**

  After extraction, the otoliths are then collected, cleaned and organized into groups for the next step. For species such as herring, the collected whole otoliths can be used directly for analysis and microscopy and the steps related to sectioning can be skipped. For other species such as cod, it is common to perform sectioning first before analysis can be done. The core or nucleus of the otolith is marked using a pencil so that upon sectioning, the slice will be made in the portion where the centre of the otolith is represented fully.
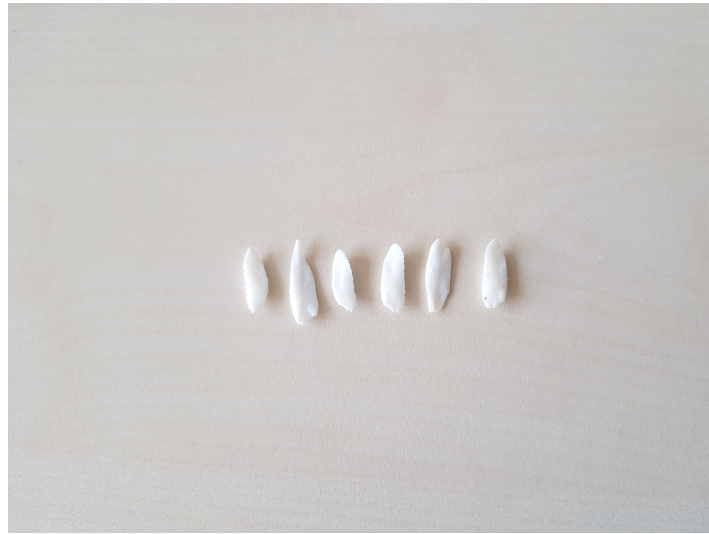
Figure 1.4: Extraction and cleaning of otoliths

- **Embedding**

  Next comes the embedding of the otoliths to a resin so that they can stay in place for easy sectioning. In this step, the otoliths are placed in a container to be filled with polyester resin to be poured in two separate stages. Extra care is needed so that the formation of bubbles will be prevented which may interfere with the microscopy and image analysis. The first batch of resin poured on the empty container takes about half an hour to solidify. Then, the collected otoliths can be placed on top of the resin with an aligned placements and uniform orientations. Afterwards, the second batch of resin is poured which takes longer (days) to solidify fully. It is important to make sure that every portion has already dried up before proceeding to the next steps.
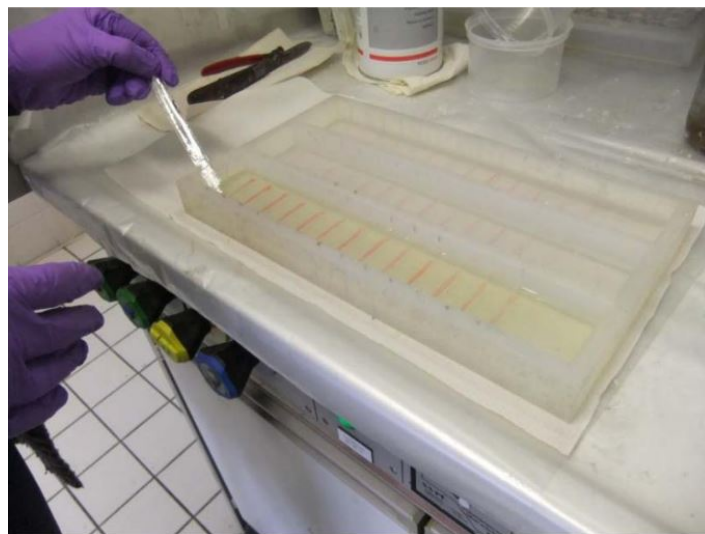
Figure 1.5: Embedding of otoliths on a polyester resin.

- **Cutting or Sectioning**

  After the resin dries up, the otoliths can now be subjected to sectioning. Using the marked portions of the otoliths, a diamond-coated double sawing blade is positioned to cut the otoliths into sections passing through the nucleus or core. The thin sections can then be mounted flat on a glass slide (i.e., in some laboratories, the glass slide is guided by a piece of spaghetti noodle which is having the same thickness as the thin otolith section). Care must be taken such that the placement of the otolith sections on the slide is secure and stable.
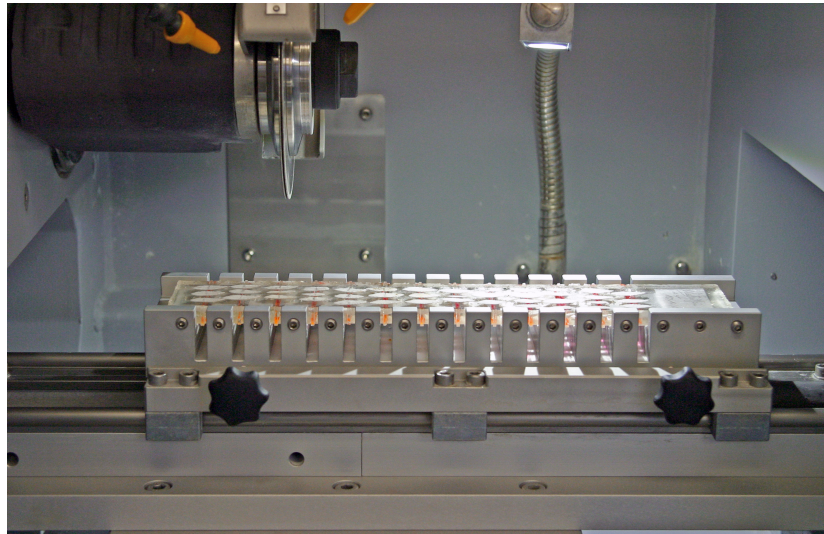
Figure 1.6: Otolith sectioning using specialized tool.

- **Viewing**

  Finally, viewing of the sectioned otoliths can then be performed with a digital camera directly connected to the microscope eyepiece/viewport that can take pictures with the help of bundled imaging software. Here, certain adjustments to the image can be done such as increasing the brightness or exposure, zooming in or out as necessary and some other image preprocessing which might be specific to the reader or the laboratory. Once the images are captured, they are saved and organized properly in a folder while at the same time, the reader records the relevant metadata such as the fish identifier, trip number and other important information. Optionally, this is also the step in which the age reader makes the manual reading which can be recorded right away on the metadata.

  At this point, it would be good to emphasize that this is the step where the AI-based methods developed so far are designed to be engaged (i.e., can be of help or assistance). As pointed by Moen et al. (2018), the bulk of the processing is actually spent on the other steps and the time-saving benefit of AI-based approaches are not really amounting to a significant proportion. Nevertheless, this is the most important part since it is the step where errors are made which have big consequences, as highlighted in the earlier sections. Hence, this will be the crucial stage in which AI-based approaches can be of great utility.
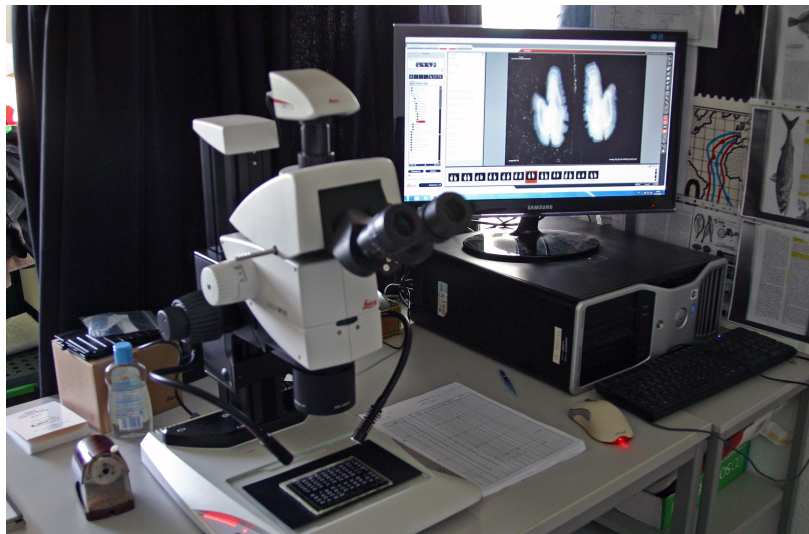
Figure 1.7: Microscopy and viewing of otolith image.

1.3 Workshops and Exchanges

In order to provide reliable estimate, extensive training of age readers is required which can take several years. Sometimes, the protocols can be highly specific to a given species and can even be contradictory to the guidelines for other species. Lastly, workshops and exchanges are needed which create an avenue for age readers to cross-check each other's methodologies and to come up with consistent and comprehensive guidelines. An example of an output from a workshop is shown in Figure 1.8.
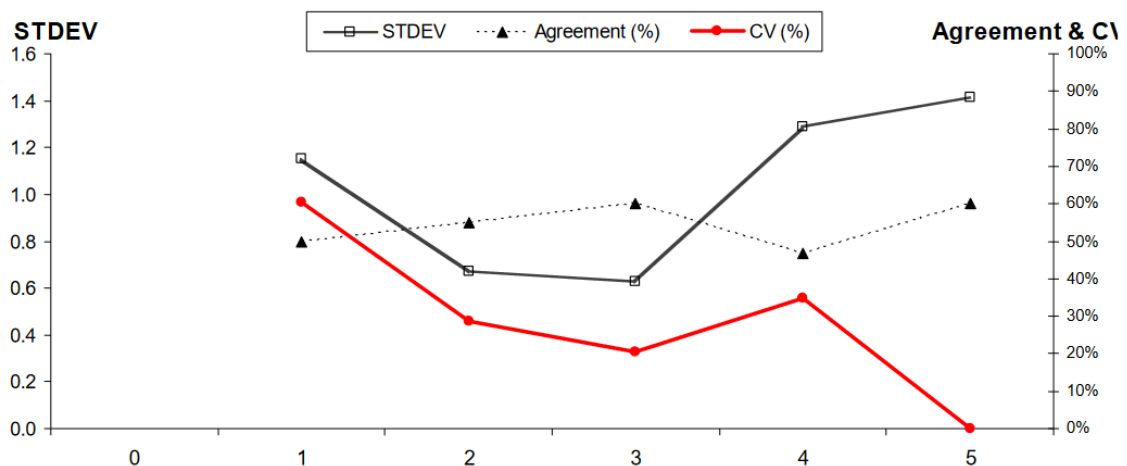
Figure 1.8: Sample result of a workshop measuring the values for percent agreement, standard deviation and coefficient of variation (CV) among age readers for North Sea cod. (extracted from ICES (2008)).

It can be seen from the sample workshop results (ICES, 2008) that there can be a wide range of disagreements among the age readers. For deciding the accuracy, the modal age (i.e., the most common) is often taken as the accepted value of the age from which evaluation can be done. In addition to the percent agreement, another important metric computed from the workshops is the so-called Coefficient of Variation (CV) which is basically a measure of the standard deviation over the mean of the age readings. This will indicate the degree in which the readings differ from each other and hence will be a good measure of the overall agreement among the age readers.

1.4   Age Validation

As mentioned in the previous section, the modal age among the readers can be considered as the proper age in the absence of actual validated age value. However, this has a major problem when it happens to be inaccurate due to some mis-interpretation of certain ring features. Hence, it is important to have a reference collection of validated images in order to check for the possibility that the guidelines for manual age reading being followed by most readers correspond to the actual age of the fish. Unfortunately, the process of validation is a very tedious process and can only be done in some selected circumstances. Also, there are several types of validation techniques that can be applied depending on the scenario and often there is no methodology that works for all fish species. These are discussed on the next subsections as derived from the review by Campana (2001).

Carbon Dating/Bomb Radiocarbon

This method is particularly useful for long-lived fish species (Campana, 2001). This method makes use of the nuclear testing that was conducted around 1958-1965 that cause the release of atmospheric ^{14}C which were eventually incorporated into the structures in fish, corals and other animals in the sea. The radioactive decay rate for this isotope can then be used to age the fish and validate whether the number of periodic patterns from otoliths correspond to this computed age value.

Radiochemical Dating

This method has some similarity to the one involved with bomb radiocarbon. In this method, however, the radioactive decay of naturally occurring radioisotopes such as ^{210}Pb, ^{226}Ra and ^{228}Th are being monitored and analyzed in order to derive the fish age (Campana, 2001). The resulting values from this analysis can then be used to validate the age values that were derived from the traditional ring counting approach involving otoliths or some other relevant structures.

Mark-Recapture

In this process, certain fish cohorts of known age are tagged with a special marker (e.g. OTC) and then are released in the wild (Campana, 2001). After certain periods of time, these are then recaptured and studied in order to verify whether the added growth rings in the otolith correspond to the time period they are released. This analysis is often done in conjunction with other analyses such as age-length analysis in order to study the growth rates and age structure of a fish stock (McQueen et al., 2018).

Marginal Increment Analysis

For this method, the underlying assumption is that, for a growth pattern to be valid, there should be some periodicity when it comes to the formation of the pattern in proportion to the time duration spent on a given partial time period (e.g. annual or even daily) Campana (2001). Then by measuring and plotting the growth of these potential patterns, a sinusoidal curve could be observed indicating that indeed the formation occurs periodically. Unfortunately, although this is the most common validation technique, it has some subjective aspects (e.g., quantifying the partial growth of a particular ring) and is also known to be affected by the microscopy aberrations inherent in the observation of the marginal edges.

1.5 Image Analysis and Automation

Researchers have been trying to automate ways to derive age estimates based from otoliths. The review by Fisher and Hunter (2018) provides an interesting overview of the different attempts for automation proposed throughout the years involving classical image processing and traditional computer vision. The fundamental working premise of these approaches is that the winter and summer rings will be manifested in the otolith image as alternating dark and light bands that can be easily quantified and analyzed.
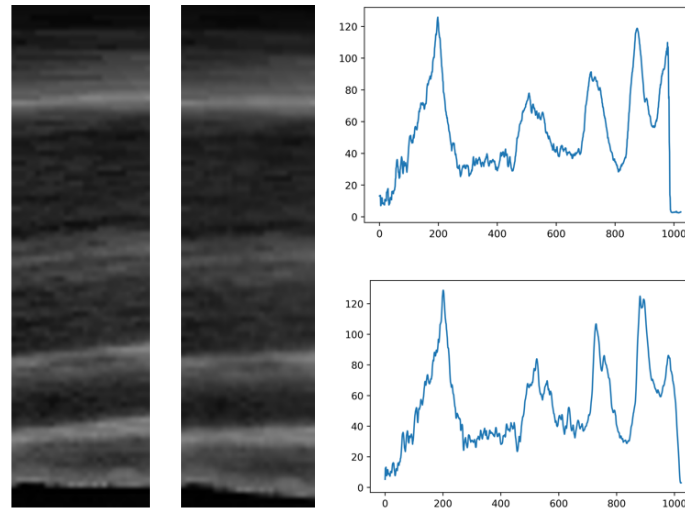
Figure 1.9: An example of image intensity bands on a small strip of otolith along with the corresponding intensity curve with peaks and troughs.

Several attempts in this direction (Troadec, 1991; Formella et al., 2007; Fisher and Hunter, 2018) have been tested which date back to the early days of the field of computer vision. In this approach, the alternating band of opaque and translucent rings can be directly measured through the changing intensity values similar to the one shown in Figure 1.9. In this figure, the peaks and troughs of the intensity curve are counted by a variety of peak counting algorithms or even transformed into a different space (e.g. Fourier space) from which the counting can be done automatically. This approach has shown promising results especially with fish species that have near consistent concentric rings such as plaice (Fablet, 2006). However, this method suffers a lot when the growth rings are not consistent throughout the otolith and when the outer otolith shape no longer aligns with the curvature of the rings.

1.6   Artificial Intelligence and Deep Learning

The field of artificial intelligence has grown substantially over the recent decades. Some of the famous classical algorithms include Support Vector Machines (SVMs), Random Forests and Artificial Neural Networks (ANNs). During this classical machine learning period, there is a separate subject area known as feature engineering (Mahony et al., 2019; Dzieżyc et al., 2020) which involves developing methods to find the best feature sets to use as inputs of the algorithms. Although this additional step seems tedious and complicated, at least, there is somehow a control on the features that the algorithms will use for the prediction, making them semi-explainable in certain ways. Also, the domain knowledge (Dzieżyc et al., 2020) of the feature engineer is highly relevant for this part as it greatly aids in designing the best feature sets.
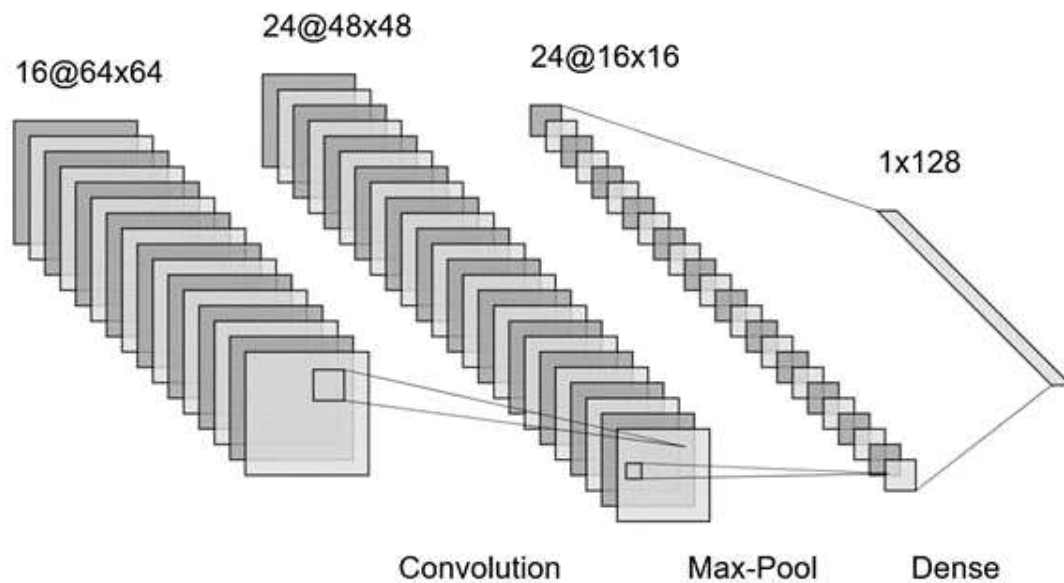
Figure 1.10: An example of convolutional neural network architecture showing the large number of layers and the multiple operations involved.

With the introduction of deep learning, the practice of feature engineering already became obsolete as the network itself can already perform the feature extraction step (Bengio et al., 2013). In addition, the network connecting inputs and outputs became enormous and mathematically complicated which again further contributed to its somewhat "black box" characteristics.

In a way, most deep learning algorithms can be fundamentally compared to the classical machine learning algorithm which is the Artificial Neural Network (ANN). In this algorithm, inputs in the form of "feature sets" are connected into nodes organized into several layers (O'Shea and Nash, 2015). The nodes from one layer are fully-connected to the other nodes of the next layer and so on. Most ANNs are designed with only around 3-4 layers as adding more layers greatly increases the processing time with minimal gains. Although effective in some problems, in the early years of computing, the computational power was too insufficient for it to work effectively with reasonable training duration.

During this classical machine learning era, there were already some ground-breaking works from LeCun et al. (2015), which are now considered pioneering studies paving the way for deep learning. In their deviation from the design of the artificial neural network, they removed certain inefficient components and created various operations that create more meaningful and computationally efficient connections and weight updates. As a result, the Convolutional Neural Network (CNN) was developed that has certain operations like Convolution and Pooling which transforms the original feature space into one where the relevant and important features can be emphasized. Figure 1.10 shows these operations and how they fit in the overall design. It can be seen that this network is "deeper" than the classical neural network formulation as it can now contain arbitrarily

large number of layers.

A lot of studies followed afterwards where it became clear how CNNs can perform excellently in the task of image classification. In 2012, Krizhevsky et al. (2012) demonstrated the capability of this architecture on ImageNet dataset and outperformed the algorithms that are considered best for the task. Since then, there were a lot of subsequent further applications of CNN on image classification that it became the state-of-the-art when it comes to such task.

1.7   Mask R-CNN

Region-based CNN got their name from focusing on a region of the image instead of the entire image (He et al., 2017). Earlier algorithms utilizing this concept have undergone incremental improvements throughout the years starting with the basic Fast R-CNN (Girshick, 2015), which was optimized further to become Faster R-CNN (Ren et al., 2015) then later improved to pave the way to Mask R-CNN. All of these algorithms work on a common framework which has two basic network branches: one to detect the bounding box of a particular region of interest (ROI) and another to predict the class (classify) of that detected region. The main addition of Mask R-CNN to this family of algorithms is that there is a third network branch that predicts the segmentation mask hence making it able to accomplish both detection and segmentation tasks simultaneously (He et al., 2017). Figure 1.11 shows the simplified architecture of Mask R-CNN.
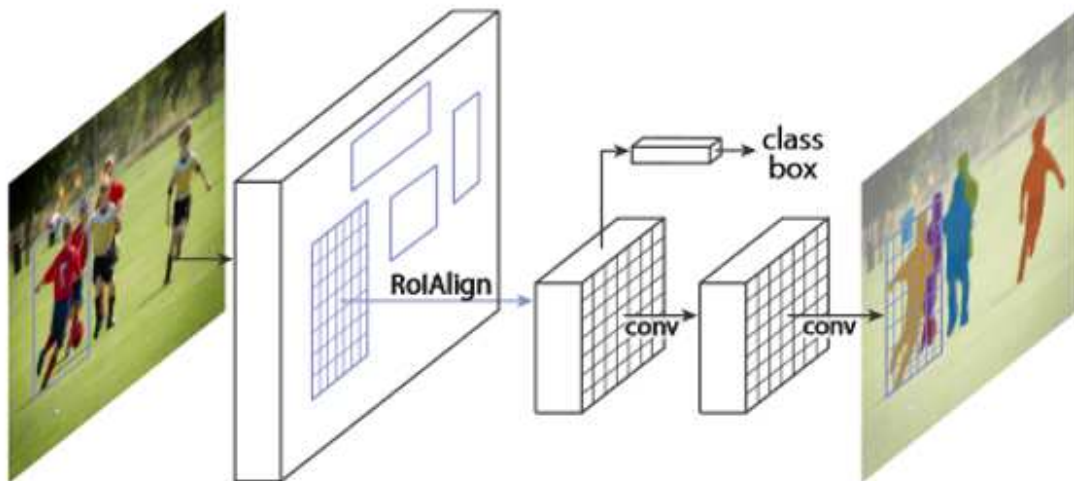
Figure 1.11: Network design of Mask R-CNN showing the multiple branches of the architecture(extracted from He et al. (2017)).

1.8   U-Net

As another widely popular algorithm based on CNN, the U-Net got its name from the
"U" shaped architecture which was designed primarily for the purpose of medical image
segmentation (Ronneberger et al., 2015). There are two distinctive phases for the algorithm.
The first phase involves a contracting path (a common CNN design) where series of
downsampling operations are applied reducing the feature dimensions but increasing
the feature channels. The second phase is the expansive path where the operations are
made for upsampling and the dimensions are increased (until it gets back to the original
image dimension) while the feature channels are reduced. Figure 1.12 shows the U-shaped
architectural design of this algorithm.
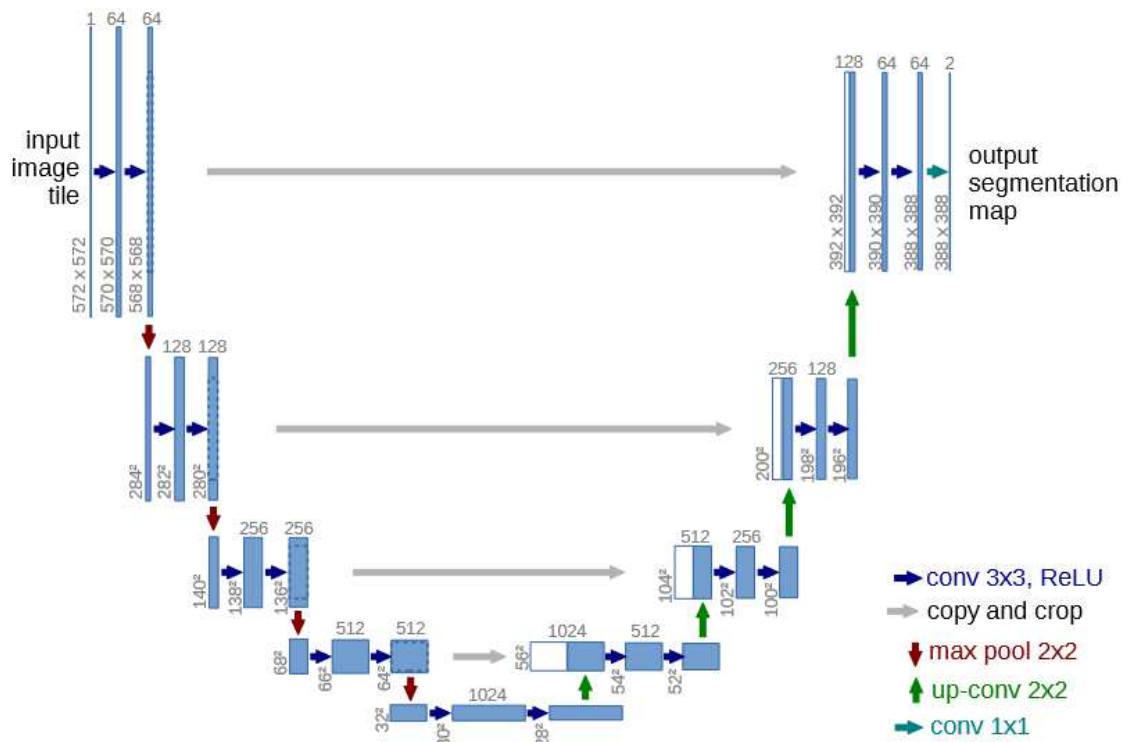
Figure 1.12: Network design of U-Net showing the typical encoder-decoder architecture
(extracted from Ronneberger et al. (2015)).

1.9   Ground-Truth Labeling

As supervised algorithms, both U-Net and Mask R-CNN require ground-truth labels for
their training. Using these labels, the algorithms can compute, during each iteration of the
training, how far their current predictions are to the expected value (i.e., the ground-truth).
In contrast to classification or regression tasks, the labels are not just categories or numeric

values. Instead, certain regions of the image are marked in a way that describes the location of the object of interest (i.e., object to be detected or segmented).

Naturally, the object of interest is in the form of otolith regions or specific portions of the otoliths that are intended to be detected. For this study, the ground-truth labeling follows the usual manual age reading practice where age readers make use of two major axes (left and right), whichever contains the complete set of rings. When there are obstructions and microscopic aberrations within these axes, the readers can sometimes use some other axes such as the central axis or even along the sulcus edges. In this thesis, however, we limit the readings only on the two major axes (i.e. dorsal (left) and ventral (right) axes), see Figure 1.13.
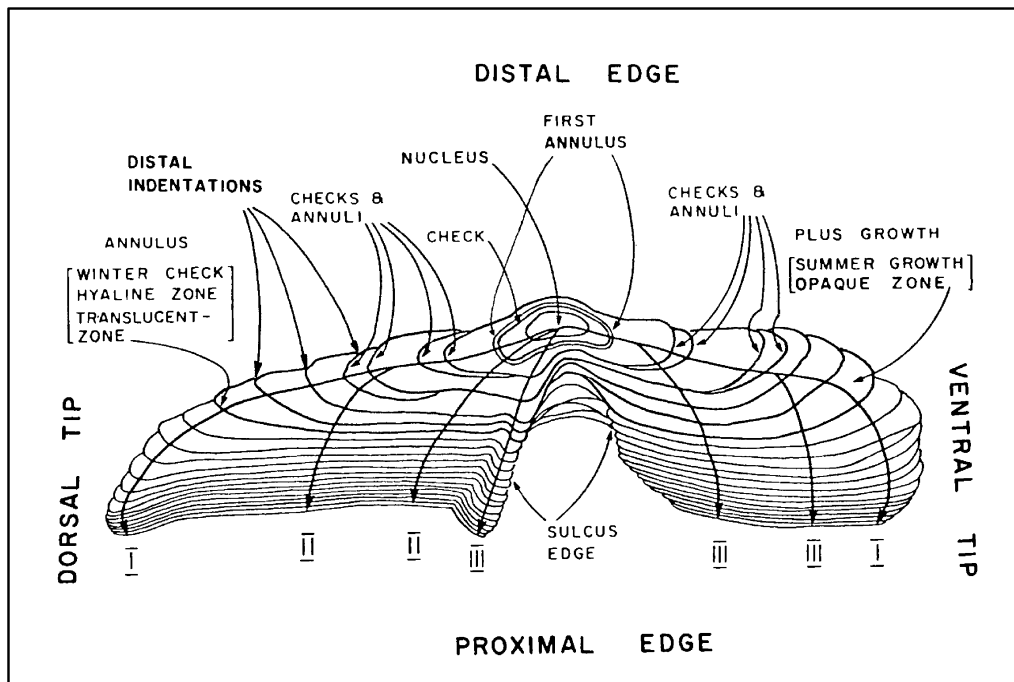
Figure 1.13: Different parts and regions of an otolith, including preferred age reading axes (extracted from MacLellan (1997)).

Visual Geometry Group Image Annotation (VIA) Tool

There are several image annotation toolkits available. Some are even commercial software tools while others are free and also open source. Despite being of the latter category, VIA (Dutta and Zisserman, 2019) contains all the necessary features needed for the study. In addition, due to its public license, there are certain modifications that can be added onto it such that the ground-truth creation becomes much simpler and faster.

By default, VIA contains several default tools that are already very convenient for annotating images. One can create bounding boxes or circles easily through its default shape options. For regional CNN algorithms that performs only bounding box detection without segmentation, these typical drawing tools are already sufficient.

In case of irregular shapes, one can use its point-by-point drawing facility to draw any arbitrary shape that covers the object of interest. This was perfectly applicable to the case of otolith rings although this step can become tedious and time-consuming. Hence, as part of the project, a new feature in the form of a brush tool was incorporated and customized into the VIA toolkit. This makes the annotation of the rings significantly easier and faster.

In addition, we edited that code so that it can be integrated into our own web-based application. As a result, apart from the efficient brush toolkit, exporting the annotations can be done without having to organize manually the folders as it is already connected to the main application's database and folder structure.

Training Loss Computation from Ground-Truth Values

During training, in simplified terms, learning is accomplished by the assignment and updating of weights on the network depending on the computed value of the loss functions implemented, a process known as backpropagation (O'Shea and Nash, 2015). In other words, fine-tuning the network weights will follow (the gradient) from the computed difference between the expected output and the predicted value.

One basic loss function is the so-called Mean Squared Error (MSE) where it computes the average of the squared difference between the predicted and expected values. This has been used in some widely used deep learning implementations for fish age determination (Moen et al., 2018; Martinsen et al., 2022).

The formulation of the loss functions for the two algorithms, Mask R-CNN and U-Net, varies significantly due to the differences in their overall designs. The Mask RCNN have multiple loss functions due to the different tasks it is trying to accomplish. It aims to detect the bounding box while also creating segmentations before finally trying to classify the detected object of interest into the appropriate class.

For the U-Net, the usual formulation is in the form of single task where the network is made to detect individual pixels of an image and classify each one whether it is a part of the background or the actual object of interest. However, depending on certain tasks and experiments, these loss functions can be customized and modified accordingly.

1.10   Dataset Overview

We obtained image datasets from two sources, namely, North Sea dataset and Baltic Sea dataset. The North Sea dataset has been captured and processed by the otolith age reading group from the Thünen Institute of Sea Fisheries. The Baltic dataset, on the other hand, was obtained from the Thünen Institute of Baltic Sea Fisheries. Below is a figure showing the sampling locations of the otoliths followed by paragraphs describing the details based on the report from ICESFIshMap (2005) regarding the species involved in the study.
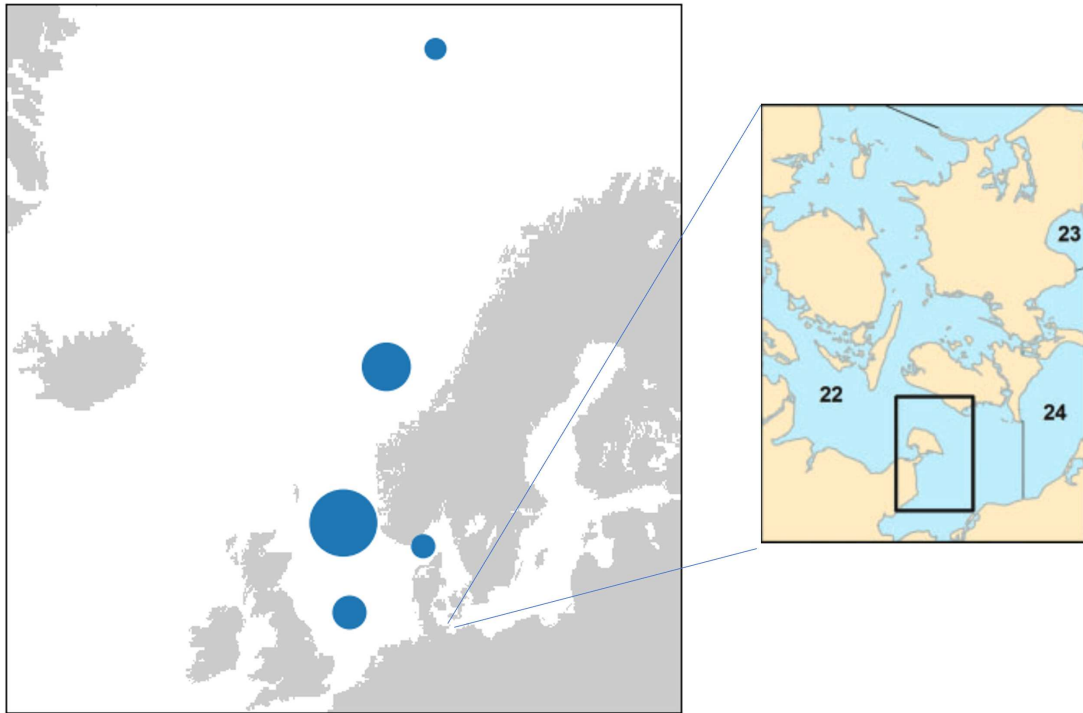
Figure 1.14: General sampling locations from which large proportions of otolith images are obtained in the study. The Baltic location is represented via zoomed-in map (extracted from McQueen et al. (2018)).

North Sea Cod

Abundant in ICES Divisions 4a, 4b, and 2a, the North Sea cod, a stock of the Atlantic cod (*Gadus morhua*), is a commercially and biologically important fish stock that comprises a large proportion of the overall catch among several surrounding countries namely Germany, Denmark, Norway, the UK, Belgium and the Netherlands (Horwood et al., 2006; ICESFIshMap, 2005). Historically, the biomass of this fish stock is very high. However, in recent years, it has undergone a substantial decline, making it a target of extensive management and monitoring efforts.

Baltic Cod

As another subgroup of the Atlantic cod, the Baltic cod is also a subject of extensive monitoring efforts due to its substantial decline in recent decades. In fact, there is a recent ICES advice indicating a total ban on targeted fishing of this fish stock (Birgersson et al., 2022; ICESFIshMap, 2005). In recent decades, in addition to high fishing pressure, the Baltic cod has also been greatly affected by several environmental and climate factors which is also somehow linked to the geography of the Baltic Sea.

Saithe

As another commercially important species in this study, saithe (*Pollachius virens*) have been one of the fish stocks that are abundant in the North Sea that also comprises a huge portion of the overall catch of the German fishing fleet (Edebohls et al., 2022). This makes it a very important fish stock to study and to properly employ accurate ageing methods.

Whiting and Haddock

Other species included in this study that are also known to be commercially important are the whiting (*Merlangius merlangus*) and the haddock (*Melanogrammus aeglefinus*). Both of these species also comprise a major part of the total catch in German fisheries (Edebohls et al., 2022). Compared to the cod and saithe, these groups are not in rapid decline and it is projected that they maintain a sustainable population (ICESFIshMap, 2005).

1.11 Research Questions

- How can the advances in the field of artificial intelligence be used to effectively perform automated fish age reading and annotation of the annuli or growth rings?

- How can the AI based approach be designed such that it can easily gain the trust and wide acceptance of the community?

- How will the approach hold when it comes to various datasets?

- How will it hold against different tasks involved in general otolith analysis?

- What are other important advantages of the AI-based methods and implications of the study that can facilitate the adoption of AI for routine age reading tasks?

1.12 Thesis Structure and Publication List

The remaining chapters of the thesis are organized in the following sequence along with the relevant publications and papers under review or in preparation.

- Chapter 2 includes the first paper (published) which involves proving the effectiveness of the proposed approaches and performing benchmarks against existing methods.

  Cayetano, A., Stransky, C., Birk, A., and Brey, T. (2024). Fish Age Reading Using Deep Learning Methods for Object Detection and Segmentation. *ICES Journal Of Marine Science* **81**: 687-700

- Chapter 3 includes the second paper (under review) where the topic is on extending the algorithms further with advanced techniques as well as looking for ways to increase the usage and adoption of the approaches for general use.

Cayetano, A., Stransky, C., Birk, A., and Brey, T. (2024). An Interactive AI-driven Platform for Fish Age Reading. *PLOS ONE (under review)*

- Chapter 4 includes the third paper (in preparation), where both methods, Mask R-CNN and U-Net, are used not only for age reading but also for some other tasks. Here we demonstrate their adaptability for different tasks under a multi-stage framework.

  Cayetano, A., Stransky, C., Birk, A., and Brey, T. (2024). Multi-stage Framework for Otolith Analysis. *in prep.*

- Chapter 5 includes the synthesis where the findings from the different papers are summarized and presented in ways that answers the underlying research questions formulated for this PhD thesis.

1.13   References

Beamish, R. and McFarlane, G. (1995). A discussion of the importance of aging errors, and an application to walleye pollock: the world's largest fishery. In *Recent Developments in Fish Otolith Research*, pages 545–565. Columbia: University of South Carolina Press.

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:1798–1828.

Bernreuther, M. and Wilhelms, I. (2013). Dünnschnitte Manual- Thünen Institute of Sea Fisheries.

Birgersson, L., Söderström, S., and Belhaj, M. (2022). The Decline of Cod in the Baltic Sea – A review of biology, fisheries and management, including recommendations for cod recovery. *The Fisheries Secretariat, Stockholm, Sweden*.

Campana, S. (1999). Chemistry and composition of fish otoliths: pathways, mechanisms and applications. *Marine Ecological Progress Series*, 188:263–297.

Campana, S. (2001). Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology*, 59:197–242.

Campana, S. and Thorrold, S. (2001). Otoliths, increments, and elements: keys to a comprehensive understanding of fish populations? *Canadian Journal of Fisheries and Aquatic Sciences*, 58:30–38.

Chilton, D. and Beamish, R. (1982). Age determination methods for fishes studied by the groundfish program at the pacific biological station. *Canadian Special Publication of Fisheries and Aquatic Sciences*, 60:1–102.

Dutta, A. and Zisserman, A. (2019). The VIA Annotation Software for Images, Audio and Video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, New York, NY, USA. ACM.

Dzieżyc, M., Gjoreski, M., Kazienko, P., Saganowski, S., and Gams, M. (2020). Can we ditch feature engineering? end-to-end deep learning for affect recognition from physiological sensor data. *Sensors*, 20(22).

Edebohls, I., Niemann, M., Berkenhagen, J., Döring, R., and Schröder, A. (2022). Steckbrief zur Meeresfischerei in Deutschland 2022. *Johann Heinrich von Thünen Institute, Germany*.

Fablet, R. (2006). Statistical learning applied to computer-assisted fish age and growth estimation from otolith images. *Fisheries Research*, 81:219–228.

Fisher, M. and Hunter, E. (2018). Digital imaging techniques in otolith data capture, analysis and interpretation. *Marine Ecology Progress Series*, 598.

Formella, A., Vázquez, J., Carrión, P., Cernadas, E., Vázquez, A., and Pérez-Gándaras, G. (2007). Age reading of cod otoliths based on image morphing, filtering and fourier analysis. *Proceedings of the 7th IASTED International Conference on Visualization, Imaging, and Image Processing*.

Girshick, R. (2015). Fast R-CNN. *IEEE International Conference on Computer Vision (ICCV)*.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988.

Horwood, J., O'Brien, C., and Darby, C. (2006). North Sea cod recovery? *ICES Journal of Marine Science*, 63(6):961–968.

ICES (2008). Report of the Workshop on Age Reading of North Sea Cod (WKARNSC). *ICES CM 2008/ACOM*, 39:71.

ICESFIshMap (2005). ICES FishMap. https://www.ices.dk/about-ICES/projects/EU-RFP/Pages/ICES-FIshMap.aspx.

Jackson, J. R. (2007). Earliest references to age determination of fishes and their early application to the study of fisheries. *Fisheries*, 32(7):321–328.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521:436–444.

Lux, F. and Service, U. S. N. M. F. (1971). *Age Determination of Fishes: (revised)*. Fishery leaflet. United States Department of Commerce.

MacLellan, S. (1997). How to age rockfish (Sebastes) using S. alutus as an example - the otolith burnt section technique. *Can. Tech. Rep. Fish. Aquat. Sci*, 2146:42.

Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G., Krpalkova, L., et al. (2019). Deep Learning vs. Traditional Computer Vision. https://arxiv.org/pdf/1910.13796.

Martin, I. and Cook, R. (1990). Combined analysis of length and age-at-length data. *ICES Journal of Marine Science*, 46:178–186.

Martinsen, I., Harbitz, A., and Bianchi, F. (2022). Age prediction by deep learning applied to Greenland halibut (*Reinhardtius hippoglossoides*) otolith images. *PLOS ONE*, 17:e0277244.

McQueen, K., Hrabowski, J., and Krumme, U. (2018). Age validation of juvenile cod in the western Baltic Sea. *ICES Journal of Marine Science*, 76:430–441.

Moen, E., Handegard, N., Allken, V., Albert, O., Harbitz, A., and Malde, K. (2018). Automatic interpretation of otoliths using deep learning. *PLOS ONE*, 13:e0204713.

Nichols, R. and DeMartini, E. (2008). *Preliminary estimates of age and growth for the endemic Hawai'ian Grouper (Hapu'upu'u, Epinephelus quernus, F. Serranidae).* Pacific Islands Fish. Sci. Cent., Natl. Mar. Fish. Serv., NOAA, Honolulu, HI 96822-2396. Pacific Islands Fish. Sci. Cent. Admin. Rep. H-08-06, 19p.

O'Shea, K. and Nash, R. (2015). An Introduction to Convolutional Neural Networks. *CoRR*, abs/1511.08458.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer. (available on arXiv:1505.04597 [cs.CV]).

Schulz-Mirbach, T., Ladich, F., Plath, M., and Heß, M. (2019). Enigmatic ear stones: what we know about the functional role and evolution of fish otoliths. *Biological reviews of the Cambridge Philosophical Society*, 94.

Smith, D., Fenton, G., Robertson, S., and Short, S. (1995). Age determination and growth of orange roughy (*Hoplostethus atlanticus*): a comparison of annulus counts with radiometric ageing. *Canadian Journal of Fisheries and Aquatic Sciences*, 52:391–401.

Thorson, J. and Prager, M. (2011). Better catch curves: Incorporating age-specific natural mortality and logistic selectivity. *Transactions of the American Fisheries Society*, 140:356–366.

Troadec, H. (1991). Frequency demodulation on otolith numerical images for the automation of fish age estimation. *Aquatic Living Resources*, 4:207–219.

van den Broek, W. (1983). Ageing deepwater fish species: report of a visit to the United Kingdom September-November 1982. *Miscellaneous Series Fisheries Research Division, Ministry of Agriculture and Fisheries (unpublished report, MAF Fisheries, Greta Pt Library, Wellington, NZ)*.

VanderKooy, S., Carroll, J., Elzey, S., Gilmore, J., and Kipp, J. (2020). A practical handbook for determining the ages of Gulf of Mexico and Atlantic Coast fishes. *Gulf States Marine Fisheries Commission Publication*, 300.

Vitale, F., L. Worsøe Clausen, L., and Chonchúir, G. N. (2019). Handbook of fish age estimation protocols and validation methods. In *ICES Cooperative Research Report No. 346. 180 pp.*

Chapter 2

Fish Age Reading Using Deep Learning Methods for Object Detection and Segmentation

Arjay Cayetano[1*], Christoph Stransky[1], Andreas Birk[2], Thomas Brey[3]

1 Thünen Institute of Sea Fisheries, Bremerhaven, Germany
2 School of Science and Engineering, Constructor University, Bremen, Germany
3 Faculty of Biology and Chemistry, University of Bremen, Bremen, Germany

2.1   Abstract

Determination of individual age is one essential step in the accurate assessment of fish stocks. In non-tropical environments, the manual count of ring-like growth patterns in fish otoliths (ear stones) is the standard method. It relies on visual means and individual judgement and thus is subject to bias and interpretation errors. The use of automated pattern recognition based on machine learning may help to overcome this problem. Here, we employ two deep learning methods based on Convolutional Neural Networks (CNNs). The first approach utilizes the Mask R-CNN algorithm to perform object detection on the major otolith reading axes. The second approach employs the U-Net architecture to perform semantic segmentation on the otolith image in order to segregate the regions of interest. For both methods, we applied a simple postprocessing to count the rings on the output masks returned which corresponds to the age prediction. Multiple benchmark tests indicate promising performance of our implemented approaches comparable to recently published methods based on classical image processing and traditional CNN implementation. Furthermore, our algorithms showed higher robustness compared to the existing methods, while also having the capacity to extrapolate missing age groups and to adapt to a new domain or data source.

Keywords: *fish age reading, otoliths, artificial intelligence, deep learning, object detection, segmentation*

2.2   Introduction

Individual age is an essential parameter in the analysis of fish population dynamics and thus a precondition for both sustainable management and a thorough understanding of the ecological role of a fish stock. The common approach in estimating the age of a fish is to make use of patterns along calcified structures such as scales and otoliths (ear stones) and observe the appearance of the annual growth zones (or annuli) (Panfili et al. (2002)). These growth zones are formed by the uneven deposition of calcium carbonate and proteins as the fish experiences seasonal changes. Correspondingly, each single alternating opaque and translucent ring formation represents a period of one year (Campana, 1999; Panfili et al., 2002). Hence, in traditional age reading, human experts perform manual counting of these ring patterns which require individual judgement especially if the rings are hardly distinguishable.

The pattern of ring formations can be distinct for each fish species, hence making the task of annual growth zone detection extremely challenging. Moreover, due to known environmental effects on otolith growth (Campana, 1999), even different stocks of the same species can also have different ring patterns (Williams et al., 2005). In some cases, false rings and double rings can occur which may lead to over-estimation of fish ages. Likewise, some rings can also be very faint and ambiguous, leading to under-estimated age values (Campana, 1999; Carbonara and Follesa, 2019).

As otolith images and age data are collected in large quantities by various institutions as part of routine stock assessment, it is necessary to make the process of age reading

scalable and less error-prone. In addition, the lack of age readers for a given species can also be a limitation due to the extensive nature of training required. Even an expert on one species needs to be trained again for another species due to the differences in guidelines and protocols. Hence, it is not surprising that over the recent decades, a lot of attempts have been made to explore the possibility of automating the process. The first approaches were based on classical image processing techniques coupled with signal processing methods (Troadec, 1991; Formella et al., 2007; Fisher and Hunter, 2018). This usually involves reading the intensity peaks within a specific sector of the otolith starting from the core (nucleus) down to the outer edge.

As the field of artificial intelligence (AI) has become more and more advanced, automation efforts shifted towards the use of approaches based on machine learning. Fablet and Josse (2005) designed one of the earliest studies utilizing machine learning algorithms to classify otolith images according to age groups. They explored the use of Support Vector Machines (SVM) and Artificial Neural Network (ANN) coupled with some elements of classical image processing as part of feature engineering. The work done by Bermejo et al. (2007) is another classical machine learning approach involving the use of hand-crafted morphological features combined with principal component analysis (PCA) and SVM.

Recently, with the emerging popularity of deep learning, the practice of feature engineering becomes obsolete due to the fact that this process is incorporated in the learning network itself (Bengio et al., 2013). Moen et al. (2018) became one of the earliest adopters of this technology when they used Convolutional Neural Network (CNN) and regression to obtain good age estimates for Greenland halibut (*Reinhardtius hippoglossoides*) otoliths.

One main issue with this existing deep learning formulation, however, is the seemingly black-box nature of the process. It is able to give age estimates but it provides no direct information on how it derived such predictions. The follow-up studies by Ordoñez et al. (2020) and Martinsen et al. (2022) aimed to find some potential clues and explanation in the form of heatmaps indicating individual pixel relevance. While they managed to show the focal regions considered by the algorithm, some doubts still remain as these highlighted parts are not the usual areas associated with manual age reading process.

Another argument against the above-mentioned traditional CNN approaches is that they are known to require a large amount of training data in order to avoid overfitting. Hence, given a limited set of image data, it is possible that the implemented deep learning algorithm can only handle datasets that are very similar to those used during training. Consequently, it is very likely that the resulting deep learning model will not be robust enough to generalize and extrapolate on seemingly unfamiliar data. Recently, there have been several new studies implementing novel methods not covered in this study such as the use of Transformers by Sigurdardóttir et al. (2023) and Ensemble Learning by Moen et al. (2023) which potentially can address the mentioned shortcomings of traditional CNN while the issues of explainability remain.

In our study, we propose to overcome these limitations by reformulating the problem and approaching it from the perspective of object detection and segmentation. That is, we directly adopt how the manual age reading process is done by explicitly performing detection and/or segmentation of annual rings which will then be automatically counted to derive the age estimates. To accomplish this, we utilize two deep learning algorithms,

namely Mask R-CNN (He et al., 2017) and U-Net (Ronneberger et al., 2015), which are known for their effectiveness in detecting or segmenting, respectively, any specified region of interest on a given image.

In this proposed reformulation of the problem, we aim to reduce the level of abstraction inherent in the process and increase the explainability of the deep learning-based approach by making the procedure directly compatible with the traditional ring counting method used by humans. Also, we hypothesize that the number of required images for training will be considerably less as each image is already composed of multiple training instances in the form of labeled annual rings which are treated as individual regions of interest. To demonstrate the plausibility of the approach, we performed several benchmarking tests that compare the overall performance of the proposed approaches against published methods based on deep learning as well as traditional signal processing. In addition, we also evaluate and compare the robustness of the methods as well as their capacity to extrapolate and adapt to new datasets.

2.3   Methods

Otolith images and their corresponding age readings were provided by the Thünen Institute of Sea Fisheries and the Thünen Institute of Baltic Sea Fisheries. The image collection can be divided into two sets: 1) the North Sea dataset (https://doi.org/10.5281/zenodo.8341092) and 2) the Baltic Sea dataset (https://doi.org/10.5281/zenodo.8341149). The North Sea dataset consists of images from several demersal species including North Sea cod (*Gadus morhua*), saithe (*Pollachius virens*), haddock (*Melanogrammus aeglefinus*) and whiting (*Merlangius merlangus*). To achieve higher statistical power, we only used the otolith images with ages 1-11, as this range contains enough data for both training and testing. For the Baltic Sea dataset, the otoliths are composed purely of Baltic cod (*Gadus morhua*). Likewise, we only used those with ages ranging from 1 to 5 as these age groups contain a sufficient number of images for the analyses. It is important to note that, in contrast to the North Sea set, the manual age readings from the Baltic Sea dataset are all validated using tetracycline markings ((Krumme et al., 2020)). For more details on both datasets, a table is included (Table 2.5) under the Supplementary Materials.

Table 2.1: A summary of the number of images available per species along with the sampling area and abbreviations used in this study. For species-wise experiments and analyses, both the N-haddock and N-whiting were not used as they have insufficient quantities.

Species	Area	Number of Images
Gadus morhua (N-cod)	North Sea	194
Pollachius virens (N-saithe)	North Sea	351
Melanogrammus aeglefinus (N-haddock)	North Sea	78
Merlangius merlangus (N-whiting)	North Sea	37
Gadus morhua (B-cod)	Baltic Sea	1155

No. of images

Figure 2.1: The number of images for each age group for both datasets. A total of 660 otolith images (ages 1-11) were included for the North Sea dataset while there were 1155 images in the Baltic Sea dataset (ages 1-5). For a detailed tabular summary of each age group, please refer to the Supplementary Material.

Preliminary manual checks were done on the two image datasets to ensure that no duplicates are taken and that all images are unambiguously named. Also, there were cases where some otolith images have artifacts that obscure a significant portion of the otolith. For our purposes, it is important that those were not included. Lastly, since the methods require at least one annual ring for the ground-truth preparation, images with age 0 were also excluded.

For obtaining the North Sea otolith images, it is a common practice to apply some image filters via an imaging software to make the rings more visible. Hence, for this dataset, all the images have already gone through some preprocessing for image enhancement. The Baltic images, on the other hand, were utilized in their raw states.

Data Preparation and Configuration of the Methods

For each dataset, we used randomized subsampling to create the training set and consequently segregate the test set with the remaining out-of-sample images. As shown in Figure 2.1, the raw number of images for each age group varies considerably. To avoid a prediction bias towards the age groups with more data, the subsampling was done such that there is a rebalancing of age groups after every randomized selection (i.e., given a certain quantity, excess training images on some age groups were removed while those with fewer images were refilled).

As shown in Table 2.2, the partitioning of the datasets was done for multiple experiments. Each age group contains the same number of training images with the exception of species-wise experiments. Lastly, apart from splitting the data into training and test set, there is also a need to select the validation set that determines the training checkpoints (i.e., for saving the model state in each epoch whenever there is an improvement in the loss computed). Instead of further dividing the training set to create the validation set, we opted to construct it via data augmentation involving horizontal flipping of the training images.

Table 2.2: The number of images used for each data split along with the number of runs or subsampling replicates done in each experiment. The validation data is derived entirely by data augmentation of training images via horizontal flipping operation; it hence has the same quantity as the training set.

Experiment Type	Training and Validation	Testing	Runs
Basic Evaluation	132-North Sea images	528-North Sea images	20
	150-Baltic Sea images	1005-Baltic Sea images	4
Robustness Test	132-North Sea images	528-North Sea images	20
	150-Baltic Sea images	1005-Baltic Sea images	4
Age Extrapolation	84-North Sea images	188-North Sea images	8
	120-Baltic Sea images	42-Baltic Sea images	4
Interchanging Domains	132-North Sea images	1155-Baltic Sea images	8
	150-Baltic Sea images	660-North Sea images	4
Trained with N-Cod	132-N-cod images	351-N-saithe images	8
		1155-B-cod images	8
Trained with N-Saithe	132-N-saithe images	194-N-cod images	8
		1155-B-cod images	8

The next step was to conduct ground-truth labeling which is required as part of the supervised learning process. In the next four subsections, we describe separately each algorithm involved in the study to highlight their differences and some simplifications adopted for our purposes. The first two algorithms, namely Classical Image Processing and CNN Regression represent the methods that are already existing in the literature and which serve as baseline for comparison. Then, we describe our proposed approaches based on Mask R-CNN and U-Net and elaborate the way these methods can perform age estimation totally compatible with traditional ring counting methods. To facilitate the understanding of the entire process, our source code (written in Python 3.8 (Van Rossum

and Drake, 2009) with machine learning libraries such as Keras 2.2.4 (Chollet et al., 2018) and Tensorflow 1.15 (Abadi et al., 2015) is available on Github (https://github.com/arj aycc/ai_otolith/tree/v1.2). Also, a schematic diagram outlining the main steps for the proposed deep learning approaches is given as a supplementary material (Figure 2.12).

Classical Image Processing

For the image processing approach, we chose to explore mainly the methods that uses intensity peak counting as this approach is quite popular and straightforward to use as reviewed by Fisher and Hunter (2018). Simplifying the ideas from the literature (Troadec, 1991; Formella et al., 2007), the method we finally implemented was to simply create a polar transformation of the sector slices from otolith images and convert them into square tiles using the relative distances of the pixels starting from the otolith nucleus or core down to the outer edge. A schematic diagram of the process is given in Figure 2.15 in the Supplementary Materials.

(a)   (b)
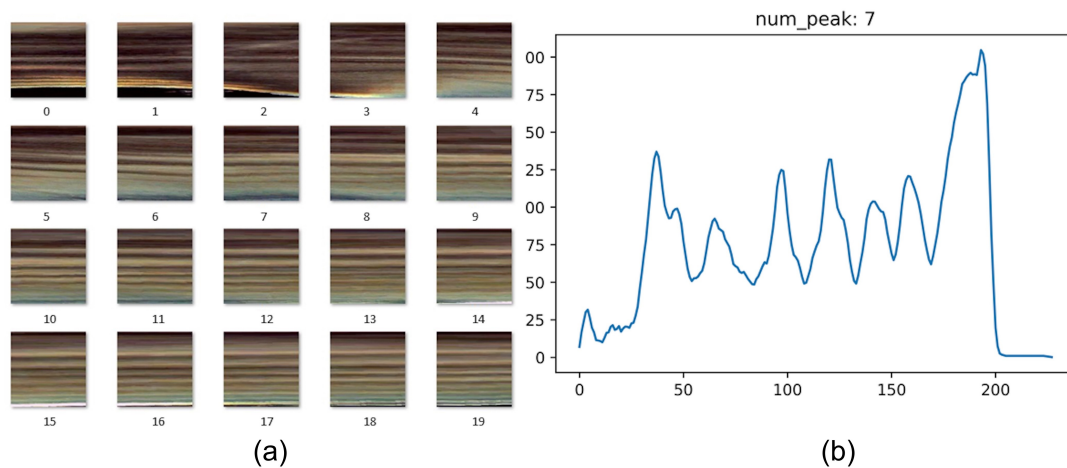
Figure 2.2: a) A set of image strips that were transformed from otolith sector slices along a reading axis. b) The resulting intensity plot when the row-wise average was taken for a single strip with the resulting peak count at the top.

As a preliminary step, we needed to first identify the outer otolith contour and the nucleus from the images. A simple application of the watershed algorithm (from the python skimage library (van der Walt et al., 2014)) isolates most otoliths from their corresponding background with great accuracy from which the outer contour can be obtained. There are few cases which appear to generate erratic contours, especially if the outer otolith edges are not clearly distinguishable. For our purposes, we simply identified and manually corrected these erratic contours by using a standard image annotation tool. We opted to use the Visual Geometry Group (VGG) Image Annotation tool abbreviated as VIA (Dutta and Zisserman, 2019) due to its simplicity and extensibility. In fact, we managed to incorporate our own code into this tool where we created a brush feature to facilitate the annotation as it is also needed for the ground-truth preparation of the other methods.

For identifying the nucleus of the otoliths, several classical image processing techniques are also widely popular (Cao and Fablet, 2006; Harbitz, 2009). We chose a simple heuristic based on ellipse approximation (Harbitz, 2009) to locate the approximate nucleus position which worked quite well for the Baltic Sea dataset. However, for North Sea images, some nucleus coordinates were missed so we had to do manual adjustments using the same annotation tool so as not to introduce another source of error and to focus only on the steps involving annual rings.

Overall, the entire process relies on the assumption that there is a proportionality among the growth of the rings on a certain local portion of the otolith (Fablet and Josse, 2005). Hence, it is expected that when the otolith sectors are sliced and divided into small enough pieces, the transformed rings will be approximately aligned (Figure 2.2-a). With these transformed images, it is straightforward to generate a good intensity signal plot by taking either the mean or median of pixel rows from top to bottom across multiple slices along the major axes (Figure 2.2-b). To derive the age reading, we performed a peak counting procedure using a peak detection algorithm based on a standard implementation available from the literature (Billauer, 2009).

CNN Regression

Convolutional Neural Networks (CNNs) are one of the most widely used algorithms to deal with image datasets (Krizhevsky et al., 2012). The core idea is roughly inspired by the biological neural network where the concept of neurons is represented using mathematical interconnected nodes (O'Shea and Nash, 2015). The information propagation is made through a process of weight updates along these interconnected nodes using intricate mathematical operations with the goal of making the predictions be as close as possible to the actual or expected value through the evaluation of one or more loss functions during each training epoch. These nodes are typically grouped into layers and each node can have multiple connections into other nodes located at the next layer. What primarily differentiates CNNs from traditional artificial neural networks is the number of layers; for the former, it is several orders of magnitude higher (i.e., the layers go deeper) than for the latter.

The most basic use of a CNN is in a supervised manner which could be formulated as either classification or regression (Martinsen et al., 2022; Moen et al., 2018; Ordoñez et al., 2020; Politikos et al., 2021). That is, a discrete or continuous value will be returned as prediction which directly corresponds to the probable category or measurement that it learned from the labeled training data. In the case of regression, a basic loss function for these types of CNNs is usually in the form of mean squared error (MSE) (Martinsen et al., 2022; Moen et al., 2018) which is given in the following equation:

$$Loss = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

For the CNN regression method used by Moen et al. (2018), they chose to use regression, where the age estimates are turned into a continuous value. Also, they used another

useful concept of CNN known as transfer learning, where a pre-trained model, primarily InceptionV3 (Szegedy et al., 2015), was reused by preloading its weight into the network prior to training.

To use this approach, mainly as a benchmark reference, we obtained the exact implementation from Moen et al. (2018) available from https://doi.org/10.21335/NMDC-1949 633559. It only involves a simple data loading step where a list of image paths and their corresponding age labels are placed in a comma-separated file. This file will then be taken by the algorithm to start the supervised training. One minor issue, however, is their use of otolith pairs (left and right otoliths) which is not applicable in our study. Therefore, as a simple workaround, we flipped each otolith image horizontally to somehow have a pseudo-pairing and make the implementation compatible.

Mask R-CNN

Likewise, as implied by its name, Mask R-CNN is also a deep learning algorithm based on Convolutional Neural Networks (He et al., 2017). The main output, however, is primarily in the form of detection masks and bounding box coordinates of the object of interest as found within the image. That is, detection masks are pixel markings that indicate the spots occupied by the object of interest while the bounding box consists of numeric coordinates within the image that contains this object of interest. This feature of the Mask R-CNN algorithm allows it to perform both object detection and instance segmentation simultaneously

For this algorithm, there is an implementation from Matterport (Abdulla, 2017) containing the entire learning workflow starting from the data loading step up to the training as well as the testing. To utilize the code, we first need to provide its needed inputs, namely the image and its ground-truth annotations. As discussed above, we selected the VIA tool (Dutta and Zisserman, 2019) for annotating the images due to its simplicity. We marked the parts of the images along the left and right major axes which represent a portion of the winter annuli to be treated as the objects of interest for detection. Figure 2.3 shows an example of the annotation using the VIA tool.

As mentioned above, there can be different ways to implement loss functions for each algorithm. For Mask R-CNN, instead of the basic MSE, it needs to have multiple loss functions in order to check how far the predicted masks are from the actual regions while also computing the errors for the predicted bounding boxes (He et al., 2017). In the study by Zimmermann and Siems (2019), they further added another loss function related to the edges of the contours generated from the predicted masks. We used this version since it was demonstrated to learn faster and more efficiently (Zimmermann and Siems, 2019). Also, for this implementation, transfer learning was involved where an existing model (Matterport, 2017) trained from the COCO dataset (Lin et al., 2014) was pre-loaded instead of training from scratch.

The output of Mask R-CNN still needs to undergo a post-processing step in order to derive the age estimates. A schematic diagram was included in the Supplementary Materials (Figure 2.16) that summarizes the process. The core idea is to scan the masks and to find their alignment towards the center which indicates that they belong in the

same reading axis. The process starts by locating the nucleus and measuring the distances of the masks to this reference point. Then, each mask is visited from the nearest to farthest to label their positions. To perform labeling, the angle (in radians with respect to the nucleus) of a mask is measured through its endpoints. Two masks are aligned if their angles overlap. To label a mask, increment by 1 the label of the most recently visited mask that aligns to it. If there is none, then label it as 1. Once all the masks are visited, sort the labels then find the highest value which will indicate the highest ring count (corresponding to the age reading).
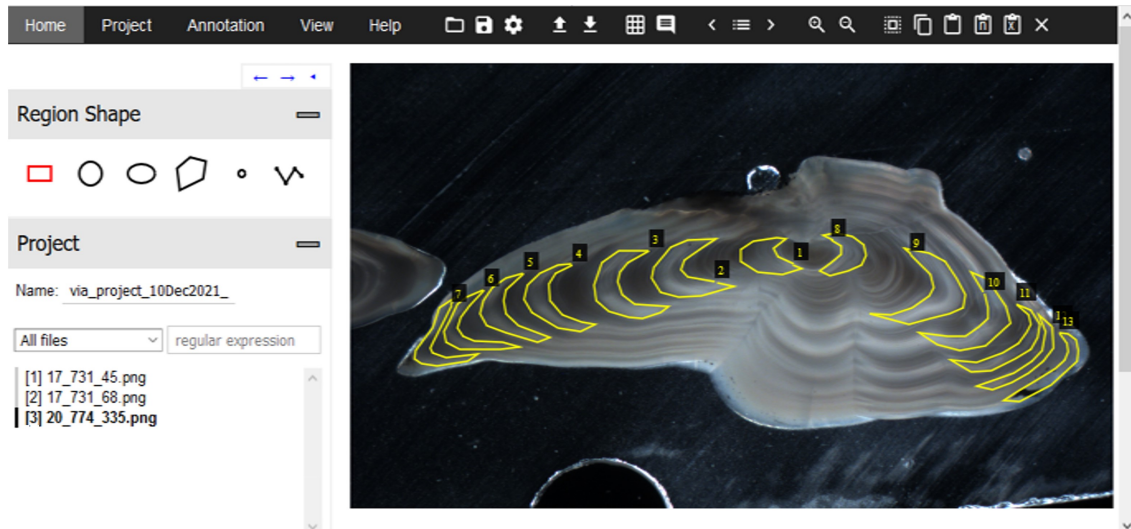
Figure 2.3: An example of ground-truth preparation made using the VIA annotation tool showing the annotations (yellow) that mark the regions of interest within an otolith image.

U-Net

U-Net also makes use of the CNN architecture (Ronneberger et al., 2015) similar to the two previously presented deep learning algorithms. The main difference, however, is that for U-Net, the final output is composed only of a segmentation mask for the entire image corresponding to the pixels detected representing the object of interest. Because of this, U-Net is usually utilized for problems involving semantic segmentation of images.

To train the algorithm, ground-truth masks are likewise needed to mark the regions to be segmented by the U-Net. In this study, two ways of ground-truth labeling were followed. One method involved masking the entire concentric annuli and the other involved masking only a certain portion of the annuli along the reading axes. For the former, new sets of ground-truthing has to be created using the VIA annotation tool. For the latter, we reused the same reading axes annotations made previously for Mask R-CNN.

In contrast to Mask R-CNN, only the segmentation masks are returned by U-Net and no bounding boxes are generated. Hence, there is usually only one loss function involved which determines whether each pixel of the image was properly marked either 1 or 0 depending on whether they are part of the object of interest or not, respectively. The basic loss function can be the MSE of these per-pixel differences but it can be modified

as needed. In fact, for this study, we used the extension proposed by Ronneberger et al. (2015) where more weights are given on the pixels in-between the objects of interest. That is, the algorithm has to be more careful when marking those pixels between each annulus because the errors from these portions weigh more than the rest. Otherwise, without weighted loss, the U-Net has the tendency to return overlapping contours, especially when the rings are near each other, which is particularly happening at the edges.

Similar to Mask R-CNN, the segmentations cannot be readily isolated from the rest of the pixels. Hence, it is therefore also necessary to perform a post-processing step in order to remove noise and easily count the proper segmentations where age estimates are derived. This process is summarized in Figure 2.16 of the Supplementary Materials. As there are no detection scores like the ones from Mask R-CNN, we applied a simple noise filter based on the size and relative position of the segmentation. That is, if a segmentation is found, it is first checked whether it is just a random noise before including it in the ring count. This is similar to the criteria also applied in peak detection methods where certain peaks are eliminated according to their relative sizes and positions.

In this study, we explored different configurations for this algorithm in order to identify the best performing variant. First, two different ground-truth methods were tested: one annotation set marks only the portion along the major axes while the other annotation covers as many annuli as visible in the image. Secondly, we also compared the performance of implementing U-Net with transfer learning using pre-trained VGG weights (Simonyan and Zisserman, 2015), similar to the implementation of Abdellatif (2021), against the default implementation, which is trained from scratch. At this point, it is worth mentioning that for all the pre-trained models used in each deep learning method explored in this study, the training set from which they were originally trained on are all composed of images from common objects and not specifically for otolith.

Benchmarking

There are three basic benchmark tests that we conducted in order to thoroughly assess and compare the overall performance of the algorithms which we measured in terms of percentage agreement. For the first test, we performed the usual training, validation and testing, using images from the same data source. This test also involved identifying initially the best hyperparameters and configurations of each algorithm that will be used for subsequent experiments. For the CNN-regression method, we used the default or suggested hyper-parameters taken from the study of Moen et al. (2018). For the other algorithms, we implemented a simplified grid-search on the different configurations and hyperparameters and evaluated their performance on a subset of the test data.

For the second test, we performed some variations of the first test to evaluate two criteria: 1) the robustness of the algorithms when slight changes/perturbations on the images are introduced and 2) the ability of the algorithms to extrapolate (higher) age groups when they are explicitly removed from the training data. The former involved simple background removal with increased brightness on the test images to see whether the algorithms have taken cues on unreliable features such as background artifacts or even the differences in lighting. The latter involved complete removal of any training data from

higher age groups (ages 8-11 for North Sea dataset, age 5 for Baltic Sea dataset) to see if the algorithms can extrapolate these higher age ranges without encountering them during training.

Lastly, for the third batch of tests, we checked for inter-dataset and inter-species performance to assess how adaptable the models are when analyzing new sets of data from a completely unfamiliar species or domain. For the basic case, we interchanged the test sets for North Sea and Baltic Sea otoliths and assessed the new performance (i.e., the models trained from North Sea images were tested against Baltic Sea test images and vice versa). For the other case, we segregated the images further into different species to see whether training them on a specific species makes the algorithms completely unable to generalize on the other species. Conversely, we also aimed at finding out whether training the algorithms on a given species allows them to handle the same species from a completely different source. For simplicity, in this experiment, we use the term inter-species loosely despite also treating the North Sea cod and Baltic Sea cod as separate groups.

Coefficient of Variation Analysis

In the context of age reading evaluation, apart from percentage agreement, another important metric is the so-called coefficient of variation (CV) which is especially useful during age reading workshops where readers from various institutions gather to cross-check the possible differences in the way they perform age readings. This value can be computed using the following formula (Campana, 2001):

$$CV = \frac{\sigma}{\mu} \cdot 100$$

where σ = standard deviation, μ = mean of age estimates from the readers

For reference, we used two separate ICES workshops- one for North Sea cod (ICES, 2008) and another for Baltic Sea cod (ICES, 2020), where participating readers performed age estimation on cod images using their own methodologies. It was reported that for both the North and Baltic Sea workshops, the readers had a significant disagreement indicated by the computed CV of about 40% (39.8% to be precise) and 15%, respectively. Optionally, for the North Sea workshop, we may exclude the values contributed by broken otoliths and refer only to the result for sectioned otoliths which is around 22.5%. Hence, for this study, similar to the formula used by Moen et al. (2018), we also computed the CV by treating the automated and manual readings as individual readers and assess whether the age estimate variations fall within the same range attained by human readers.

Statistical Analysis

To check for the statistical significance of the comparisons, we used the standard pairwise t-test available in the R programming language (R Core Team, 2023) along with the correction proposed by Nadeau and Bengio (2003) which is implemented in the correctR package (Henderson, 2023). We carefully considered the fact that some assumptions

of the standard *t*-test are violated by the data partitioning used to create the training and test splits. As mentioned earlier, we employed a small variation of the randomized subsampling for creating the training and test sets which means that the images used for each run are not completely independent (i.e., the training and test sets of one run could have images that were also included in the other runs). This leads to a high probability of type I error causing a problematic rejection of the null hypothesis in pairwise comparison of algorithms (Dietterich, 1998). Fortunately, the ground-breaking study made by Nadeau and Bengio (2003) suggests that a simple correction of the standard *t*-test can overcome this limitation. Therefore, for the main test involving general performance comparisons, this corrected resampled *t*-test is used as it satisfies the conditions needed for the statistical analysis. For the other test cases which deviate greatly from standard randomized subsampling (e.g., age-wise and species-wise test), we used the standard *t*-test while taking into account the potential pitfalls mentioned.

2.4   Results

One straightforward advantage of the CNN-regression algorithm used by Moen et al. (2018) is that the age readings are readily available and directly outputted in the model predictions. For all the other methods, however, an intermediate output has to be generated first before the actual age reading can be derived.

For the classical image processing approach, the intermediate results are in the form of signals that indicate the image intensity values from the nucleus to the outer edge of the otoliths as shown in Figure 2.2-b.

For the Mask R-CNN, the final detections need to be post-processed first as described in the Methods Section in order to directly appear on the image as shown in Figure 2.4. Apart from the colored masks, it can be seen that there are bounding boxes that are also depicted containing the prediction scores. These values range from 0.0 to 1.0 and directly correlate with the model's confidence on the predictions.

For the U-Net algorithm, the intermediate result also needs to undergo post-processing before the age estimates can be derived. Figure 2.5 shows an example of a raw mask output of the U-Net as well as the resulting image masks after the post-processing procedure similar to the one performed for the Mask R-CNN output.
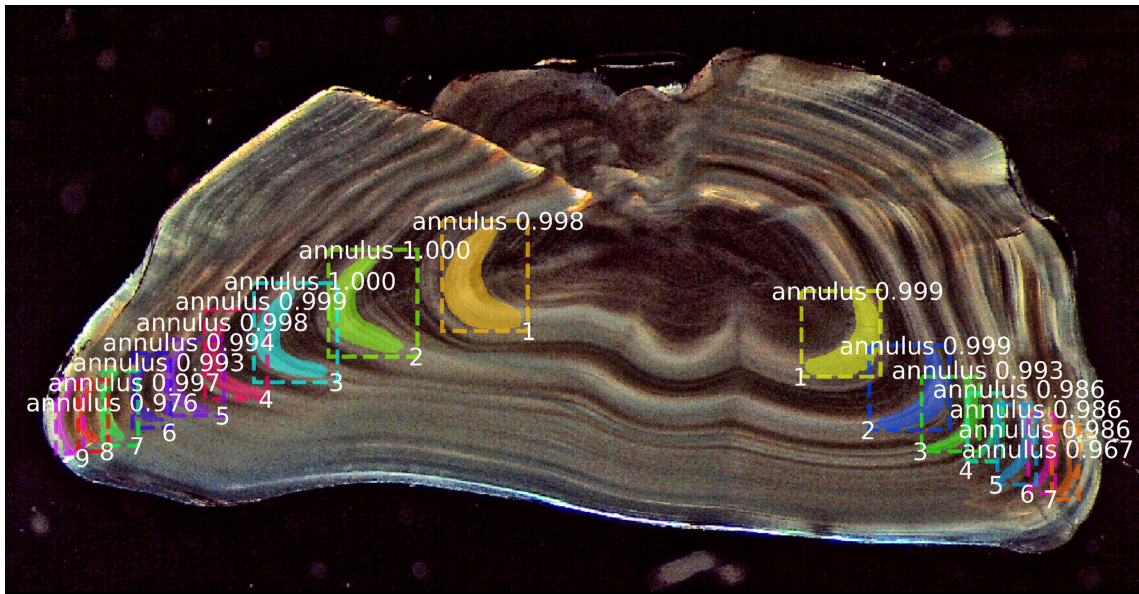
Figure 2.4: Mask R-CNN object detections with the corresponding bounding boxes and scores. Higher scores indicate higher model confidence which can be used to filter out those predictions which do not surpass a certain detection threshold (i.e., a hyperparameter that can be adjusted as needed).
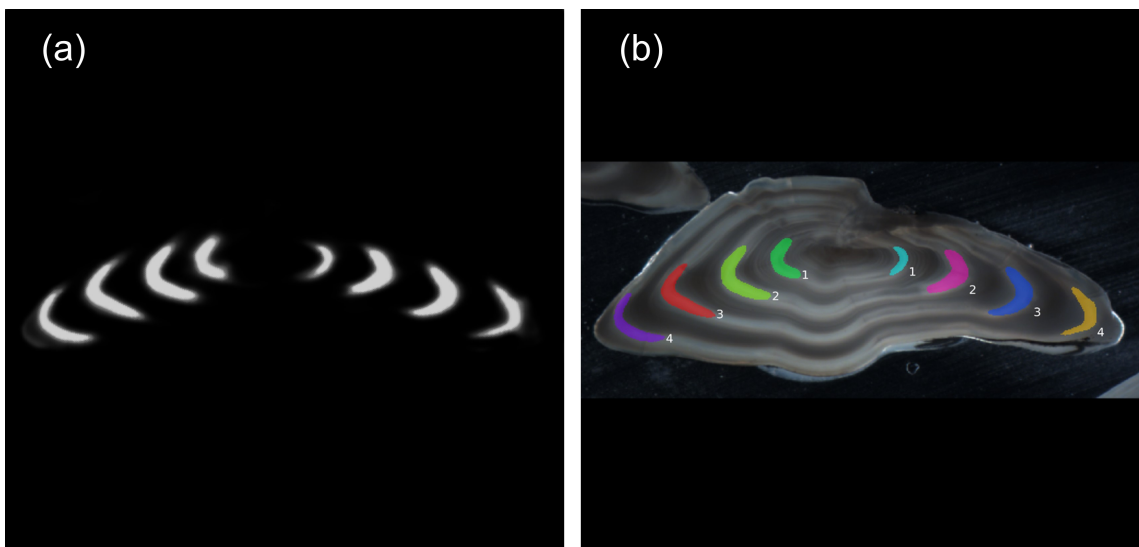
Figure 2.5: The raw U-Net output alongside a sample end-result after the post-processing step. The direct output of a U-Net model is a mask indicating the regions it segmented (a) that can be post-processed to generate the ring count (b).

After the post-processing stage for each algorithm, the derived age estimates are then plotted against the manual age readings, as shown in Figure 2.6. It can be seen that there is a diagonal trend that becomes apparent with these plots indicating the relative agreement between the automated and the manual readings. The plot also shows how far

the under- and over-estimates are from the diagonal indicating the biases of each method. For illustration, only the test results of a single run with North Sea images are shown in the figure. For the plots of all the runs including those of the Baltic Sea images, refer to the Supplementary Material.
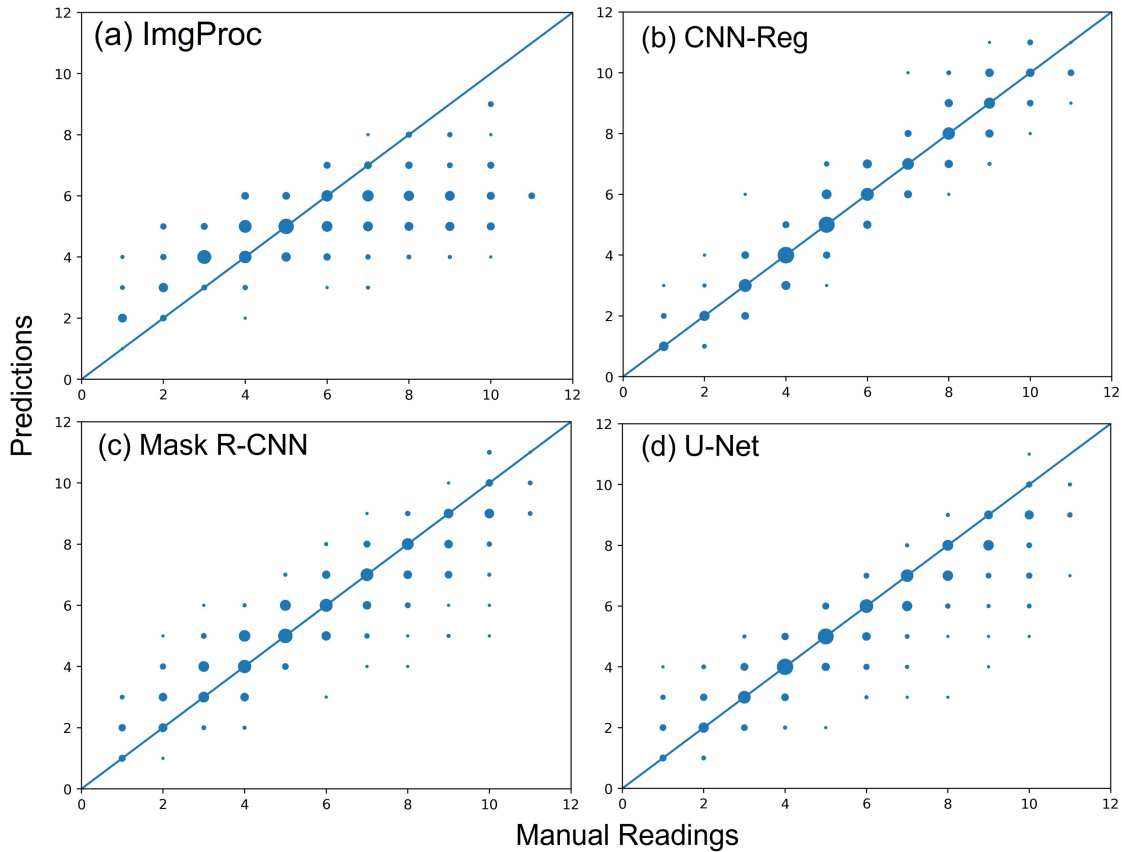
Figure 2.6: The plots of automated age estimates against the manual age readings on a test set involving North Sea images using the various approaches, namely a) Image Processing, b) CNN-Regression (rounded off), c) Mask R-CNN and d) U-Net.

Figure 2.7 provides a clearer comparison of the performance of the different algorithms tested. The resulting trend is different for the North Sea dataset and the Baltic Sea dataset. The CNN-regression has a clear edge with 55% and 87% mean accuracy for North Sea and Baltic Sea images, respectively. The Mask R-CNN has a slightly poorer performance on North Sea images (46%) but it has a decent mean accuracy on Baltic Sea images (72%). On the other hand, the U-Net algorithm manages to be competitive with 54% mean accuracy on the North Sea dataset and 72% mean accuracy for the Baltic Sea dataset. Lastly, the traditional automation method using classical image processing attains the poorest performance, showing only 26% and 54% mean accuracy for North Sea and Baltic Sea datasets, respectively. Hence, this approach was no longer used for further analysis to focus more on the deep learning algorithms.

Table 2.3: The coefficient of variation (CV) of the different methods against the manual readings. The reference value is 40% for the North Sea dataset and 15% for the Baltic Sea dataset, which correspond to the CVs from a group of readers during two ICES workshops on cod otoliths.

Method	North Sea Dataset	Baltic Sea Dataset
ImgProc	19.1%	16.4%
CNN-Reg	7.4%	3.8%
M-RCNN	10.9%	10.1%
U-Net	10.5%	9.6%
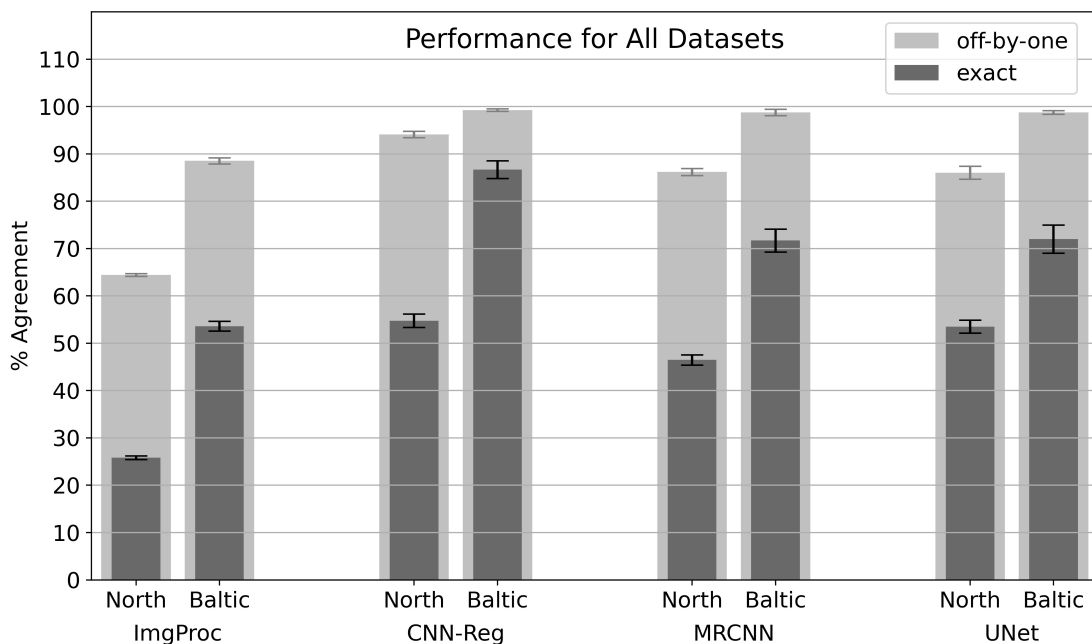
Figure 2.7: Overall performance of the different algorithms on North Sea and Baltic Sea datasets across multiple runs with randomly subsampled test sets (n=20 for North Sea Dataset, n=4 for Baltic Sea Dataset). Applying the corrected resampled t-test to compare each proposed deep-learning method (M-RCNN and U-Net) to the published CNN-regression method yields corresponding p-values = 0.14 and 0.43 (> 0.05) for North Sea images and p-values = 0.003 and 0.048 (< 0.05) for Baltic Sea images.

Using the corrected resampled t-test, the null hypothesis that CNN-regression results do not differ from the results of both the proposed methods has failed to be rejected in the North Sea dataset (p-values > 0.05), while it was rejected for the Baltic Sea dataset (p-values < 0.05). This indicates that the proposed methods have a similar performance to the CNN-regression on the North Sea images but fail to attain the same competence on the Baltic Sea images where the CNN-regression shows its clear advantage.

To assess if an automated method is good enough to be treated like an individual human reader, we also computed the CV for each method as shown in Table 2.3. With a reference

value of 40% and 15% taken from the North Sea and Baltic Sea workshops, respectively, it can be seen that the computed CVs for the deep learning methods fall significantly below these thresholds indicating that they are indeed already at the level of human readers. It is important to note that for the North Sea workshop, we may only consider the result for sectioned otoliths and ignore the values for broken otoliths which is not relevant in this study. Hence, even if the reference value is adjusted to 22.5%, the same conclusion is still valid. That is, the CV results from this study still fall below the workshop reference values. This means that theoretically, if the AI-based methods are included in a workshop with human readers, the readings they provide will deviate within the same range as the ones from the human readers.

The next set of experiments evaluates the robustness of the different methods when the test images are subjected to slight variations (i.e., involving background removal and increased brightness). Figure 2.8 reveals one surprising disadvantage of the published CNN-regression method compared to the proposed methods. Just with the mentioned image perturbations, a very drastic change in performance is seen for the CNN-regression in both North and Baltic datasets. Only slight degradation of performance is observed for Mask R-CNN and U-Net.
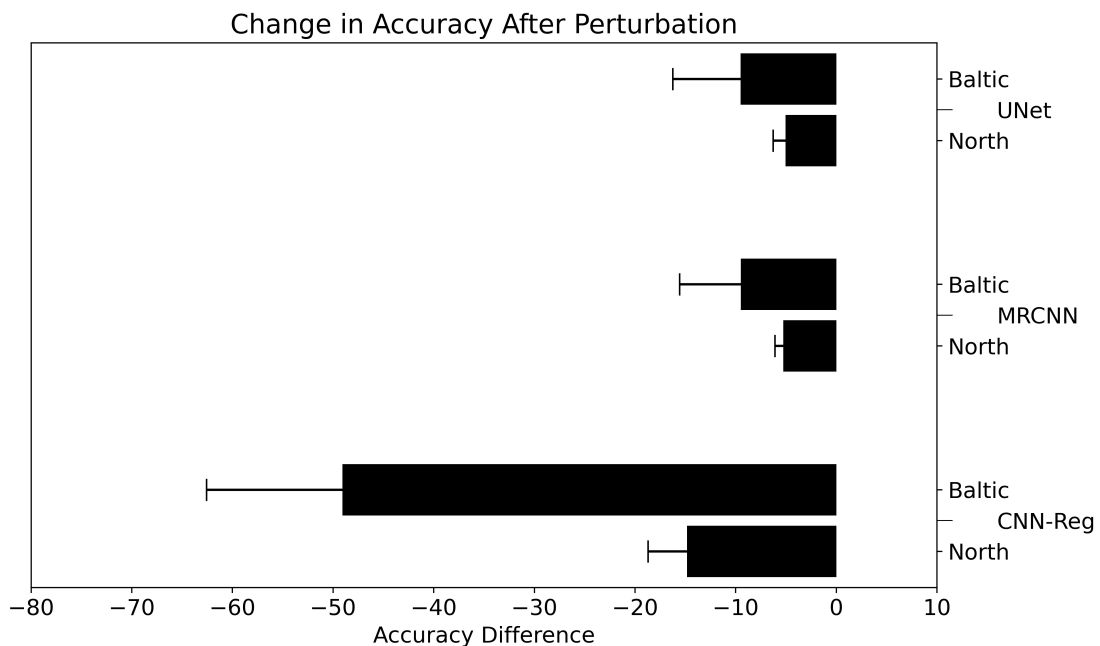
Figure 2.8: Degradation of predictive performance of each algorithm when the background of the otoliths on the test images is removed while subsequently increasing the image brightness. Comparing the changes in accuracy of the proposed methods against that of the CNN-regression yields p-values < 0.05 using standard *t*-test (n=20 for North Sea, n=4 for Baltic Sea).

Another interesting experimental setup was designed to measure the ability of the methods to extrapolate on data they have not encountered before. In this experiment,

we removed the training images with high age values and limited the range to ages 1-7 for North Sea dataset and ages 1-4 for Baltic Sea dataset. Then, we tested the resulting models on a test set containing only images with age values greater than those used during training. Figure 2.9 summarizes the result and demonstrates the extrapolation ability of the different methods.
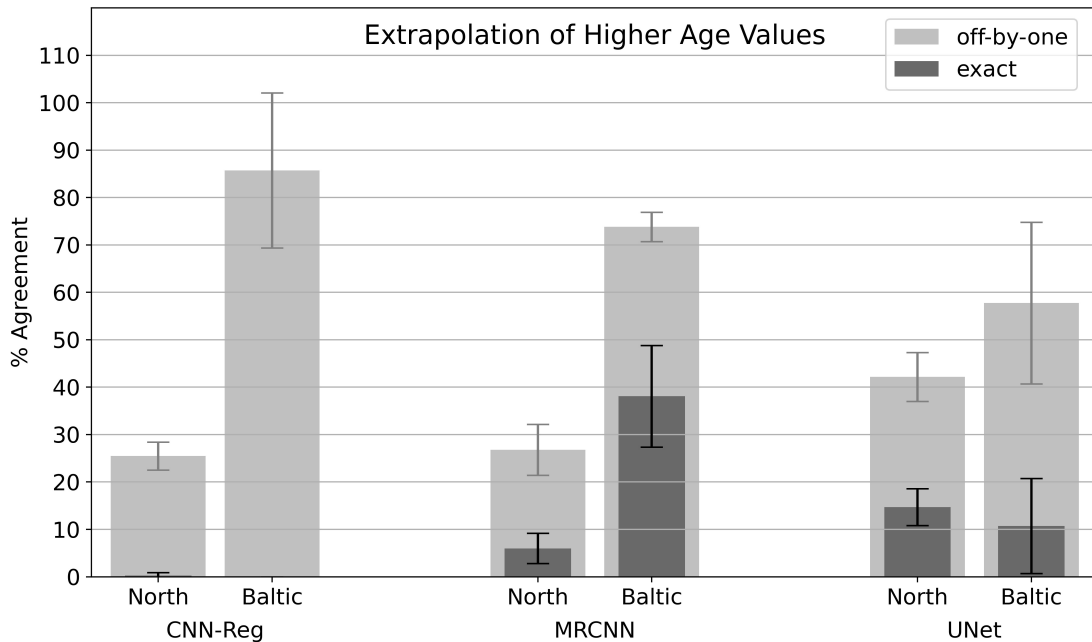
Figure 2.9: Performance of each deep learning algorithm on higher age groups that were excluded during training. For the North Sea runs (n=8), images with ages 8-11 were used for testing as they were excluded from training. For the Baltic Sea runs (n=4), only age 5 images were left out during training and were consequently used for testing. The standard *t*-test gives p-values < 0.05 for the pair-wise comparison against CNN-regression.

It can be immediately seen that the published CNN-regression fails almost completely in getting any correct estimate for higher age groups that were not included during training. In contrast, both the proposed algorithms manage to attain a decent accuracy level, showing their ability to extrapolate on unknown data.

For the last test, we further highlighted this capacity of each algorithm to handle datasets that were not introduced during training. For the first case, we interchanged the test images of both datasets and re-tested the previously trained models without re-training on the new set. That is, the existing models trained from the North Sea dataset were tested on the Baltic Sea test images and vice versa. Figure 2.10 demonstrates yet another advantage of our proposed algorithms compared to the published CNN-regression method.
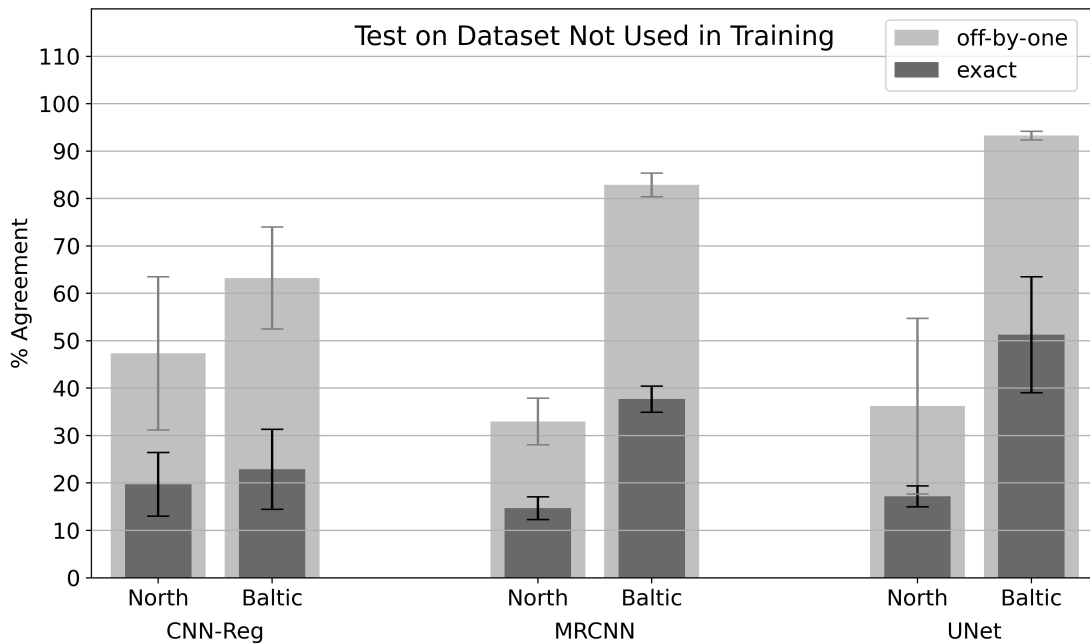
Figure 2.10: Performance of the deep learning models trained on one dataset and tested against the other dataset and vice versa. For the Baltic Sea test case (n=20), standard *t*-test show significant difference (p < 0.05) when comparing the CNN-regression against the proposed methods.

Overall, it can be seen that the CNN-regression algorithm attains the worst performance when given a new and unfamiliar data source or domain. This means that it learned features too specific on the dataset it was trained on resulting to its failure to generalize on the other dataset with seemingly new otolith characteristics, different microscopy lighting and image capture technique. In fact, this concept, referred to as domain adaptation, has also been explored in the study by Ordoñez et al. (2022), where they also evaluated this capacity on a similar standard CNN implementation but with classification instead of regression. They used images of the same species (Greenland halibut) from two different sources: one dataset came from the Norwegian laboratory while the other dataset was taken from their counterpart in Iceland. Similar to what we have observed, they also reported that this standard CNN formulation performed poorly when tested across the two different data sources. Hence, they proposed certain modifications to the default implementation, but this is beyond the scope of our study.

To elaborate on this observation further, we conducted another test focusing mainly on inter-species performance. For this setup, we explicitly trained the algorithms using only one specific species and performed tests on the other species. Figure 2.11-a shows the comparison of test performance across species when the training involves only North Sea cod images while Figure 2.11-b shows the results if only North Sea saithe images were included.
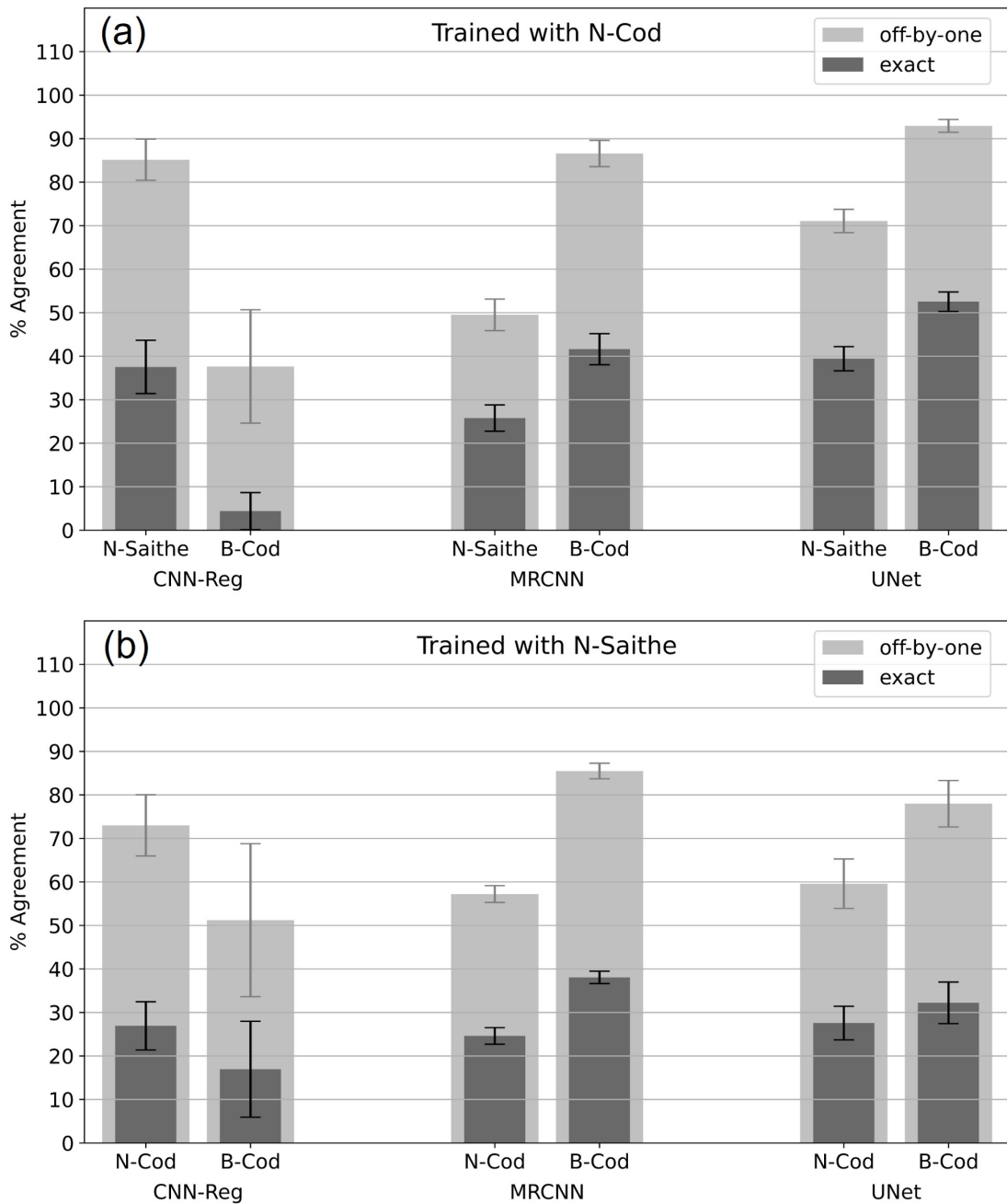
Figure 2.11: Performance of the methods across species (and stock) when the training involves a) only North Sea cod images. b) only North Sea saithe images. The standard *t*-test (n=8) shows high significance (p < 0.05) on the Baltic cod test case for both Mask R-CNN and U-Net after pairwise comparison against the previously published method.

There are some interesting observations worth emphasizing for this batch of results. First, it can be immediately seen from both plots that the overall inter-species accuracy of the proposed methods surpasses that of the previously published CNN-regression method indicating that the proposed methods have more generalization capacity. Specifically,

the performance discrepancy is quite large when it comes to the Baltic test images. This is somehow surprising when comparing to the result from the previous experiment. It seems that purely using North Sea cod images for training makes the performance of the published CNN-regression method to become even worse compared to using a mixed set (Figure 2.9) or even pure North Sea saithe images (Figure 2.11-b). This result is directly in contrast to the results from the two proposed methods where the accuracy values for predicting a new set of images coming from a different source (e.g. Baltic dataset of purely cod) becomes higher when the training set involved the same species (i.e. North Sea Cod in Figure 2.11-a) compared to a completely difference species (i.e. North Sea saithe in Figure 2.11-b). This implies that there could be species-specific patterns utilized by the proposed algorithms to help in the prediction of a new set of the same species.

In summary, from Figures 2.9, 2.10 and 2.11, it can be concluded that the CNN-regression method exhibited the least adaptability when it is subjected to a completely unfamiliar dataset. This means that to use this algorithm for each new species or even just a new age group, a new batch of training has to be performed to update the model or in the worst case, a complete retraining has to be conducted to create a totally different model. In contrast, for the two new algorithms proposed, the previous knowledge they had on one species can potentially still be usable for another species.

2.5   Discussion

Various studies have already shown that the standard CNN classification or regression performs satisfactorily when it comes to age estimation of various fish species (Martinsen et al., 2022; Moen et al., 2018; Politikos et al., 2021). Apart from the predictive power, another big advantage of their approach is the training simplicity where minimal ground truth preparation is needed. However, to be widely accepted, this formulation has one big issue and that is, its black-box nature. The follow-up study done by Ordoñez et al. (2022) tried to find a way to explain the decisions for this type of CNN but it still leads to more questions and counter-intuitive observations.

In the work presented here, we have shown that the use of object detection and segmentation algorithms can be a good alternative formulation when it comes to automating the fish age reading process. In addition to having a comparable performance on multiple test sets, we demonstrated that it also has several advantages compared to multiple methods that can be found in the literature. In particular, we showed that the resulting models are more robust even when some perturbations are introduced into the images. Also, we demonstrated its ability to extrapolate and generalize on datasets which were not introduced during the training phase especially those coming from a completely different source. Lastly and maybe most importantly, this new way of applying deep learning on automated age reading makes the overall process more explainable due to its direct compatibility with traditional manual methods.

One major drawback is the seemingly tedious process of doing data preparations especially the ground-truth labeling. While this may be true, it is important to note that this will only be the case if we need to train a new model with each new dataset that

we obtain. However, as demonstrated by the results, there is a potential for the object detection and segmentation models to be reusable with a completely new dataset. This means that the ground-truth preparation will eventually become less and less required as retraining becomes unnecessary in some instances. In contrast, the standard CNN regression formulation will always need to be trained with each new dataset due to its lack of adaptability.

It is important to note, however, that all these observations involving the CNN regression formulation is only tested using the implementation from the study conducted by Moen et al. (2018). It is possible that with newer designs and architecture, these limitations may no longer be true. Also, there are already novel approaches that exist in literature which seem promising when it comes to handling the known limitations of older deep learning designs such as the use of Transformers (Sigurdardóttir et al., 2023) and Ensemble Learning (Moen et al., 2023). It will indeed be interesting to conduct further benchmarking with these new approaches to see if the advantages of our proposed methods remain valid. Also, it is worth mentioning that the statistical tests performed in this study namely standard *t*-test and corrected resampled *t*-test, have limitations with respect to reducing type I and type II statistical errors (Bouckaert and Frank, 2004; Nadeau and Bengio, 2003) so more repetitions are needed to make stronger claims. It is hence an option to explore other statistical methods apart from a *t*-test which will ensure that both the type I and type II errors are minimized during benchmarking.

Lastly, one important concept of CNN which is widely used in this study is the concept of transfer learning. For all the deep learning approaches we tested, we took advantage of this facility and preloaded some pre-trained models. Therefore, there is an apparent future direction where the process of reusing a newly trained model can be improved further and training can be done using a base pre-trained otolith model (instead of VGG16 or InceptionV3). Also, for U-Net and Mask R-CNN, this base model can possibly aid on generating new ground-truth labels for future datasets and then enables a self-sustaining loop where each updated model will be reused to generate annotations for newer datasets and so on. In this way, creation of annotations will be AI-assisted and not entirely done from scratch needing only a simple manual correction if necessary.

2.6   Conclusion and Future Outlook

With the growing size of the otolith image datasets that are being collected and processed by various institutions, it is becoming apparent that the advances in the field of big data analytics, computer vision and machine learning can be of great use. This study is another step towards scalable otolith analysis and it successfully demonstrated how one can utilize the well-known techniques in object detection and segmentation to automatically perform age reading on otolith images.

As the age estimates of AI-based methods match closer and closer to those from manual age readings, it becomes clearer that the predictive performance is not the only criterion towards their general acceptance. Features such as robustness, adaptability, and in particular explainability are also important considerations which were all exhibited by the

proposed approaches in this study.

With an automated system for age estimation, the process of analyzing a large number of images can be highly efficient, scalable and less susceptible to logistic and subjective limitations. Using the proposed algorithms, we aim to create a framework or a system (i.e., a web application) that can be used as a platform for high-speed processing of large datasets. As a general toolkit for otolith image analysis, it can be made to provide not only age information but also other relevant measurements such as otolith radius and annulus distances, which are useful parameters for certain biological and ecological models. Lastly, we also hope that this future framework can be an avenue for a more collaborative effort within the community where models, images and even annotation data can be shared efficiently and even allow continuous enhancements of existing models and techniques.

2.7   Acknowledgement

We thank the members of the Otolith Age Reading Group at Thünen Institute of Sea Fisheries for the detailed discussions and walkthrough of the process of manual age reading. We thank Friederike Beußel and Hendrik Brückner for the collection of North Sea otolith images. Likewise, we thank Dr. Uwe Krumme from the Thünen Institute of Baltic Sea Fisheries for providing the Baltic Sea dataset. Lastly, we thank Marianne Camoying for the helpful insights on manuscript writing.

2.8   Supplementary Materials

Table 2.4: The number of images for each age group in the North Sea dataset as well as the Baltic Sea Dataset.

Age Group	North Sea	Baltic Sea
Age 1	26	226
Age 2	37	277
Age 3	56	343
Age 4	63	267
Age 5	74	42
Age 6	58	-
Age 7	53	-
Age 8	57	-
Age 9	47	-
Age 10	41	-
Age 11	16	-
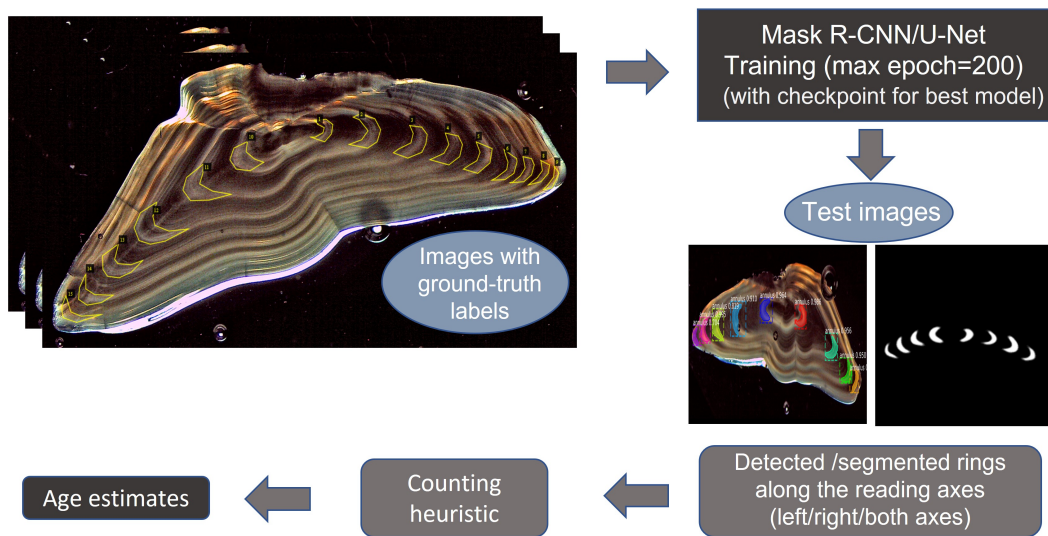
Figure 2.12: Schematic diagram of the entire process involved for the two proposed deep learning methods. Despite their differences, the same ground-truth labels can be used for both algorithms depending on the configuration being tested.
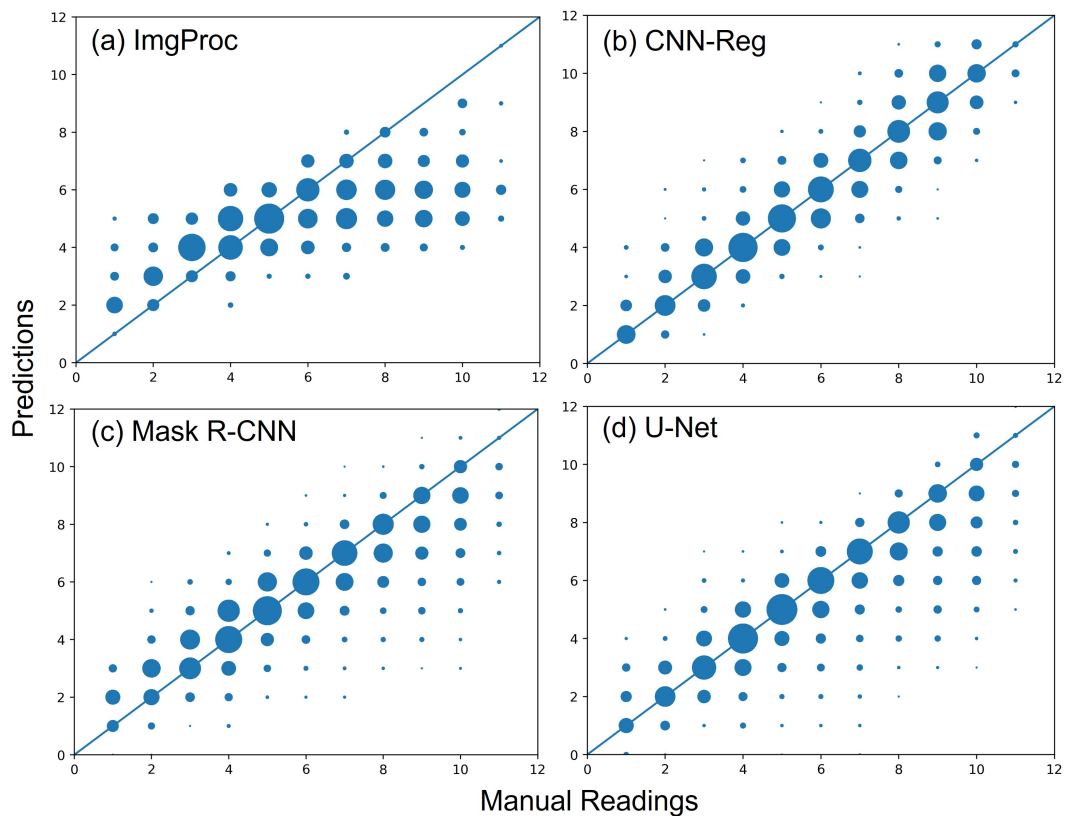
Figure 2.13: Plot of the automated age estimates against the manual readings for all runs (n=20) using the North Sea dataset.
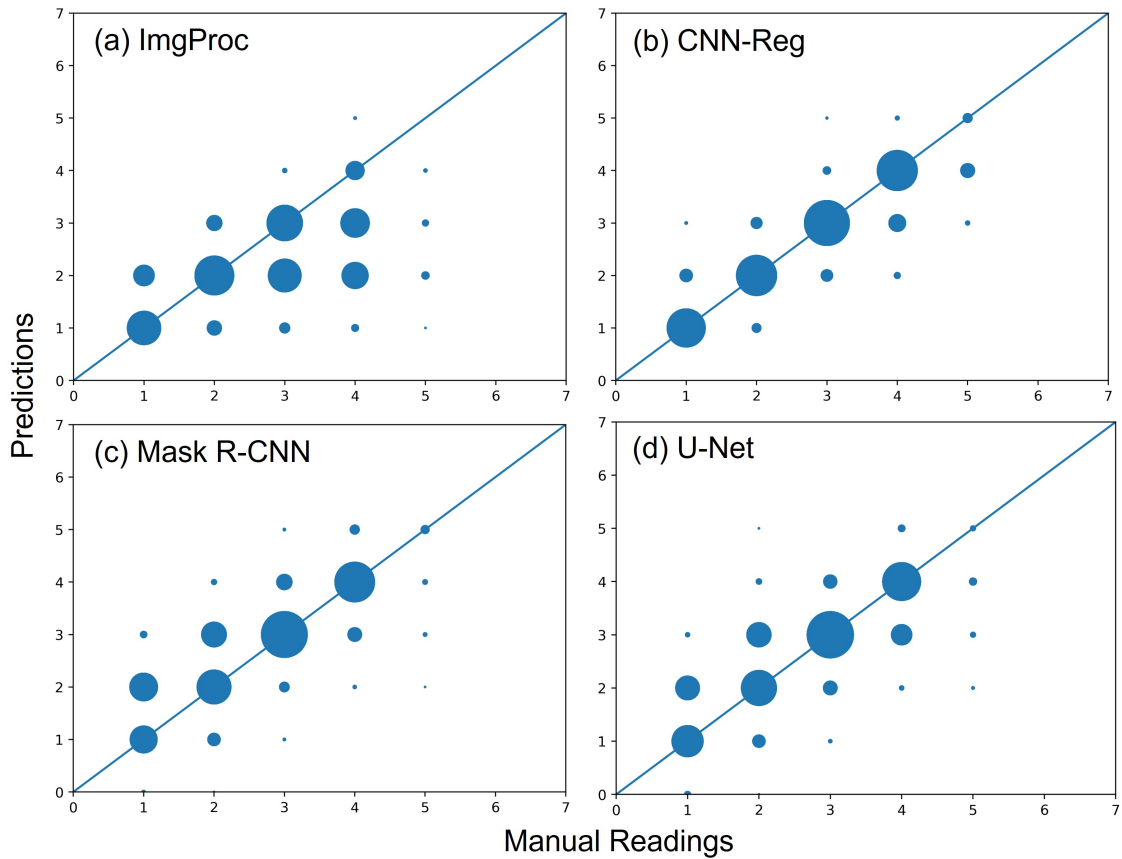
Figure 2.14: Plot of the automated age estimates against the manual readings for all runs (n=4) using the Baltic Sea dataset.

Table 2.5: Details regarding the images contained in both North Sea and Baltic Sea datasets. It is important to note that despite the name, the North Sea dataset contains few images of otoliths sampled outside the North Sea. They are, however, prepared in the same manner as all other North Sea otolith images justifying their inclusion in the set.

Imaging Information	North Sea	Baltic Sea
Otolith Preparation	Sectioned	Sectioned
Lighting	Transmitted	Transmitted
Magnification	Varying/Scaled-to-Fit	Varying/Scaled-to-Fit
Microscope and Camera	Leica system	Olympus + Zeiss
File Formats	TIFF -> PNG	CZI -> PNG
Sampling Dates	mostly 2016 to 2021	2016 to 2020
Sampling Locations	mostly 4a, 4b, 2a, 2b, 3a	3.c.22
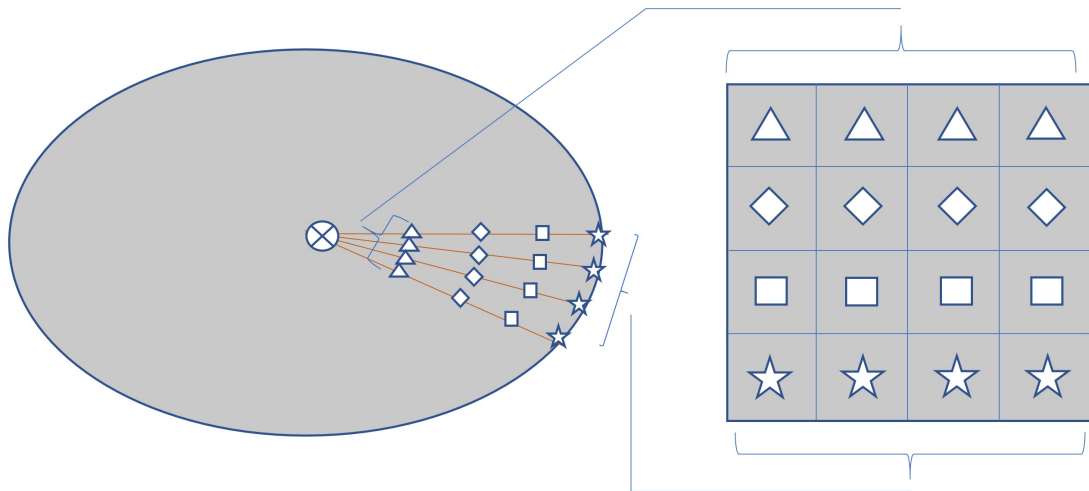
Figure 2.15: The process of polar-to-Cartesian mapping to create a square image (4x4) based from an otolith radial slice. If we specify higher values for the height/width of the resulting square image, then the number of corresponding points from the otolith also increases creating a transformed image with higher quality.
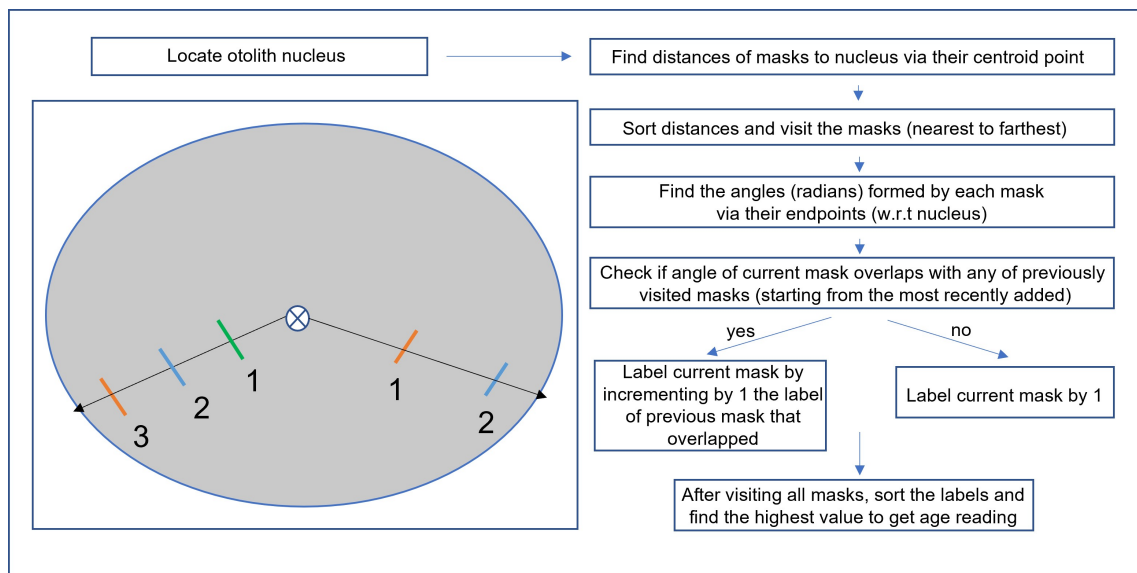
Figure 2.16: Schematic diagram for the post-processing step for both Mask R-CNN and U-Net.

2.9 References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). Tensor-Flow: Large-scale machine learning on heterogeneous systems. *Software available from tensorflow.org*.

Abdellatif, A. (2021). vgg16-u-net. https://www.kaggle.com/code/aithammadiabdellatif/vgg16-u-net.

Abdulla, W. (2017). Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. *Available from https://github.com/matterport/Mask_RCNN*.

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:1798–1828.

Bermejo, S., Monegal, B., and Cabestany, J. (2007). Fish age categorization from otolith images using multi-class support vector machines. *Fisheries Research*, 84:247–253.

Billauer, E. (2009). peakdet: peak detection using MATLAB (non-derivative local extremum, maximum, minimum). http://billauer.co.il/peakdet.html.

Bouckaert, R. and Frank, E. (2004). Evaluating the replicability of significance tests for comparing learning algorithms. *Advances in Knowledge Discovery and Data Mining. PAKDD 2004. Lecture Notes in Computer Science*, 3056.

Campana, S. (1999). Chemistry and composition of fish otoliths: pathways, mechanisms and applications. *Marine Ecological Progress Series*, 188:263–297.

Cao, F. and Fablet, R. (2006). Automatic morphological detection of otolith nucleus. *Pattern Recognition Letters*, 27:658–666.

Carbonara, P. and Follesa, M. (2019). Handbook on fish age determination: A Mediterranean experience. *General Fisheries Commission for the Mediterranean. Studies and Reviews*, 98:1–179.

Chollet, F. et al. (2018). Keras. Github repository. *Available from https://github.com/fchollet/keras*.

Dietterich, T. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10:1895–1923.

Dutta, A. and Zisserman, A. (2019). The VIA Annotation Software for Images, Audio and Video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, New York, NY, USA. ACM.

Fablet, R. and Josse, N. L. (2005). Automated fish age estimation from otolith images using statistical learning. *Fisheries Research*, 72:279–290.

Fisher, M. and Hunter, E. (2018). Digital imaging techniques in otolith data capture, analysis and interpretation. *Marine Ecology Progress Series*, 598.

Formella, A., Vázquez, J., Carrión, P., Cernadas, E., Vázquez, A., and Pérez-Gándaras, G. (2007). Age reading of cod otoliths based on image morphing, filtering and fourier analysis. *Proceedings of the 7th IASTED International Conference on Visualization, Imaging, and Image Processing*.

Harbitz, A. (2009). A Generic Ad-Hoc Algorithm for Automatic Nucleus Detection from the Otolith Contour. *4th International Otolith Symposium*.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988.

Henderson, T. (2023). correctR: corrected test statistics for comparing machine learning models on correlated samples. R package version 0.1.3. https://CRAN.R-project.org/package=correctR.

ICES (2008). Report of the Workshop on Age Reading of North Sea Cod (WKARNSC). *ICES CM 2008/ACOM*, 39:71.

ICES (2020). Report of the spring 2019 Western Baltic cod (*Gadus morhua*) age reading exchange – SD 22. *ICES. 2020*.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.

Krumme, U., Stötera, S., McQueen, K., and Pahlke, E. (2020). Age validation of age 0-3 cod *Gadus morhua* in the western Baltic Sea through mark-recapture and tetracycline marking of otoliths. *Marine Ecology Progress Series*, 645:141–158.

Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). Microsoft coco: Common objects in context. *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, Proceedings, Part V*, 13:740–755.

Martinsen, I., Harbitz, A., and Bianchi, F. (2022). Age prediction by deep learning applied to Greenland halibut (*Reinhardtius hippoglossoides*) otolith images. *PLOS ONE*, 17:e0277244.

Matterport (2017). Mask R-CNN. https://github.com/matterport/Mask_RCNN.

Moen, E., Handegard, N., Allken, V., Albert, O., Harbitz, A., and Malde, K. (2018). Automatic interpretation of otoliths using deep learning. *PLOS ONE*, 13:e0204713.

Moen, E., Vabø, R., Smoliński, S., Denechaud, C., Handegard, N., and Malde, K. (2023). Age interpretation of cod otoliths using deep learning. *Ecological Informatics*, 78.

Nadeau, C. and Bengio, Y. (2003). Inference for the generalization error. *Machine Learning*, 52:239–281.

Ordoñez, A., Eikvil, L., Salberg, A., Harbitz, A., and Elvarsson, B. (2022). Automatic fish age determination across different otolith image labs using domain adaptation. *Fishes*, 7:71.

Ordoñez, A., Eikvil, L., Salberg, A., Harbitz, A., Murray, S., and Kampffmeyer, M. (2020). Explaining decisions of deep neural networks used for fish age prediction. *PLOS ONE*, 15:e.0235013.

O'Shea, K. and Nash, R. (2015). An Introduction to Convolutional Neural Networks. *CoRR*, abs/1511.08458.

Panfili, J., de Pontual, H., Troadec, H., and Wright, P. (2002). Manual of fish sclerochronology. *Ifremer-IRD coedition*.

Politikos, D., Petasis, G., Chatzispyrou, A., Mytilineou, C., and Anastasopoulou, A. (2021). Automating fish age estimation combining otolith images and deep learning: The role of multitask learning. *Fisheries Research*, 242:106033.

R Core Team (2023). R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer. (available on arXiv:1505.04597 [cs.CV]).

Sigurdardóttir, A., Sverrisson, P., Jónsdóttir, A., Gudjónsdóttir, M., Elvarsson, B., and Einarsson, H. (2023). Otolith age determination with a simple computer vision based few-shot learning method. *Ecological Informatics*, 76:102046.

Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. *CoRR*, abs/1512.00567.

Troadec, H. (1991). Frequency demodulation on otolith numerical images for the automation of fish age estimation. *Aquatic Living Resources*, 4:207–219.

van der Walt, S., Schönberger, J., Nunez-Iglesias, J., Boulogne, F., Warner, J., Yager, N., Gouillart, E., Yu, T., et al. (2014). scikit-image: image processing in Python. *PeerJ*, 2:e453.

Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.

Williams, A., Davies, C., and Mapstone, B. (2005). Variations in the periodicity and timing of increment formation in red throat emperor (*Lethrinus miniatus*) otoliths. *Marine Freshwater Research*, 56:529–538.

Zimmermann, R. S. and Siems, J. N. (2019). Faster training of Mask R-CNN by focusing on instance boundaries. *Computer Vision and Image Understanding*, 188:102795.

Chapter 3

An Interactive AI-driven Platform for Fish Age Reading

Arjay Cayetano[1*], Christoph Stransky[1], Andreas Birk[2], Thomas Brey[3]

1 Thünen Institute of Sea Fisheries, Bremerhaven, Germany
2 School of Science and Engineering, Constructor University, Bremen, Germany
3 Faculty of Biology and Chemistry, University of Bremen, Bremen, Germany

3.1   Abstract

Fish age is an important biological variable required as part of routine stock assessment and analysis of fish population dynamics. Age estimates are traditionally obtained by human experts from the count of ring-like patterns along calcified structures such as otoliths. To automate the process and minimize human bias, modern methods have been designed utilizing the advances in the field of artificial intelligence (AI). While many AI-based methods have been shown to attain satisfactory accuracy, there are concerns regarding the lack of explainability of some early implementations. Recently, we have developed explainable AI-based approaches based on U-Net and Mask R-CNN having direct compatibility with traditional ring counting procedures. We further extend this approach by creating an interactive website housing these explainable AI methods allowing age readers to be directly involved in the AI training and development. An important aspect of the platform presented in this article is that it allows the additional use of different advanced concepts of Machine Learning (ML) such as transfer learning, ensemble learning and continual learning, which are all shown to be effective in this study.

Keywords: *fish age reading, otoliths, artificial intelligence, deep learning, object detection, segmentation*

3.2   Introduction

Computer-assisted annotation of image data is a standard tool in marine biology since quite a while (Schlining and Stout, 2006) with increasing amounts of automated processing over the years (Gomes-Pereira et al., 2016). The progress in Artificial Intelligence (AI), or more precisely in Machine Learning (ML) in the form of deep neural networks (LeCun et al., 2015) has provided a further boost to this trend (Rubbens et al., 2023; Radeta et al., 2015; Katija et al., 2022). This also holds for the application area which motivates the work presented in this article, namely the field of fish age reading.

Stock assessments rely heavily on fish age data which is primarily derived from the so-called otoliths or ear stones. These are calcified structures that form ring-like patterns influenced by the changing seasons (Campana, 1999) similar to the pattern of ring formation found in tree trunks as used for dendrochronological studies (VanderKooy et al., 2020). Hence, in the same manner, age readers count the number of such otolith rings (annuli) in order to derive the fish age estimates (Panfili et al., 2002).

Often, otolith rings are not straightforward to identify or to detect. Consequently, human readers need to undergo extensive training and even attend a number of workshops with other age readers with the goal of standardizing and minimizing the subjective aspects of the process. As part of these workshops, inter-reader agreements are measured by allowing them to read a selected set of otolith images. Based on these workshops (ICES, 2008, 2020), it was observed that there can be a concerning degree of disagreement among age estimates even for otoliths with simple patterns such as those of cod. It is therefore important to find a definitive and unbiased solution to this problem in order to prevent miscalculations involving such important biological parameter.

Hence, there has been a growing interest in recent years regarding the use of AI in age reading. Some early attempts in this endeavor dates back from the years where classical machine learning algorithms were popular along with the practice of feature engineering. These include the works of Fablet and Josse (2005) as well as from Bermejo et al. (2007) where classical neural networks and support vector machines were explored.

With the progress in deep learning, the number of studies applying AI-based methods for otolith age reading increased substantially. A lot of these studies made use of Convolutional Neural Networks (CNN) with either classification or regression formulation (Moen et al., 2018; Ordoñez et al., 2020, 2022; Martinsen et al., 2022; Politikos et al., 2021). Recently, a new batch of approaches emerged using more recent concepts such as Transformers (Sigurdardóttir et al., 2023) and Ensemble Learning (Moen et al., 2023), indicating the continued pursuit to further improve AI-based approaches for otolith age reading.

There are some concerns, however, when it comes to the black-box nature of many of these implementations. This is primarily due to their lack of compatibility with the traditional manual methodology in which the rings are explicitly counted to derive the age values. Hence, newer designs were recently developed with the goal of making the process as compatible as possible with the ring-counting procedure. Bojesen et al. (2024) applied a novel method involving generative models to specifically mimic the ring annotations often made by readers especially during workshops serving as visual guides for easier ring identification. That is, certain markings (dots) are placed by the AI-method within the image indicating the ring portions it detected. Likewise, in own work, Cayetano et al. (2024) successfully created automated ring annotations using Mask R-CNN (He et al., 2017) and U-Net (Ronneberger et al., 2015), which are popular methods for object detection and segmentation. These methods are shown to satisfactorily create masks of the rings at certain reading axes from which age estimates are derived through a simple automated counting procedure. In this way, a human can easily comprehend the way the fish age was determined and validate the result if needed.

To further increase the acceptance and trust in AI-based approaches for age reading, we extend the use of explainable methods by creating a platform which enables the readers to directly participate in the development and training of such AI models, similar to the idea behind DeepOtolith (Politikos et al., 2022). In this study, we developed an interactive web-based application that houses both Mask R-CNN and U-Net algorithms and makes them accessible for age readers via an intuitive user interface. An important scientific aspect introduced in this work is that this approach allows using additional, more advanced Machine Learning concepts, which boost the performance as shown in this article.

From the perspective of an AI-based fish-age-reading tool, the idea is to provide a mixed collection of AI models composed of user-trained models, along with the models we developed (Cayetano et al., 2024) as well as those generic models trained from common objects such as VGG (Simonyan and Zisserman, 2015) and Mrcnn-Coco (Abdulla, 2017; Lin et al., 2014). This opens up additional options to use advanced ML concepts to further improve the performance. Concretely, the following three contributions are presented in this article. First, transfer learning is incorporated where existing models are reused to initiate a new round of training on a new set of otolith images. Second, multiple models can be consolidated during the testing/prediction stage using an ensemble method to

create an aggregate prediction, which is significantly better than the individual model predictions. Third, continual learning is used, i.e., the user can build an initial model from a specific dataset which will be trained repeatedly each time a new dataset comes without completely forgetting the original dataset it was trained on.

3.3   Materials and Methods

Implementation and Design

We provide here a short overview of the technical details of the platform as background information. This also includes information on the basic ML methods that we build upon before the main conceptual contributions of this article are presented in the succeeding sections.

The platform is implemented as a web application which is currently hosted at Thünen Institute on a Linux server with high-performance graphics cards for deep learning and using a docker setup for Tensorflow-GPU/Keras library (Abadi et al., 2015; Chollet et al., 2018). Currently, the website access is limited within the Thünen local network but a portable standalone version (Windows only) can be downloaded for testing, exploring and trying out some selected useful features. The implementation (DOI 10.5281/zenodo.8341297) is based on Python and the Django web framework (Django Software Foundation, 2019). The portable version (DOI 10.5281/zenodo.10954470) is packaged in a zip file, containing all the necessary libraries for starting the webserver and can even be configured to allow other local machines to connect.

Similar to the previous work (Cayetano et al., 2024), preliminary image processing steps were included such as outer contour detection and selected image adjustments. For instance, the watershed algorithm (Pedregosa et al., 2011) can be employed for automatically segmenting the otoliths from its background. In case of errors in the contour detection, an annotation toolkit is integrated for the end-user, namely VIA or Visual Geometry Group Image Annotation tool (Dutta and Zisserman, 2019).

Likewise, the web application also provides the two deep learning methods for fish age reading that we previously developed (Cayetano et al., 2024), which are based on Mask R-CNN (He et al., 2017) and U-Net (Ronneberger et al., 2015). These two methods were demonstrated to be accurate and robust in estimating the fish age while at the same time, creating image annotations of the otolith rings (i.e., image masks or markings showing the identified otolith annuli). Both algorithms are supervised methods requiring ground truth labels. The labels are in the form of mask annotations within an image indicating the region of interest to be detected or segmented, respectively, by Mask R-CNN and U-Net.

As mentioned before, VIA is integrated, which can be used to create ground-truth annotations using its drawing tool sets. Furthermore, we augmented VIA and integrated our own custom brush tool, which makes the creation of annotations for this specific application case faster and easier. Fig 3.1 shows an example of an irregular shape created using the brush tool which will be otherwise tedious to draw using the default VIA drawing toolkits.
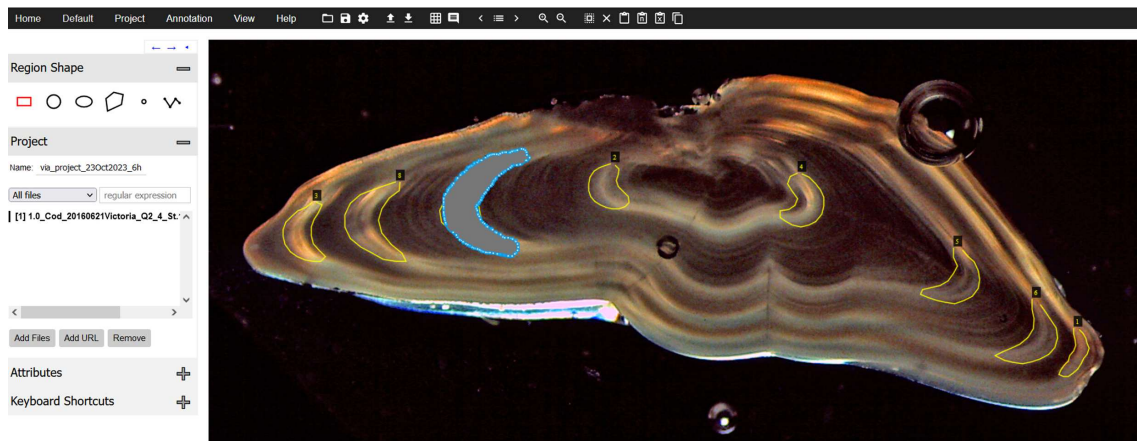
Figure 3.1: Improved VIA annotation toolset. Sample annotation accomplished using the brush tool we implemented within the integrated VIA toolkit.

As another modification of the VIA tool to ease the annotation process, we also utilize existing pre-trained otolith models from own work (Cayetano et al., 2024) that can perform reasonable initial annotations on new sets of images to be edited/corrected by end-users as necessary. That is, this feature serves as an AI-assistant that helps the users create the ground-truth labels for new images they upload. The process is illustrated in Fig 3.2.

Training Functionalities and Transfer Learning

After creating the annotations, the next step is to use them for training. First, the validation data needs to be created which will be used for measuring the performance of the model during training so that the best-so-far state can be saved and retrieved accordingly. In this study, the validation data was derived simply by using horizontal flipping of the training data. The test data, however, will be different for each set of training and testing experiment we conduct which is illustrated in Fig 3.3.

The next step is to choose the type of training to initiate which can either be from scratch or from reusing previous models. The former trains a model from the ground up using the default random weight initializations of the CNN. The latter takes advantage of transfer learning to load existing model weights during network initialization from which the training can continue. This will be elaborated in the following subsection.

Transfer Learning

There are essentially three ways by which the user can select existing models for transfer learning. First is the basic case involving existing models published by the AI community which are trained from common objects such as VGG (Simonyan and Zisserman, 2015) and Mrcnn-Coco (Abdulla, 2017; Lin et al., 2014). The second case involves reusing models trained from application specific data, here otolith images, such as the ones we already provided (Cayetano et al., 2024) which can be reused for any otolith image dataset even with different species or domains. The last case is the use of highly domain-specific
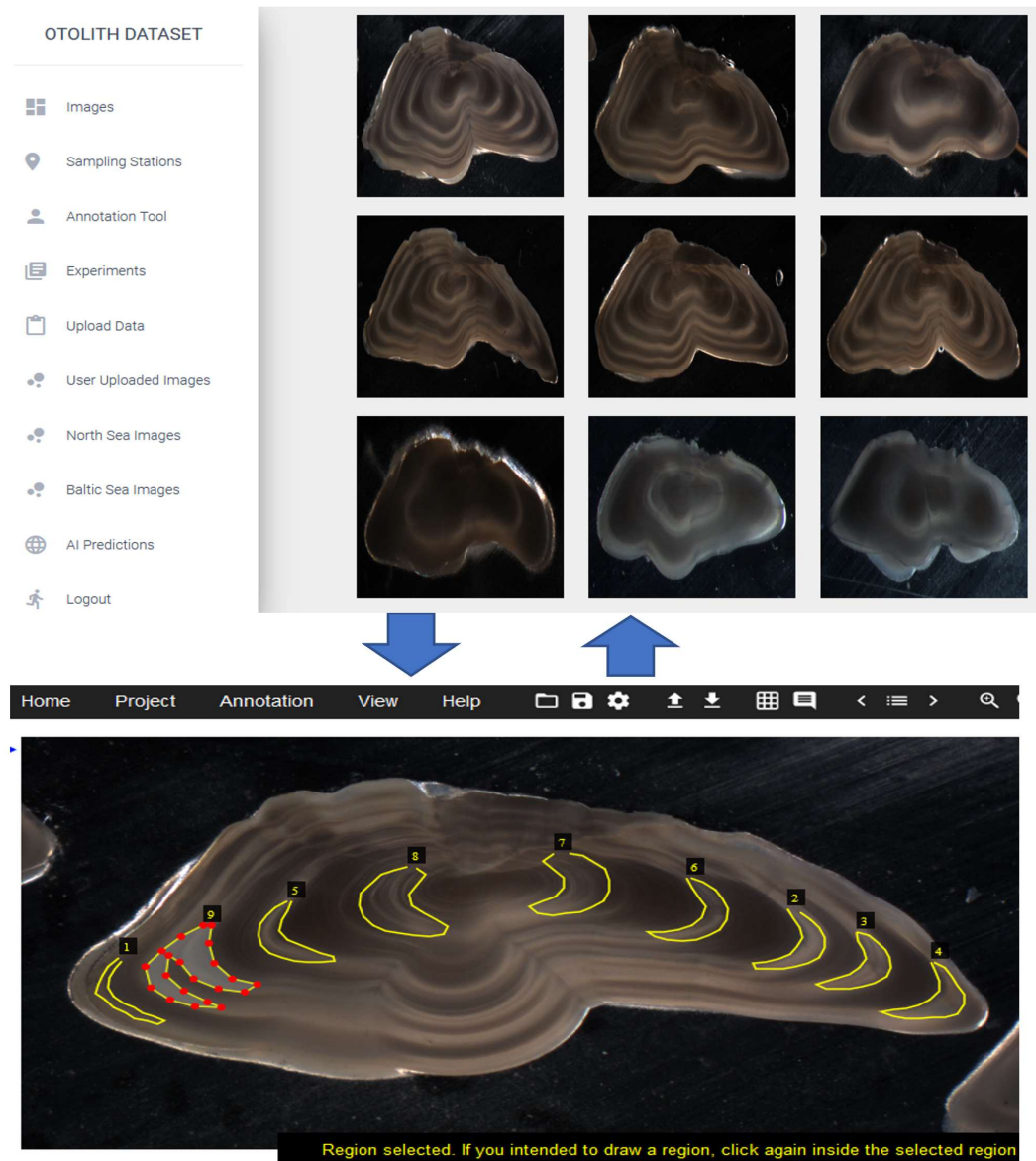
Figure 3.2: With AI assistance. The AI-assisted annotation we implemented where the users can edit the initial annotation given by the AI which can be saved and utilized in the platform.

models, i.e., training involves images with the same application specific data and the same characteristics (e.g. lighting, orientation), which hence potentially requires only a simple model update.

Fig 3.3-B shows how we divided the current datasets in order to explore the concept of transfer learning and utilize the existing otolith models which are taken from the previous study we conducted (Cayetano et al., 2024). The details of our datasets (North Sea and Baltic Sea datasets) along with additional information on how we trained our existing
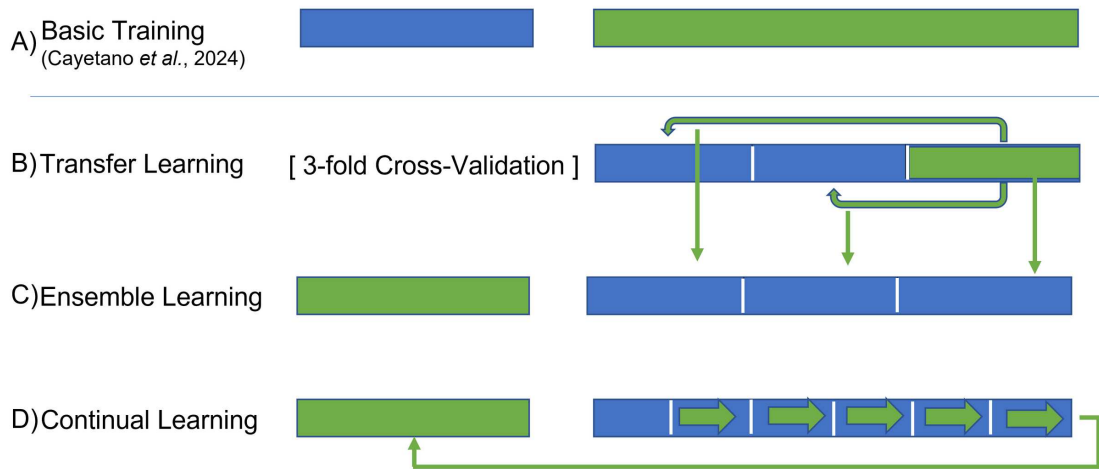
Figure 3.3: **Training-testing splits.** The partitioning of the dataset for the different experiments we conducted namely (A) basic training done from the previous study (B) transfer learning (C) ensemble learning (D) continual learning. The training set is marked in blue while the test set is marked in green. In addition, the green colored arrows point to the previous test sets that will eventually become training sets for the succeeding stages of the relevant experiments.

models are described in our previous article and summarized in Fig 3.3-A.

As depicted in the figure, a small subset of the images (1/5 for the North Sea dataset and around 1/10 for the Baltic Sea dataset) was used for training our existing models and a larger subset was used for testing. These models are domain-specific which means that there are separate models created for the North Sea dataset and the Baltic Sea dataset.

To fully utilize our image collection, we have to reuse the larger subset (i.e., the previous test set shown in green in Fig 3.3-A) in order to conduct new rounds of training with the different scenarios mentioned in this study. To evaluate the use of transfer learning, we conduct a 3-fold cross-validation experiment (Fig 3.3-B) where this subset is divided into three parts. Then, following the usual procedure, each round of cross-validation uses one of the three folds as test set from which the performance of the scenarios can be evaluated.

Testing Functionalities and Ensemble Learning

In this section, we present the options for performing predictions on a separate set of images. In the basic case, the trained model created by the user is tested individually on a different set of images prepared for testing. The application's output are the predicted annotations within the image itself along with the estimate of the age value.

In the more advanced scenario, the predictions of the models can be combined to create one aggregate prediction using ensemble learning. This will be elaborated in the next section.

Ensemble Learning

Ensemble Learning is a machine learning technique in which multiple models are combined into a single aggregate model having a prediction which is computed or even learned from the individual predictions of its constituents (Ganaie et al., 2022). For simplicity, we can describe it as a form of meta-analysis or meta-learning where patterns are derived from the individual model predictions. For this experiment, the constituent models we used are the models trained in the previous experiment involving transfer learning. The following three types are considered here:

- **Model Averaging**

  This is the simplest method for creating an ensemble model. It only involves simple averaging of the predictions of the individual models. As reviewed by Ganaie et al. (2022), there are cases in which this can be a reasonable choice compared to some other more complex ensemble approaches.

- **Linear Regression**

  In this method, an additional round of training is required using linear regression as the meta-learner and using the deep learning model predictions as the inputs. As illustrated in Fig 3.3-C, this training is performed on the out-of-sample predictions from the 3-fold cross-validation conducted in the transfer learning experiment. Then, the testing of the ensemble can be performed on the original training set from the previous study (Cayetano et al., 2024), which we now designate as the test set. This means, however, that for this experiment, we have to exclude the use of some transfer learning models having base weights that were trained previously on this specified test set.

- **Random Forest**

  Random forest is a classical ML algorithm which by definition, also functions as an ensemble model for classification and regression (Breiman, 2001). For our purposes, we treat is as a typical classifier that can learn patterns from the predictions of the deep learning models. Likewise, as shown in Fig 3.3-C, we train it from the test predictions of the individual models from the transfer learning experiment and test it on the original training images used in our previous study (Cayetano et al., 2024). Hence, we will also exclude any model containing pretrained weights that has been derived from this currently designated test set.

Continual Learning

In this section, we present the use of continual learning. As mentioned in the previous section on AI-assisted annotation, this is a more advanced feature that can be used for creating ground-truth annotations where the user can choose to retrain the existing model that performs the initial annotation with the new annotated images assisted by the model itself.

In this functionality, users will be able to perform training in batches of data and with each batch, the model from the previous batch will be reloaded to annotate the new batch (with corrections from the user if necessary) and then subsequently, conduct training using the same set. This workflow loop is summarized in Fig 3.4.
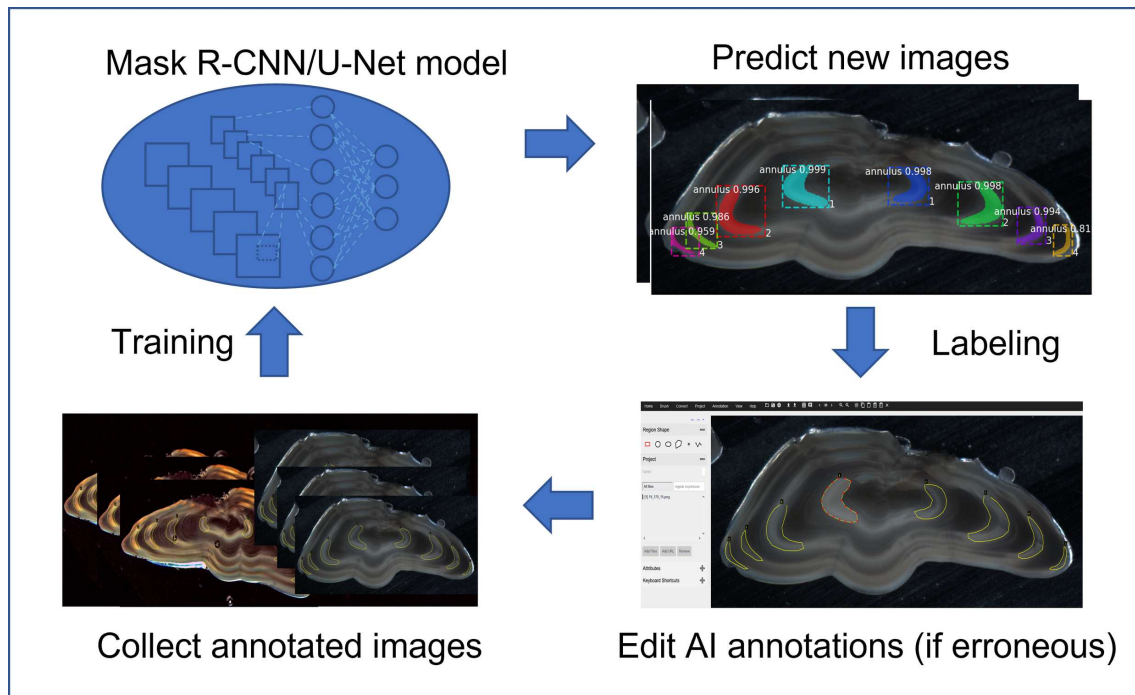
Figure 3.4: Continual learning workflow. The self-sustaining loop for continual learning where the model predictions on a new dataset can be edited, corrected and approved by the user so that it can be used to conduct retraining to utilize the new labeled images.

The main hurdle, however, is the so-called catastrophic forgetting or catastrophic interference (French, 1999; Parisi et al., 2019) where the model will eventually not recognize the previous dataset it has learned while it is currently being trained on a completely new dataset. Here, we investigate this phenomenon by observing the deterioration of prediction on the original dataset or source domain (North Sea dataset, in this case), as the model undergoes retraining using different training sets containing images from a new domain (i.e., Baltic Sea dataset). In addition, we explore the commonly proposed solution for catastrophic forgetting which involves rehearsal (Ratcliff, 1990; Robins, 1995). We employ the basic rehearsal method in which we include some images from the original domain and mix it on each new batch of images from the new domain to be used for further training. In this manner, the model has the chance to "review" or "rehearse" its original domain while also learning from the new set.

3.4   Results and Discussion

Fig 3.5 shows the website page containing a sample set of images with the corresponding predictions and annotations of the AI models that can be created using the web application. As shown in the figure, color markings of otolith rings are directly placed within the image making it easy to interpret and verify by the age reader.
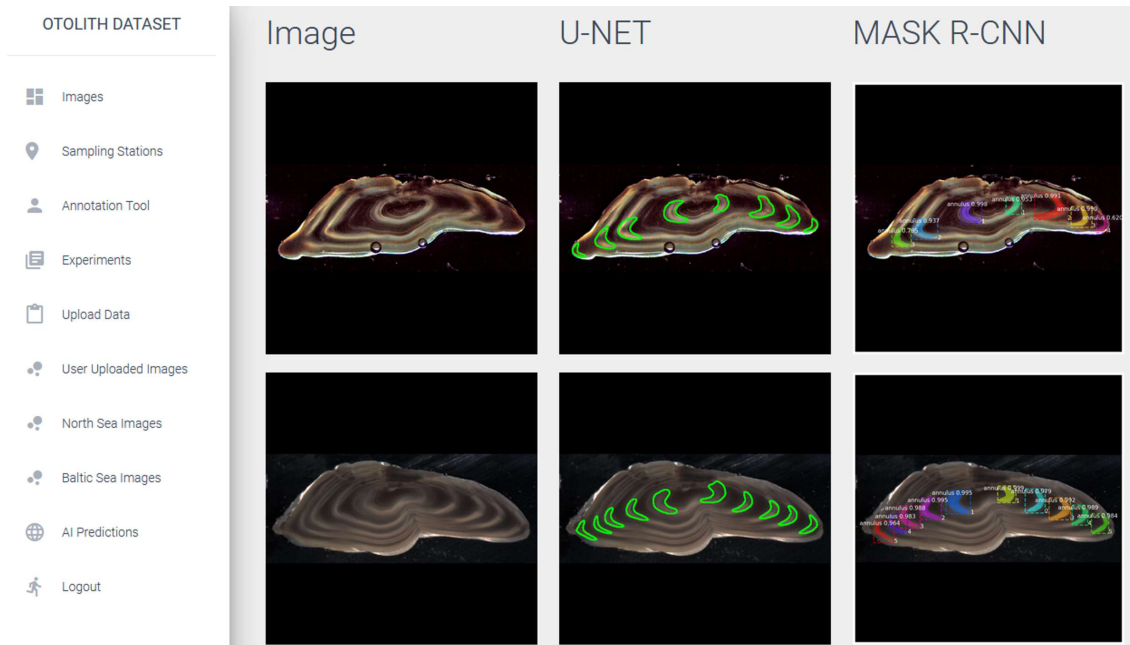
Figure 3.5: Webpage showing predictions. Automated image annotations created by the two different AI methods, U-Net and Mask R-CNN, as shown on the web application.

Transfer Learning

In Fig 3.6, the result of the experiments with the different types of transfer learning is shown. The evaluation of performance was conducted via 3-fold cross-validation as described in the Methods section.

It can be seen that for both algorithms (Mask R-CNN and U-Net), the model without any type of transfer learning has consistently attained the least accuracy. In addition, it can be seen that, in the case of the North Sea dataset, the model with pretrained weights based on North Sea images attained the highest accuracy for both Mask R-CNN (53.2%) and U-Net (62.5%). In the case of the Baltic Sea dataset, the model with pretrained weights based on Baltic Sea images also performed excellently attaining the second highest accuracy (73.3%) for Mask R-CNN and the highest accuracy (77.8%) for U-Net.

In addition, we further elaborate the performance of the best models created with transfer learning to see the distribution of their predictions on each age classes as shown in Fig 3.6(C-D). It can be seen from the violin plot, that the predictions of the best models are excellent from ages 1-8 and start to decline from ages 9 onwards. Nevertheless, the overall
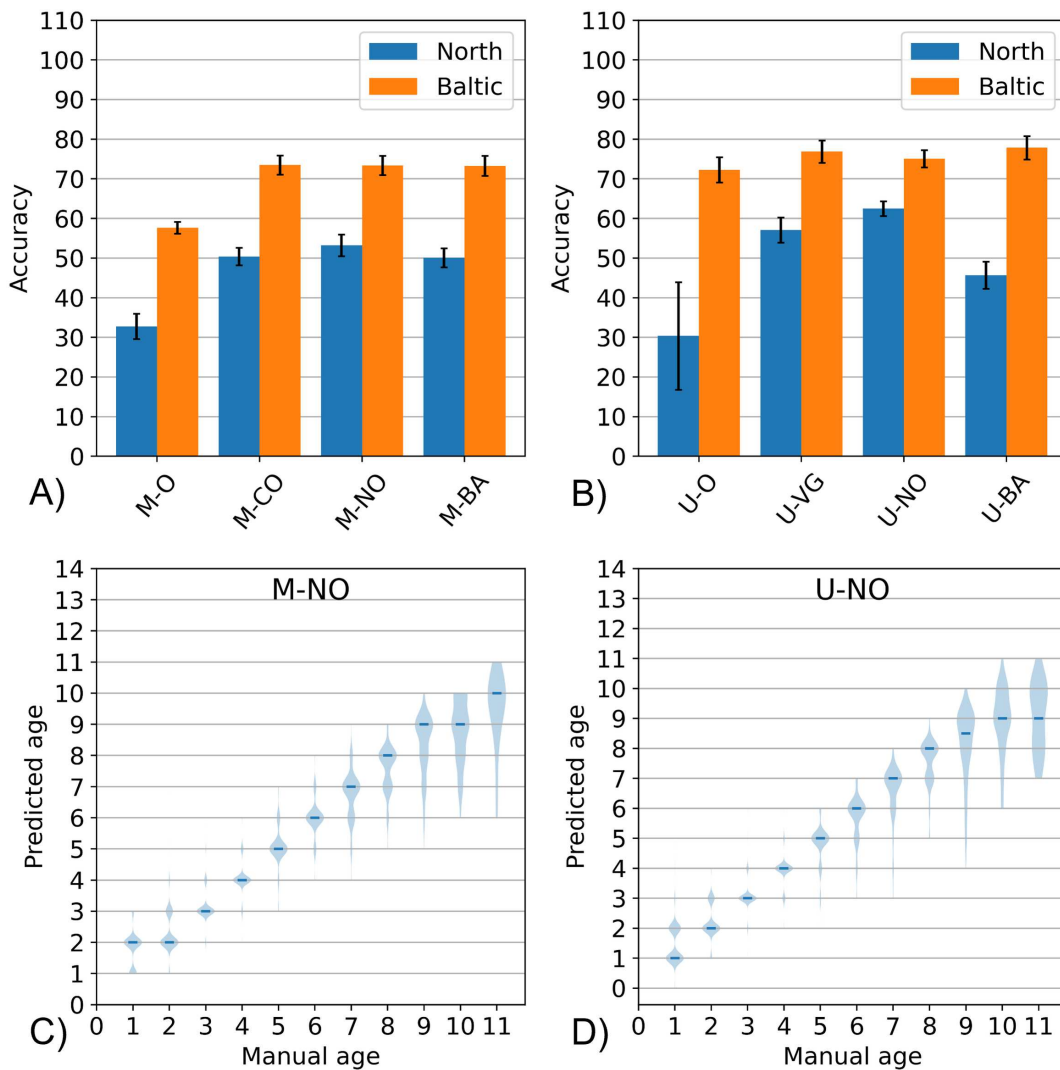
Figure 3.6: Cross-validation performance. The performance of the different models tested using 3-fold cross-validation (3 repeats using 3 different otolith base weights creating a total of 9 runs). All Mask R-CNN models start with "M" while all U-Net models start with "U". The models tagged with "-O" refer to those without transfer learning while the rest indicates the identifier of the pre-existing base models (-CO refers to the Coco model, -NO refers to the North Sea-trained model, -BA refers to the Baltic-trained models, and -VG refers to the VGG model).

performance can still be considered satisfactory despite using an imbalanced training set, which is a consequence of cross-validation, in contrast to the one used in the previous study (Cayetano et al., 2024) where each age group contains the same number of images (random sampling with rebalancing).

With these results, the advantage of transfer learning has been demonstrated and shown to be superior compared to training from scratch without any pretrained weights. This aligns with the results of various AI-based age reading studies such as the works of Moen

et al. (2018), Martinsen et al. (2022) and Ordoñez et al. (2020), where they previously demonstrated the effectiveness of pretrained Inception, Xception and VGG, respectively. In addition, our results also indicate that, in some cases, the use of base models pretrained from otolith images obtains better performance compared to using typical generic models pretrained from common objects. Hence, in this study, we present another future perspective to obtain further improvements by encouraging the development of more otolith base models that can be reused by the community using the concept of transfer learning.

Ensemble Learning

As mentioned, the ensemble approach is a meta-learning process requiring the model predictions as inputs for training and testing. Hence, we require sufficiently diverse models that can be combined accordingly. For this, we have reused the models trained using the transfer learning experiment from the previous section.

The training set of the ensemble would be the predictions of each individual model on the out-of-sample test set created via 3-fold cross-validation from the previous section. Then, for testing and proper analysis, we need to evaluate the performance of the ensemble on the training set used in the previously published study (Cayetano et al., 2024), which is now currently designated as the test set. In order for this to work, however, it was necessary to exclude the models containing weights trained from that set because they will naturally be familiar with it, making the results too optimistic. Hence, in Fig 3.7, it can be seen that we evaluate six models per dataset instead of eight as done in the previous section. The accuracy values for these selected models are plotted in a bar chart which we will use to compare with the combined/ensemble models.

In Fig 3.8, the three types of ensemble approaches are evaluated. Fig 3.8(A-C) show the distribution of predictions on each age classes produced by the three ensemble methods. It can be observed that the ensemble using Model Averaging produced the worst distribution among the three methods. When it comes to the distribution across all age groups, the Linear Regression produced the best trend where the age-wise median values were located on the expected position with the exception of age 11. For the Random Forest method, both age 5 and age 11 otolith images were heavily under-estimated but the rest of the age groups are excellently predicted. To know which one is superior, we also plotted their corresponding accuracy values (Fig 3.8-D). It can be seen that Random Forest (RF) attains the highest accuracy values for both North Sea and Baltic Sea images. Moreover, for both datasets, it surpasses the performance of its best constituent model which is depicted in the plot as dashed blue and orange lines, respectively. This means that the use of an ensemble, particularly in form of a Random Forest, can be used to further improve predictions. These results complement the findings from Moen et al. (2023), where they showed how simple averaging ensemble already provides a decent increase in accuracy. In our study, we explored additional ensemble types apart from model averaging and showed that further improvements can be obtained with other more sophisticated ensemble methods.
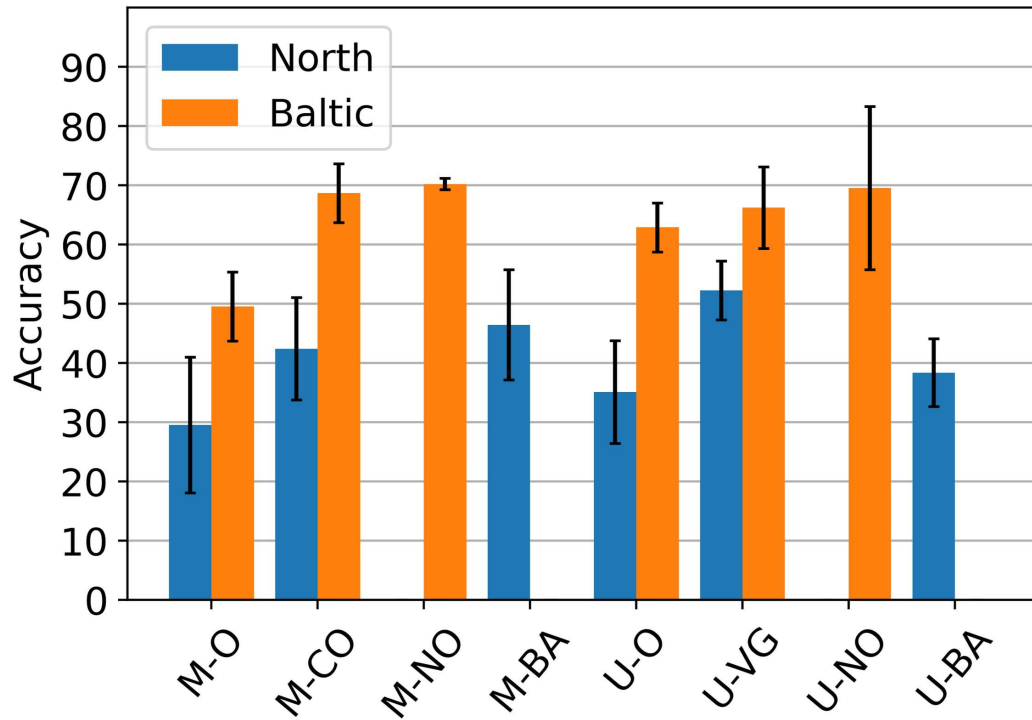
Figure 3.7: Performance of constituent models. Performance of each individual model used to create the ensemble models (n=3). In compliance with standard practice, all models from the cross-validation experiments are discarded. To measure the performance of the methods, i.e., those starting with M- represents Mask R-CNN and U- for U-Net, a final training was done where all images from the three folds of cross-validation were included. Then, evaluation was done on an unused set, i.e., the set which was used for training in the previous study (Cayetano et al., 2024) which is now designated as the test set.

Continual Learning

In this section, we first investigate the phenomenon known as catastrophic forgetting in which a trained model tends to forget its original training set once retrained with a completely new dataset. In Fig 3.9(A-D), we illustrate this behavior by reusing an existing U-Net model previously trained from the North Sea images and retraining it in four different ways.

To see how well it remembers the original training set, we use a validation set composed of the exact images it was previously trained on. We then plot the training loss (blue) and the validation loss (orange) as the model undergoes retraining under four different scenarios. In Fig 3.9-A, we run a baseline scenario where the training set is composed of the exact images from the original training without introducing new images. This represents the case where the model is still completely familiar with the validation set resulting to the expected highly optimistic loss plot. Afterwards, we perform additional
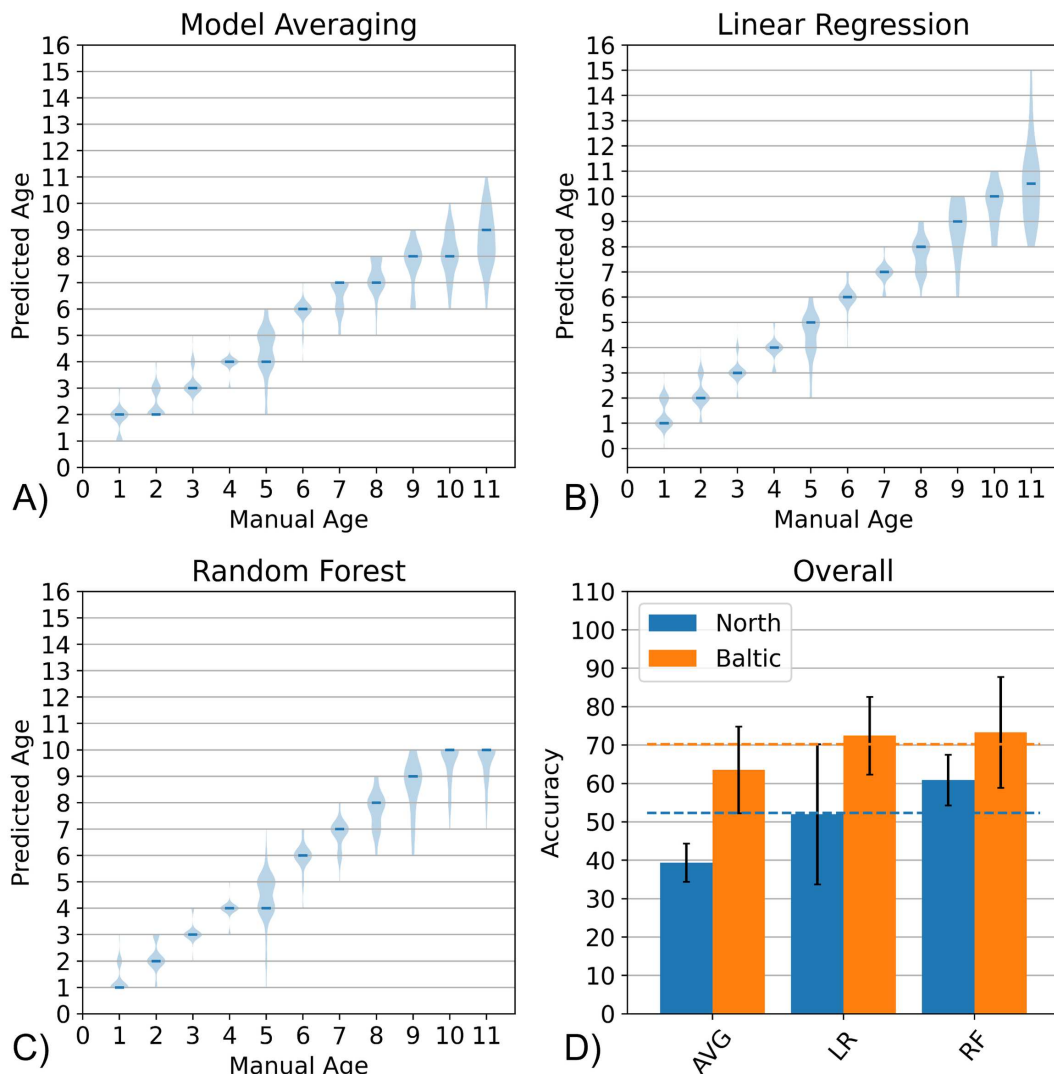
Figure 3.8: **Ensemble performance.** Distribution of age predictions (n=3) using the three ensemble methods namely (A) Model Averaging (AVG) (B) Linear Regression (LR) (C) Random Forest (RF). The overall accuracy of the three methods is summarized in a bar chart (D) along with the accuracy of the best constituent model (from previous experiment) indicated by dashed lines.

experiments where we add new domain images into the training set. In Fig 3.9-B, we create a mixed training set composed of 33.3% new images and 66.7% old images. It can be seen how the model still manages to have improvements in both training and validation loss albeit in a more unstable manner compared to the baseline scenario. Next, we increase the proportion of the new images (66.7%) for training and plot the resulting loss in Fig 3.9-C. Here, the improvement is now heavily leaning towards the training while the validation loss is maintained (i.e., neither improves nor deteriorates). Finally, we explore the case where all training images are from the new domain. In Fig 3.9-D, it can be seen that the
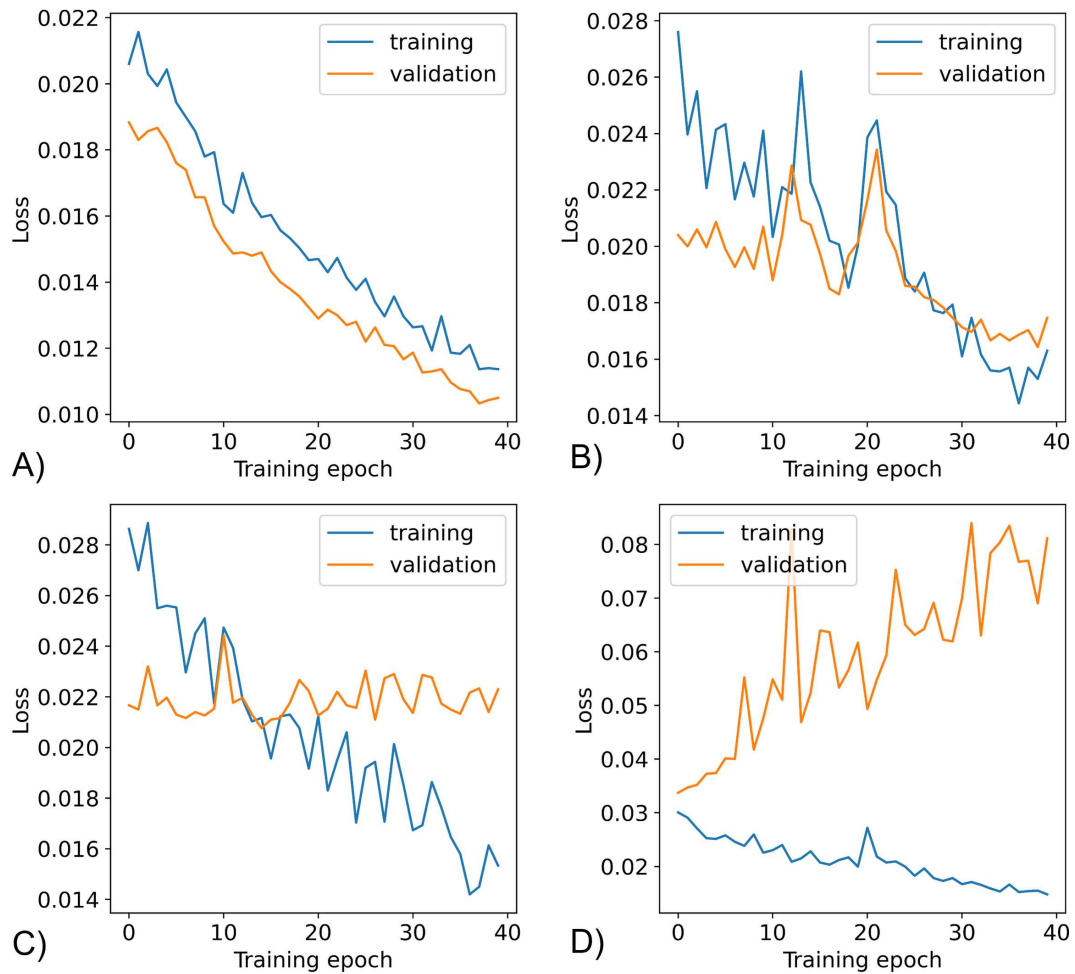
Figure 3.9: Catastrophic forgetting. The effect of catastrophic forgetting investigated (n=3) using four different training scenarios involving images from original domain (North Sea) and the new domain (Baltic Sea). (A) training set with purely old domain images (B) mixed training set with 33.3% new and 66.7% old images (C) mixed set with 66.7% new and 33.3% old images (D) training set consisting purely of new domain images.

validation loss gets worse while the training loss improves indicating that the model's familiarity with the old training images deteriorates as the model learns the new domain images. This deterioration of the validation loss represents the catastrophic forgetting phenomenon as the model encounters increasing proportion of unfamiliar images during training.

In the next experiment, we investigate the most common solution proposed for handling catastrophic forgetting which is the rehearsal method (Ratcliff, 1990; Robins, 1995). For this, we evaluate a scenario with and without rehearsal. This is done separately for Mask R-CNN and U-Net using pretrained weights based on North Sea images. Hence, similar to the previous experiment, we consider the North Sea dataset as the original domain while the Baltic Sea dataset serves as the new domain. For rehearsal, we use a mixed training set

where the number of original domain images is 50% of the new domain images. These old images are reintroduced in each subsequent training stage as described in the Methods section (Fig 3.3-D).

The accuracy plots in Fig 3.10 show the effectiveness of the rehearsal approach in preventing catastrophic forgetting. The models with rehearsal, shown in green lines, are able to maintain good accuracy values for the original domain as well as for the new domain that they are currently learning from. This observation is true for both Mask R-CNN and U-Net. The models without rehearsal are only good for the current domain and they become completely unfamiliar with the original domain they were previously trained on. Similar to the previous experiment, the performance of the models on the original domain deteriorates drastically as they learn from the new domain without rehearsing on the old images.
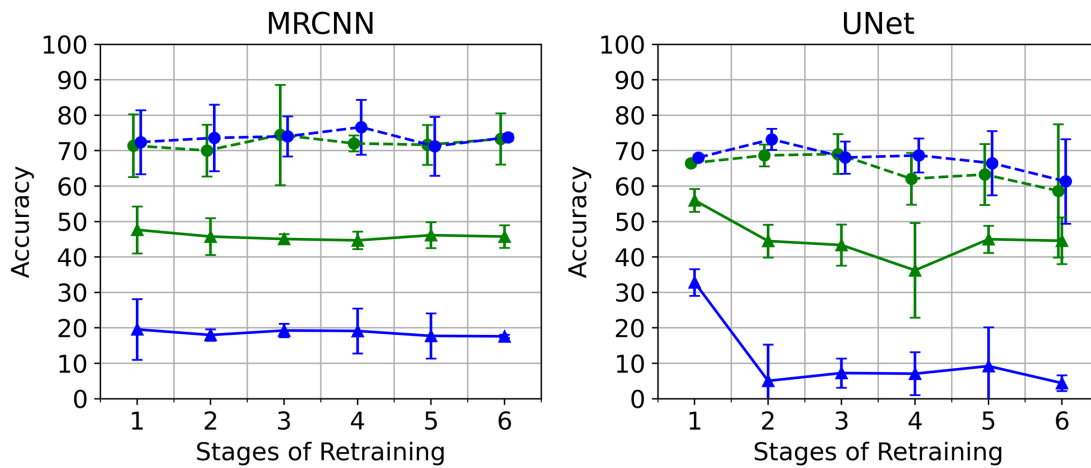
Figure 3.10: Performance with- and without rehearsal. The accuracy values for the continuously learning models at six stages of retraining. The green lines indicate the models with rehearsal while the blue lines indicate those without rehearsal (normal training). The continuous lines (with triangle markers) represent the accuracy for the original domain while the dashed lines (with circular markers) represent the accuracy for the current/new domain. In both algorithms, the models with rehearsal are able to maintain good accuracy for both the original domain and the current domain. The models without rehearsal only performed satisfactorily for the current domain and drastically became unfamiliar with the original domain it was previously trained on.

As a final comparison, we present confusion matrices showing the prediction accuracy of the models with- and without rehearsal evaluated on the current domain. As seen from Fig 3.11, all models have comparable performance, i.e., they are all excellent in predicting ages 2-4 but performed slightly worse for ages 1 and 5. For each algorithm, the two variations of training (with- and without rehearsal) performed similarly as seen by the age-wise accuracy values along the diagonal of the corresponding matrix. This observation indicates that with rehearsal, the performance of the models on the new domain remains unaffected even when the training is conducted on a mixed set composed of both the old and the new images.
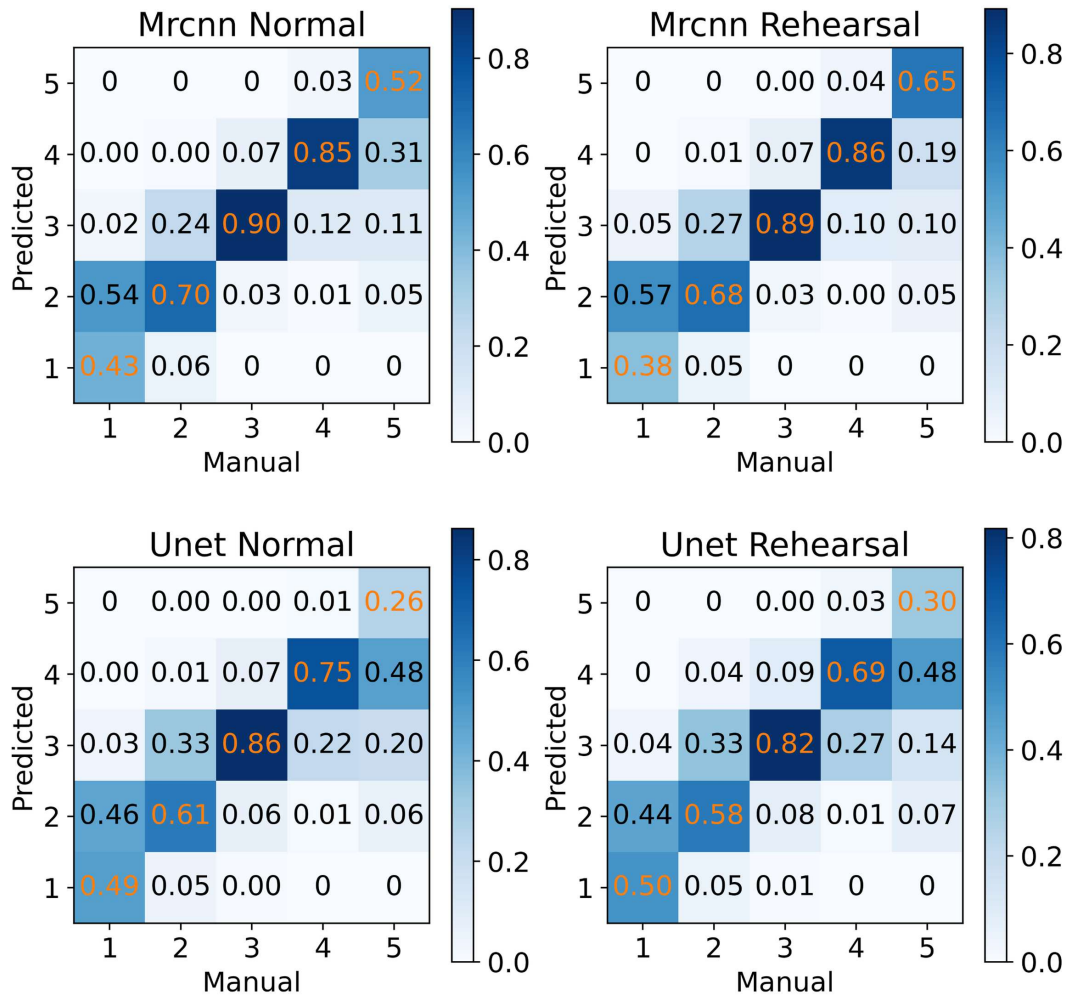
Figure 3.11: Age-wise accuracy. Confusion matrices containing the age-wise accuracy values of the models without rehearsal (Normal) and with rehearsal (Rehearsal) when evaluated on the current domain. Despite the inclusion of old images from the original domain via rehearsal, the performances of the rehearsal models are not affected for the new domain; this is indicated by the similar accuracy values attained by both model variants. The negligible percentage of predictions outside age range 1-5 are not shown.

3.5 Conclusions

We presented work on using higher-level Machine Learning strategies in the context of automated fish age reading based on images of otoliths. This is a highly relevant scientific and economical problem as fish age is an important biological variable required as part of routine stock assessment and analysis of fish population dynamics.

The presented work is grounded in an easy to use, web-based tool-chain targeted at human age-readers as end-users without a background in Artificial Intelligence. The tool-chain provides methods the end-users are already familiar with, e.g., standard image

processing and annotation tools, but also explainable AI in form of ring segmentations and automated counting.

An important aspect of the presented approach is that higher-level ML concepts such as transfer learning, ensemble learning and continual learning can be employed. It has been shown in this article that these techniques lead to improved performance which can be highly beneficial and timely especially with the recent growing interest in the use of AI for fish age reading. First, it was demonstrated here how transfer learning increased the overall accuracy especially when otolith base models are used instead of the generic base models trained from common objects. Second, it has been shown that, with ensemble learning especially using a Random Forest meta-learner, there is a further improvement in accuracy compared to just the individual performance of the constituent models. Lastly, using continual learning coupled with a basic rehearsal approach, it has been shown that one can create a single model which has an excellent accuracy for different datasets (or domains) making it unnecessary to train separate models for each set. In this study, not only that these concepts were explored in detail but also, they were packaged into a web-based platform where they can be easily utilized by end-users, even those without coding or programming background.

3.6 Acknowledgments

3.7 References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). Tensor-Flow: Large-scale machine learning on heterogeneous systems. *Software available from tensorflow.org*.

Abdulla, W. (2017). Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. *Available from https://github.com/matterport/Mask_RCNN*.

Bermejo, S., Monegal, B., and Cabestany, J. (2007). Fish age categorization from otolith images using multi-class support vector machines. *Fisheries Research*, 84:247–253.

Bojesen, T., Denechaud, C., and Malde, K. (2024). Annotating otoliths with a deep generative model. *ICES Journal of Marine Science*, 81:55–65.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

Campana, S. (1999). Chemistry and composition of fish otoliths: pathways, mechanisms and applications. *Marine Ecological Progress Series*, 188:263–297.

Cayetano, A., Stransky, C., Birk, A., and Brey, T. (2024). Fish age reading using deep learning methods for object-detection and segmentation. *ICES Journal of Marine Science*, fsae020.

Chollet, F. et al. (2018). Keras. Github repository. *Available from https://github.com/fchollet/keras.*

Django Software Foundation (2019). Django. *Available from https://djangoproject.com.*

Dutta, A. and Zisserman, A. (2019). The VIA Annotation Software for Images, Audio and Video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, New York, NY, USA. ACM.

Fablet, R. and Josse, N. L. (2005). Automated fish age estimation from otolith images using statistical learning. *Fisheries Research*, 72:279–290.

French, R. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Science*, 3:128–135.

Ganaie, M., Hu, M., Malik, A., Tanveer, M., and Suganthan, P. (2022). Ensemble deep learning: a review. *Engineering Applications of Artificial Intelligence*, 115:105151.

Gomes-Pereira, J., Auger, V., Beisiegel, K., Benjamin, R., Bergmann, M., Bowden, D., et al. (2016). Current and future trends in marine image annotation software. *Progress in Oceanography*, 149.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988.

ICES (2008). Report of the Workshop on Age Reading of North Sea Cod (WKARNSC). *ICES CM 2008/ACOM*, 39:71.

ICES (2020). Report of the spring 2019 Western Baltic cod (*Gadus morhua*) age reading exchange – SD 22. *ICES. 2020.*

Katija, K., Orenstein, E., Schlining, B., Lundsten, L., Barnard, K., Sainz, G., et al. (2022). FathomNet: A global image database for enabling artificial intelligence in the ocean. *Scientific Reports*, 12:15914.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521:436–444.

Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). Microsoft coco: Common objects in context. *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, Proceedings, Part V*, 13:740–755.

Martinsen, I., Harbitz, A., and Bianchi, F. (2022). Age prediction by deep learning applied to Greenland halibut (*Reinhardtius hippoglossoides*) otolith images. *PLOS ONE*, 17:e0277244.

Moen, E., Handegard, N., Allken, V., Albert, O., Harbitz, A., and Malde, K. (2018). Automatic interpretation of otoliths using deep learning. *PLOS ONE*, 13:e0204713.

Moen, E., Vabø, R., Smoliński, S., Denechaud, C., Handegard, N., and Malde, K. (2023). Age interpretation of cod otoliths using deep learning. *Ecological Informatics*, 78.

Ordoñez, A., Eikvil, L., Salberg, A., Harbitz, A., and Elvarsson, B. (2022). Automatic fish age determination across different otolith image labs using domain adaptation. *Fishes*, 7:71.

Ordoñez, A., Eikvil, L., Salberg, A., Harbitz, A., Murray, S., and Kampffmeyer, M. (2020). Explaining decisions of deep neural networks used for fish age prediction. *PLOS ONE*, 15:e.0235013.

Panfili, J., de Pontual, H., Troadec, H., and Wright, P. (2002). Manual of fish sclerochronology. *Ifremer-IRD coedition*.

Parisi, G., Kemker, R., Part, J., Kanan, C., and Wermter, S. (2019). Continual lifelong learning with neural networks: a review. *Neural Networks*, 113:54–71.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12:2825–30.

Politikos, D., Petasis, G., Chatzispyrou, A., Mytilineou, C., and Anastasopoulou, A. (2021). Automating fish age estimation combining otolith images and deep learning: The role of multitask learning. *Fisheries Research*, 242:106033.

Politikos, D., Sykiniotis, N., Petasis, G., Dedousis, P., Ordoñez, A., Vabø, R., et al. (2022). DeepOtolith v1.0: An Open-Source AI Platform for Automating Fish Age Reading from Otolith or Scale Images. *Fishes*, 7:121.

Radeta, M., Zuniga, A., Motlagh, N., Liyanage, M., Freitas, R., Youssef, M., et al. (2015). Deep Learning and the Oceans. *IEEE Computer*, 55:39–50.

Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review*, 97:285–308.

Robins, A. (1995). Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Science*, 7:123–146.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer. (available on arXiv:1505.04597 [cs.CV]).

Rubbens, P., Brodie, S., Cordier, T., Barcellos, D. D., Devos, P., Fernandes-Salvador, J., et al. (2023). Machine learning in marine ecology: an overview of techniques and applications. *ICES Journal of Marine Science*, 80:1829–1853.

Schlining, B. and Stout, N. (2006). MBARI's Video Annotation and Reference System. *IEEE/MTS OCEANS*.

Sigurdardóttir, A., Sverrisson, P., Jónsdóttir, A., Gudjónsdóttir, M., Elvarsson, B., and Einarsson, H. (2023). Otolith age determination with a simple computer vision based few-shot learning method. *Ecological Informatics*, 76:102046.

Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations*.

VanderKooy, S., Carroll, J., Elzey, S., Gilmore, J., and Kipp, J. (2020). A practical handbook for determining the ages of Gulf of Mexico and Atlantic Coast fishes. *Gulf States Marine Fisheries Commission Publication*, 300.

Chapter 4

Multi-stage Framework for Otolith Analysis

Arjay Cayetano[1*], Christoph Stransky[1], Andreas Birk[2], Thomas Brey[3]

1 Thünen Institute of Sea Fisheries, Bremerhaven, Germany
2 School of Science and Engineering, Constructor University, Bremen, Germany
3 Faculty of Biology and Chemistry, University of Bremen, Bremen, Germany

4.1   Abstract

Fish otolith is a structure that provides a wide array of biological information having great importance in fisheries science. Primarily, it has been used extensively for fish age determination where the patterns of its ring (or annulus) formation are traditionally counted manually to derive the fish age. Moreover, various studies involving the general otolith morphology have been conducted giving clues regarding the fish stock along with some essential biological details. Recently, the use of artificial intelligence (AI) for otolith analysis has increased substantially. As a step in this direction, we previously showed how deep learning algorithms for object detection and segmentation, namely U-Net and Mask R-CNN, can be used to automatically estimate the fish age and create annotations of the annuli. We then packaged the approach into a web-based application where advanced concepts such as transfer learning, ensemble learning and continual learning can be used in conjunction with the base methods. In this study, we present a multi-stage framework in which both U-Net and Mask R-CNN can be utilized further for the overall otolith analysis to obtain not only the age information but also other properties and measurements such as the otolith contour, nucleus position, and annuli distances. Here we elucidate the effectiveness of these methods to perform these tasks while creating more ways to make the methods accessible to the age reading community.

Keywords: *fish age reading, otoliths, artificial intelligence, deep learning, object detection, segmentation*

4.2   Introduction

Otolith is the structure responsible for the balance and sound detection of fish (Campana, 1999). It is formed by the deposition of calcium carbonate which has been found to be highly linked to the geographic and environmental conditions (Stransky, 2005; VanderKooy et al., 2020). Specifically, it was observed that the ring-like formations within the otolith correspond to the seasonal variations in non-tropical environments (i.e., winter and summer seasons produce different ring patterns within the otolith). Consequently, the manual identification and count of these rings have been a valuable source of information for determining the fish age (Proctor et al., 2021).

In addition, various studies have shown that other morphological properties of the otolith can be utilized for further analyses. For instance, the works of Campana and Casselman (1993), Stransky et al. (2008), and Hüssy et al. (2016) showed how the outer otolith shape can be used for distinguishing between different stocks. This has a particular importance, for instance, in determining the fish diet of sea mammals (Enoksen et al., 2016) and sea birds (Polito et al., 2011), where mixed otoliths are obtained from their gastrointestinal contents. Also, the study by Denechaud et al. (2020) has shown that the growth patterns inherent with the annuli formation can serve as proxy for studying environmental conditions affecting specific fish stocks.

Due to its biological importance, the otolith has been a subject of many studies involving computer vision and image processing techniques. For instance, there are studies that

have utilized classical image processing to determine the fish age as reviewed by Fisher and Hunter (2018) along with the extraction of the outer shape (Stransky, 2014) and the localization of the nucleus (Harbitz, 2009; Cao and Fablet, 2006).

Recently, with the progress in AI, particularly deep learning, otolith datasets have been subjected to various studies utilizing different sophisticated AI algorithms. For instance, the study of Stock et al. (2021) showed how one can utilize Convolutional Neural Networks (CNNs) and raw otolith images for species identification which is traditionally done with shape analysis. Furthermore, a large of number of studies have also shown the effectiveness of different AI approaches to determine the fish age (Moen et al., 2018; Ordoñez et al., 2022; Politikos et al., 2021; Sigurdardóttir et al., 2023; Moen et al., 2023). In response to the criticism regarding the lack of explainability of initial approaches, some studies even showed techniques to demystify the inner workings or mechanisms of AI predictions (Ordoñez et al., 2020; Martinsen et al., 2022). Lastly, other studies created web applications (Politikos et al., 2022) and frameworks (İşgüzar et al., 2024) that facilitate the wider useability of these modern approaches.

More recently, newer studies focused on ways to make the age reading process directly compatible with traditional ring counting. For instance, Bojesen et al. (2024) showed the use of generative models for creating dot-based annotations of the rings. Likewise, in own work, Cayetano et al. (2024b) showed the capabilities of U-Net (Ronneberger et al., 2015) and Mask R-CNN (He et al., 2017) to create another form of annotations within the rings from which age can be derived. As another step towards wider acceptance, the follow-up study by Cayetano et al. (2024a) packaged the methods into a web-based toolchain that enables easy-to-use access while at the same time, integrating advanced ML concepts that can be used to boost the performance further.

The importance of software tools to help in otolith analysis cannot be overstated. For instance, the ICES (2023) SmartDots have played a very important role for the age reading community especially during otolith workshops and exchanges. The tool provides an intuitive visual aid to the age readers in order to mark or annotate the rings forming the basis of their estimates that can be easily verified and studied to identify erroneous assessments (Pinto et al., 2018). This reduces the ambiguity of the process and lessens the subjective aspect which will eventually contribute to its standardization. Some other essential features are also included in the ICES SmartDots such as a measuring tool along with an image intensity visualization which facilitate efficient otolith analysis (ICES, 2023).

In this study, we present a multi-stage framework embedded within the existing web-based application we developed where both the AI methods presented can be utilized further, not only for the task of fish age reading but also as a toolkit for general otolith analysis. First, we show how the methods can be employed to perform morphological processing such as the detection of outer otolith shape or contour as well as the localization of the otolith core or nucleus. Next, we show how the annotation toolsets added into the existing web application can complement software tools such as ICES SmartDots by allowing the use of dot-based annotations for both the input and output of the software. Then, with these annotations, we provide a way to perform measurements of annuli distances in order to detect patterns of growth among different species in our datasets. This is in addition to the usual process in which the annotations are used to derive age

estimates. Lastly, we compile all the approaches in order to perform end-to-end analysis using a completely new set of unprocessed otolith images to highlight its effectiveness to perform the different tasks involved.

4.3   Methods

Multi-stage Framework Design and Implementation

A summary of the overall stages of the framework is depicted in Figure 4.1. It starts with otolith contour extraction which isolates the otolith of interest from its background. Next comes the otolith nucleus localization where the approximate area surrounding the core is highlighted by the appropriate method. Lastly, the third stage follows where the otolith rings/annuli are identified and marked. For this last part, there are two post-processing final steps that can be done: one in which the markings are counted to obtain the age information (Cayetano et al., 2024b) and another involving measuring the distances between the marked annuli to check for growth patterns which will be explored in this study.

The platform is coded in python and the codebase is available in Github (https://github.com/arjaycc/ai_otolith/). The application can either be deployed as a webserver where multiple machines can connect or can also be used as local software tool with the available standalone portable version (https://doi.org/10.5281/zenodo.10954471). For the webserver running at Thuenen Institute of Sea Fisheries, a docker setup is involved which is deployed on a Linux machine containing the libraries for Tensorflow-GPU/Keras (Abadi et al., 2015; Chollet et al., 2018). As the extension of the previous two studies, further details of the setup can be found elsewhere which will not be elaborated here.

Outer Contour Detection

In this section, we present the facility that allows contour detection using deep learning. This step is particularly important for three main reasons: 1) when age readers take otolith images, multiple otoliths are sometimes present in the same slide along with the one being captured; 2) to make AI training more robust, it is generally helpful to remove unreliable information such as the apparent sizes of the otoliths which can depend on the manner of image capture (i.e., readers can have their own custom magnification which may not be consistent); 3) detecting the otolith outer shape is the preliminary input of many studies involving stock discrimination using shape analysis (Stransky, 2014).

First, we need to determine which algorithm is better suited for this task. We utilize all the images of the Baltic Sea dataset used in the previous studies and conduct 3-fold cross validation to evaluate the performance of the two methods for detecting the otolith outer shape or contour. For the ground-truth labels, we utilize a semi-automated approach where we initially apply the watershed algorithm from scikit-learn (Pedregosa et al., 2011) on the images, in conjunction with the opencv library (Bradski, 2000). This algorithm is a well-known classical computer vision technique for background isolation. After this
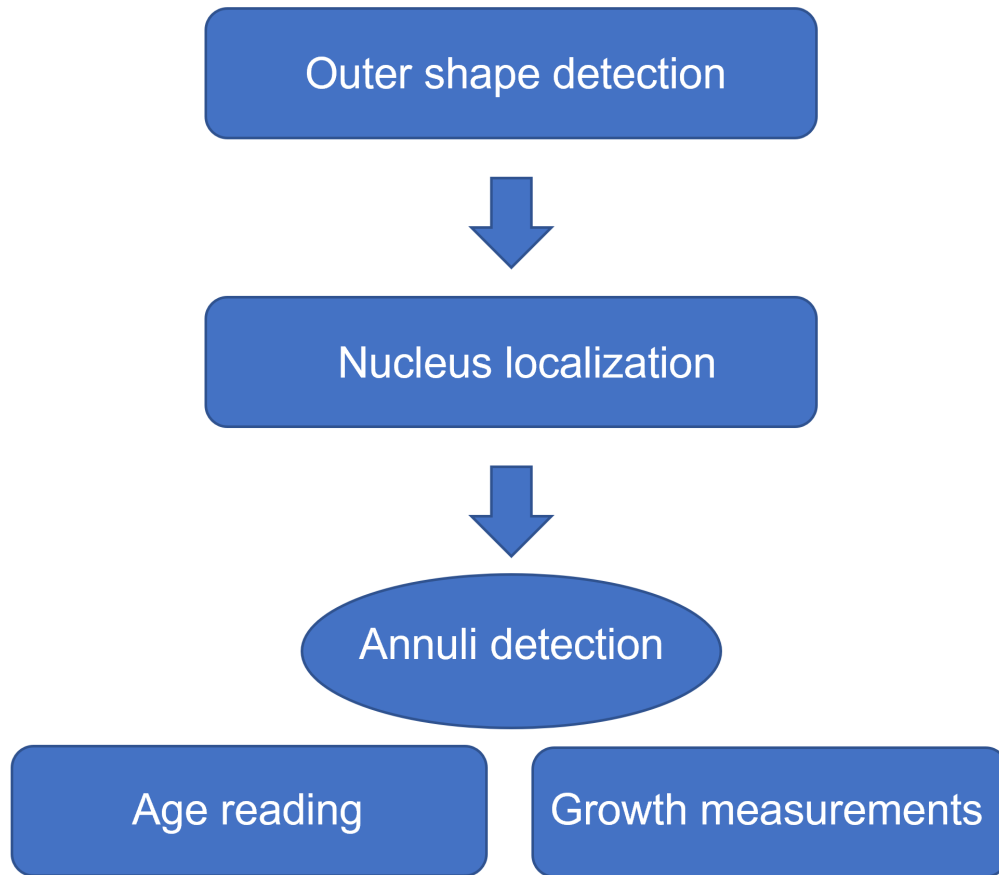
Figure 4.1: Overview of the multi-stage framework.

approach, there will be a set of proper contours for most of the images. Similar to the previous study (Cayetano et al., 2024b), in the case of erratic contours, we perform manual contour corrections with the use of an annotation toolkit known as VIA or Visual Geometry Group Image Annotation (Dutta and Zisserman, 2019) software.

As a way to simplify the use of VIA, we previously integrated it in our web application so that it can be called as needed and that its output can be saved following our file-naming and database structure. In addition, we also previously modified the codebase of this tool such that certain features like brush tool can be utilized as part of it.

To evaluate the models, we need to define certain metrics utilized in this study. In the context of image segmentation, where predictions are evaluated pixel-wise, we define four prediction categories namely true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). These are summarized in Table 4.1.

With these four values, we can compute the following metrics (Alouini, 2021; Huynh, 2023) and create a tally of the testing instances in which one algorithm is better than the other. These five metrics are popular evaluation methods for comparing and benchmarking algorithms that perform image segmentation.

Table 4.1: Different types of per-pixel prediction outcomes/categories (against the ground-truth labels) as the image is segmented by the AI-methods.

		Ground-Truth	
		Otolith Region	Background
AI Prediction	Otolith Region	TP	FP
(pixel-wise)	Background	FN	TN

$$Precision = \frac{TP}{(TP + FN)}$$

$$Recall = \frac{TP}{(TP + FP)}$$

$$Pixel\ Accuracy = \frac{(TP + TN)}{(TP + TN + FN + FP)}$$

$$F1\ Score = \frac{2TP}{(2TP + FP + FN)}$$

$$Jaccard\ Score/IOU = \frac{TP}{(TP + FP + FN)}$$

Nucleus Detection

For this stage, we also need to determine which of the two algorithms is best suited for the task. Likewise, we perform a 3-fold cross-validation experiment where we utilize all the images from the North Sea dataset used in the previous studies we conducted.

Similarly, the ground-truth labeling of the nucleus is done using the integrated VIA toolkit which is made even easier with the brush tool we implemented. Also, we perform the evaluation using the metrics mentioned in the previous section and create a tally comparing the two algorithms.

It is important to note that, for our purposes, we characterize the nucleus as the central region of the otoliths, inside the first annulus ring and under the sulcus. Figure 4.2 shows an illustration of this relative location along with some other important terminologies regarding otolith regions and axes.

Annuli Annotation

In own work (Cayetano et al., 2024b), we showed how the use of two AI methods, namely Mask R-CNN and U-Net, can be an effective approach to perform age reading and to
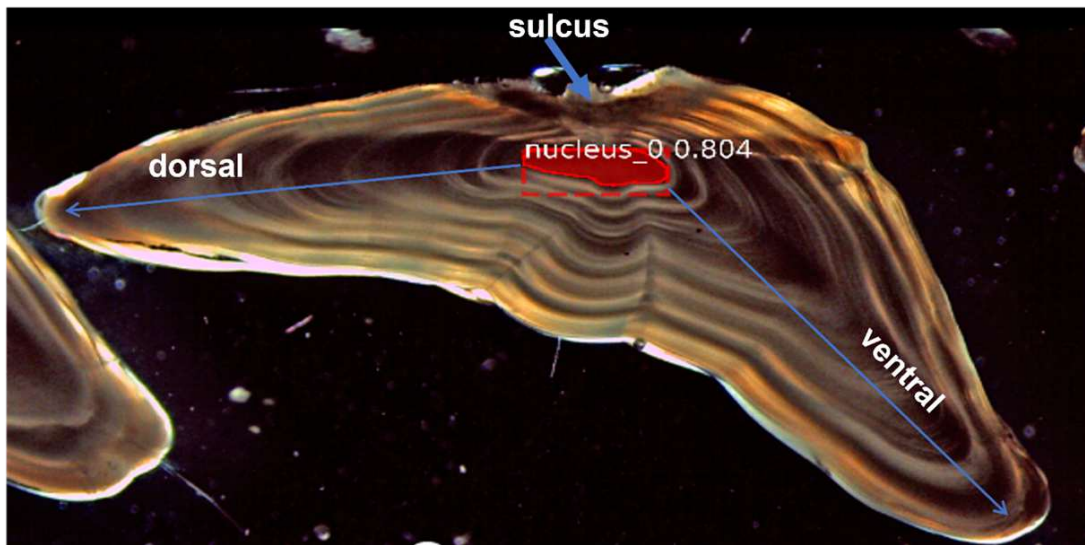
Figure 4.2: Some of the important regions/axes of the otolith relevant for this study along with a sample annotation of the nucleus location.

generate annotations of the rings. This makes the process explainable and can be directly verified by the age readers.

In this section, we briefly discussed how the old ground-truth annuli annotations were created in the previous studies and how the process differs to the one we explored here, which will be elaborated on the next section.

Similar to the other stages, we utilized the VIA toolkit to mark the portions of the annuli along the major reading axes. In addition, in certain subsets, we also employed AI-assisted annotation where we used existing models to create preliminary annotations which were manually corrected as needed. In this way, ground-truth labeling does not start from scratch and we only need to work on the erroneous initial annotations made by the AI method.

ICES SmartDots Compatibility of the New Method of Annotation

As mentioned earlier, ICES (2023) SmartDots is a popular software commonly used by age readers for annotating the otolith rings which is especially useful during age reading workshops and exchanges (Pinto et al., 2018). Using this tool, they are able to create dot annotations of the annuli as a visual aid for the manual age reading process. It is therefore necessary that the tools we are developing are compatible with this standard practice so that it can be easily adapted by the community. Also, with this approach, there is also a potential to reuse the data from otolith workshops and exchanges for training the AI-based methods.

In line with this, we further increase the useability of our own toolchain by incorporating a way to represent the annotations with dots instead of the wider default masks of the annuli that we initially implemented. There are two ways by which we supported the ICES SmartDots concept. For the first one, we simply take the centroid of the output/prediction

masks of the AI methods from which the central point can be eventually derived which is represented by a dot. For the second adaptation, we create a new annotation tool method where, instead of the usual ground-truth masks created from our previous studies, we allow the users to simply mark the rings using dots in VIA. Then, we perform background processing of these dots in order to create the appropriate masks needed by the algorithms. We investigate the feasibility of this approach by comparing the performance using this ground-truth labeling approach to the results from our previous study.

Measurements of Growth Patterns

Another advantage of age reading methods that generate ring annotations is that further downstream analyses can still be done. For instance, we perform a completely new experiment in which we measure the annuli distances in order to study the different growth patterns that may exist among the selected species and stocks in this study. For species-wise experiments, we focus on distinguishing the growth patterns between saithe (*Pollachius virens*) and cod (*Gadus morhua*). For the stock-wise analysis, we discriminate between the North Sea cod (abbrev. as N-Cod) and the Baltic Sea cod (abbrev. as B-Cod).

Multi-stage Otolith Analysis

We perform a final experiment using a completely new set of unprocessed images which were not used in any way in the previous studies. This new set is a randomly segregated subset of the total collected Baltic Sea images from surveys which we specifically allocate for this multi-stage experiment. In contrast with the previously used Baltic Sea dataset, the images are in their unprocessed and uncropped state which means that they are at the original dimension and magnification. This also means that there could be the presence of other surrounding otoliths and background artifacts which we aim to handle and overcome with the multi-stage framework.

Using the appropriate methods selected by the previous experiments, we subject this new set of images following a stage-by-stage procedure. First, we segment the otoliths and create the outer contours which will be used to automatically re-scale and crop the images in order to remove the surrounding artifacts. Then, we locate the otolith core or nucleus region and mark the central point accordingly. Lastly, the otolith rings are annotated which will serve as the input for two postprocessing steps namely, age estimation and annuli distance measurements.

4.4   Results

For the first stage, we elucidate the capabilities of the two algorithms to generate the outer shape or contours of the otoliths. Some examples of the masks generated by the two methods for this task are shown in Figure 4.3-a (Mask R-CNN) and Figure 4.3-b (U-Net).

To evaluate their performance, we obtained the values of the different metrics we discussed in the methods section namely pixel accuracy, recall, precision, F1 score and
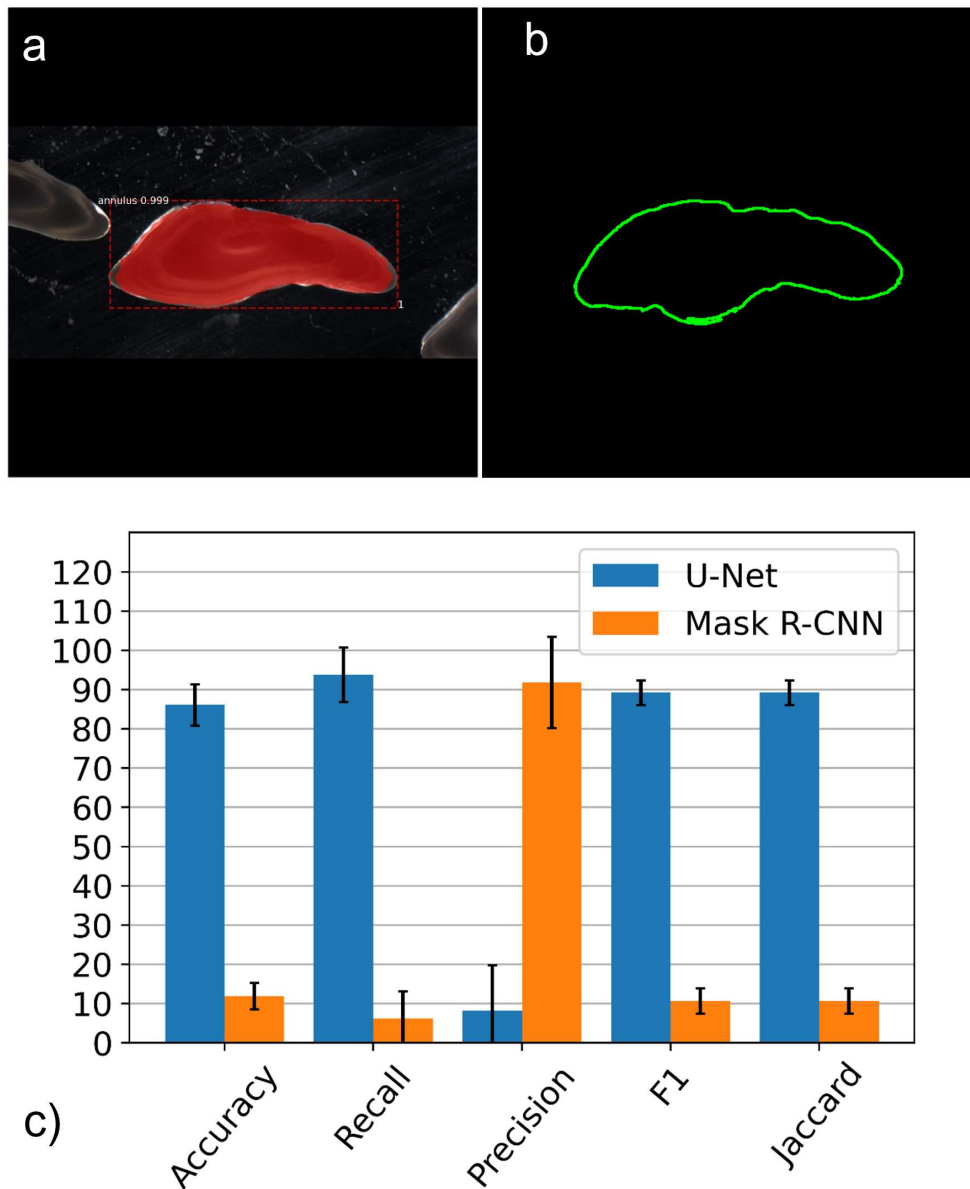
Figure 4.3: a) Mask R-CNN detection of the otolith. b) U-Net segmentation and extraction of the contour. c) Percentage of instances in which one method is better than the other and vice versa.

Jaccard score. To avoid any outlier test image from bringing down the overall values of the metrics, we simply tallied each test image instance where one method performs better than the other and vice versa. The resulting plot (in percentage of all test instances) is shown in Figure 4.3-c. In can be seen how the U-Net outperforms the Mask R-CNN for this task.

For the second stage, some examples of the output of the two algorithms are also shown in Figure 4.4-a (Mask R-CNN) and Figure 4.4-b (U-Net). Likewise, we perform evaluation

using the metrics and tallied the total image samples where one method surpasses the other and vice versa. This is shown in Figure 4.4-c. For this stage, it can be seen that Mask R-CNN has the advantage compared to the U-Net.
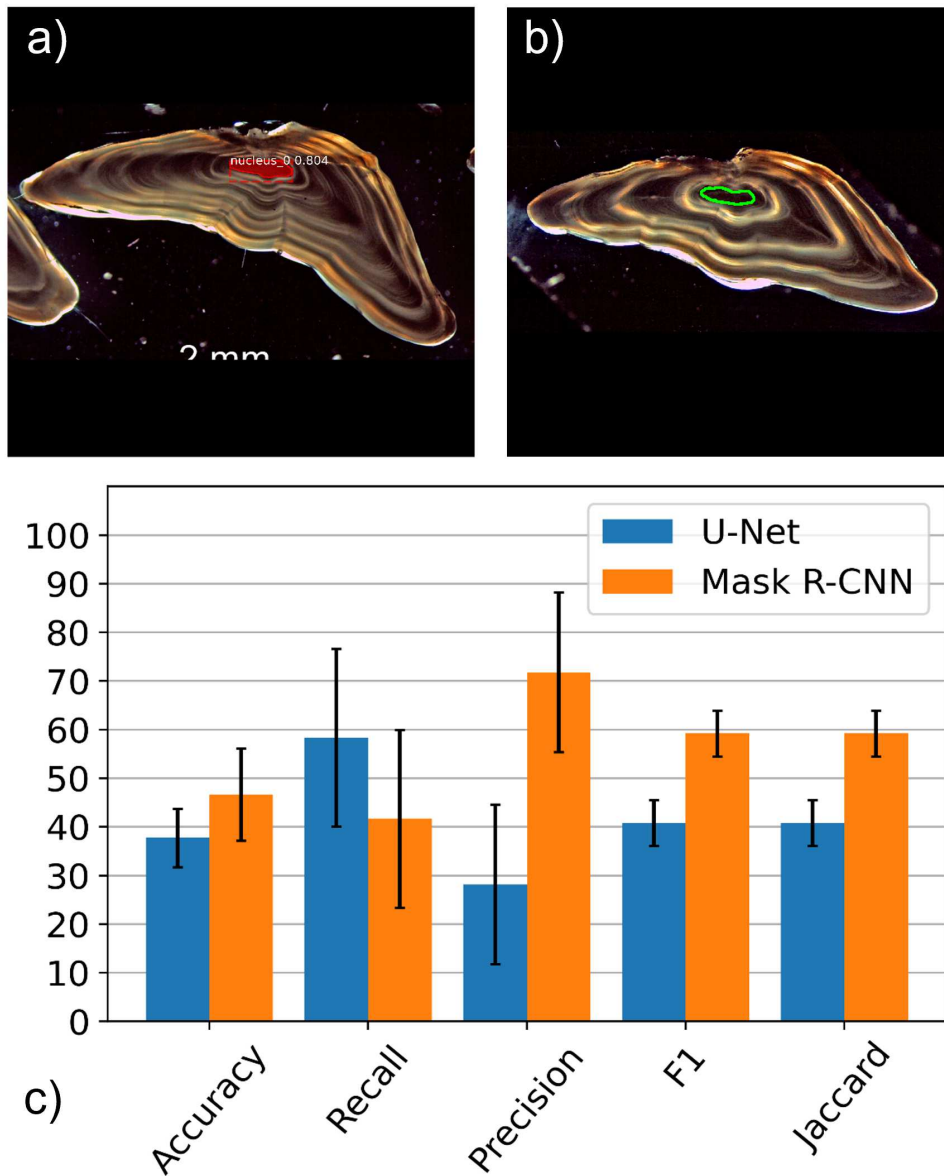
Figure 4.4: a) Mask R-CNN detection of the nucleus b) U-Net segmentation of the nucleus. c) Comparison of the two algorithms based on the metrics.

For the third stage, we first need to check the effectiveness of dot annotations as potential replacements of the wider and more complex mask annotations from previous study. Naturally, these dots cannot be immediately used as inputs of the algorithm due to the fact that each dot covers only one pixel within the entire image dimension. Hence,

we create a pre-processing step using a simple heuristic that converts these dots into wider representations which satisfy the requirements of the algorithms. Figure 4.5 shows the difference in performance with these dot-based starting annotations compared to the original ground-truth masks from previous studies we conducted.
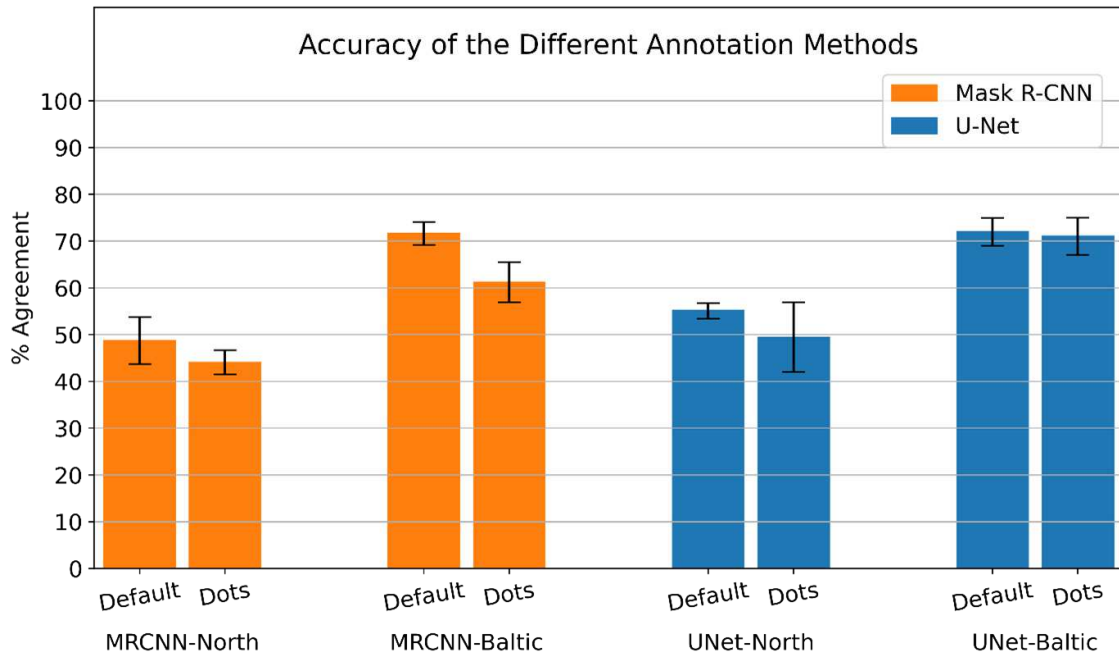
Figure 4.5: The accuracy of age estimates when using the different methods of annotations. Normal refers to the old method used in previous study (Cayetano et al., 2024b) while Dots annotations are the ones explored here.

After establishing the feasibility of the dot-based annotations, we then proceed to investigate whether there is indeed a trend regarding the growth of the annuli among the different species in our dataset. To do this, we reuse the dot annotations from the previous experiment serving as the central point representation of each ring having a single pixel coordinate. Using this point/dot, we measure the distances of the annuli focusing along a single reading axis, i.e., the axis which, most of the time, contains the complete set of rings. Figure 4.6 shows the resulting trend.

Finally, we conduct the multi-stage analysis by using a different set of unprocessed images which are all from the Baltic Sea dataset. The trained models from each stage performs their corresponding tasks which will be taken as inputs to the next stage and so on. Finally, the final performance evaluation is done by taking accuracy of the age predictions on this new set of images along with measuring the annuli distances to see if they align with the expected trend discussed above. The results are presented in Figure 4.7 and Table 4.2.

It can be seen that for the age reading task, the two methods managed to successfully attain satisfactory agreement with the manual age reading values as shown in Figure 4.7-a and Table 4.2. The box plots show useful information regarding the manner by which the predictions are distributed for each age class. The orange marker represents the median
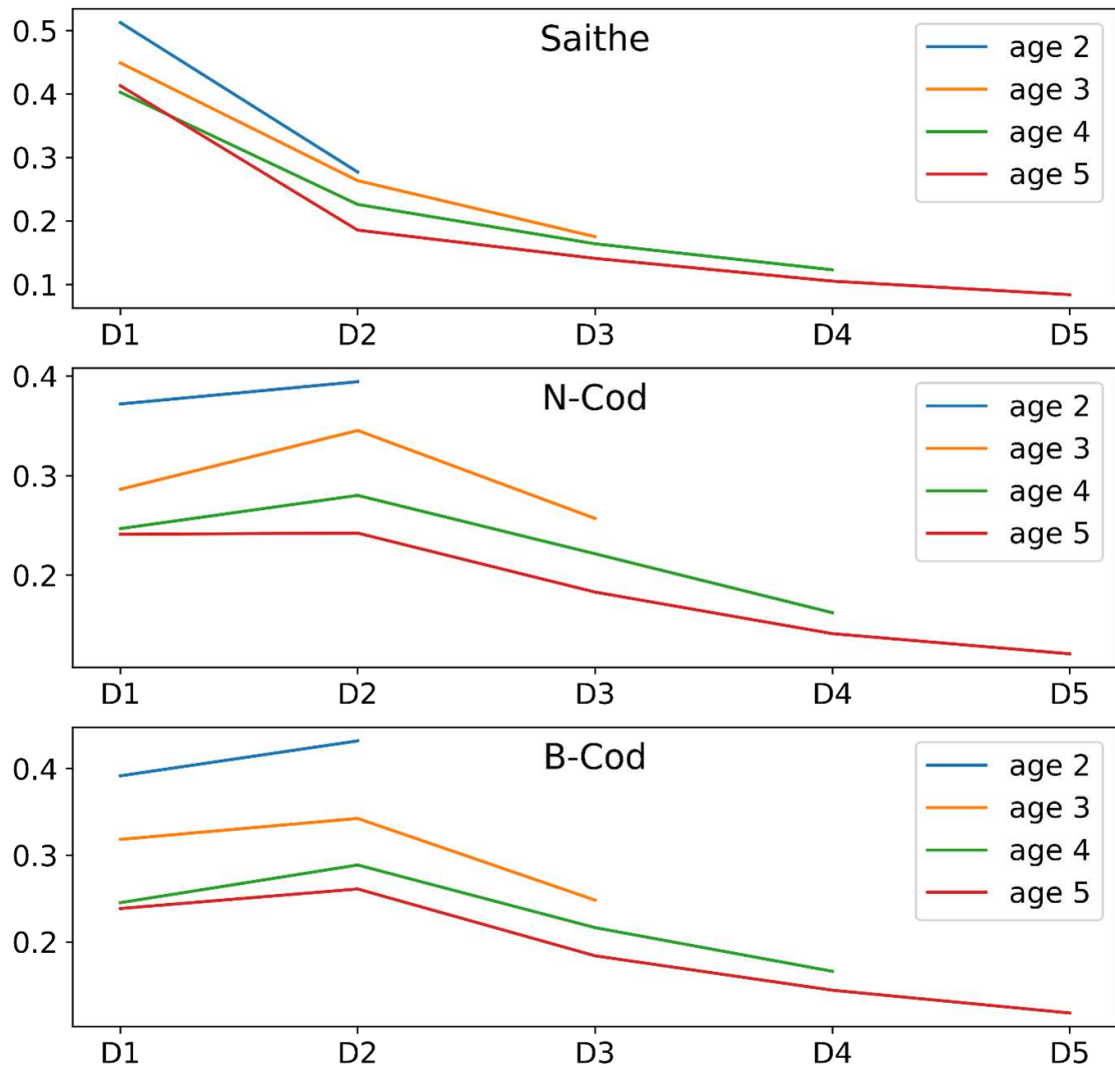
Figure 4.6: The annuli distances (% w.r.t. otolith radius) among the different species and stocks explored in this study indicating the differences in their growth patterns. For ring 2 to 5 (i.e., D2-D5), the distances are with respect to the preceding ring. For ring 1 (D1), the distance is measured from nucleus.

values of the predictions while the boxes indicate the predictions within the 1st and 3rd Quartile (Inter-Quartile Range or IQR). The apparent small sizes of these boxes indicate that the predictions of both methods are within a satisfactory narrow range. Moreover, in Table 4.2, the mean overall accuracy values in terms of percent agreement with the manual readings are also worth highlighting. Both Mask R-CNN and U-Net attained satisfactory agreement percentages with values equal to 70.8% and 68.6%, respectively. In addition, if we consider predictions which are off by just one unit, the percentages become more impressive reaching as high as 97.8% and 96.6%, respectively.

For the task of finding growth patterns via annuli distance measurements, the line plots in Figure 4.7-b also show some interesting trends. For both methods, a distinctive pattern
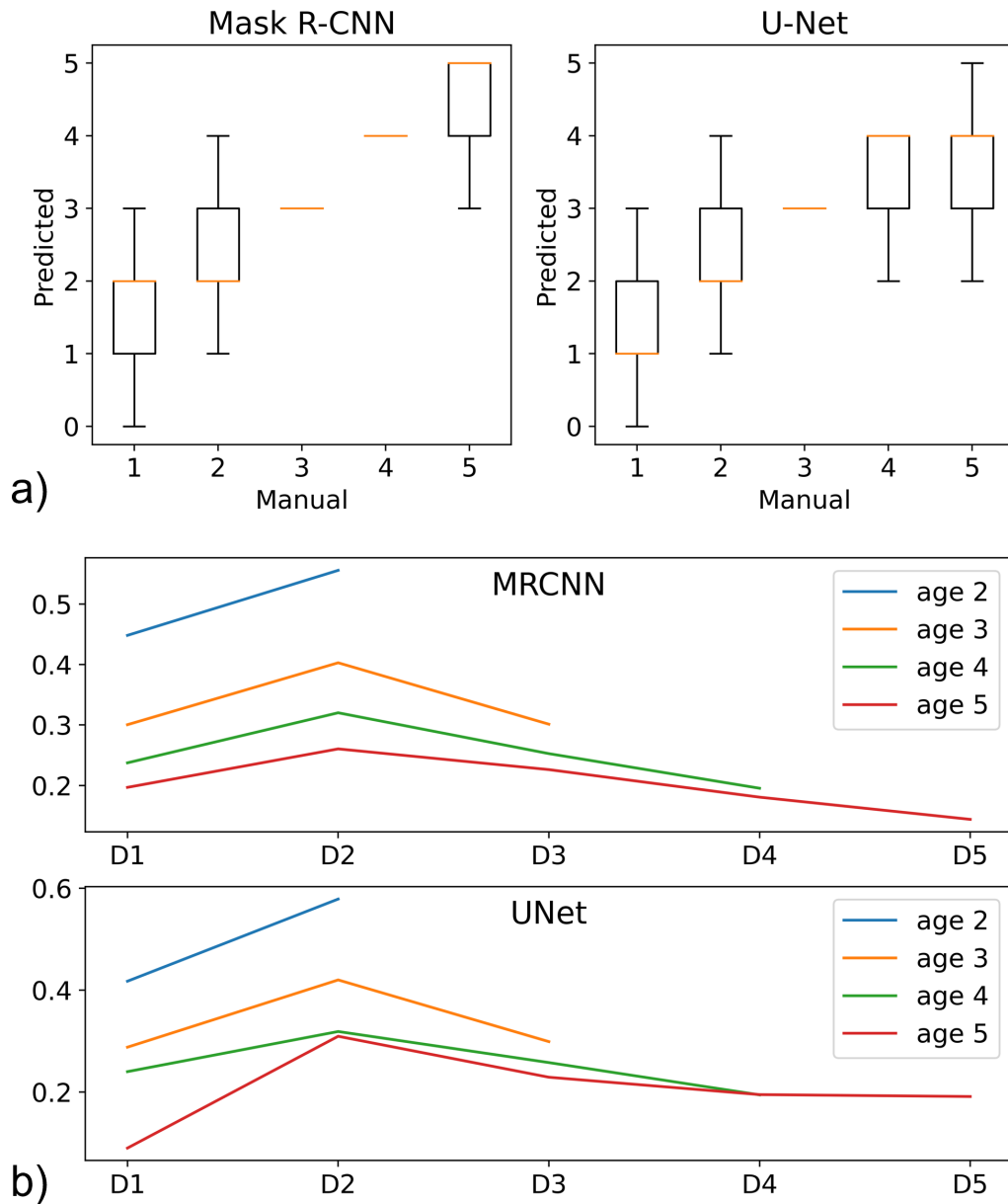
Figure 4.7: Result of the final stage for the multi-stage experiment involving a completely new batch of Baltic cod otolith images. a) Age-wise distribution of predictions of Mask R-CNN and U-Net b) Growth patterns measured using the distances of the centroid points of the predicted annulus rings generated by the two methods.

became apparent in which the distance initially increases (D1-D2) then goes downwards after the second annulus (D2-D5). This is the same growth trend that was observed when manual dot annotations were measured and shown in Figure 4.6. Hence, this means that the pattern remains consistent even with the purely automated AI-based approach and with an entire new set of Baltic cod images. Given that the pattern for North Sea saithe

(*Pollachius virens*) shown in Figure 4.6 is completely different, the AI-generated growth patterns for Baltic cod (*Gadus morhua*) can be theoretically used for distinguishing between the two species. This is particularly useful for applications involving analysis of the diet of sea birds and mammals where mixed sets of otoliths are obtained from their stomach contents or faeces samples (Enoksen et al., 2016; Polito et al., 2011).

Table 4.2: Overall accuracy of the age estimates generated by the two methods on a new set of Baltic cod images. In addition to the exact agreement, we also included the percentages when predictions are within 1 unit of error.

	Mask R-CNN	U-Net
Exact Agreement	70.8%	68.8%
Off-by-one	97.8%	96.6%

4.5   Discussion

In this study, we managed to show how the different methods are capable of performing other important tasks apart from their previously published effectiveness in age estimation. The results give a strong case of employing a multi-stage approach when it comes to otolith analysis especially when handling raw and unprocessed images. Such multi-stage approach has already been elucidated by the work of İşgüzar et al. (2024) when they managed to use Faster R-CNN to isolate the otoliths before employing Gaussian process regression model to estimate the age. Here we provide another perspective regarding the multi-stage approach where we used the same set of algorithms for both the preliminary stages (contour extraction and nucleus detection) and the main stage (age determination and annuli distance measurement).

For the first stage, the task of contour detection has been handled well by the two algorithms. In Figure 4.3, we showed that U-Net is generally better than Mask RCN for this task as far as the overall tally of effectiveness is concern. However, it might also be good to highlight at this point that there seems to be a negligible difference in values when the absolute values of the metrics are considered. Hence, using the Mask R-CNN for this stage could still work.

In the context of this study, the use of this contour extraction is limited to the image pre-processing steps such as background isolation as well as in cropping the otolith away from other surrounding otoliths and artifacts. However, for the purpose of stock discrimination using the otolith shape/contour (Stransky, 2014), having images of sectioned otoliths limits this possibility since the shape can be affected by the angle or manner of otolith cutting which cannot be guaranteed consistently. Nevertheless, the results still serve as valid showcasing of the capability of the algorithm for this type of task. The exploration of its usage in shape analysis will be an excellent future direction especially with the use of the proper type of otolith images (i.e., whole otoliths).

Similarly, for the second stage, we compare the two methods by relying on the overall tally of the instances where one method is better than the other. The results (Figure

4.4) show that Mask R-CNN outperforms the U-Net for the task of nucleus localization. However, in terms of the score values, the difference is actually minimal such that even the use of U-Net for this task will be sufficient.

As noted by Cao and Fablet (2006), detecting the nucleus is a pre-requisite for several downstream morphological analysis in which a central reference point is needed. In our study, the position of the nucleus is highly relevant to the other steps performed on the next stage of the framework (i.e., age estimation and measurement of annuli distances). In both cases, the central point of the reading axes has to be located first before the post-processing steps can be initiated. In Cayetano et al. (2024b), the nucleus coordinates were obtained using the heuristic proposed by Harbitz (2009) which involves approximation of the coordinate based on the ellipse fitted following the otolith shape. In this study, we obtain the nucleus coordinates without the need for other algorithms and we have shown that the same methods for annuli detection can also work on the task of localizing the nucleus or even other otolith regions (i.e., as a future direction for this study).

On the final stage, the creation of annuli annotations take place where there are two different post-processing steps that can be done. For the first one, the goal is to derive the age estimates which has been already highlighted in our previous works (Cayetano et al., 2024b). It was previously elucidated that both the methods perform this task effectively and that, with the help of advanced concepts such as transfer learning, ensemble learning and continual learning, the performance of the algorithms can be boosted further. In this study, we give another perspective to this task by showing that even with the use of simplified dot annotations, high accuracy of age prediction can still be attained. This was elucidated in Figure 4.5 where the difference in accuracy for dot-based and normal (wider) annotation methods is minimal. Therefore, this is another step towards making AI development and training even easier. Also, as the otolith workshops and exchanges already make use of these dot annotations via ICES (2023) SmartDots, there is also another possibility that the datasets from such events can be reused to train the relevant AI algorithms.

For the second post-processing in the final stage, the alternative objective is to measure the distances between the annuli to see if there are growth patterns that can be extracted. As discussed in the study by Denechaud et al. (2020), such patterns can serve as proxy for studying the fish growth which can reflect the underlying environmental conditions of the fish habitat. In this study, we utilize the annuli growth patterns to distinguish between species and stocks. It was observed that for discriminating between species (saithe vs cod), such patterns can indeed be useful and effective as shown by the very distinct trend in Figure 4.6. However, using it for stock discrimination will not be possible due to the similarity between the observed growth patterns for the North Sea cod and Baltic cod.

To summarize, we enumerate below the general contributions and findings of this study:

1. **Contour/Outer Shape Extraction**

   U-Net is excellent in 4 out of 5 evaluation metrics for the task of outer contour extraction which can be used to isolate the otolith from its background especially when handling images captured from a slide of otoliths adjacent to each other.

2. **Nucleus Detection**

   Mask R-CNN performs better on 4 out of 5 metrics for the task of nucleus localization which is a pre-requisite of the post-processing steps to be performed on the next stages of the framework.

3. **Annuli Growth Measurements**

   There is a noticeable trend when it comes to the patterns of annuli growth within the otolith which enables differentiation of species (i.e., saithe and cod) but not sufficient for discriminating between stocks of the same species (i.e., North Sea cod and Baltic Sea cod).

4. **ICES SmartDots Compatibility**

   It is highly feasible to use dots as initial ring annotation which can be easily converted to the proper masks needed by the AI methods allowing the possibility that annotation data from existing tools such ICES SmartDots can be reused.

5. **Multi-stage Age Reading**

   The use of multi-stage approach is highly effective in handling unprocessed and uncropped images and that, both the two methods are sufficient to utilize for all the stages starting from contour extraction, nucleus localization as well as annuli detection and segmentation.

4.6   Conclusion

The study is another step towards the continued pursuit to improve the overall process of otolith analysis and age reading while making the AI-based methods accessible to the age reading community. The multi-stage framework and platform we developed is designed to be compatible with existing tools such as the ICES SmartDots making it familiar and intuitive to use. In addition, the output of the AI methods we presented, namely U-Net and Mask R-CNN, is in the form which is compatible with the traditional ring counting methodology contributing greatly to its explainability and trustworthiness.

Aside from the web-based tool we developed, there is also a future direction or next stage in this endeavor where we aim to connect the AI-based toolchain directly into the imaging software for microscopy such that it can be initiated on-the-fly as the otolith image is being captured. In this way, it can easily assist the age reader regarding the imaging aspect in real-time contributing further to its useability.

Ultimately, as the use of these software toolchains becomes popular and widely adapted, there will more collection of training datasets (i.e., images with annotations) which will further increase the accuracy and effectiveness of the AI-based approaches for otolith age reading and analysis.

4.7 References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). Tensor-Flow: Large-scale machine learning on heterogeneous systems. *Software available from tensorflow.org*.

Alouini, Y. (2021). All segmentation metrics! https://yassinealouini.medium.com/all-segmentation-metrics-be65e0653529.

Bojesen, T., Denechaud, C., and Malde, K. (2024). Annotating otoliths with a deep generative model. *ICES Journal of Marine Science*, 81:55–65.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Campana, S. (1999). Chemistry and composition of fish otoliths: pathways, mechanisms and applications. *Marine Ecological Progress Series*, 188:263–297.

Campana, S. and Casselman, J. (1993). Stock discrimination using otolith shape analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 58:30–38.

Cao, F. and Fablet, R. (2006). Automatic morphological detection of otolith nucleus. *Pattern Recognition Letters*, 27:658–666.

Cayetano, A., Stransky, C., Birk, A., and Brey, T. (2024a). An Interactive AI-driven Platform for Fish Age Reading. *Manuscript submitted for publication*.

Cayetano, A., Stransky, C., Birk, A., and Brey, T. (2024b). Fish age reading using deep learning methods for object-detection and segmentation. *ICES Journal of Marine Science*, fsae020.

Chollet, F. et al. (2018). Keras. Github repository. *Available from https://github.com/fchollet/keras*.

Denechaud, C., Smoliński, S., Geffen, A., Godiksen, J., and Campana, S. (2020). A century of fish growth in relation to climate change, population dynamics and exploitation. *Global Change Biology*, 26:5661–5678.

Dutta, A. and Zisserman, A. (2019). The VIA Annotation Software for Images, Audio and Video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, New York, NY, USA. ACM.

Enoksen, S., Haug, T., Lindstrøm, U., and Nilssen, K. (2016). Recent summer diet of hooded *Cystophora cristata* and harp *Pagophilus groenlandicus* seals in the drift ice of the Greenland Sea. *Polar Biology*, 40:931–937.

Fisher, M. and Hunter, E. (2018). Digital imaging techniques in otolith data capture, analysis and interpretation. *Marine Ecology Progress Series*, 598.

İşgüzar, S., Türkoğlu, M., Ateşşahin, T., and Ömerhan Dürrani (2024). FishAgePredictioNet: A multi-stage fish age prediction framework based on segmentation, deep convolution network, and Gaussian process regression with otolith images. *Fisheries Research*, 271:106916.

Harbitz, A. (2009). A Generic Ad-Hoc Algorithm for Automatic Nucleus Detection from the Otolith Contour. *4th International Otolith Symposium*.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988.

Hüssy, K., Mosegaard, H., Albertsen, C., Nielsen, E. E., Hansen, J., and Eero, M. (2016). Evaluation of otolith shape as a tool for stock discrimination in marine fishes using Baltic Sea cod as a case study. *Fisheries Research*, 174:210–218.

Huynh, N. (2023). Understanding Evaluation Metrics in Segmentation. https://www.kaggle.com/code/nghihuynh/understanding-evaluation-metrics-in-segmentation.

ICES (2023). SmartDots manual: How to annotate. Version 06. ICES User Handbooks. https://doi.org/10.17895/ices.pub.22810604.

Martinsen, I., Harbitz, A., and Bianchi, F. (2022). Age prediction by deep learning applied to Greenland halibut (*Reinhardtius hippoglossoides*) otolith images. *PLOS ONE*, 17:e0277244.

Moen, E., Handegard, N., Allken, V., Albert, O., Harbitz, A., and Malde, K. (2018). Automatic interpretation of otoliths using deep learning. *PLOS ONE*, 13:e0204713.

Moen, E., Vabø, R., Smoliński, S., Denechaud, C., Handegard, N., and Malde, K. (2023). Age interpretation of cod otoliths using deep learning. *Ecological Informatics*, 78.

Ordoñez, A., Eikvil, L., Salberg, A., Harbitz, A., and Elvarsson, B. (2022). Automatic fish age determination across different otolith image labs using domain adaptation. *Fishes*, 7:71.

Ordoñez, A., Eikvil, L., Salberg, A., Harbitz, A., Murray, S., and Kampffmeyer, M. (2020). Explaining decisions of deep neural networks used for fish age prediction. *PLOS ONE*, 15:e.0235013.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12:2825–30.

Pinto, C., Davies, J. O., Coster, K. D., Milar, C., Allegaert, W., Osypchuk, A., Catarino, R., Godiksen, J., Bekaert, K., Holdsworth, N., and Torreele, E. (2018). SmartDots: an online international platform for age reading workshops and calibrations. *International Conference on Marine Data and Information Systems, Barcelona, Spain*, page 148.

Politikos, D., Petasis, G., Chatzispyrou, A., Mytilineou, C., and Anastasopoulou, A. (2021). Automating fish age estimation combining otolith images and deep learning: The role of multitask learning. *Fisheries Research*, 242:106033.

Politikos, D., Sykiniotis, N., Petasis, G., Dedousis, P., Ordoñez, A., Vabø, R., et al. (2022). DeepOtolith v1.0: An Open-Source AI Platform for Automating Fish Age Reading from Otolith or Scale Images. *Fishes*, 7:121.

Polito, M., Trivelpiece, W., Karnovsky, N., Ng, E., Patterson, W., and Emslie, S. (2011). Integrating Stomach Content and Stable Isotope Analyses to Quantify the Diets of Pygoscelid Penguins. *PLoS ONE*, 6:e26642.

Proctor, C., Robertson, S., Jatmiko, I., and Clear, N. (2021). An introductory manual to fish ageing using otoliths. *ISBN 978-0-646-83421-4*, page 41.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer. (available on arXiv:1505.04597 [cs.CV]).

Sigurdardóttir, A., Sverrisson, P., Jónsdóttir, A., Gudjónsdóttir, M., Elvarsson, B., and Einarsson, H. (2023). Otolith age determination with a simple computer vision based few-shot learning method. *Ecological Informatics*, 76:102046.

Stock, M., Nguyen, B., Courtens, W., Verstraete, H., Stienen, E., and De Baets, B. (2021). Otolith identification using a deep hierarchical classification model. *Computers and Electronics in Agriculture*, 180:105883.

Stransky, C. (2005). Geographic variation of golden redfish (*Sebastes marinus*) and deep-sea redfish (s. mentella) in the north atlantic based on otolith shape analysis. *ICES Journal of Marine Science*, 62:1691–1698.

Stransky, C. (2014). Chapter seven - morphometric outlines. In Cadrin, S. X., Kerr, L. A., and Mariani, S., editors, *Stock Identification Methods (Second Edition)*, pages 129–140. Academic Press, San Diego, second edition edition.

Stransky, C., Baumann, H., Fevolden, S., Harbitz, A., Høie, H., Nedreaas, K., Salberg, A., and Skarstei, T. (2008). Separation of Norwegian coastal cod and Northeast Arctic cod by outer otolith shape analysis. *Fisheries Research*, 90:26–35.

VanderKooy, S., Carroll, J., Elzey, S., Gilmore, J., and Kipp, J. (2020). A practical handbook for determining the ages of Gulf of Mexico and Atlantic Coast fishes. *Gulf States Marine Fisheries Commission Publication*, 300.

Chapter 5

Synthesis

5.1 Summary and Overview

This section summarizes and recapitulates the papers included as different chapters of the thesis while the next sections deal with the specific ways in which the research questions were resolved by the relevant results in each paper.

In the first paper (Chapter 2), the fish age reading capabilities of the two algorithms, Mask R-CNN and U-Net, were demonstrated and shown to be effective. Several benchmarking tests were done in order to show how the algorithms compare against the state-of-the-art CNN-regression formulation from Moen et al. (2018) as well as against the classical image analysis approach. Apart from excellent overall accuracy, the methods were also shown to be robust and adaptive even with a heavily modified or completely new dataset under a new domain. More importantly, the methods were shown to be highly explainable as they are compatible with the traditional age reading process involving annuli counts.

For the second paper (Chapter 3), advanced machine learning techniques namely transfer learning, ensemble learning and continual learning, were implemented on top of the two methods and were demonstrated to further boost their accuracy and even their ability to handle changing datasets. In addition, a web-based platform housing the algorithms was developed with the goal of making the adoption or wider usage easier for the age reading community. The web-based application is aimed to serve as an AI-assistant where readers can use it for image analysis with its proven capacity to automatically mark the growth rings, serving as visual guides or even providing a second opinion regarding some grey areas involving the identification of growth rings.

Lastly, in the third paper (Chapter 4), it has been shown that both methods can also be used effectively for general otolith analysis which is not only limited to fish age reading. A multi-stage framework was presented such that both methods are utilized in three different stages namely contour extraction, nucleus localization and annuli counting and distance measurements.

5.2   Fish Age Reading Effectiveness

- How can the advances in the field of artificial intelligence be used to effectively perform automated fish age reading and annotation of the annuli or growth rings?

This research question deals with demonstrating the overall effectiveness of the two approaches employed in the study on the task of fish age reading. All three papers managed to answer this question and the next paragraphs present the relevant results and findings that elaborate in detail this capability.

First, in Chapter 2, Figures 2.7 to 2.11 showed how it is accurate, robust, and adaptive to different datasets, in addition to its advantage regarding the explainability aspect. Figure 2.7 compared the performance of the methods with the widely referenced published method (Moen et al., 2018) along with the typical formulation of classical image processing approach. In addition, the coefficient of variation (CV) of the methods was compared to some of the results of some age reading workshops (Table 2.3). Based on these, the overall performance of the readings can be considered satisfactory and within the range expected for human error. In fact, in most workshops, the discrepancy among the readers could even be greater as evident in the larger coefficient of variation reported indicating the subjective limitations of the traditional approach.

In Chapter 3, the base methods were extended further by using even more advanced concepts in order to boost the performance and accuracy of the AI models. Figures 3.6, 3.8 and 3.10 showed the effectiveness of transfer learning, ensemble learning and continual learning to handle domain shift scenarios. The results indicate that there could still be further improvements in the overall performance by employing these advanced techniques on the base models.

Lastly, in Chapter 4, fish age reading was again performed effectively which is elucidated using a box-plot in Figure 4.7. In this study, the context is under a multi-stage framework where images are coming from their raw states that were sequentially processed and taken into the final state where they can be ready for age reading. In practice, such multi-stage framework will be handy in actual applications as the workflow normally starts from the basic raw images without the pre-processing.

5.3   General Acceptance and AI Assistant

- How can the AI based approach be designed such that it can easily gain the trust and wide acceptance of the community?

This research question is another important consideration that was heavily taken into account in this PhD thesis, particularly during the conceptualization phase and the selection of the appropriate algorithm to use. It is not enough to have a working and accurate method for automated fish age reading. It must also be easy to explain and be as close as possible to the traditional approach (i.e. ring counting) in order to easily build the trust and acceptance of the age reading community.

Initially, one of the earlier approaches designed for the thesis is a hybrid of the modern and classical machine learning. As mentioned in the General Introduction, one advantage of the classical machine learning is the use of feature engineering (Mahony et al., 2019; Dzieżyc et al., 2020) where one has a degree of control on the feature set to use making it inherently partially explainable. Hence, the initial thought process is to utilize this concept of feature engineering but use deep learning algorithms instead of classical ones such as SVMs or fully-connected artificial neural networks (ANNs). Figure 5.1 shows the basic idea of the feature engineering we initially implemented.
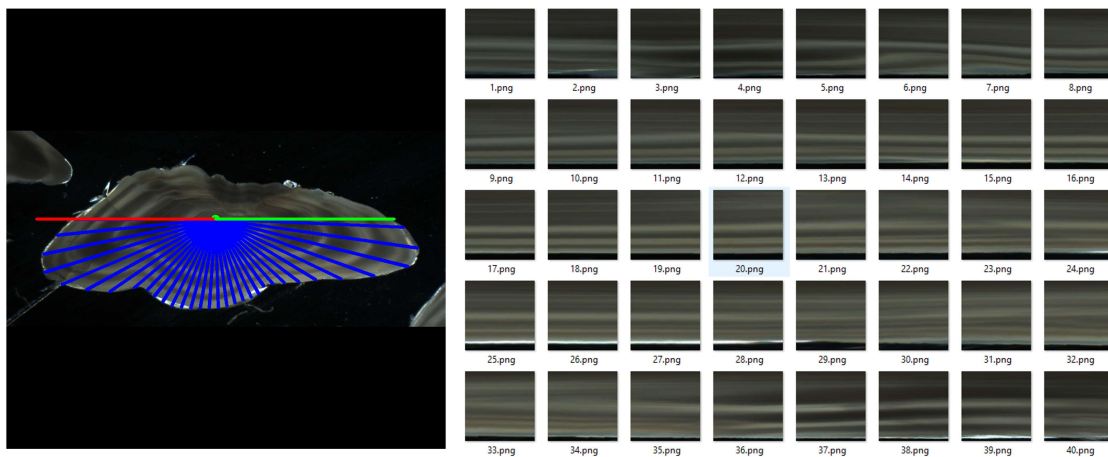
Figure 5.1: Initial experiments involving feature engineering of otolith radial strips to be used by CNN.

In the figure, the alternating intensity bands of the growth rings are used as inputs to the CNN which will be used to perform regression to estimate the fish age. This method, however, was eventually discarded as there can still be difficulties explaining the predictions based from the input strips and that it was not able to attain convincing performance.

As discussed in Chapter 2, the state-of-the-art approach by Moen et al. (2018) also suffered from this problem regarding explainability. Their model can only give results in the form of numeric estimates hence will not be able to provide clarifications on the predictions making it hard to verify. Also, despite the further extensions (Ordoñez et al., 2020; Martinsen et al., 2022) to give light to the mechanism of this typical CNN formulation, more questions and doubts are still left to be answered. For instance, the Layer-wise Relevance Propagation (LRP) method (Alber et al., 2018) they used gave potential explanation in the form of heatmaps (Ordoñez et al., 2020) but it can be seen that the focal regions shown are not the relevant regions considered by age readers during traditional age reading.

In Bojesen et al. (2024), it was implied that there is a general lack of trust and acceptance in the community for any approach that is not compatible with the traditional ring counting methods. The ultimate goal, therefore, is to map the approaches in such a way that it can be directly attributed to the existing knowledge about otolith structure and ring patterns. In their study, they were indeed able to develop a method that annotates and

counts the rings which have been verified by age readers within their group.

Similar to the method from Bojesen et al. (2024), this thesis project formulated the problem in a way that has direct compatibility with the traditional approach. The ring patterns give immediate clues on how the age readings are derived making it explainable and verifiable, which in turn, increases the trust and general acceptance by the age reading community.

Moreover, in Chapter 3, another innovation was presented for easier adoption of the AI-based approaches. We created a web-application that gives importance to the expertise of the age readers by providing a facility for them to also take part in the development of the AI models. That is, the website makes it easier for them to get involved in the training of the AI models via the interactive interface that takes their inputs and the corrections they make on the current model predictions in order to conduct further iterations of training and model development.

As discussed in Fisher and Hunter (2018), Mahe et al. (2009) as well as in Moen et al. (2018), the bulk of the processing in any age reading workflow is on the preparation of the otoliths. Therefore, human age readers will still be involved in the process. In the last step of the workflow, however, AI could definitely provide some assistance especially when dealing with species having ambiguous protocols.

5.4   Effectiveness on other domains/data sources

- How will the approach hold when it comes to various datasets?

This research question concerns the usual problem observed in deep learning studies wherein certain AI models are only good for the specific domain they are exposed to. This behavior of AI systems can be attributed to the problem of domain adaptation (Ben-David et al., 2010; Ordoñez et al., 2022). The effect can be manifested even with just simple change in the image characteristics e.g., change in image tone or general lighting.

In this study, it was demonstrated that both approaches are less susceptible to the problem of domain shift and that there are effective ways to further handle the changes on the image domains. Specifically, in the first paper (Chapter 2), both methods were shown to have less performance deterioration when tested against data sources or domains which are different from the initial domain they were trained on. Moreover, even without higher age values in the training set, the AI methods are observed to be capable of extrapolating the missing age groups, a feature that was completely missing from the standard formulation adopted from the works of Moen et al. (2018).

In the second paper (Chapter 3), this facility was elaborated further and there are several ways provided to handle the potential dataset changes. First, it was shown that using otolith-derived base weights for transfer learning, the performance of the AI models on a new dataset is better compared to the ones using generic object weights such as VGG or Coco or even those Imagenet weights such as ResNet, Inception and Xception, which is a typical design of many AI-based otolith studies (Moen et al., 2018; Politikos et al., 2021; Ordoñez et al., 2020; Martinsen et al., 2022; Moen et al., 2023). Secondly, with the use of

continual learning, we can further effectively handle a change in dataset. With the use of basic rehearsal approaches, the training of AI methods can be done in a way that prevents the "catastrophic forgetting" of the old/original dataset while undergoing training on a new dataset.

In the study of Ordoñez et al. (2022), they tried a simple experiment to test the possible effects of this domain shift on the predictions of existing AI models. In their study, they showed how the simple CNN formulation (i.e., classification) was not able to adapt properly with the different nature of the other dataset they used. Without modifications, both Mask R-CNN and U-Net methods are already adaptive and hence are less affected by the problem of changing domains or data sources.

From a completely different perspective, the creation of the web-based application can also be considered an alternative indirect solution to the problem of ever-changing datasets. With the web application, there will be a repertoire of trained models which are collected and stored for easier retrieval. When one of the models is familiar with the current dataset being analyzed (i.e. it was previously trained on that stock or species), one can easily reload and use it accordingly for that particular context. Furthermore, the selection of appropriate model to reload can be done automatically given certain datasets without needing user inputs. Figure 5.2 shows the web-based retrieval of specific models which will facilitate the processing and analysis of different datasets.
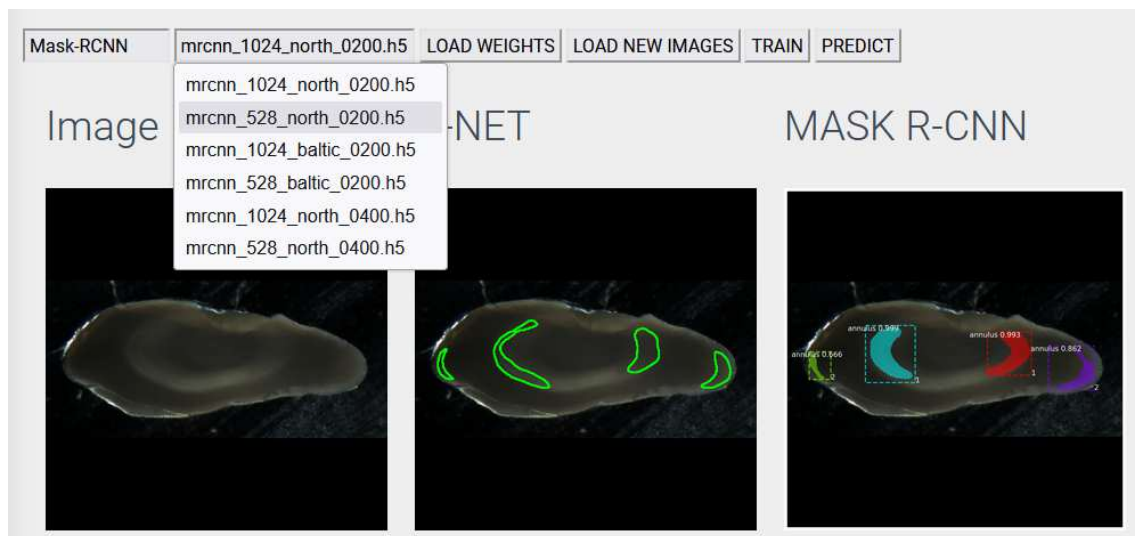
Figure 5.2: Retrieval of different models via web-based interface.

5.5   General Otolith Analysis

- How will it hold against different tasks involved in general otolith analysis?

This is an even broader objective that encompasses not only the task of age reading but also other tasks such as contour detection, nucleus localization and other necessary

processing performed as part of general otolith analysis. In this PhD thesis, this objective has been addressed in several ways.

The first non-age-reading task performed successfully by both algorithms is the task of contour detection. Its most basic use is on removing the background of the otolith (background subtraction) depending on the requirement of an experiment. More importantly, obtaining the outer shape or contour of an otolith is a pre-requisite of another related field which is otolith shape analysis (Campana and Casselman, 1993; Stransky et al., 2008; Hüssy et al., 2016). A lot of studies have been conducted in this area such as those that are able to classify or discriminate the fish species or even fish stocks based on the otolith shape (Stransky, 2014).

In most cases, contour extraction is done using some proprietary software and with dedicated staff performing image analysis. In this study, it was shown that there is a potential for this step to be done automatically and even on a per-batch basis. In this way, the tedious manual process can be skipped and the results can be immediately be generated for the whole set containing large quantities of otolith images.

Similarly, for the task of nucleus localization, the two methods are also capable of correctly identifying the location of otolith core. Identifying the central region of the otolith is a very important initial step in many image-based otolith analyses (Harbitz, 2009; Cao and Fablet, 2006). For instance, to measure the annuli radius, one needs to start measuring from a central reference point (i.e., due to the irregular shape of otoliths, taking the half of the diameter will not be accurate). Moreover, for the image processing-based automated approaches, the central region will dictate the starting point of any intensity-based measurements used for peak-counting. Alternatively, one can also localize the sulcus instead of the nucleus and based on its geometry, derive the location of the core. This is depicted in Figure 5.3-a.
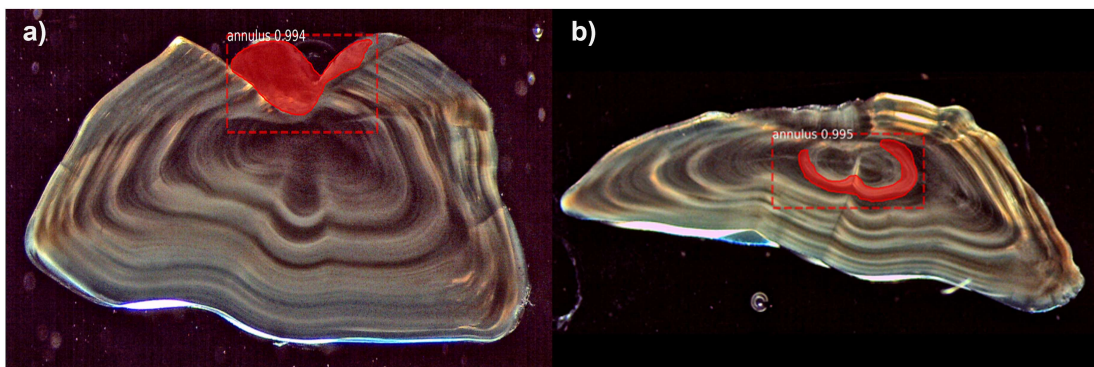
Figure 5.3: a) A sample AI annotation of the sulcus region b) A specialized AI model that detects only the first-annulus.

The flexibility of the ground-truth labels for training allows the methods to handle many task variations. For instance, one might be curious to know the effectiveness of using central reading axes, or even "full" ring annotations which covers as much portion of annuli as possible. In some cases, it is also beneficial to mark the central axis or even just the first annulus (i.e., the ring that often causes issues due to occurrence of false or double

rings). An example of this highly specific or special-purpose detection task is shown in Figure 5.3-b.

In other words, the users are free to explore and try out different labels that they think would be relevant for the particular fish species that they are currently working on. All of these give a strong case that the methods used in the study have the capacity to handle many different tasks.

5.6   Other Important Implications

- What are other important advantages of the AI-based methods and implications of the study that can facilitate the adoption of AI for routine age reading tasks?

Avenue for Collaboration

The study aims to promote further the collaboration among the different institutes in order to tackle the problem of fish age reading. The exchanges and workshops are already an existing manifestation of the willingness of the community to work together to improve the process. In this "big data" era, collaboration is all the more relevant and beneficial in so many ways. Firstly, the datasets collected by various institutions can be used to further enhance the training of the different AI models. Secondly, the models can also be shared to everyone in the community so that concepts such as transfer learning and continual learning can be utilized whenever applicable. Lastly, sharing of knowledge and expertise is essential as the progress in AI accelerates at a pace where not everybody can keep up right away.

With the findings in this study, particularly in Chapter 3, the use of an otolith base model for transfer learning was found to boost the accuracy further, even when used on a completely different dataset. Hence, collaboration will be a big advantage to properly conduct age reading studies with higher accuracy, especially with the useful concept of transfer learning, where one does not have to start from scratch (i.e., there is no need to re-invent the wheel).

In addition, the idea of creating a web-based application is to eventually provide wide access to the different models developed and shared by the community. Of course, certain licensing and copyright should be clarified first and there should be a collaborative effort to maintain the server, in order for this idea to work. At the moment, the current webserver is hosted with limited access within the Thünen Institute of Sea Fisheries but it is readily deployable for public access if granted the permission and the resources.

Compatibility with ICES SmartDots

The introduction of the ICES SmartDots tool (Pinto et al., 2018; ICES, 2023) has been a game-changer when it comes to conducting the otolith workshops and exchanges. Since then, the age reading community has become familiar with its use and the different functionalities it can offer. It is therefore ideal that for any new software tool developed

for otolith analysis, it should be somehow compatible with the ICES SmartDots platform. In this way, the immediate usage and adoption of the new toolkit can be facilitated.

Hence, for this thesis, the compatibility to the ICES SmartDots platform has been one of the most important design considerations. As the ability to mark the rings using dots has been the central feature of ICES SmartDots, the web-based application was conceptualized and developed in a way that can take advantage of this facility.

In Chapter 4, the effectiveness of this compatibility consideration has been carefully elucidated. As input, the dots can be used to create the ground-truth labels which will be used to train the AI models. As output, the predicted otolith masks can be converted to dots from which certain typical processes (e.g., distance measurements) can be performed easily. At this stage, the compatibility of the webapp to the ICES SmartDots is still on the prototype phase. It is hope that, as future direction, the connection between the web-based application developed in this thesis project and the ICES SmartDots will be seamless and that they can interact with each other smoothly.

It would really be helpful for AI developers to be able to directly use the annotations from ICES SmartDots in order to improve their models further and which can also be shared with everyone. The ultimate goal is for ICES SmartDots to be the go-to software for otolith analysis and other relevant biological parameters. As such, it should be updated with all the latest progress in AI by allowing contributions from AI developers directly which can facilitate further the exchange of data (images + annotations), models and even methods between the different participating institutions. In this way, the progress in AI-based otolith analysis can be greatly accelerated and will eventually be at par with the sophisticated technologies applied on other fields such as medicine or even that of natural language.

Downstream Applications

One major advantage of a method that generates the mask of annual rings is that several downstream applications can be conducted using the model outputs. For instance, we showed in Chapter 4 that there can be some useful patterns embedded on the growth of annuli that can be used to distinguish between fish species. Such distance measurements are even applicable in ecological and environmental studies. For instance, in the study by Denechaud et al. (2020), it was observed that the growth of the rings can be used as proxy for studying the growth of fish given certain environmental conditions. Hence, the ability to mark the annulus rings and measure their distance will be very handy for this purpose. In other studies, the otolith has been found to record certain biochronologies (Smoliński and Gutkowska, 2024; Campana, 2001). In fact, one of the age validation methods make use of the occurrence of these certain events which will be indirectly manifested on the annuli or growth rings (Campana, 2001).

In addition, from the multi-stage perspective, having the capability to detect and measure/quantify any part of the otolith will also be valuable for data archiving efforts where the images are stored is a consistent manner for easier documentations. Figure 5.4 is an example of multiple tasks done in one-go which can be useful in some miscellaneous purposes. For instance, using this capacity, an AI system can be programmed to auto-

matically detect the orientation of otolith images and perform the necessary adjustments before saving and archiving the image. Aside from that, this will also allow the removal of certain imaging artifacts like air bubbles and other surrounding otoliths or even purposely zoom in-and-out in order to make the view uniform for easier analysis.
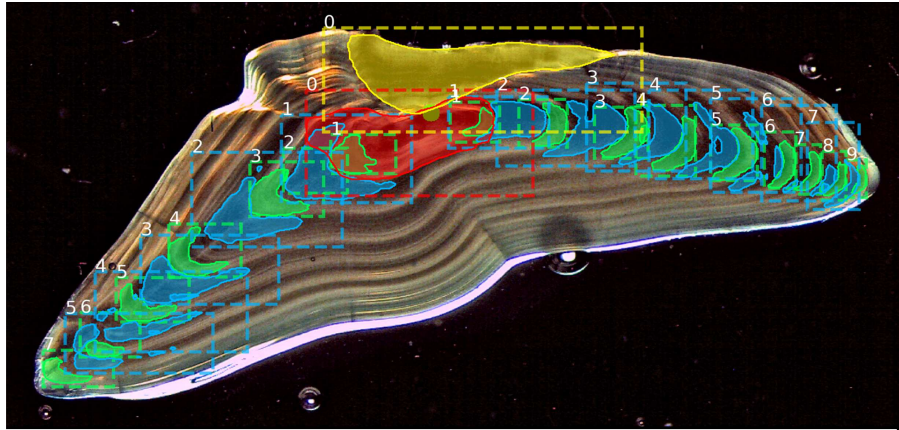
Figure 5.4: Multiple different tasks for different applications.

Lastly, there is an increasing number of studies making use of annuli measurements to perform certain categorizations. For instance, in the study of McQueen et al. (2018), annuli diameters were used as a solution to the problem of detecting lower age fish groups. In their study, they performed measurements on validated cod otoliths and elucidated that the annuli diameters can be used to distinguish juvenile cod images. This is another area where the automatic annuli detection can be of great use especially for bulk processing of these types of images.

5.7   Limitations and Future Outlook

Here the limitations of the study and common errors encountered are presented along with the potential solutions which are beyond the scope of the thesis. Hence, these further steps are listed as future directions in which the project can be continued forward.

Margin Errors and False or Double Rings

Upon closer inspection of the most common types of errors by the algorithms, it became immediately apparent that a great part of it has something to do with the marginal edge as illustrated in Figure 5.5. That is, if there is an indication of an annulus at the outer edge of the otolith, then it is automatically counted by the AI models. In theory, this is the desired behavior as the number of annuli corresponds roughly to the age of a fish. However, certain guidelines that have been formulated through the decades worth of validation efforts have suggested that in some cases, it is not accurate to count the last visible ring on the otolith due to the findings regarding marginal annulus (Proctor et al.,

2021). There is a threshold annulus width to consider and there is also a need to take into account the date of capture (Proctor et al., 2021; McQueen et al., 2018). Figure 5.6 summarizes this margin rule.
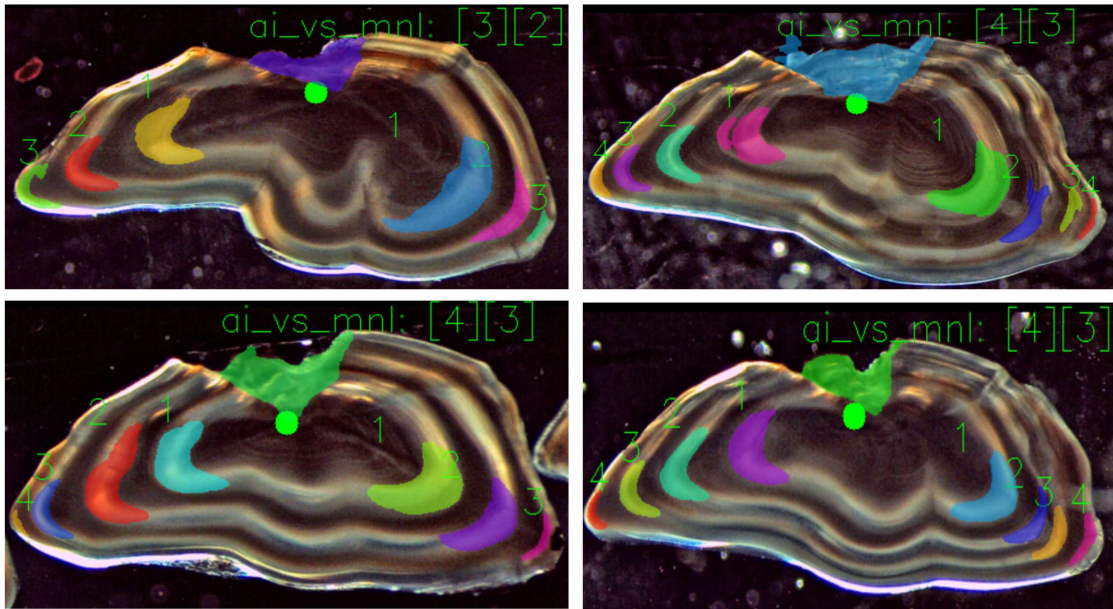
Figure 5.5: Common error involving the marginal annuli along the otolith edge.
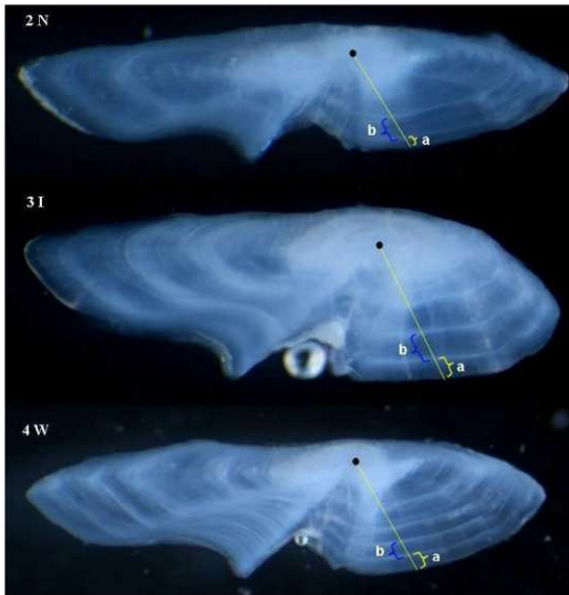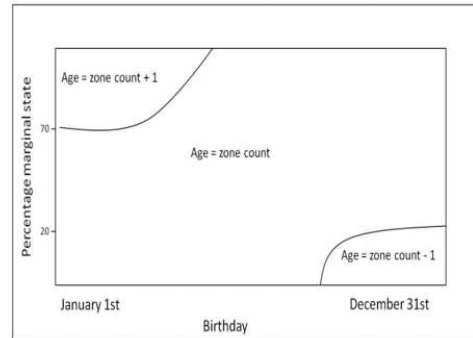

Figure 5.6: An age reading guideline for marginal rings which was not implemented in this study (extracted from Proctor et al. (2021)).

In this thesis, the date of capture of the otoliths has not been considered due to the

absence of this metadata in certain subsets of the images. Moreover, to measure the given threshold of annulus width, there should also be a zoom/scale metadata which is unfortunately absent in some sets of images. Hence, as a simplification, we assume that majority of the time, the age estimate will correspond to the count of annuli.

As a future step, however, especially with new batch of images, it would be better in the long run to have full support regarding this margin rule (both manual and AI-based) and that the required metadata to implement this will be consistently available and taken into consideration during the entire process.

Another common error in estimating the age of the otolith images is the occurrence of split and double rings (Panfili et al., 2002). This type of error is also a common error even with manual readings. There are corresponding guidelines (Carbonara and Follesa, 2019) for figuring this out systematically which are not elaborately configured in this study. Therefore, there is a future direction where the use of these special guidelines can be incorporated into the AI-based approaches so that age estimates can be adjusted accordingly.

Too Specific Imaging Requirements

Another primary limitation of the thesis project is that the focus is in the use of sectioned otolith images with transmitted light where the winter rings are shown as light zones and the summer rings are shown as dark zones. There are few initial experiments where the ground-truth labels are focused on the dark zones but due to the convention used for the manual process, these were not considered further. Figure 5.7 is an example showing the initial results from such experiments. Such preliminary result is a good indication that if one has to use reflected light (the light and dark bands are reversed), then potentially, the algorithms can still work satisfactorily.
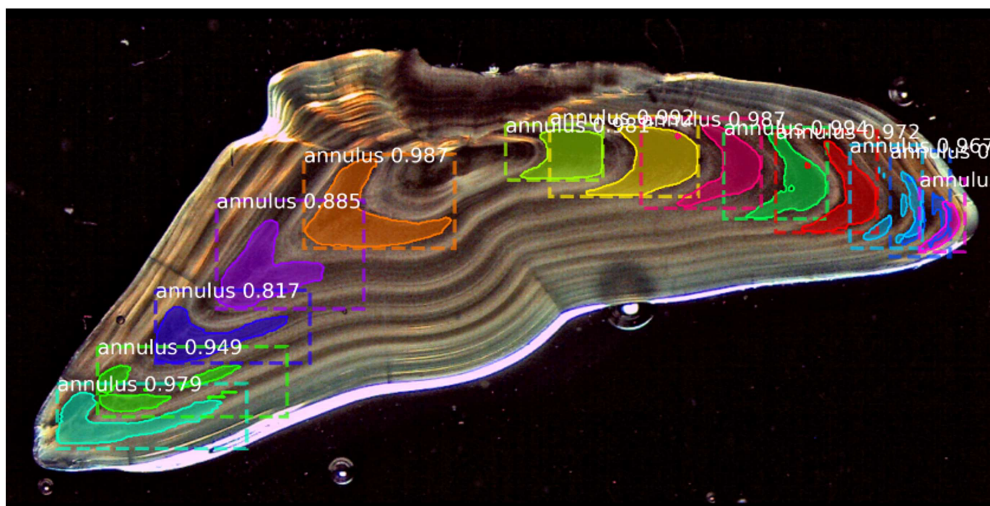
**Figure 5.7:** Prediction of the AI (Mask R-CNN) on the summer rings.

As another showcasing of this capability, in Figure 5.8-a, a sample result of an initial

experiment using reflected light on a whole otolith, was shown. It would be good to explore more on this type of study to see if there are changes needed to the methods or if it would be sufficient to just provide a different kind of ground-truth labels.

These proto-typical results provide a good starting point for the future direction involving the further extension of the application to accommodate a different light condition and even an entirely different type of otolith preparation (e.g., whole otoliths and broken otoliths).

The other limitation of the study is related to having the need to orient the image in certain upright way. The solution to this was partly covered in the third paper (Chapter 4). For instance, one can detect the nucleus or even sulcus and re-orient the image accordingly as part of the multi-stage workflow. The actual implementation of such heuristic is not covered in the study and could be a good future work.

Lastly, there is also a limitation regarding the use of only the major axes for counting and marking the rings. To handle this, there are initial attempts conducted such the one shown in Figure 5.8-b that marks the full rings which worked quite well as an alternative configuration. Similarly, another potential solution initially designed was to create a new set of ground-truth labels along an alternative axis (e.g., central axis or along the sulcus edges). Within the scope of the study, however, both of these methods were not thoroughly explored.
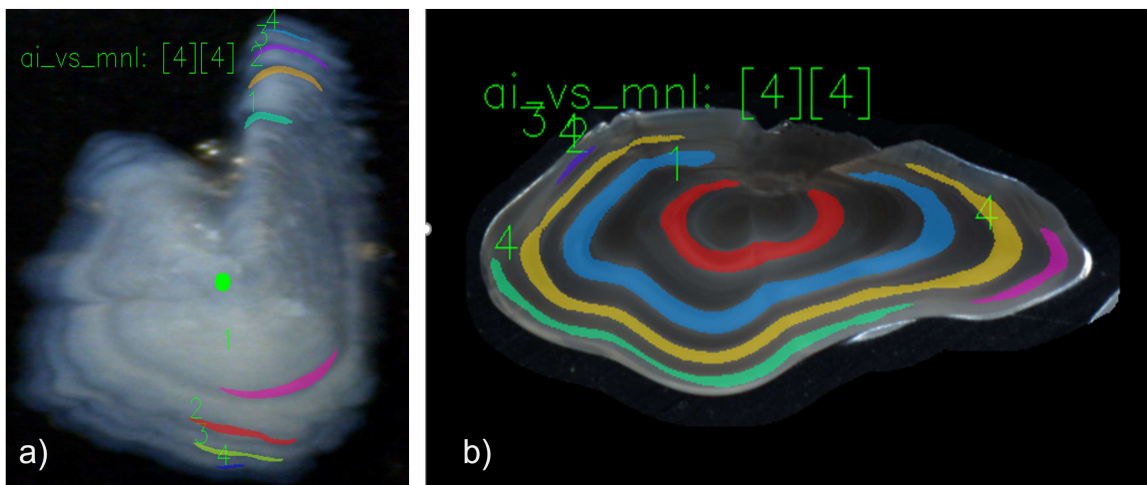
Figure 5.8: Sample AI annotations on whole (a) otoliths and the annotation involving full (b) rings.

Detecting Age 0 Otoliths

One of the major drawbacks of algorithms relying on region-based detection or segmentation is that if a certain subset of the training images does not have the object of interest, then that subset can no longer be used for training. Such is the case for otolith images of age 0. It is not possible to effectively use the "empty" images as training images for the

algorithm as it tends to dilute the number of pixels with actual positive labels. As a result, the model will lean towards marking every pixel as zero or background as it gets heavily penalized once it tries to create masks on age 0 images.

With that said, there are certain modifications to the design of the algorithms such that we can do as future steps to take into account the existence of these "empty" images. In the version that was partially explored in this study, the design of the network (U-Net only) was changed to accommodate two simultaneous tasks, one to classify images into age categories (including age 0) and another task to perform the segmentation in the case the image has been classified as having non-zero age. This multi-task formulation has actually been explored in the study of Politikos et al. (2021) where they used similar multi-task approach for improving the accuracy of the AI models by incorporating another task involving predicting the fish length (i.e., in addition to the task of predicting the fish age). In some studies, such multi-task design has been known to even reduce overfitting and increases the generalization capacity of the AI models (Chen et al., 2019).

As a future direction, one can explore further the same multi-task approach and create a new task for the algorithm to first detect the existence of the object of interest before allowing the algorithm to mark any region within the image. In this way, only those predicted to contain annuli will undergo the detection stage. Otherwise, there is a tendency of the algorithm to mark parts of the images or to have all regions marked as background depending on the proportion of these age 0 images used as part of the training.
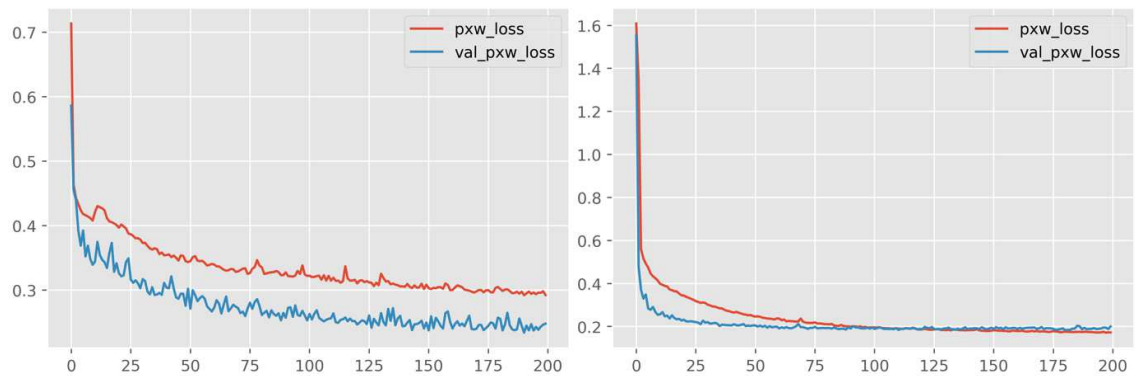
Further Exploration of Hyperparameters

During the development, in addition to the different ground-truth labeling possibilities, there are also certain algorithm hyper-parameter choices that need to be resolved before proceeding to one experiment and the next. Some of these hyper-parameters can be easily investigated through the use of systematic approaches such as grid-search as conducted in Chapter 2. There are some hyperparameters, however, that can be better resolved through manual inspection of certain "trial" runs to make it easier to assess whether it would be worth it to spend some time optimizing such parameters.

One of the hyperparameters that can be further explored is the number of epochs for the training. Figure 5.9 shows the typical curve of the loss values (both training and validation) as the training epoch increases. In the first case, it can be observed that the training epoch has been cut prematurely and that further improvement could have been attained if the training was allowed to continue. For the second case, the optimal validation loss has been reached early and that no matter how long the training is conducted, it can no longer be improved further. In this thesis, there is a balancing that was done to asses properly the appropriate number of epochs. That is, the priority is to achieve a good enough loss value but at the same time, it is also necessary that runs will not consume considerably longer time for a miniscule improvement. Hence, for simplicity, few trial runs have been conducted just to balance this trade-off between computing time and loss improvement. Through these initial runs, it was decided to go with a consistent epoch equal to 200. Naturally, for future directions, it would be nice to have a more systematic way of deciding this hyper-parameter which could be more optimal than the one used in

this study.

Figure 5.9: The plot of training and validation loss with increasing number of epochs.

Updated Implementations and Data Augmentation

In line with the exploration of hyper-parameters and algorithm configurations, there is also a future direction to explore the new updates regarding the implementation of certain algorithms.

The recent boom in AI resulted to many new types of algorithms and implementations that further increases the performance of the base methods. In Chapter 3, we somehow had a glimpse on these improvements with the use of advanced techniques such as ensemble learning and continual learning. There are still a whole lot of new add-ons to the base methods that were not explored in this study.

With the progress in the field of deep learning, in short periods of time, there are many modifications of the base algorithms developed that can make it even more accurate. This is the case for both Mask RCNN and U-Net. For instance, there exists already the so-called Attention- and Transformer-based versions (Vaswani et al., 2017) of these two algorithms that are shown to be effective in certain benchmarks. Also, a new batch of alternative network representations are available with the recent studies regarding the use of Kolmogorov-Arnold Networks or KANs (Liu et al., 2024).

Another modification that can bring out potential improvements would be the implementation of improved data augmentations. For the thesis project, the current data augmentations used is based from the default implementation from Keras library (Chollet et al., 2018) which is sufficient for most purposes. For more improvements in this aspect, there are several far advanced methods to use recently that make use of the latest progress in the field of artificial intelligence, particularly in the field of generative AI (Ramesh et al., 2021; Saharia et al., 2022). For example, the generative models have become the state of the art in terms of image generation and even natural language processing. Hence, one can leverage these new approaches in order to create even better synthetic images for training the algorithms especially when there is not enough actual otolith image data available.

Improved Useability and Microscopy Integration

The web-based application is only the start of the ongoing endeavor to make the use of AI-based approaches in fish age reading be widely accepted. Having it setup as a webserver will allow connection to other machines in a seamless manner. One can connect from a client machine using a browser to upload the images for processing using the AI methods hosted in the webserver.

Alternatively, as a future step, there will also be client-side application that can load directly the images from microscopes via appropriate libraries then programmatically connect to the webserver on the background to finally return an annotated image. In this way, there is no need to take an extra step creating folders of processed images. The software is intended to already perform the necessary saving and pre-processing such that it can be directly uploaded into the server via some background processes.

Lastly, for those that are intending to use the stand-alone application (i.e. portable app presented in Chapter 3), there is also a plan to improve the useability further. Currently, the models are utilizing CPU which will take considerably slower training time. As future steps, further instructions and setups will be available such that if the machine contains the necessary graphics card specifications, the app can be upgraded by the user accordingly such that the GPU-based version will be in use.

The ultimate goal is to make the use of AI be as easy as any other software tool such as ICES SmartDots. Hence, it should be simplified as much as possible. Creating intuitive user interface takes considerable effort and will require constant back and forth between the end-users and the developers. Hence this is planned as part of the next steps which will involve many user feedback and interactions.

5.8   References

Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., Samek, W., Müller, K.-R., Dähne, S., and Kindermans, P.-J. (2018). innvestigate neural networks!

Ben-David, S., Blitzer, J., Crammer, K., et al. (2010). A theory of learning from different domains. *Machine Learning*, 79:151–175.

Bojesen, T., Denechaud, C., and Malde, K. (2024). Annotating otoliths with a deep generative model. *ICES Journal of Marine Science*, 81:55–65.

Campana, S. (2001). Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology*, 59:197–242.

Campana, S. and Casselman, J. (1993). Stock discrimination using otolith shape analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 58:30–38.

Cao, F. and Fablet, R. (2006). Automatic morphological detection of otolith nucleus. *Pattern Recognition Letters*, 27:658–666.

Carbonara, P. and Follesa, M. (2019). Handbook on fish age determination: A Mediterranean experience. *General Fisheries Commission for the Mediterranean. Studies and Reviews*, 98:1–179.

Chen, Q., Peng, Y., Keenan, T., Dharssi, S., Agro´n, E., Wong, W. T., Chew, E. Y., and Lu, Z. (2019). A multi-task deep learning model for the classification of age-related macular degeneration. *AMIA Jt Summits Transl Sci Proc. 2019 May 6;2019:505-514. PMID: 31259005; PMCID: PMC6568069.*

Chollet, F. et al. (2018). Keras. Github repository. *Available from https://github.com/fchollet/keras.*

Denechaud, C., Smoliński, S., Geffen, A., Godiksen, J., and Campana, S. (2020). A century of fish growth in relation to climate change, population dynamics and exploitation. *Global Change Biology*, 26:5661–5678.

Dzieżyc, M., Gjoreski, M., Kazienko, P., Saganowski, S., and Gams, M. (2020). Can we ditch feature engineering? end-to-end deep learning for affect recognition from physiological sensor data. *Sensors*, 20(22).

Fisher, M. and Hunter, E. (2018). Digital imaging techniques in otolith data capture, analysis and interpretation. *Marine Ecology Progress Series*, 598.

Harbitz, A. (2009). A Generic Ad-Hoc Algorithm for Automatic Nucleus Detection from the Otolith Contour. *4th International Otolith Symposium*.

Hüssy, K., Mosegaard, H., Albertsen, C., Nielsen, E. E., Hansen, J., and Eero, M. (2016). Evaluation of otolith shape as a tool for stock discrimination in marine fishes using Baltic Sea cod as a case study. *Fisheries Research*, 174:210–218.

ICES (2023). SmartDots manual: How to annotate. Version 06. ICES User Handbooks. https://doi.org/10.17895/ices.pub.22810604.

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., and Tegmark, M. (2024). KAN: Kolmogorov-Arnold Networks. https://arxiv.org/abs/2404.19756.

Mahe, K., Ogor, A., Hüssy, K., Christensen, A., Mosegaard, H., Warnes, S., De Pontual, H., Harbitz, A., Gudmundsson, E., Thordarson, G., Parisi, V., Cotano, A., and Carbini, S. (2009). *Automated Fish Ageing (AFISA)*. (AFISA, proj. nr. 2243). FP6 project, Specifically targeted research project of innovation project - Referece no 044132.

Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G., Krpalkova, L., et al. (2019). Deep Learning vs. Traditional Computer Vision. https://arxiv.org/pdf/19 10.13796.

Martinsen, I., Harbitz, A., and Bianchi, F. (2022). Age prediction by deep learning applied to Greenland halibut (*Reinhardtius hippoglossoides*) otolith images. *PLOS ONE*, 17:e0277244.

McQueen, K., Hrabowski, J., and Krumme, U. (2018). Age validation of juvenile cod in the western Baltic Sea. *ICES Journal of Marine Science*, 76:430–441.

Moen, E., Handegard, N., Allken, V., Albert, O., Harbitz, A., and Malde, K. (2018). Automatic interpretation of otoliths using deep learning. *PLOS ONE*, 13:e0204713.

Moen, E., Vabø, R., Smoliński, S., Denechaud, C., Handegard, N., and Malde, K. (2023). Age interpretation of cod otoliths using deep learning. *Ecological Informatics*, 78.

Ordoñez, A., Eikvil, L., Salberg, A., Harbitz, A., and Elvarsson, B. (2022). Automatic fish age determination across different otolith image labs using domain adaptation. *Fishes*, 7:71.

Ordoñez, A., Eikvil, L., Salberg, A., Harbitz, A., Murray, S., and Kampffmeyer, M. (2020). Explaining decisions of deep neural networks used for fish age prediction. *PLOS ONE*, 15:e.0235013.

Panfili, J., de Pontual, H., Troadec, H., and Wright, P. (2002). Manual of fish sclerochronology. *Ifremer-IRD coedition*.

Pinto, C., Davies, J. O., Coster, K. D., Milar, C., Allegaert, W., Osypchuk, A., Catarino, R., Godiksen, J., Bekaert, K., Holdsworth, N., and Torreele, E. (2018). SmartDots: an online international platform for age reading workshops and calibrations. *International Conference on Marine Data and Information Systems, Barcelona, Spain*, page 148.

Politikos, D., Petasis, G., Chatzispyrou, A., Mytilineou, C., and Anastasopoulou, A. (2021). Automating fish age estimation combining otolith images and deep learning: The role of multitask learning. *Fisheries Research*, 242:106033.

Proctor, C., Robertson, S., Jatmiko, I., and Clear, N. (2021). An introductory manual to fish ageing using otoliths. *ISBN 978-0-646-83421-4*, page 41.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., and Sutskever, I. (2021). Zero-shot text-to-image generation.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., Salimans, T., Ho, J., Fleet, D. J., and Norouzi, M. (2022). Photorealistic text-to-image diffusion models with deep language understanding.

Smoliński, S. and Gutkowska, J. (2024). Otolith biochronology for the long-term reconstruction of growth and stock dynamics of fish. *Reviews in Fish Biology and Fisheries*, 34:405–419.

Stransky, C. (2014). Chapter seven - morphometric outlines. In Cadrin, S. X., Kerr, L. A., and Mariani, S., editors, *Stock Identification Methods (Second Edition)*, pages 129–140. Academic Press, San Diego, second edition edition.

Stransky, C., Baumann, H., Fevolden, S., Harbitz, A., Høie, H., Nedreaas, K., Salberg, A., and Skarstei, T. (2008). Separation of Norwegian coastal cod and Northeast Arctic cod by outer otolith shape analysis. *Fisheries Research*, 90:26–35.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA. Curran Associates Inc.

Acknowledgement

I will forever be indebted to those who accompanied and supported me for my PhD and those who dedicated time and effort to make this all possible.

First, I would like to thank Dr. Christoph Stransky, for his unwavering guidance, support and trust throughout this journey. His supervison and leadership have enabled me to reach the finish line of this PhD marathon and to go even further.

Also, I would like to thank Prof. Dr. Thomas Brey, for accepting me as a PhD advisee and for his teachings and encouragements that have brought out the best out of my capabilities.

I also extend my gratitude To Prof. Dr. Andreas Birk, for sharing his expertise and knowledge in order to polish the gem out of this research project.

I'm extending my thanks as well to the members of my thesis committee, Prof. Dr. Marko Rohlfs and Victor Escalona.

My heartfelt thanks to the Thünen Institute and KIDA Funding and everyone who helped in many other ways:

- to the institute director, Dr. Gerd Kraus, who accepted and trusted me for this role

- to Frieda, Dr. Uwe Krumme and all the members of the Otolith Age Reading Group, for the datasets and the detailed walk-through of the process

- to Hendrik and Karl, who is always ready to help even with language translations

- to Sabrina, Chen, Valeria, Bernhard, Miriam, Kim, Serra, Katrine, Jonas and all the early career scientists from TI, for the crucial suggestions and insights

- to Côme, Alf, Ketil and all the otolith scientists who have provided important feedback

- to Borris, Nicole, Monika, Wencke, Sandra, Marcellus, Annika and all members of Thünen Institute

I will always be thankful to my family and friends, who essentially made me the person I am today.

Finally, I would like to highlight my never-ending appreciation to my wife, Marianne, who always loves and believes in me through ups and downs, celebrates and cheers for me like a proud *gwin* during triumphant moments, lifts me up and comforts me like a warm cocoon when things go south and turbulent.

Declaration on the contribution of the candidate to a multi-author article/manuscript which is included as a chapter in the submitted doctoral thesis

Chapter 2: Fish Age Reading Using Deep Learning Methods for Object Detection and Segmentation

Contribution of the candidate in % of the total work load (up to 100% for each of the following categories):

Experimental concept and design:	ca. 80 %
Experimental work and/or acquisition of (experimental) data:	ca. 100 %
Data analysis and interpretation:	ca. 90 %
Preparation of Figures and Tables:	ca. 100 %
Drafting of the manuscript:	ca. 90 %

Chapter 3: An Interactive AI-driven Platform for Fish Age Reading

Contribution of the candidate in % of the total work load (up to 100% for each of the following categories):

Experimental concept and design:	ca. 80 %
Experimental work and/or acquisition of (experimental) data:	ca. 100 %
Data analysis and interpretation:	ca. 90 %
Preparation of Figures and Tables:	ca. 100 %
Drafting of the manuscript:	ca. 90 %

Chapter 4: Multi-stage Framework for Otolith Analysis

Contribution of the candidate in % of the total work load (up to 100% for each of the following categories):

Experimental concept and design:	ca. 80 %
Experimental work and/or acquisition of (experimental) data:	ca. 100 %
Data analysis and interpretation:	ca. 90 %
Preparation of Figures and Tables:	ca. 100 %
Drafting of the manuscript:	ca. 95 %

Date:

Signatures:

Universität Bremen
Fachbereich 2
Prüfungsamt Chemie
z. Hd. Frau Frauke Ernst
Leobener Straße

28359 Bremen
Deutschland

Prüfungsamt
Chemie

Frauke Ernst
Geschäftsstelle
Fachbereich 02
Leobener Str. / NW2
D-28359 Bremen

Verwaltungspavillon 06
Tel. 0421 218-62802
Fax 0421 218-9862802
frauke.ernst@uni-bremen.de
www.fb2.uni-bremen.de

Versicherung an Eides Statt

Name, Vorname	Cayetano, Arjay
Matrikel-Nr.	6093939
Straße	
Ort, PLZ	

Ich,   Arjay Cayetano   (Vorname, Name)

versichere an Eides Statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.
Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.
Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Ort, Datum / Unterschrift

EXAMINATION COMMITTEE

Examiner/Chairperson: **Prof. Dr. Matthew Nielsen**

University of Bremen

Fachbereich 2 Biologie/Chemie

Bremen, Germany

Examiner/Reviewer 1: **Prof. Dr. Thomas Brey (supervisor)**

University of Bremen

Fachbereich 2 Biologie/Chemie

Bremen, Germany

Examiner/Reviewer 2: **Prof. Dr. Andreas Birk**

Constructor University

School of Science and Engineering

Bremen, Germany

Examiner/Reviewer 3: **Dr. Christoph Stransky**

Thünen Institute of Sea Fisheries

Deputy Director

Bremerhaven, Germany

Date of colloquium: 29 August 2024