Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

Magnus Wiik

Assessing Re-Identification Capabilities of Salmon Body Parts Using AI-Based Computer Vision Methods

Master's thesis in Cybernetics and Robotics
Supervisor: Annette Stahl
Co-supervisor: Christian Schellewald, Rudolf Mester and Espen Berntzen Høgstedt
June 2024

NTNU
Norwegian University of Science and Technology

Magnus Wiik

Assessing Re-Identification Capabilities of Salmon Body Parts Using AI-Based Computer Vision Methods

Master's thesis in Cybernetics and Robotics
Supervisor: Annette Stahl
Co-supervisor: Christian Schellewald, Rudolf Mester and Espen Berntzen Høgstedt
June 2024

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of Science and Technology

Preface

This document presents the work of a Master´s thesis on computer vision for salmon re-identification. It is part of the MTTK — Cybernetics and Robotics study program at the Department of Engineering Cybernetics at the Norwegian University of Science and Technology (NTNU). The work was carried out during the spring semester of 2024 in collaboration with Sintef Ocean. Supervision was performed by Annette Stahl, Rudolf Mester, and Espen Berntzen Høgstedt from NTNU and Christian Schellewald from Sintef Ocean. The thesis is written for people with an engineering background interested in aquaculture and computer vision.

Trondheim, 2024-06-10

Magnus Wiik

Acknowledgment

I want to thank the following people for their help during the project: Annette Stahl, Rudolf Mester, and Espen Berntzen Høgstedt from NTNU, and Christian Schellewald from Sintef Ocean; thank you for your valuable guidance during our meetings.

M.W.

Abstract

Technological advancements in the aquaculture industry have accelerated, and it is now possible to use technology to analyze salmon's welfare. A wealth of information can be extracted from image and video data through machine learning methods that utilize deep learning principles. To automate the welfare estimation of individual salmon in the aquaculture industry, it is necessary to identify each salmon. In this Master´s thesis, a salmon re-identification pipeline is developed. The pipeline is used to analyze the body parts of the salmon to find which are the most informative. The Deep learning models within the re-identification pipeline are trained using three datasets that are constructed during the project. Video recordings of salmon are provided by Sintef Ocean. The re-identification pipeline contains modules for salmon detection, body part detection, and individual re-identification. By evaluating the pipeline, the re-identification accuracy using images of each body part is as follows: Thorax (87.7%), dorsal fin (86.3%), eye (50.0%), pectoral fin (43.1%), and caudal fin (49.3%).

Sammendrag

Den teknologiske utviklingen i oppdrettsnæringen har skutt fart, og det er nå mulig å bruke teknologi for å analysere velferden til laks. Man kan hente ut mye informasjon gjennom bilde og video data gjennom maskinlæringsmetoder som utnytter dyplæringsprinsippet. For å kunne automatisere velferdsestimeringen av laks i oppdrettsnæringen, må man kunne identifisere laksen for å gi hvert enkelt individ et veldferdsestimat. I denne masteroppgaven er det laget en programvare som bruker dyp læring for å identifiserer laks basert på bilder av kroppsdeler. Programvaren er brukt til å finne ut hvilke kroppsdeler som er mest informative og best egnet til re-identifisering. Arbeidet er gjort med videoer av laks i tanker, som er levert av Sintef Ocean. For å trene modellene som inngår i programvaren, er tre datasett laget. I programvaren inngår modeller for deteksjon av laks, deteksjons av kroppsdeler og individuell re-identifisering. Evalueringen av programvaren viser at re-identifiseringsnøyaktigheten for bilder av hver kroppsdel er som følger: Bryst (87.7%), ryggfinne (86.3%), øyne (50.0%), brystfinne (43.1%), og halefinne (49.3%).

Contents

Chapter 1

Introduction

1.1 Background

The seafood industry is important in global food security, amongst other things, due to its high feed conversion ratio [1]. Farmed Atlantic salmon is among the most sustainable meat products in the world as it grows efficiently, and there is high demand for by-products of salmon production [2]. But the industry has its problems. As most of the salmon are produced in open sea systems, they are prone to parasites such as salmon lice [3]. These parasites are lethal for salmon, and treatments to counteract infestations are also causing harm to the salmon themselves [4]. Gill disease is an emerging problem for farmed salmon during its sea phase, [5]. Many of the problems can become easier to manage by shifting the industry towards precision fish farming, which aims to improve the fish farmer´s ability to monitor, control, and document biological processes in the fish farm, [6]. By monitoring individual fish, one can assess the welfare status of the population from the ground up, which can lead to more precise welfare estimates. Individual re-identification opens the possibility of following individual salmon and logging their welfare status over time. One can then assess the severity of diseases and take action if needed. The result can be better welfare for the salmon and increased income for fish farmers, as mortality rates within the farm can decrease. A non-invasive option to monitor fish is by using cameras to record videos and deep learning techniques to extract welfare information. Camera technology using Deep learning has been successfully implemented in tasks, such as identifying loser fish [7], fish feeding systems [8], fish counting and biomass estimation [9]. For these algorithms to accurately estimate welfare, growth, and biomass, one needs to identify the salmon individuals. We know that the melanin spot pattern on the salmon body is a way of identifying a salmon [10]. However, no one has yet analyzed and compared the re-identification capabilities of several body parts of the salmon. In this thesis, we build a deep learning pipeline to assess the re-identification capabilities of salmon body parts and compare their informativeness.

1.2 Objectives

The main objectives of this Master's thesis are to:

1. Develop a pipeline using deep learning and computer vision methods for assessing the re-identification capabilities of salmon body parts.
2. Find which salmon body parts are best for re-identification using the developed re-identification pipeline.

1.3 Approach

The thesis consists of two parts. First, a structured literature review is conducted. The field of salmon biometrics and salmon re-identification algorithms is summarized. Following this, we will apply the methods from the salmon re-identification literature to analyze the salmon body parts and find out how suitable each body part is for salmon re-identification.

1.4 Contributions

In this project, three datasets are constructed. The datasets are made for salmon detection, body part and point detection, and salmon body part re-identification. A re-identification pipeline containing three Deep learning methods is developed. The pipeline and the datasets are used to find the most capable body parts on the salmon for re-identification.

1.5 Limitations

This study analyzes the thorax, dorsal fin, eye, pectoral fin, and caudal fin. Technologies from the Deep learning library PyTorch are used to develop the re-identification pipeline. This study does not analyze the stability of each body part over a time period.

1.6 Sustainability

The United Nations (UN) has created a list of sustainable development goals that define the collaborative work the nations in the UN have agreed upon. The work in this thesis falls under goals 12 and 14. Goal number 12 is called Responsible consumption and production, and goal number 14 is Life below water. The UN Sustainability Goal 12 states we should ensure responsible consumption and production. The UN Sustainability Goal 14 states that we should conserve and use the oceans, seas, and marine resources for sustainable development. Salmon

re-identification covers both goals as it leads to more responsible production of farmed salmon. It improves the salmon's life by facilitating the evaluation of individual welfare. Re-identification of farmed salmon ensures that the farmer gains insight into each individual and can assess their welfare and growth. As one can re-identify salmon individuals, tracking how each salmon's health is developing is possible. This technology's possible future outcomes include the assessment of wounds, scale loss, behavior, and eating patterns. This positively contributes to the responsible production of Atlantic salmon, as the welfare of the animals is more prioritized, [11].

1.7 Outline

- Chapter 1: Introduction to the Master´s thesis.
- Chapter 2: Literature review. Discussion about salmon welfare and the state of the art within salmon re-identification technologies.
- Chapter 3: Theoretical background for computer vision and AI. Presents ResNet, Faster R-CNN, and Integrated gradients.
- Chapter 4: Salmon dataset construction process. Presents the salmon datasets, augmentation techniques, and challenges.
- Chapter 5: Salmon re-identification approach. Presents the salmon re-identification pipeline. Furthermore, it discusses each module in detail and the training and evaluation process.
- Chapter 6: Results from the salmon re-identification pipeline and discussion. Presents the results from evaluating the re-identification pipeline and discusses its results in a bigger context.
- Chapter 7: Conclusion and recommendations for future work.
- Bibliography
- Appendix A: Acronyms.
- Appendix B: Confusion matrices from pipeline evaluation and training data for the re-identification module.

Chapter 2

Literature review

This chapter covers the literature review abbout salmon welfare and salmon re-identification technologies. First, the definition for salmon welfare is presented. It then presents salmon needs for good welfare and indicators for assessing welfare. Finally, it presents visual salmon re-identification methods in the literature.

2.1 Salmon welfare

Animal welfare is a term that describes how an animal perceives its quality of life, [12]. The animal welfare definition is quite vague and is not a tangible state that can be measured. However, welfare indicators can be measured and give insights about the welfare state of animals, but not reveal it entirely. Before presenting a salmon welfare framework, we present the salmon welfare needs.

2.1.1 Salmon welfare needs

In the general case, salmon welfare needs are about providing the necessities for its immediate survival, which in the literature is termed ultimate needs, and needs to sustain in the future, proximate needs, [13]. An animal's ultimate needs encapsulate Nutrition, Respiration, Thermoregulation, Maintaining osmotic balance and body integrity, [14]. When a need is not met, the emotional reward system causes the individual to experience emotions that guide the animal to meet its needs, [15]. When the animal cannot meet its needs, the experienced welfare of the individual is reduced [16]. Several of these needs can be monitored using visual techniques such as computer vision. But to measure the welfare of salmon, we need protocols that tell us what information to focus on.

2.1.2 Individual welfare assessment

There exist several frameworks for evaluating the welfare of salmon, but in the following, we use the framework made by [14]. To estimate the welfare of salmon individuals, there are two frameworks for measuring the needs of the salmon. The focus here will be on animal-based operational indicators, as this is the framework to measure an individual's welfare in a fish farm, [12]. This framework is important, as it gives fish farmers an indication of whether they should act to improve the welfare status of the salmon individuals during production.

Animal-based Operational Welfare Indicators

- Eye roll (VER)
- Sea lice
- Condition factor
- Hepato-somatic index
- Cardio-somatic index
- Handling trauma
- Rigor mortis time
- Emaciation state
- Sexual maturity state
- Vertebral deformation
- Fin damage and fin status
- Scale loss and skin cond.
- Snout jaw wound
- Eye haemor. and status
- Opercula deformation

Figure 2.1: This list is adapted from [14], and shows animal-based welfare indicators that may be estimated on a fish farm without the presence of a fish health expert.

In figure 2.1, there is a list of animal-based operational welfare indicators that can be estimated without fish health expertise. One example of using computer vision in fish farms is by estimating the number of sea lice in a fish farm. This would yield great value for fish farmers as this is one of the main problems for open aquaculture systems today. Assigning a sea lice number to every individual in the farm would be even better as you get more precise estimates for the welfare scoring of the population. Assigning a welfare score to each individual in the fish farm can be done for other welfare indicators, such as fin damage or scale loss. To achieve this, there is a need to develop fast and robust re-identification systems that enable salmon welfare estimation for each individual. In the future, each salmon in the fish farm can have its own health diary, which is a big step towards optimizing the fish welfare. There are multiple ways to identify a salmon individual, and how it is possible to do so is presented next.

2.2 Salmon re-identification methods

Using visual traits on salmon individuals to identify individuals is good for welfare as these methods are non-invasive. Here, we discuss several traits that are used for salmon re-identification today. First, we address melanin spots on the thorax (chest) and operculum (head), before we move on to iris.

2.2.1 Thorax melanin spots

Salmon individuals have been re-identified using images of their thorax (chest). In Espen Høgstedt´s Master thesis, he identified 7 salmon based on thorax images of both sides on each individual. He developed a re-identification algorithm using a Keypoint R-CNN to detect the front half of the fish together with a set of points. He used the set of points to change the perspective of the detected salmon, and focus on the center of the body, called thorax. These images were fed into various Deep learning image classification models such as AlexNet, EfficientNet, and ResNet. The highest accuracy was obtained using ResNet101 to classify the salmon individuals, where he obtained a Rank-1 accuracy of 99.51% on unseen data, [17].

In [18], they identified salmon using the dot pattern on the salmon´s thorax. They showed that it is possible to identify a salmon using a dot localization algorithm to determine the exact position of every dot on the body. This approach yielded a 100% accuracy for images taken out of the water and a lower accuracy for in-water images. They also performed identification using a Histogram of Gradients (HOG) to analyze both the dots and the rest of the salmon´s skin. The approach yielded, in general, a lower accuracy both for in-water and out-of-water images. This suggests that a method using the dot position only is a more robust approach to identify salmon. One reason for this may be that using the gradients of the salmon skin is prone to lighting changes between the images, which can result in a slightly different embedding across examples. In real aquaculture conditions, there is additional debris, algae, and feces in the water, which can cause an additional drop in accuracy for the HOG approach, further suggesting that a dot localization approach may be more robust in real-world conditions, [18].

2.2.2 Operculum melanin spots

Stien showed that salmon can be identified using only the melanin spots on the operculum (head). They used a group of three fish biologists that classified images of 246 salmon. There was one set of images when the salmon were 12 months old and one set of images when they were 22 months old. The task was to match the individuals between the two image sets. They found that it was enough for salmon with 4 or more spots at 12 months to identify the individual later with certainty. However, only 32% of the salmon had 4 spots this early. This suggests that if

a third set of images was taken when all the salmon had 4 or more spots, then the identification could be certain for all. A point-pattern-matching algorithm was tested on the image set, using the position of the melanin spots in relation to each other as an identification method. The algorithm was tested by computing the distance between one image to other images in a set and achieved a Rank-1 accuracy of 85%. The match was the image pair with the least distance compared to other image pairs. This suggests that if images of salmon are taken in a controlled environment, it is possible to automatically identify individual salmon using the distances between the spots. Taking the uniqueness of each melanin spot by taking the shape and location of the salmon body into consideration should make the algorithm more robust against misclassification, [10].

Mathisen developed a dataset containing images of salmon heads in an aquaculture environment and tested it on an image classification model called FishNet, which is adapted from FaceNet [19]. They trained the FishNet model by using a training loss function called Triplet loss, which minimizes the distance between images of the same individual and maximizes the distance between images of different individuals, [20]. A Deep learning model such as the one used here will find the optimal relation between the input image and the output identity with regard to a loss function. It should focus on discriminating parts of the fish head, such as the melanin spots, the iris, or a combination of several things on the head. However, one loses control over what the model will focus on to make its predictions. Therefore, a way to monitor that the Deep learning model is using biometric features is to use an Explainable AI technique that can describe the image regions that the model uses to make its predictions, such as Integrated Gradients [21]. The authors also inform that the images have been taken over a short period of time, which means that the images for each individual look very similar. Using many similar images to train a neural network will increase the chance of overfitting the model and ultimately lead to worse performance in a different environment or video. To counteract this effect, one can obtain more data, annotate images using different environmental conditions, or use augmentation techniques.

2.2.3 Iris

Foldvik found that using the iris of a salmon is viable for identification during the middle part of the life stage. Images were taken of 1286 individuals over a time period of 533 days. In total, 6 images were taken per individual, with more frequent image capturing in the early life stage. This was due to the considerable changes to the body during this life stage of the salmon. Identification of the iris was performed using the out-of-the-box software VeriEye 2.10 Standard SDK by Neurotechnology. For salmon in the middle life stage, the identification accuracy was 100% using iris. This suggests that the salmon iris´s uniqueness is comparable to the human iris. No matches were made using the images from the earliest life stage of the salmon to any other life

stage. This shows that the big changes to the salmon body in its early life caused the stability of the iris not to be as good as that of humans, [22].

Schraml used Deep learning techniques to identify individual salmon using iris images. They investigated if a Deep learning approach would improve the long-term identification stability of the iris, as opposed to a texture feature-based approach, which shows weak stability in the long term. The texture feature-based approach explicitly utilizes texture information of the iris band (frequency information) and compares this information to textures from another iris to find its match, [23]. This technology is used in these studies [24] and [22], where it has shown strong short-term identification but weak long-term stability. The Deep learning approach showed even weaker long-term stability using iris images for individual identification. However, the dataset contains a small number of images, 4 images per fish taken two months apart. This suggests that there is more to gain from using Deep learning if the dataset is increased in size by sampling the fish iris more frequently.

Chapter 3

Theoretical background for computer vision and AI

This chapter covers the necessary theory to understand the basics of how the Computer vision and AI models in the re-identification pipeline work. We start by briefly explaining what a neural network is before we present the image classification model, ResNet. Then, the detection model Faster R-CNN is presented, and finally, the theory behind the explainable AI method, Integrated gradients, is explained.

3.1 Neural networks

Neural networks are a central part of the salmon re-identification pipeline, and we, therefore, start with Aggarwal's explanation of the term.

"Artificial neural networks are popular machine learning techniques that simulate the mechanism of learning in biological organisms. The human nervous system contains cells, which are referred to as neurons. The neurons are connected to one another with the use of axons and dendrites, and the connecting regions between axons and dendrites are referred to as synapses. [...] The strengths of synaptic connections often change in response to external stimuli. This change is how learning takes place in living organisms."[25]

For salmon re-identification using computer vision, convolutional neural networks are a natural choice. A convolutional neural network takes images as input and performs convolution operations using learned weights to extract information from the image. This keeps the structural information intact, which is essential for tasks like detection, pose estimation, and segmentation. The filters are learned through training the network on a training set, where you show the network examples and the correct prediction.

3.2 ResNet

ResNet is a convolutional neural network that can learn complex features from images and is, therefore, part of the re-identification pipeline. We come back to this in section 5.4. Now, the model architecture will be presented.

The residual network (ResNet) is a deep convolutional neural network (DCNN) used mainly for computer vision tasks. Deep learning has been successful due to stacking layers of neurons on top of each other to create network structures that are able to generate complex nonlinear relationships between input and output layers. However, when a network reaches a certain depth, the solvers are unable to find good or better solutions than similar networks with shorter depths. In [26], they call this the degradation problem, and it is addressed by reformulating the network structure to learn residual (error) functions. The residual functions are generated for a number of stacked layers at a time, where the input is subtracted from the output of the stacked layers. Performing this reformulation of the network yields easier optimization for a deeper neural network and higher performance than other networks of similar depth, [26]. This neural network is chosen as the large depth of the network enables it to learn highly complex features, which can prove to be useful for reidentifying individuals.

Feature Extraction is a key part of Deep learning re-identification systems. It is the process of finding and extracting important information from input data. A deep learning network (e.g. ResNet) performs feature extraction by performing convolution operations on the input image at multiple stages through the network. By performing these convolution operations in a sequence, it extracts features of increasing complexity. For example, when an image of a salmon goes into a ResNet model, it is convolved with a kernel to extract simple structures in the image, like edges. In the next layers, information like corners, which can resemble the tips of fins on the salmon body, is extracted from the image. As we go deeper, patterns like the scales on the salmon are extracted. Next, parts like the eyes, snout, or fins are extracted from the image. In the final layers, objects like the salmon are extracted. As the data goes deeper into the network, more complex features are extracted, which are based on the simpler features from earlier in the network. In the final layers, there are fully connected layers that compress the features into vector numbers that contain the extracted information. Extracting information from data such as images is one of the big strengths of a deep learning model compared to more traditional methods. Choosing the right information to take into consideration when reidentifying a salmon is very important. However, it is crucial to input quality data into the deep learning model so it learns the right distinguishing features between salmon individuals.

Multi-class cross-entropy Loss

The multi-class cross-entropy loss is a function for evaluating the loss of prediction from a neural network in a multi-classification task. A multi-classification task is where a neural network predicts a class in a set of possible classes. The loss from the multi-class cross-entropy is used to update the neural network´s belief about the relationship between input data and output class.

$$l(x,y) = L = (l_1,...,l_N)^T, l_n = -\sum_{c=1}^{C} w_{y_n} log \frac{\exp x_{n,y_n}}{\sum_{c=1}^{C} \exp x_{n,c}} y_{n,c} \tag{3.1}$$

Equation 3.1 is from [27]. In our case x is the input image, and y is the true class. C is the number of classes to choose from. N is a value that represents the number of images in an update step for the neural network. We come back to the application equation 3.1 in section 5.4.1.

3.3 Intersection over union

Intersection over union (IoU) is an important metric within Computer vision. The metric is shown in equation 3.2. The IoU of two bounding boxes A and B is the correctly predicted overlap between the boxes divided by the area spanned by both boxes.

$$IoU = \frac{A \cap B}{A \cup B} = \frac{true\ positives}{false\ negatives + true\ positives + true\ negatives} \tag{3.2}$$

3.4 Faster R-CNN

Faster R-CNN is an object detection model, [28]. Faster R-CNN takes images as input and localizes and classifies objects within the image using a Deep learning network, commonly referred to as the backbone when part of the larger model. In their paper, the backbone is called a Region Proposal Network (RPN), since it proposes regions of interest that are fine-tuned and classified in two fully connected layers. The two fully connected layers are called the bounding box regressor and bounding box classifier. A bounding box is a box that represents the region of interest within an image. The bounding box regressor refines the bounding box proposals from the RPN. The bounding box classifier predicts the class indicator of the region of interest. A class indicator is a label that describes the contents of a region of interest. Earlier, State-of-the-art detection models used Selective search for region proposal generation, but this was their main bottleneck regarding time efficiency, [29]. Faster R-CNN introduced the sharing of Deep learning network layers between RPN and the fully connected layers, which increased time efficiency. The region proposals from the RPN are generated much more rapidly and have reduced the inference time from seconds to fractions of a second.

To train the RPN, a label is assigned to every region proposal, which is either negative or positive. The conditions for label assignment are as follows:

- The region proposal(s) with the highest IoU overlap with the true bounding box is assigned a positive label.
- A region proposal with a higher IoU than 0.7 with any true bounding box is assigned a positive label.
- A region proposal is assigned a negative label if the IoU is below 0.3 with any true bounding box.
- Region proposals that are neither qualified for a negative or positive label are considered neutral and do not contribute to training.

The labeled predictions are fed into a loss function, which determines how to update the model with respect to regression and classification. The loss function for the RPN is as follows:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*) \tag{3.3}$$

Equation 3.3 represents the loss evaluated on the predictions from the bounding box regressor $\{p_i\}$ and bounding box classifier $\{t_i\}$. Each term in the loss function is a sum of loss functions for each prediction from the regressor and classifier. N represents the number of predictions. p_i is the probability that region proposal i is an object, in our case, a salmon or a specific body part. t_i is a vector representing the coordinates for bounding box i. λ is a weighting parameter that changes the importance of each task, bounding box regression, or classification. The loss functions for the regressor and classifier are as follows:

$$L_{reg}(t_i, t_i^*) = \begin{cases} 0.5(t_i - t_i^*)^2, & \text{if } |t_i - t_i^*| < 1 \\ |t_i - t_i^*| - 0.5, & \text{otherwise} \end{cases} \tag{3.4}$$

Equation 3.4 represents the loss function for the bounding box regressor, commonly known as the smooth L1 loss function. If the difference between the predicted bounding box and the true bounding box is less than 1, the loss is determined by evaluating a second-degree polynomial. If the difference between prediction and truth is greater than 1, the loss is determined by evaluating an absolute error.

$$L_{cls}(p_i, p_i^*) = -\frac{1}{N_{cls}}\sum_i p_i^* \log(p_i) + (1 - p_i^*)\log(1 - p_i) \tag{3.5}$$

Equation 3.5 represents the loss function for the bounding box classifier, commonly known as Log loss. If the true class indicator is 1, the loss is determined by taking the logarithm of the

predicted class indicator. If the true class indicator is 0, the loss is determined by taking the logarithm of 1 minus the predicted class indicator. In the case of salmon detection, 1 is equal to the salmon, and 0 is equal to the background.

For more information about Faster R-CNN, see the paper by [28]. The application of Faster R-CNN is presented in chapter 5.

3.5 Integrated gradients

Integrated gradients is an Explainable Artificial Intelligence (XAI) method in the class of feature attribution. It generates a visualization of the importance of each pixel in an input image to a Deep learning network. In the following, we present why it is useful to apply Integrated gradients and the theory behind the method, [21].

The motivation behind using Integrated gradients is to find the importance of each pixel in an image with respect to the prediction made by a Deep learning network. For salmon re-identification, it is interesting to apply Integrated gradients to model predictions, as they show how the biometric information is used to arrive at a certain prediction. Consider the case where the Integrated gradients method shows that the Deep learning network applies importance to the background in an image, during re-identification using the dorsal fin on the salmon. In this case, one knows the prediction is made by focusing on irrelevant information. Therefore, one can discard this prediction and initiate actions to guide the Deep learning network towards focusing on the biometric information instead. Examples of actions may be to add more images of salmon in different environments or from multiple angles. The Deep learning network can then find that the dorsal fin is the common factor between images in the dataset, as the background of the images has variation. XAI methods are valuable as they generate insight into how a Deep learning network generates predictions, which leads to increased trust in predictions. In the following, the background theory for the method is presented.

Consider a function F that represents a Deep learning network, $F : R^n \rightarrow [0,1]$. Let an input image be $x \in R^n$. A baseline (reference) image is defined as xt' $\in R^n$. The baseline image is a black image. To calculate the importance of pixels in the input image, we consider a straight path from the input image to the baseline image. The integrated gradients are obtained by summing the gradients from points on the path between the images. One can compute the importance of any pixel in the input image as follows:

$$IntegratedGrads_i(x) = (x_i - x_i^{'}) \int_{\alpha=0}^{1} \frac{\partial F(x^{'} + \alpha(x - x^{'}))}{\partial x_i} \partial \alpha \tag{3.6}$$

Equation 3.6 represents the accumulated gradients along one path *i*. The path is the range of values the pixel has from zero to the value in the input image, where $\alpha = 0$ represents a

zero-valued pixel and $\alpha = 1$ represents the pixel with its value in the input image. At each step, $(x^{'} + \alpha(x - x^{'}))$, on the path i, the image is fed into the Deep learning model, and the gradients of the output are accumulated, $\frac{\partial F(x^{'} + \alpha(x - x^{'}))}{\partial x_i}$. The gradients of each step are integrated with a step of $\partial\alpha$, an infinitely small step. Each pixel´s Integrated gradients are scaled to reflect the actual importance of the pixel, and this is done by multiplying the integral by the difference between the baseline and the input image, $(x_i - x_i^{'})$. In practice, the integral from Equation 3.6 is approximated by summing the gradients in steps along the path from the input image to the baseline image as follows:

$$IntegratedGrads_i^{approx}(x) = (x_i - x_i^{'}) \frac{1}{m} \sum_{k=1}^{m} \frac{\partial F(x^{'} + \frac{k}{m}(x - x^{'}))}{\partial x_i} \tag{3.7}$$

Equation 3.6 presents the Integrated gradients approximation for any pixel in the input image of a Deep learning network. The approximation of the Integrated gradients is done by summing the gradients between the baseline image and the input image in m steps.

The Integrated gradients for each pixel in the input image are summed to create a mask that visualizes the pixel's importance. For more information about Integrated gradients, see the paper by [21]. We come back to the application of the method in section 5.4.3.

Chapter 4

Salmon dataset construction

This chapter presents the salmon dataset creation process. It describes the annotation for the three datasets and shows examples of annotated images. Then, the dataset augmentation pipeline is presented before discussing the challenges encountered along the way.

4.1 Salmon datasets

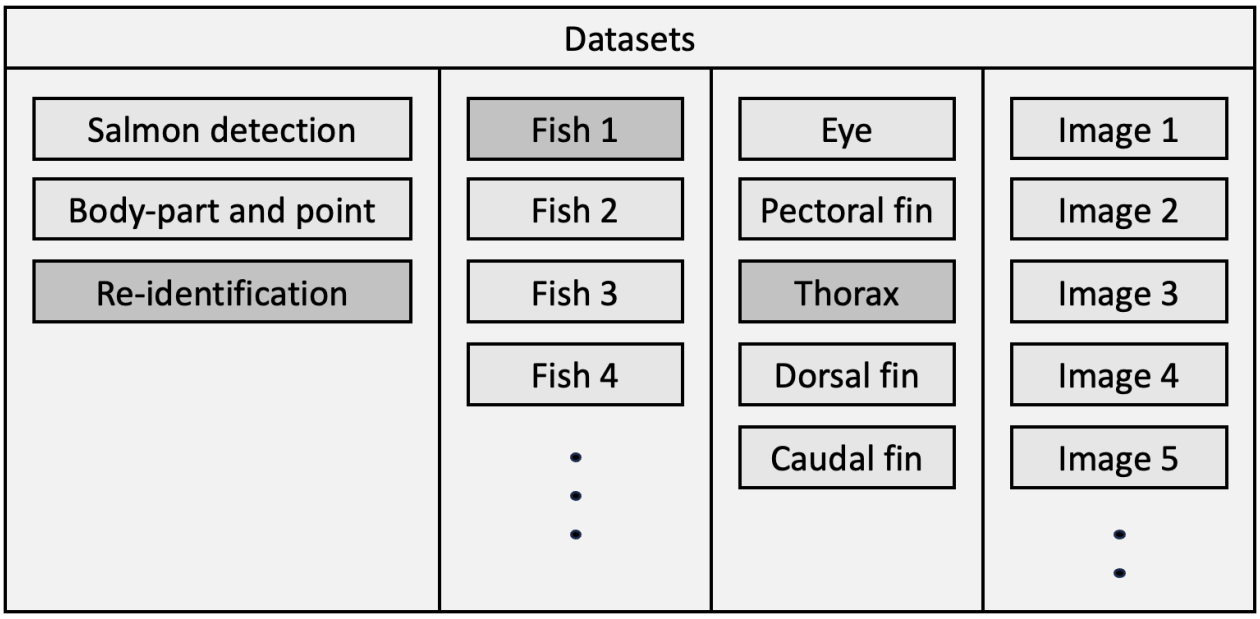


Figure 4.1: Overview of the datasets developed during the project. The Salmon detection dataset is used to train the model in Module 1 of the pipeline. The body part and points dataset are used to train the model in Module 2. The Re-identification dataset is used for the final classification task in Module 3.

The data material used in this project includes videos from experiments performed within a

project named "IDE 4" at Sintef Ocean. In this experiment (see also [30]), they investigated how water flow and oxygenation influenced the water quality and the stress levels of salmon. They recorded videos of salmon in tanks both under stress and in normal conditions. This project used the recordings under normal conditions, where oxygenation and water flow were present, as the salmon swam more in front of the camera. The salmon also swam with less effort under normal conditions so that the fins of the fish were more visible in the recordings. The circumstances described above made it easier to annotate and construct datasets with satisfactory quality. In total, three datasets were constructed. Each dataset is made for training and evaluating parts of a salmon re-identification pipeline. Figure 4.1 shows the overview of the datasets: Salmon detection, Body part and point, and re-identification. The datasets are described in more detail in the following sections.

Dataset 1: Salmon

The salmon detection dataset contains images of one or more fully visible salmon, along with bounding boxes and class indicators describing a salmon´s position and size. The class indicator describes what is inside the bounding box. Figure 4.2 shows examples of four annotated salmon individuals.

Dataset 2: Body part and point

Table 4.1: Annotated salmon body parts and points. The points are not used for training the modules in the pipeline but can be used in further work on re-identification by implementing, e.g., perspective transformation, which is discussed more in-depth in section 7

Point	**Body part**
Snout top	Eye
Snout bottom	Pectoral fin
Iris	Thorax
Pectoral fin root	Dorsal fin
Dorsal fin root front	Caudal fin
Dorsal fin root back	
Caudal fin top	
Caudal fin bottom	

The body part and point dataset contains images of at most one whole salmon individual. Annotations for the images include bounding boxes, points, and class indicators. The bounding

Figure 4.2: Example images from dataset 1. The class indicator for each bounding box is displayed with the format "Fish, Individual".

boxes describe the position and size of salmon body parts. The points describe the position of anatomical marks on the salmon. Body parts and points of interest are listed in table 4.1. It includes body parts that have been analyzed in the literature prior to this project, e.g., eye and thorax. However, some parts have not been analyzed in the literature, such as the pectoral fin, dorsal fin, and caudal fin. The background for investigating fins for re-identification is that the video recordings have shown that individuals often have unique traits in their fins due to size variations, deformation, or wounds. Figure 4.3 shows examples of annotations from the body-part and point dataset. During the project, the annotated points were experimented with to change the perspective of salmon using a perspective transformation, but it was aborted due to tuning problems and time limitations in the project plan.

Figure 4.3: Salmon body parts and points dataset example with annotations from the recorded footage from [30]. Here are five salmon individuals displayed with bounding boxes in blue and points in yellow. As seen here, the dataset contains images with various distances and perspectives to the camera.

Dataset 3: Re-identification

The re-identification dataset contains images of salmon body parts and class indicators specifying what individual a body part belongs to. Figure 4.4 shows examples of body part images belonging to the individuals Casper, Novak, and Holger. The figure shows that image shapes for a body part can vary greatly due to changes in the perspective and distance from the camera. This is shown by comparing the thorax images of Casper and Holger. When the images are fed to the re-identification models, they are converted to a standard shape, which the models expect, and this is discussed further in section 4.2.

4.2 Dataset augmentation

Data augmentation is a methodology that changes the appearance of data, in this case, images, in the dataset to increase variation and make it more difficult for the Deep learning model to overfit on unimportant information in the data. Variation within a dataset is important, as the model becomes more robust to small changes in unseen data. A data augmentation pipeline is developed to increase variation in the re-identification dataset, so that when the model sees im-

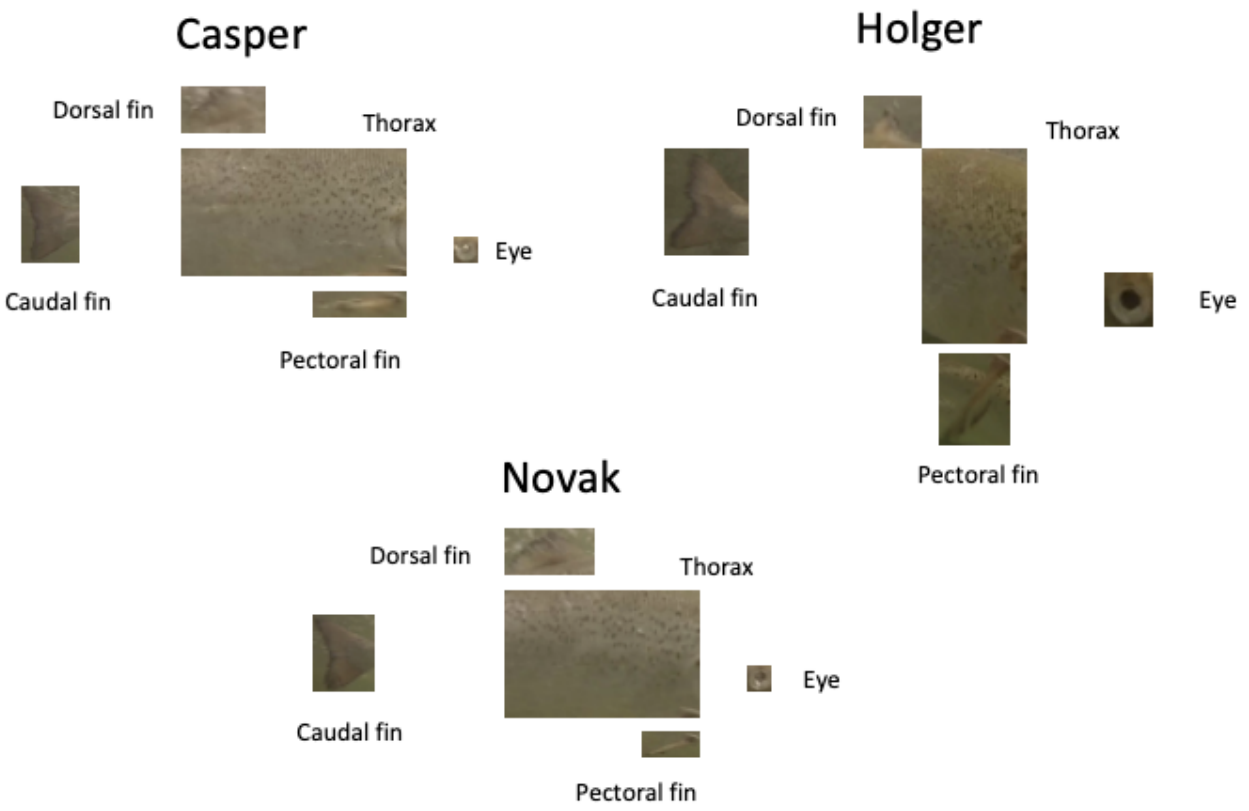


Figure 4.4: Re-identification dataset example annotations. Here, the body parts of three individuals are shown with their corresponding names. Image shapes of body part images can vary greatly, as seen when comparing Casper/Novak and Holger.

ages from a new video in the same tank, it is able to still achieve satisfying accuracy. In figure 4.5, the data augmentation pipeline is shown. To demonstrate the functionality of the data augmentation pipeline, a dorsal fin image is shown before and after data augmentation. The augmented image is exposed to a Color Jitter module and a Random Resized Crop module. The Color jitter module changes the saturation, hue, brightness, and contrast of the input image. This helps the model to become more robust against environmental changes due to illumination or particles in the scene. The Random resized crop module chooses a random position in the image and crops 224x224 pixels, which it outputs. This is because Deep learning models expect images of a certain size. In this case, the re-identification models expect the images to be 224x224 pixels, as the architecture is built around this specific image size. Feeding images of other sizes to the re-identification models would yield non-optimal results. The random cropping adds a layer of complication, which increases the models´ robustness against fluctuations in bounding box prediction.

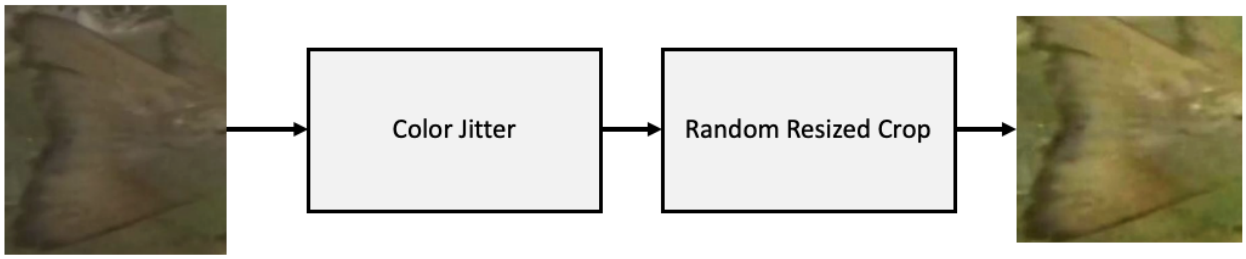


Figure 4.5: The functionality of the dataset augmentation pipeline is presented with an example of a dorsal fin image. An image is put into a Color Jitter module. In the Color Jitter module, the image brightness, hue, saturation, and contrast are altered. The image is further input to the Random Resized Crop module, which crops the image using a random position with a size of 224x224 pixels. The output image is ready to be used in model training for Module 3 of the re-identification pipeline.

4.3 Dataset Challenges

Using video recordings from a dynamic environment introduces multiple challenges. As the fish are crowded in the tank, it makes it difficult to collect data. It is important to be able to see the body parts clearly as only then can you properly determine which individual it belongs to. The conditions inside the tank are made to mimic the environment in a commercial fish farm. Therefore, the fish density inside the tank is 54 $\mathrm{kg\,m^{-3}}$, [30]. The density makes it difficult for the camera to view the full body of fish individuals over a periods of time. It was decided that images of fish are captured and analyzed individually. By first, detecting a single salmon, then detecting its body parts, and finally analyzing them for re-identification. While working with the datasets, it has been experienced that it is difficult to capture quality images for all individuals in the tank. The constant change in position for each fish makes it difficult to choose which frames to annotate. We chose only to annotate images where the whole fish and its body parts are totally visible and discard any images where at least one body part is occluded. This was decided as occlusion from body parts of other fish introduces additional complications. Because the images are labeled as one individual, an image containing multiple individuals can lead to undesirable learning of the Deep learning network. Such a scenario may be when an image contains dorsal fins of Casper and Novak, and the image is labeled Casper. From this image a Deep learning network might learn that the dorsal fins for Casper and Novak belong to Casper. To avoid these complications, the occluded images were discarded. This introduces both positive and negative effects. On the positive side, the quality of the images in the dataset remains high, as the dataset exclusively contains body part images with one correct label. On the negative side, the number of images in the video recordings that hold this quality is few. This means that a smaller dataset size must be accepted as image quality is prioritized.

To efficiently annotate images from the video recordings, AI annotation software was devel-

oped. The objective was to create large datasets semi-automatically, where the detection model does the groundwork by predicting the location of salmon in the entire video recordings. Then, one can manually look through the model predictions and choose to adjust or discard them. However, the AI annotation software was not trained well enough for the semi-automatic annotation process to be efficient. The model's accuracy was too poor, which led to much time being invested in cleaning false or imprecise predictions. The semi-automatic annotation process was aborted due to time constraints in the project, and the focus was on creating quality annotations by hand. The manual annotations were performed using the annotation software LabelMe, [31]. The video recordings were filtered so that every 5th frame was removed from the recordings. This created more variation in the datasets as images of fish with translational displacement were annotated. It also sped up the manual annotation process as the number of images was reduced. In retrospect, we think it should be easier to semi-automatically annotate salmon body parts as these images have stable quality, as the groundwork is done in dataset 1.

To get satisfactory variation in the datasets is challenging, as the salmon individuals usually prefer to swim in the same location within the tank over long periods of time. In the datasets, many individuals are captured with variations in location and perspective, but this is not the case for all. So if a new video with entirely new perspectives of the salmon individuals were to be analyzed, the re-identification pipeline might have a difficult time predicting their identities with high accuracies as the models understand the individuals based on information about their appearance, but also the context, e.g., the location, perspective, illumination, background, water clarity and so on. However, as the camera placement for each video recording is slightly different, this adds some variation to the data.

Chapter 5

Salmon re-identification approach

This chapter covers the salmon re-identification pipeline. First, an overview of the pipeline is introduced, and thereafter, a description of the model training and evaluation for each module of the pipeline. In each module section, preliminary results are presented and discussed. For the re-identification module, common false predictions for each body part model is presented, as well as a discussion surrounding the possible implications. Finally, the XAI method, Integrated gradients, is used to explain how the thorax re-identification models evaluate images of the thorax for two salmon individuals.

5.1 Re-identification pipeline

In this section, an overview of the pipeline is presented. A visualization of the salmon re-identification pipeline is shown in figure 5.1. The input to the model is images from the datasets presented in the previous chapter. The images are processed through five steps. Three of them are Deep learning models, which we call the three modules of the pipeline. First, the images are processed in Module 1, where salmon are detected. The detection module localizes the salmon. Further, the salmon are cropped out of the image and fed into Module 2, the body part detection module. The body part detection module localizes body parts. Each body part is cropped and fed into one of five salmon re-identification models that are part of Module 3, the salmon re-identification module. Each module in the pipeline is trained separately using their respective dataset. The training process was carried out on the high-performance computing platform IDUN at NTNU. In addition, an XAI method is used to investigate what information in the body part images the re-identification models focus on. In the following, a detailed explanation of each module is presented.

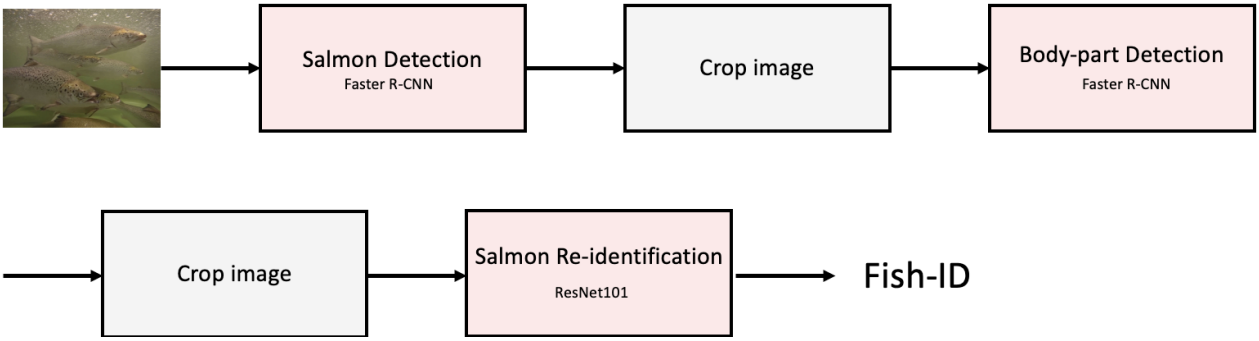


Figure 5.1: The overview of the re-identification pipeline is presented. An image is fed to a salmon detection module, which localizes salmon. The salmon detection module is a Faster R-CNN model with a ResNet50 backbone network. The localized salmon is cropped from the original image and fed to a body part detection module. The body part detection module uses the same detection model as Module 1. It localizes the body parts of a salmon, and they are cropped out into smaller images. Each body part image is converted to a standard size of 224x244 and fed into the salmon re-identification module, which contains five body part models that perform re-identification. There is one body part model for each body part of interest: Thorax, dorsal fin, eye, pectoral fin, and caudal fin. Each model is a ResNet101 convolutional neural network trained to re-identify an individual based on images of one body part.

5.2 Module 1: Salmon detection

Module 1 of the pipeline is salmon detection. This is the process of locating the position of a salmon within an image and drawing a box around its body. One of the objectives of the thesis is to assess the re-identification capabilities of salmon body parts. It is, therefore, important to provide accurate regions of interest to the salmon re-identification module. This is because the accuracy of the salmon re-identification module might get worse if the images only show fractions of body parts. The Deep learning detection model Faster R-CNN is chosen due to its two-stage detection architecture, which increases the accuracy of the predictions but sacrifices some time efficiency, as opposed to one-stage detection models [28]. Pytorch delivers Faster R-CNN with several backbones, but ResNet50 is chosen because it has the most learnable parameters, which has shown the best results for benchmarking tests on ImageNet1k [32]. ResNet50 is a 50-layered version of the image classification model discussed in the previous section. The ResNet-50 is chosen as a backbone for the Faster R-CNN as PyTorch delivers pre-trained weights for it. To add more robustness to the Faster R-CNN model, a version of the detection model with Feature-Pyramid network (FPN) was chosen [33]. This is relevant because we are dealing with data from a complex environment where salmon swim at multiple distances from the camera. By using the FPN, one can obtain good predictions even though there are salmon close and far away from the camera. The model is pre-trained on ImageNet1k. This dataset contains around 3.2 million images, which gives a general foundation for object detection [34]. The salmon de-

tection model is trained on dataset 1 from section 4.1 to specialize the model towards detecting salmon in an aquaculture environment.

5.2.1 Model training

Table 5.1: Optimal model configuration found during training the salmon detector.

Configuration	Value	Description
Backbone	ResNet50	Feature extraction backbone network
Pre-trained	True	Model is pre-trained on ImageNet1k
Batch size	5	Number of images per optimization cycle
Optimizer	SGD	Optimization algorithm
Learning rate	0.001	Initial learning rate for training
Momentum	0.9	Momentum for SGD
Weight decay	0.0005	L2 regularization penalty
Learning rate scheduler	Step LR	Gamma = 0.2 and Step size = 20 epochs
Epochs	25	Total number of training set iterations

The salmon detection model is trained using dataset 1, which contains 1181 images from multiple video recordings. The dataset is divided into training, validation, and test datasets for the model training and performance evaluation. 80% of the dataset is assigned to training and validation. Between training and validation, there is an 80-20 data split. The majority of the data goes into the training set as it is used to teach the salmon detection task to the Deep learning model. The validation set provides unseen data to evaluate the model during training, to find the best learner.

A data loader is implemented to determine how data is put into the model. For every optimization cycle, the data loader samples data randomly into batches. The data batches are used to estimate the direction of the optimal model update. This model optimization approach is called mini-batch stochastic gradient descent and is an efficient approach for model optimization since random sampling using a small set of data points often leads to a good approximation of the gradient [35]. After experimenting with different hyperparameter settings, we found a setting that provided good results on the validation set. These parameters are presented in table 5.1. The parameters were found by were found by changing parameters through trial and error. Early stopping is a regularization method that stops training the Deep learning model before it overfits to the training data. The model was trained for 100 epochs on the training data but performed at its best after training for 25 epochs.

Table 5.2: Test set results for salmon detection.

Metric	Value
mAP@50:95	0.9387
mAP@0.50	0.9965
mAP@0.75	0.9965
mAP (large objects)	0.9387
mAP (medium objects)	-1.0
mAP (small objects)	-1.0

5.2.2 Model evaluation

The salmon detection model is evaluated on a test set of 236 images, which is 20% of dataset 1. Figure 5.2 shows the test set results from the model evaluation. The model mean-average precision (mAP) on predictions within the IoU range 50:95 is 93.87% on the test set. This is a high performance and an improvement from the specialization project, where we achieved a mAP of 88.6% for front-half salmon detection, [36]. This improvement is most likely because the detection set has doubled in size, from 619 images in the specialization project to 1181 in this project. Dataset 1 also contains more individuals and multiple video recordings, which increases the dataset variation and leads to a more robust model. The model performs better when the IoU is lower, as a IoU threshold is less strict. This indicates that the class indicator prediction is performing well even if the predicted bounding box only has a 50% overlap with the actual salmon annotation. Further, the mAP for large objects is equal to the overall mAP which shows that dataset 1 only contains images that are regarded as large, which is 96^2 pixels and above. This suggests that the salmon detection model may perform worse on salmon further away from the camera.

5.2.3 Salmon detection examples

Figure 5.2 shows examples of erroneous predictions from the salmon detector. The detection model has correctly predicted the annotated bounding box, where the red and green boxes overlap, for all images in the figure with high confidence scores. However, the first three images show that additional fish are detected with low confidence. It is a question of definition whether these predictions are correct, but in this project, we only consider fish that are not occluded. As the fish from figure 5.2 are occluded and the confidence for the predictions is low, the predictions can be filtered out using a confidence threshold. However, in the last image, an occluded fish is predicted with a higher confidence than the correct prediction. A confidence threshold cannot filter out this prediction. The model may have more confidently predicted the occluded fish as its body stands more out from the tank's background.

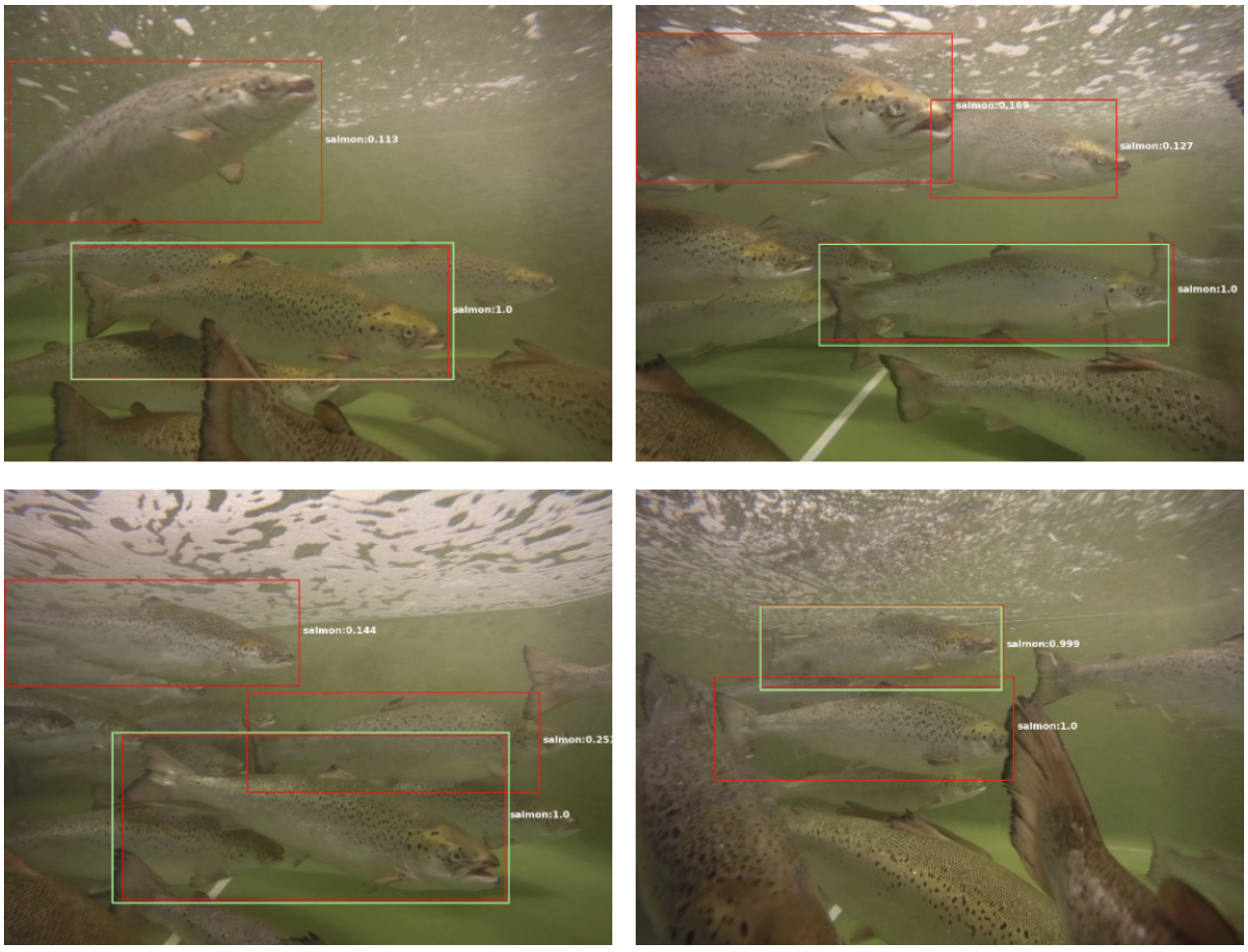


Figure 5.2: Examples of most common erroneous/imprecise predictions from the salmon detector. The green boxes are the annotated bounding boxes. The red boxes are predicted by the model. The predicted class indicator and confidence score are written in white.

5.3 Module 2: Body part detection

The second Module of the pipeline detects salmon body parts. For reasons described in the previous section, body parts are detected using a faster R-CNN model. Like in Module 1, the body part detection model has been pre-trained on Imagenet1k to make a foundation for object detection. The following provides a more detailed description of the model training and evaluation.

5.3.1 Model training

Table 5.3: Optimal model configuration found during training the body part detector.

Configuration	Value	Description
Backbone	ResNet-50	Feature extraction backbone network
Pre-trained	True	Model is pre-trained on ImageNet1k
Batch size	10	Number of images per optimization cycle
Optimizer	SGD	Optimization algorithm
Learning rate	0.001	Initial learning rate for training
Momentum	0.9	Momentum for SGD
Weight decay	0.0005	L2 regularization penalty
Learning rate scheduler	Step LR	Step size = 20. Gamma = 0.5
Epochs	58	Total number of training set iterations

The body part detector is trained on dataset 2, which contains 481 images. The same percentage division from section 5.2 is used for splitting the dataset into training-, validation- and test sets for the body part detector. The model parameters from Module 1 were used as a starting point for training the body part detector. Figure 5.3 shows the model´s configuration after training and validation. The batch size parameter is turned up as dataset 2 contains smaller images. The GPU can, therefore, handle more images in memory simultaneously. After 58 epochs, the model reached its best performance on the validation set, and the training was therefore cut off.

5.3.2 Model evaluation

The body part detector is evaluated on a test set containing 97 images. Table 5.4 shows the results from the model evaluation on the test set. It achieves an mAP@50:95 of 87%. This is lower in comparison with Module 1 of the pipeline, but still a high performance. The lower performance may be explained by the size differences between datasets 1 and 2. Fewer images in the training set lead to worse model training and, therefore, a lower mAP. The lower dataset size for

Table 5.4: Test set results for body part detection.

Metric	Value
mAP@50:95	0.8697
mAP@0.50	0.9958
mAP@0.75	0.9689
mAP (large objects)	0.8829
mAP (medium objects)	0.8394
mAP (small objects)	0.7763

the evaluation set leads to a more uncertain estimated mAP, which can impact the mAP estimate in any direction. The detector achieves a high mAP for large body parts, and it decreases with size. mAP for small objects reflects the performance of eye detection since this is the only body part that covers less than 32^2 pixels for many of the examples. The body part detector is worse at predicting the presence and location of the salmon eyes. The eyes are harder to predict in general as they are small and cover fewer pixels and, therefore, express less information than bigger body parts.

5.3.3 Body part detection examples

Figure 5.3 shows the most common erroneous predictions from the body part detector. The top image shows correct predictions for all body parts except the pectoral fin. Also, the eye is predicted twice in this image, where one prediction is from another individual. The prediction on the other individual has a very high confidence, but in this case, it can be filtered out by selecting the prediction with the highest confidence value for each body part. In the bottom image, the body part predictions have a high IoU except for the dorsal fin. It has predicted the dorsal fin to be larger than it is, likely due to the foam on the water surface behind the fin. This may indicate that the body part detector has not been trained well enough on images of fish near the water's surface.

5.4 Module 3: Salmon re-identification

The last Module of the pipeline is re-identification. Each body part of the salmon is put into an image classification model. The image classification model links a body part to its individual. ResNet101 is chosen as it is a very deep network, which results in the capacity to extract many highly complex features. A deep instead of a wide network leads to a higher model capacity with fewer trainable parameters due to shortcut connections. Shortcut connections lead to easier training of deep neural networks and higher accuracy than networks with the same

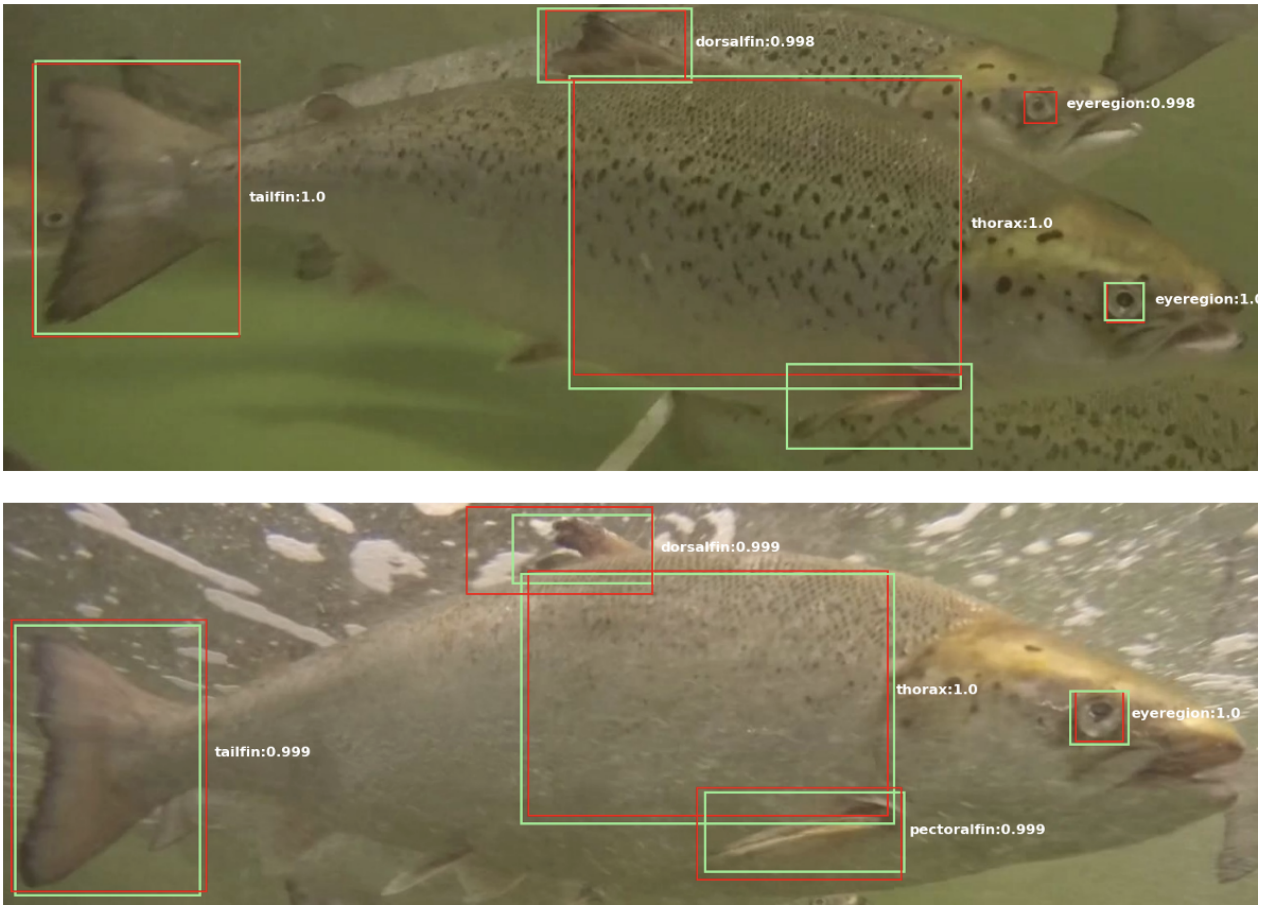


Figure 5.3: Examples of most common erroneous/imprecise predictions from body part detector. The green boxes are the annotated bounding boxes. The red boxes are predicted by the model. The predicted class indicator and confidence score are written in white.

depth, [26]. The model is trained on dataset 3, which contains 330 images per body part and, in total, 1650 images. The model is trained with a cross-entropy loss function 3.1, as this is a multi-classification problem.

5.4.1 Model training

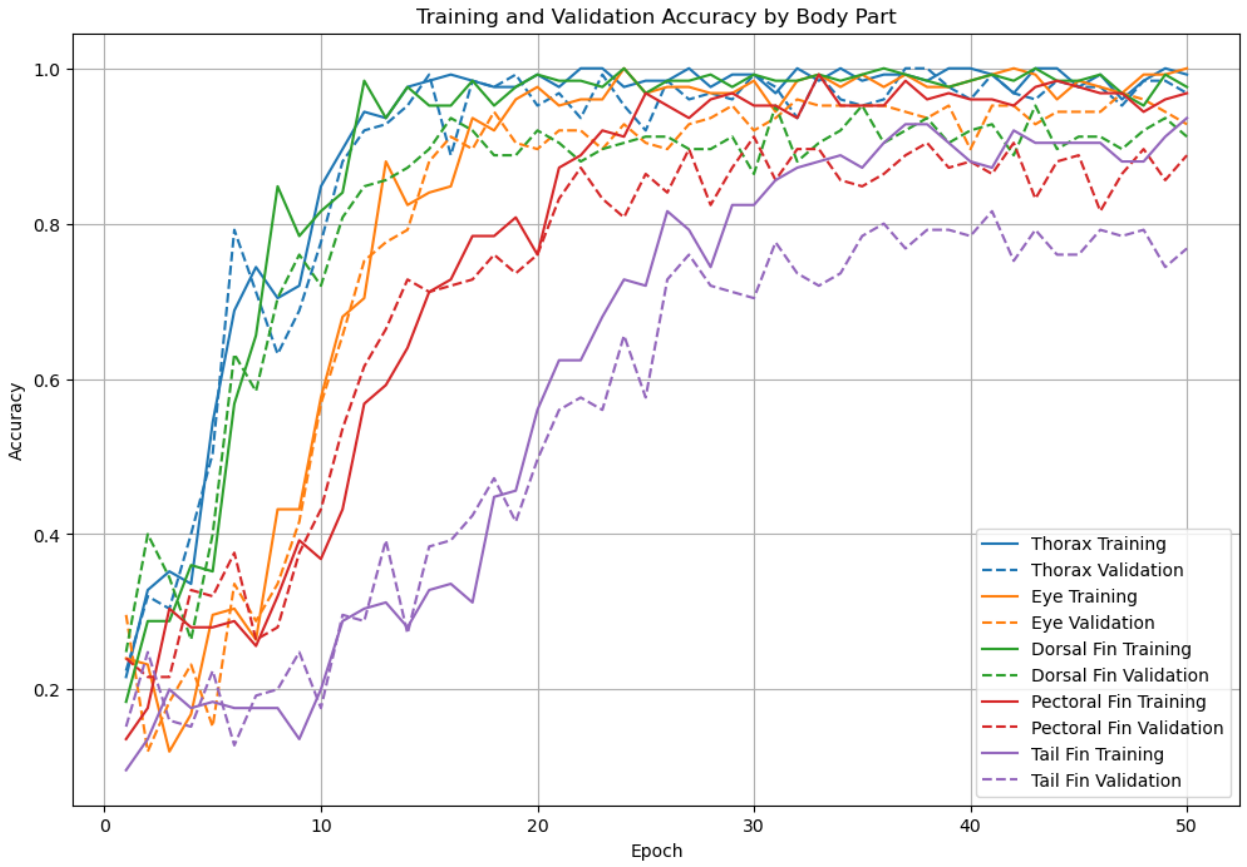


Figure 5.4: The figure shows the training process for each body part re-identification model using Rank-1 accuracy as the performance metric. Each body part model is assigned a unique color for the training- and validation curves.

There are in total five re-identification models. Every body part from the salmon has its re-identification model. The training and validation set for each of the models contains 125 images. To avoid class imbalances between individuals, the number of training images per individual is set to the individual with the lowest occurrences in the dataset. Figure 5.4 shows the accuracy for each re-identification model during training. The solid lines show the training accuracy, and each model is assigned a color. The validation accuracy is shown by the dotted lines. For the thorax re-identification model, the training and validation accuracy are very close over the training period, which indicates that the sets are similar. The re-identification models for thorax,

eye, and dorsal fin converge quickly to top accuracy on the training set when comparing them to the models for pectoral fin and tail fin. This indicates that their training sets are easier to learn from or that these body parts are more informative. However, to get a more realistic view of how the models perform, we need to review the test results.

The model for tail fin performs worse on the training and validation sets than the model for dorsal fin. This is interesting because these body parts are quite similar. Despite this, it seems like there is more biometric information in the dorsal fin than in the tail fin. By comparing the dorsal fin training dataset B.2 to the tail fin training dataset B.5, it is clear that the dorsal fins have more prominent scars and deformations than the tail fins. The scars make it easier to distinguish between individuals using the dorsal fin. Because of this, it might be more suitable for re-identification than the tail fin.

The model for the pectoral fin performs better on the training data than the model for the tail fin. This is surprising, as it is quite thin and small. Therefore, distinctiveness from, e.g., scars is not as prominent. Appendix B.3 shows that there are slight differences between individuals in fin shape. However, the fin angle is stable and unique for each individual. Jannik has its pectoral fin in a more downward position than others. Carlos has a more upward fin than others. Roger is varying the fin position between down and up. Looking into the tail fin training dataset B.5, you can not see the same variation in angle between the individuals, and this may be why images of the pectoral fin are better at distinguishing individuals from each other than images of the tail fin.

5.4.2 Model evaluation

The body part models are evaluated on test sets for each body part. The test set evaluation results for the thorax- and dorsal fin model show the corresponding accuracies: 89% and 87.7%. The eye-, pectoral fin-, and caudal fin models perform worse. Their respective accuracies are 54.8%, 52.1%, and 49.3%. These results indicate that thorax and dorsal fin are best suited for use in a re-identification system.

Thorax

Figure 5.5 shows the results of using the thorax for salmon re-identification. The diagonal of the matrix contains high values in almost every cell, which is good since this is where the predictions and actual values overlap, giving a correct prediction from the model. It has low values for the off-diagonal cells, which is a good result as they represent false predictions. For the thorax model, the most common error is predicting the individual Daniil when shown images of Alexander. Casper is being confused with Stefanos as well. This can be seen in figure 5.5 by locating the off-diagonal cell, which coincides with the individuals of interest. Overall, the re-

identification accuracy using the thorax is 89%, which is a fairly good result. Figure 5.6 presents images of the most common errors for the thorax model. Image 1 shows an example where the model predicted Daniil when shown an image of Alexander. Image 2 shows an image of Daniil correctly predicted by the model. Comparing these images, one can see that the colors of their bodies are similar. Furthermore, there are white colored debris in both images. The melanin spot pattern, however, is not very similar, and the angles of the bodies are not the same. This can indicate that the prediction is based on the color of their bodies and the debris in the images rather than the biometric information from the melanin spot pattern. Image 3 shows an example where the model predicted Casper when shown an image of Stefanos. Image 4 shows an image of Casper correctly predicted by the model. When comparing the two images, one can see that the melanin spot pattern is different. However, they both have small dark scales in the top portion of the body and have their melanin spot pattern mostly beneath these scales. Their melanin spot pattern combines big and small spots, making the patterns seem equal at first glance. This can indicate that the model has based its predictions on the combinations of small dots on the top body and larger melanin spots below.

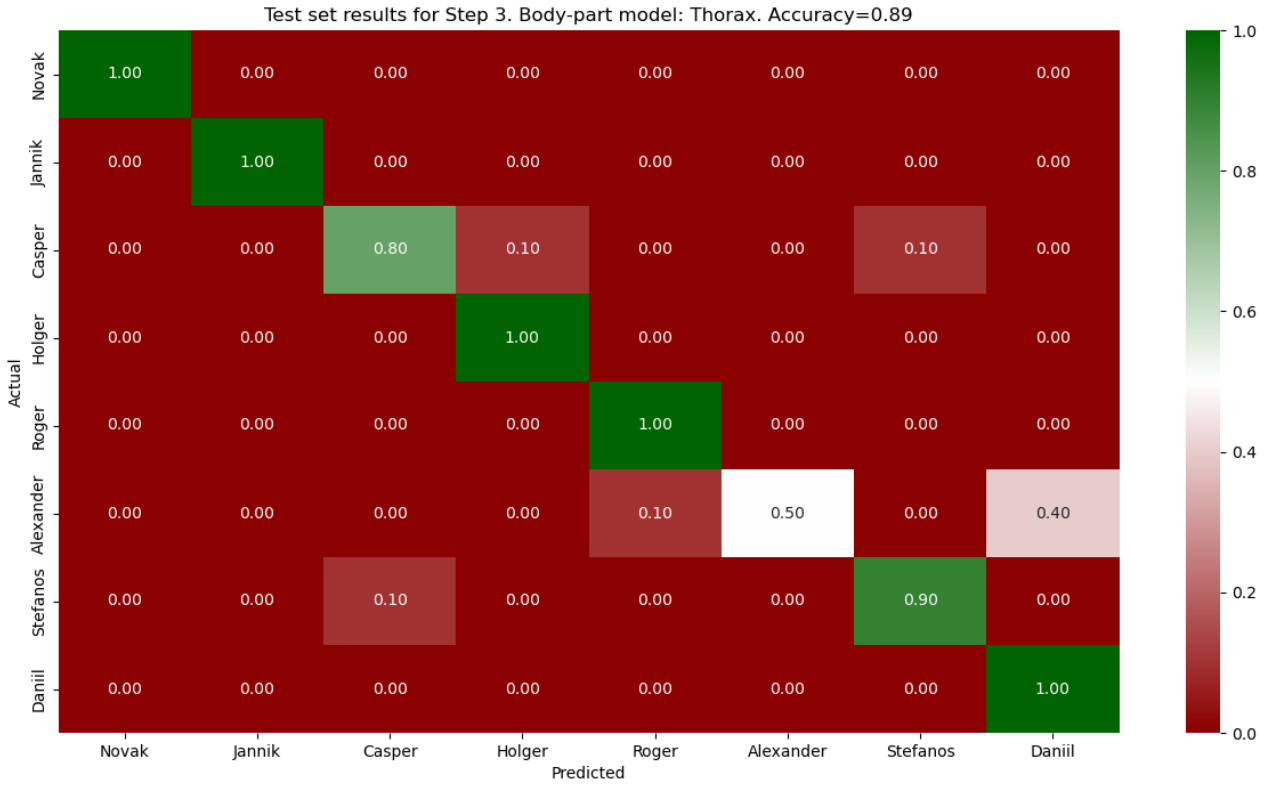


Figure 5.5: Module 3: Salmon re-identification test set results for thorax model. A confusion matrix is shown with a color bar indicating the numeric value for each cell in the matrix.

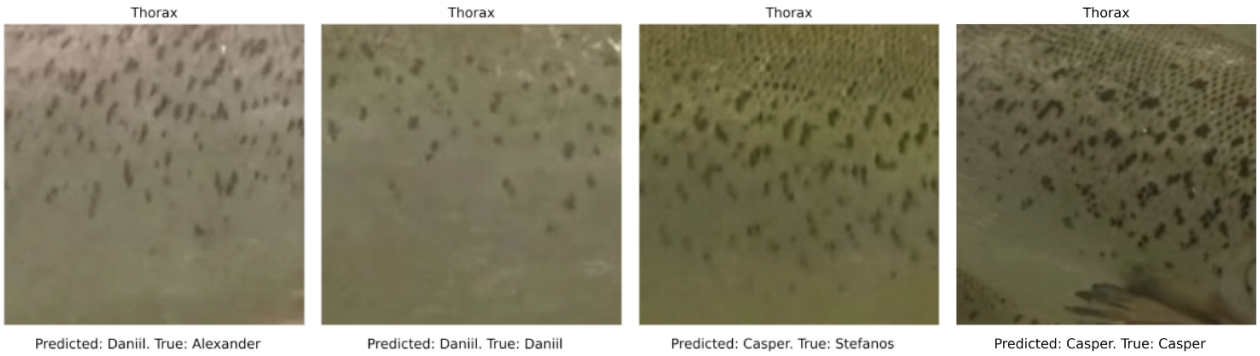


Figure 5.6: Common erroneous predictions from thorax model. Images from left to right: 1, 2, 3, 4

Dorsal fin

Figure 5.7 shows the test results of the re-identification model for the dorsal fin. The most common errors are predicting the individual Holger when shown images of Daniil. Another common is predicting error Stefanos when shown images of Holger. The model achieves an accuracy of 44% on images of Daniil, which is the lowest of all individuals from the evaluation. The model predicts the individual Holger in most cases where the actual individual is Daniil. It achieves an accuracy of 60% on images of Holger, which is the second lowest. The model predicts Holger in most cases where the actual individual is Stefanos. Figure 5.8 shows images of common errors from the dorsal fin model. Image 1 shows an example where the model predicted Holger when shown an image of Daniil. Image 2 shows an image of Holger correctly predicted by the model. By comparing the image, one can see that the image of Daniil is quite blurry, which reduces the texture information on the fin. However, it is possible to view the silhouette of the fin, which looks similar to Holger's in image 2. This indicates that the model focuses on the silhouette of the fins when making this prediction. Image 3 shows an example where the model predicted Stefanos when shown an image of Holger. Image 4 shows an image of Stefanos correctly predicted by the model. By comparing the images, one can see that the silhouettes of the fins are very similar. The color patterns on the fins are also similar, as the top half of the fin is black, and the bottom half is green/brown. There is also a white tint on the narrowest part of the fin, which increases the similarity between Stefanos and Holger. However, the image of Holger is blurry, and more emphasis is likely given to the traits like the silhouette and color pattern.

Eye

Figure 5.9 shows the test results of the re-identification model for the eye. This confusion matrix is more chaotic than the two previous body part models as there are more off-diagonal non-zero cells. The most common error is predicting the individual Novak shows images of Alexander.

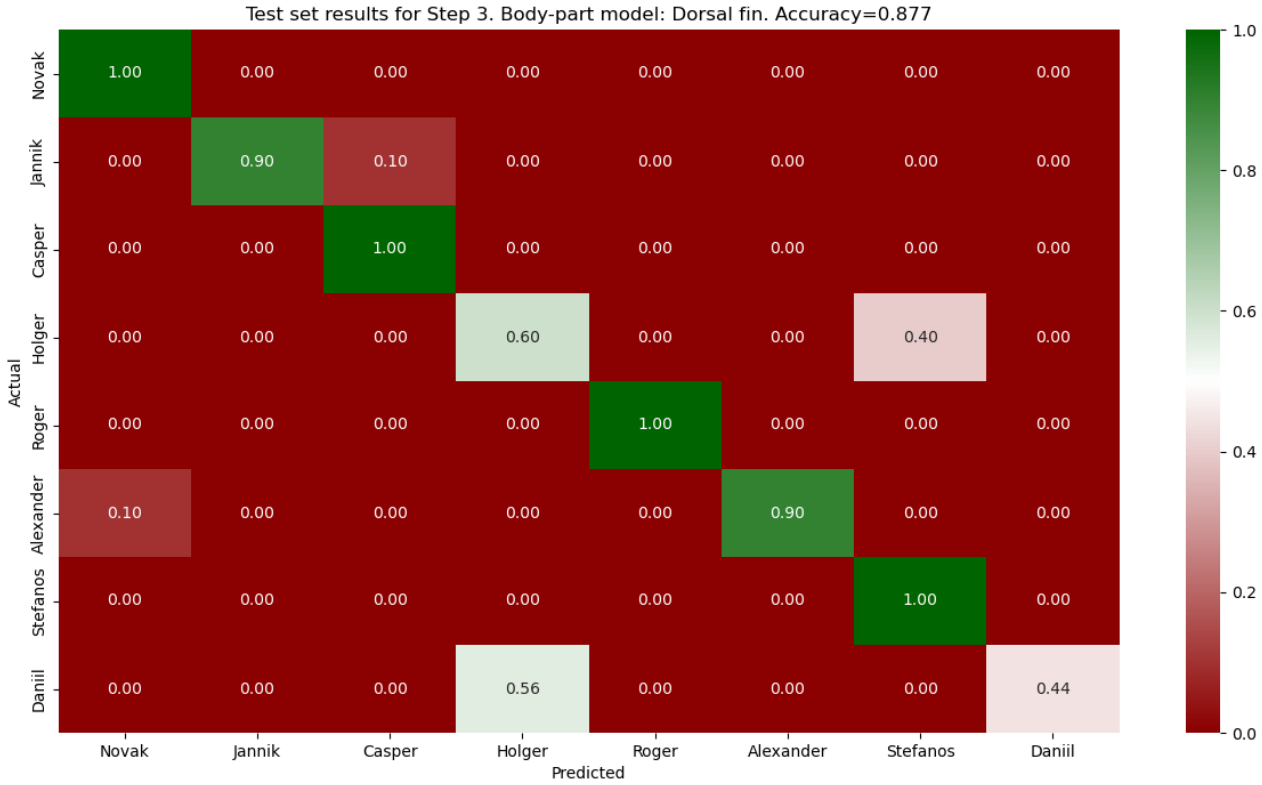


Figure 5.7: Module 3: Salmon re-identification test set results for dorsal fin model. A confusion matrix is shown with a color bar indicating the numeric value for each cell in the matrix.

Figure 5.8: Common erroneous predictions from dorsal fin model. Images from left to right: 1, 2, 3, 4

The model achieves an accuracy of 20% on images of Alexander, which is the lowest in the confusion matrix. The model predicts Novak in most cases where the actual individual is Alexander. It achieves an accuracy of 22% on images of Novak, which is the second lowest accuracy. The model achieves 100% on images of Casper and Holger. Figure 5.10 shows images of common errors for the eye model. Image 1 shows an example where the model predicted Novak when shown an image of Alexander. Image 2 shows an image of Novak correctly predicted by the

model. By comparing the images, one can see that Alexander´s eye is clearly visible as opposed to Novak´s. The size of the eyes is similar, but the iris from Alexander is a bit bigger and more defined than Novaks´s. The iris of Alexander has a white tint on the top half, but it seems that debris covers large parts of the left of Novak´s eye. This indicates that the model has focused on the shape of the eye, and its position in the image to make this prediction. These factors are influenced by the annotation process, as the bounding box is not equally tight on all eye images. Image 3 shows an example where the model predicted Holger when shown an image of Novak. Image 4 shows an image of Holger correctly predicted by the model. By comparing the images, one can see that the iris is ambiguously defined for both individuals, likely due to environmental factors. However, the image of Holger is darker than Novak´s. This indicates that the model focuses on the iris to make the prediction.

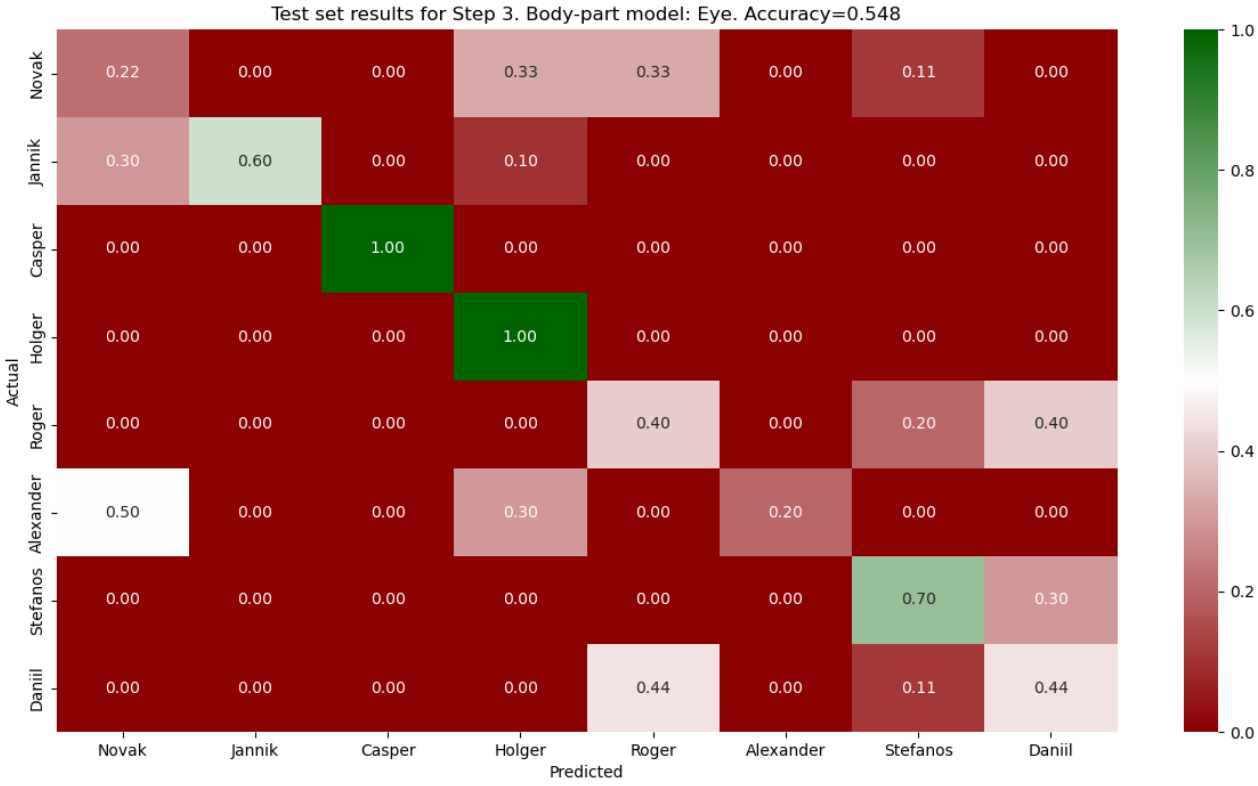


Figure 5.9: Module 3: Salmon re-identification test set results for the eye model. A confusion matrix is shown with a color bar indicating the numeric value for each cell in the matrix.

Pectoral fin

Figure 5.11 shows the test results of the re-identification model for the pectoral fin. The most common error is prediciting the individual Daniil when shown images of Holger. Another common error is predicting the individual Stefanos when shown images of Casper. The model

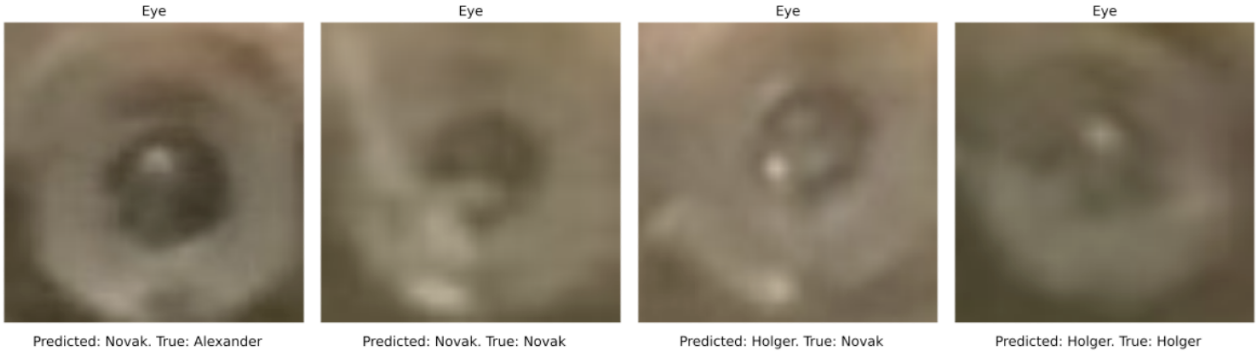


Figure 5.10: Common erroneous predictions from eye model. Images from left to right: 1, 2, 3, 4

achieves an accuracy of 20% on images of Roger and Jannik, which is the lowest in this confusion matrix. It also achieves an accuracy of 100% on images of Daniil. Figure 5.12 shows images of the most common error for the pectoral fin model. Image 1 shows an example where the model predicted Daniil when shown an image of Holger. Image 2 shows an image of Daniil correctly predicted by the model. By comparing the images, one can see that the fins have the same angle from the body. The images are blurry, so no textural information can be used here. This indicates that the model focuses on the fin angle in this prediction. Image 3 shows an example where the model predicted Stefanos when shown an image of Casper. Image 4 shows an image of Stefanos correctly predicted by the model. By comparing these images, one can see that they have the same angle from their bodies; both are facing upwards. However, as the image of Casper´s fin is clearer, it is shown that his fin is wider and darker than Stefanos´ fin. This indicates that the model focuses on the width and color of the fin and its angle to the body in this prediction.

Caudal fin

Figure 5.13 shows the test results for the re-identification model of the caudal fin. The model achieves an accuracy of 11% on images of Daniil, which is the lowest for any individual regardless of body part. It also achieves an accuracy of 22% on images of Novak, which is the second lowest accuracy for the dorsal fin model. It achieves an accuracy of 100% on images of Casper. The most common error is predicting the individual Jannik when shown images of Roger. Another common error is predicting the individual Roger when shown images of Daniil. Figure 5.14 shows images of the most common errors. Image 1 shows an example where the model predicted Jannik when shown an image of Roger. Image 2 shows an image of Jannik correctly predicted by the model. By comparing the images, one can see that both have black regions on the edge of the fin. Jannik has a rough edge, while Roger has a smooth edge. It seems like with more optimization, the model can be able to distinguish between Jannik and Roger correctly, as the roughness of the fin may be good enough biometric information. However, other than the

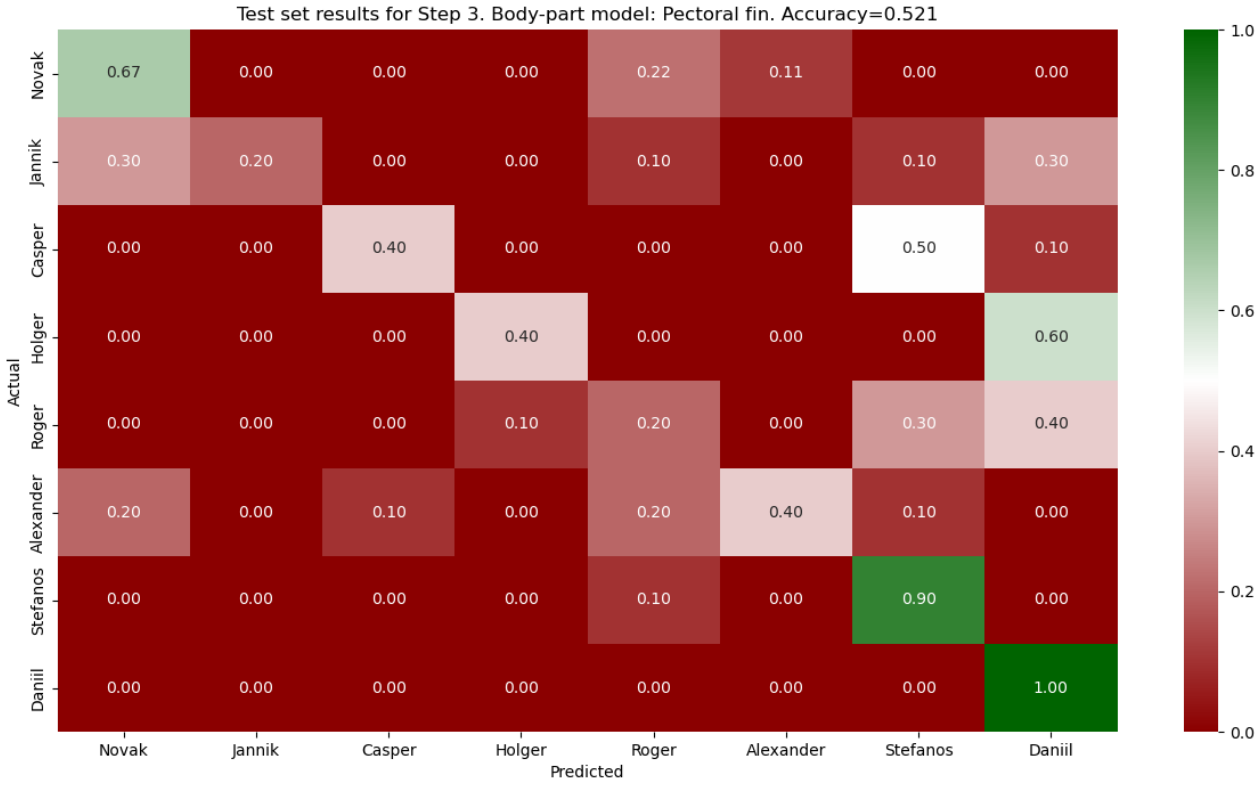


Figure 5.11: Module 3: Salmon re-identification test set results for pectoral fin model. A confusion matrix is shown with a color bar indicating the numeric value for each cell in the matrix.

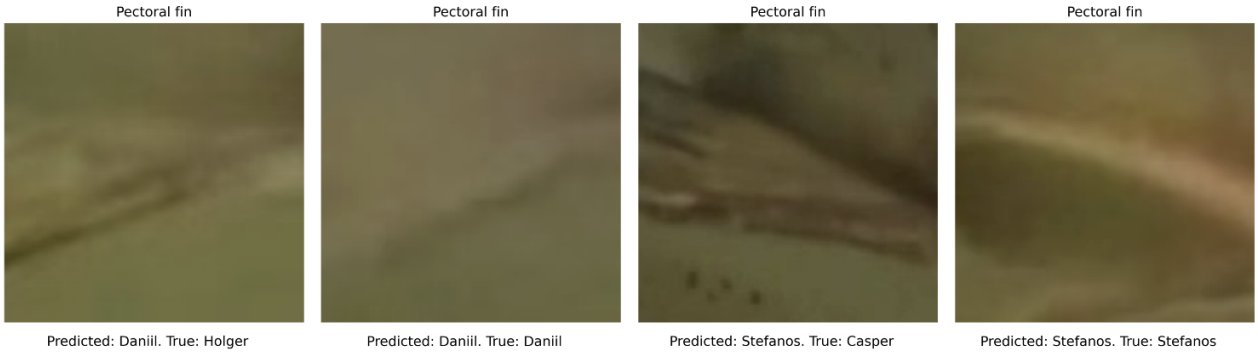


Figure 5.12: Common erroneous predictions from pectoral fin model. Images from left to right: 1, 2, 3, 4

roughness of the fin, the fins are highly similar, making it difficult to distinguish these individuals by using the caudal fin. Image 3 shows an example where the model predicted Roger when shown an image of Daniil. Image 4 shows an image of Holger correctly predicted by the model. By comparing these images, one can see a difference between the edges of the fins. Daniil´s fin is rougher and has a deep cut, which seems to hold enough information for re-identification purposes. Roger´s fin seems darker than Daniil´s, but this may be because of the illumination difference between the images. Roger also has a more pointy fin than Daniil. Here, one can

also imagine that the model can do better with more optimization. Model optimization was not prioritized in this work because of time limitations and there were other more pressing tasks for accomplishing the objectives for the thesis.

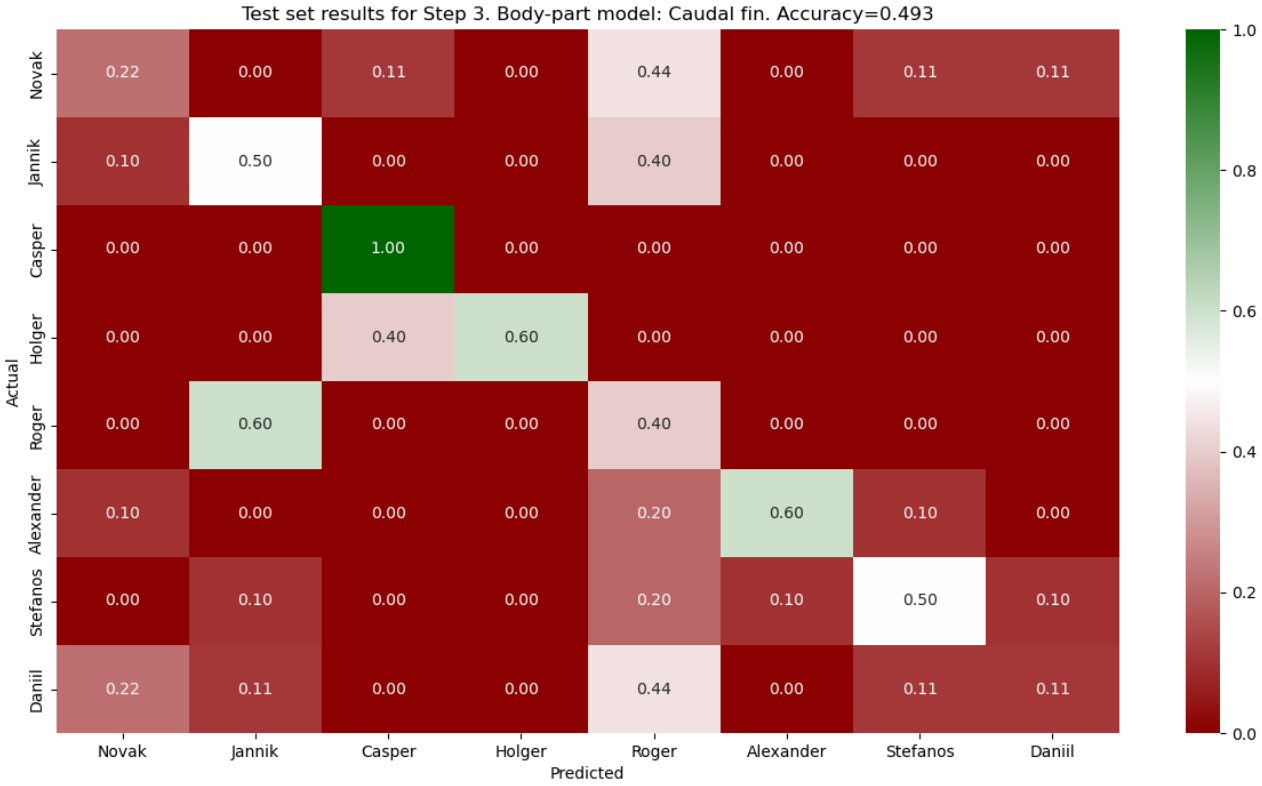


Figure 5.13: Module 3: Salmon re-identification test set results for caudal fin model. A confusion matrix is shown with a color bar indicating the numeric value for each cell in the matrix.

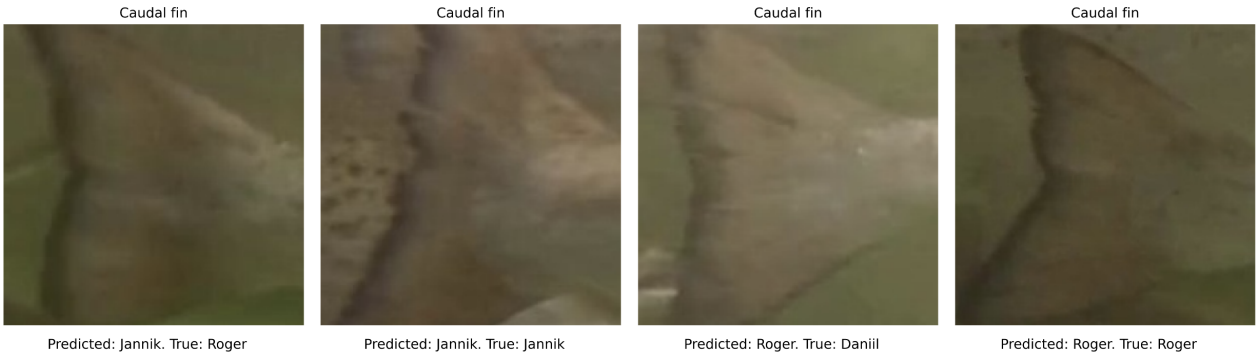


Figure 5.14: Common erroneous predictions from caudal fin model. Images from left to right: 1, 2, 3, 4

5.4.3 Explainable AI on model predictions

Integrated gradients is used during the project to see how the model performs on a detailed level, [21]. This is an XAI method that shows how the model arrived at a prediction. It highlights the most important pixels in the input image. This is useful as it becomes evident if the model uses informative parts of the image, such as the fish, or if it uses unimportant parts of the image, such as the background. By seeing this, you can take a more informed action to improve the model performance, such as adding specific images to the dataset or adjusting data augmentation techniques or parameters. We have used this method and decided to add images to the dataset from different perspectives and locations in the tank, improving model performance. Below, interesting cases for the thorax are shown with the Integrated Gradient masks overlapped.

Figure 5.15 shows eight images of the same salmon individual, Novak. The images are taken from the test set, and one can see that they show different perspectives. Comparing the images, one can see that the model finds the same spot on the melanin pattern in all of the images except for the last one. In the images where it locates this spot, the model confidence is above 80%. However, when it does not manage to find this spot, the confidence drops to around 70%. This tells us that one spot is important for identifying Novak. However, as the confidence is still high when it does not find this spot, it seems like other parts of the pattern are also important. When the angle is different, other parts of the melanin pattern are more focused on. This may indicate that the model has a feature hierarchy, where a combination of several features describes an individual, and some are more important than others.

Figure 5.16 shows images of Alexander overlapped with the Integrated Gradients mask. The top row shows correct predictions, and the bottom row shows false predictions. Across most images, one can see that the model has overfitted to the pectoral fin in the bottom right corner of the images. The pectoral fin is absent in the last two images, resulting in a confidence of 0.5% and 0%. In all the correctly predicted images in the top row, the model clearly highlights the pectoral fin. This shows that the model mostly focuses on parts of the fish that do not have much biometric value, the base of the pectoral fin. It also considers the melanin pattern, but it seems like the fin is the distinguishing factor from other individuals. The model would likely have performed better if the base of the pectoral fin had been cropped out of the images, as it could have focused more on the melanin pattern.

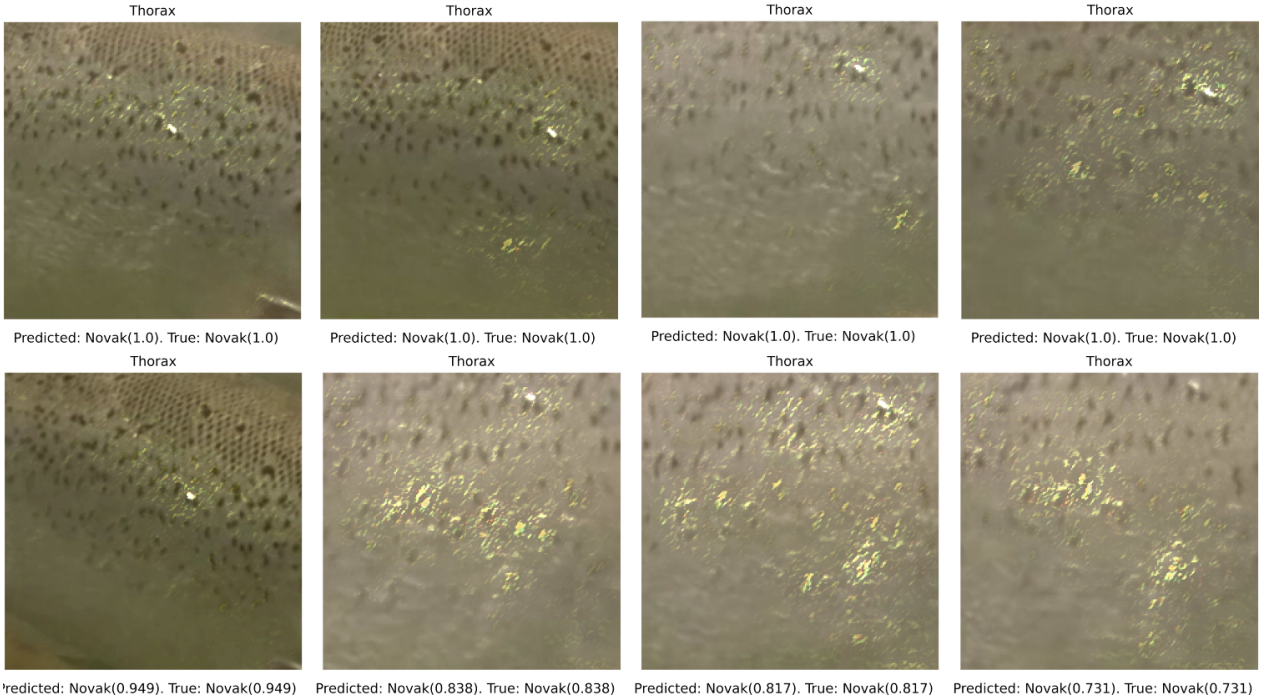


Figure 5.15: An interesting example of the use of the Integrated Gradients method to show which parts of the image are important for the prediction of the model. The input image is shown with overlaps of the Integrated gradients masks, which highlight the important pixels of the image.

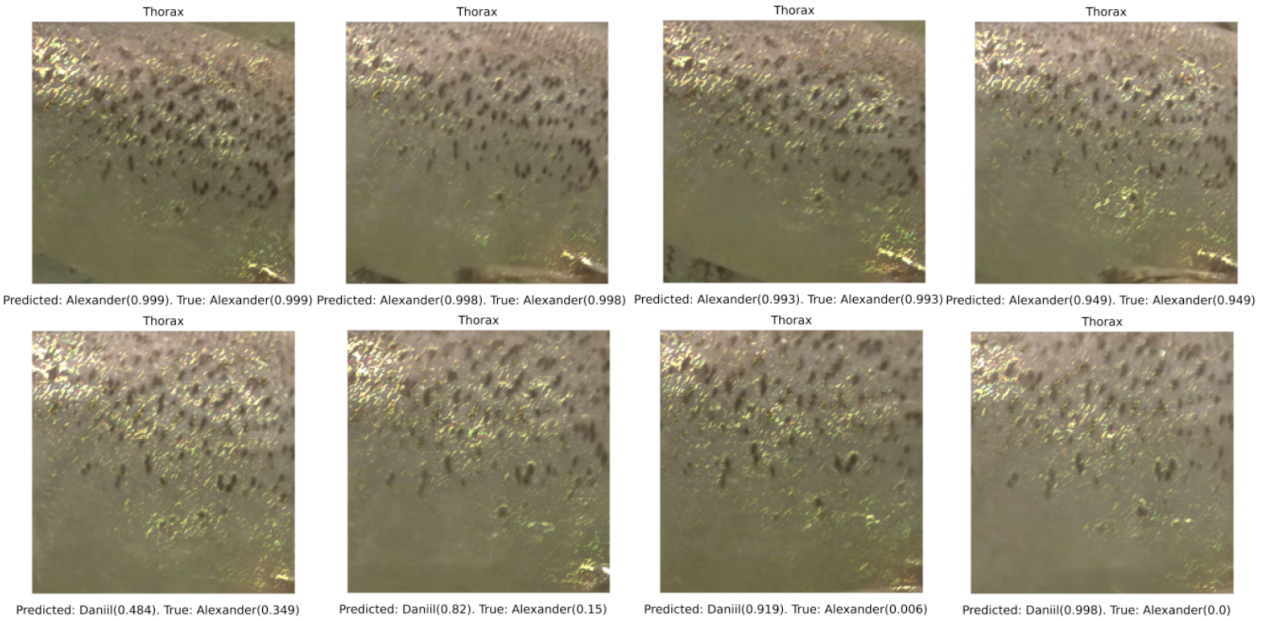


Figure 5.16: An interesting example of the use of the Integrated Gradients method to show which parts of the image are important for the prediction of the model. The input image is shown with overlaps of the Integrated gradients masks, which highlight the important pixels of the image.

Chapter 6

Results from salmon re-identification pipeline and discussion

This chapter presents the evaluation results for the salmon re-identification pipeline and discusses the results of the re-identification of salmon using body parts in a bigger context.

6.1 Pipeline evaluation

To test the body-part re-identification pipeline and compare its results to the evaluation of Module 3 in section 5.4.2, assumptions about the incoming data and predictions are made. The pipeline is tested with the same test set as in Module 3.

Assumptions

1. The predicted salmon with the highest confidence belongs to the annotated individual.
2. For each body part, the prediction with the highest confidence belongs to the annotated individual.
3. Calculating the accuracy of input-output in Module 3 is equal to calculating the accuracy of the input-output in the pipeline.

Assumption 1 is made because it eases the implementation of the pipeline. Datasets 1, 2, and 3 do not include annotations for all individuals in the video recordings, and it is necessary to make some assumptions to evaluate the pipeline. It is assumed that the detection model in Module 1 always predicts the individual annotated for each image. For some images, there is more than one individual that is annotated, but this is only for a small set of images, and the error from this is, therefore, small compared to the dataset size. A false prediction in Module 1 can carry errors through the rest of the pipeline. Considering the performance of Module 1,

which is mAP@50:95 = 93.87% on the test set, assumption 1 does not always hold, but it is a good assumption for most images. Assumption 2 is made because it eases the implementation of the pipeline. Dataset 2 contains images of, at most, a single whole salmon. However, this does not guarantee that parts of other salmon are not present within these images. Considering the performance of Module 2, which is mAP@50:95 = 86.97% on the test set, assumption 2 does not always hold, but it is a good assumption for most images. A false prediction in Module 2 can lead to imprecise accuracy calculation in Module 3. There are also cases where Module 2 fails to predict a body part. Consequently, the test set becomes smaller as Module 3 gets less input data, and the resulting accuracy estimate becomes more uncertain as it is based on fewer examples. Assumption 3 is made because it simplifies the calculation of accuracy for the pipeline. It is assumed that the accuracy of Module 3 with input from Modules 1 and 2 is equal to the total accuracy of the pipeline. This implicitly assumes that Modules 1 and 2 perfectly predict the presence of an individual. In general, this is not a good assumption. However, with our dataset, the assumption is good as the fish and its body parts are visible in each image. By accepting these assumptions, we are estimating the accuracy of Module 3 with bounding box fluctuations given by Modules 1 and 2. Keeping the assumptions in mind, the accuracy estimate is calculated based on the input and output of Module 3. This way, we can compare the accuracy estimate in section 5.4.2, and see how fluctuations in bounding box predictions from Modules 1 and 2 affect the re-identification accuracy in Module 3.

Figure 6.1 shows the test set accuracies achieved by the body part re-identification pipeline. The pipeline achieves both high accuracies for images of the thorax and dorsal fin with 87.7% and 86.3%, respectively. The accuracy of Module 3 on the thorax and dorsal fin from section 5.4.2 is 89% and 87.7%, respectively. Comparing the results from figure 6.1 to results from the evaluation of Module 3 in figures 5.5 and 5.7, the re-identification accuracies on images of the thorax and dorsal fin have dropped by 1.3% and 1.4%, respectively. The re-identification accuracies on images of the Eye, pectoral fin, and caudal fin perform worse than for images of the thorax and dorsal fin. Re-identification on images of the eye, pectoral fin, and caudal fin performs with an accuracy of 50%, 43.1%, and 49.3%, respectively. The accuracy of Module 3 on the eye, pectoral fin, and dorsal fin from section 5.4.2 is 54.8%, 52.1%, and 49.3%, respectively. Comparing the results from images of the eye, pectoral fin, and the caudal fin from figure 6.1 to the evaluation of Module 3 in figures 5.9, 5.11, and 5.13, the accuracies have dropped 4.8%, 9%, and 0% respectively. The re-identification accuracy for images of the pectoral fin has dropped below the caudal fin, making it the worst-performing body part for re-identification. For images of the eye and pectoral fin, the re-identification accuracy has dropped considerably. This indicates that images of these body parts are inaccurately located in Module 2. On the other side, using images of the caudal fin, the re-identification accuracy has not dropped at all. This indicates that the caudal fin is precisely located in Module 2.

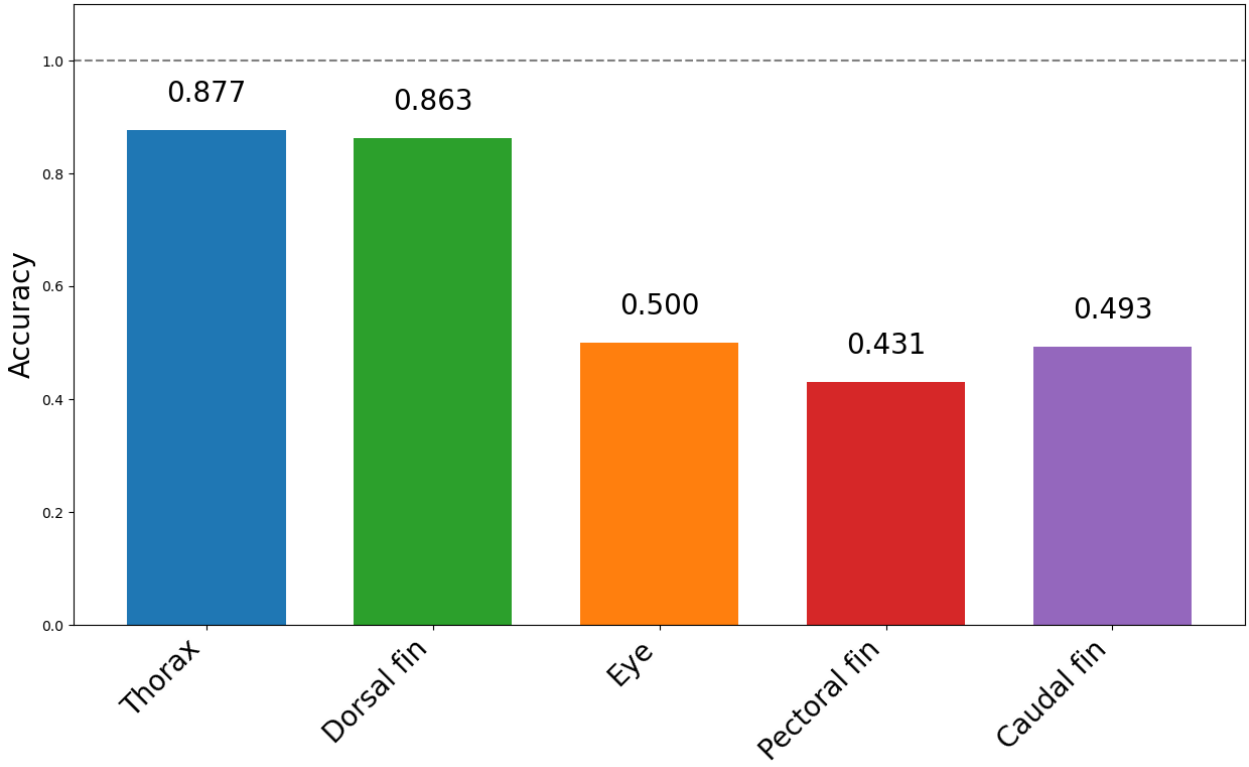


Figure 6.1: Test set results for salmon re-identification pipeline. This shows the accuracy of the salmon re-identification model using images of multiple body parts.

6.2 Discussion

The results presented in the previous section show that using images of the thorax and dorsal fin achieved the highest re-identification accuracy on the test dataset constructed in this thesis. As the thorax contains a melanin spot pattern, which is unique for every salmon, this body part is expected to be highly informative. Therefore, it is no surprise that using the thorax for re-identification yields the most accurate results. However, to our knowledge, no publications exist using images of the dorsal fin for re-identification. The dorsal fin does not exhibit any unique color pattern like the thorax. However, it is prone to damage and deformation, as seen in the video recordings. The dorsal fin, therefore, has a unique look for the individuals in this experiment. However, scars usually regenerate over time. Therefore, models using images of the dorsal fin or any other fin may need to update their database continuously to maintain re-identification accuracies over time.

The re-identification models using the eye, pectoral fin, and caudal fin show weak performance compared to the models that utilize images of the thorax and dorsal fin. Some publications use images of the iris for re-identification, as seen in section 2.2. They achieve higher accuracy in their work than what we observe with our data. One has to keep in mind that their data is

very different from ours. The data used here may not be appropriate for eye/iris re-identification as the salmon swim in an occluded environment, and the distances from the camera reach far. From figure 5.10, one can see that the depicted salmon body part images have a varying degree of obscurity due to varying distances, illumination, and debris in the water. Furthermore, the resolution of the eye images is low, and the pattern of the iris is, therefore, harder to differentiate for a neural network. No publications on using images of the pectoral and caudal fin for re-identification exist to our knowledge. Similar to the dorsal fin, these fins are unique regarding scars and deformities, but to a lower degree. Also, the pectoral fin is usually angled in such a way that it is difficult to capture the whole fin from where the camera is located, as can be seen in figure 5.12. The caudal fin is continuously moving, thereby changing its shape. This makes it difficult to differentiate between pose variability and individual variability, which is likely why the caudal fin re-identification models perform so poorly.

The number of individual fish analyzed in this project is a limiting factor. Out of 30 individuals, 8 were analyzed using the re-identification pipeline. A higher number of individuals will give a more accurate estimation of the re-identification accuracy, as a large set of individuals will give more confidence in how re-identification models will perform in general for the Atlantic salmon species. 16 individuals were annotated, but due to the large image imbalances between individuals, the 8 individuals with the most annotated images were analyzed. 250 images per individual are used for training, and 10 from a new video are used for evaluating the models and pipeline. More data is always desired; however, as we achieve high accuracies for some body parts, it seems that one can also manage to train deep learning models for re-identification with limited data.

The results from this thesis help build insight into how certain parts of the salmon are better suited for re-identification than others. With this knowledge, one can build faster and more robust algorithms in the future. By focusing on the fish's most informative parts, one can manage with less data than using other parts of the fish. It allows the use of multiple body parts simultaneously to determine an individual's identity. A voting system, where each body part votes for an individual, can improve re-identification accuracy, as relying on more than one trait can provide more stable results. For example, one can make the most informative body parts, such as the thorax and dorsal fin, carry more weight in the decision.

The physical re-identification methods like tagging and fin clipping, which hurt the fish, may soon be replaced by more salmon welfare-friendly methods based on, e.g., computer vision. The results presented here are a step towards using computer vision as the main re-identification method in research and production. This is important also due to the costs and time of physically tagging fish. With computer vision, the cost is potentially lower, as placing a few cameras in a farming facility can cover a large volume inside the cage. Innovation in re-identification algorithms can save fish farmers cost and time in the future.

Chapter 7

Conclusion and Recommendations for Further Work

In this Master´s thesis, a salmon re-identification pipeline is successfully developed. The pipeline takes images from video material and predicts the individuals based on body part images. The salmon detection module achieves an mAP@50:95 of 93.87%. The salmon body-part detection module achieves an mAP@50:95 of 86.97%. The salmon re-identification module is split into models for each body part of the salmon. Evaluating the re-identification module on a separate test set yields the following accuracies for each body part: Thorax (89%), dorsal fin (87.7%), eye (54.8%), pectoral fin (52.1%), and caudal fin (49.3%). By evaluating the pipeline as a whole, the re-identification accuracy on images for each body part is the following: Thorax (87.7%), dorsal fin (86.3%), eye (50.0%), pectoral fin (43.1%), and caudal fin (49.3%).

Objective 1 is to develop a salmon re-identification pipeline with Deep learning models using body parts to re-identify an individual. As seen in chapter 5, this objective is accomplished. The pipeline takes underwater images of salmon as input. Individual salmon are detected in Module 1 with a Faster R-CNN model. They are cropped out of the image and sent into a body part detection model in Module 2 with a Faster R-CNN model. Finally, in Module 3, the body parts are sent into body part re-identification models that assign body parts to individuals. Objective 2 is to analyze the body parts using the re-identification pipeline to find the body parts with the most biometric information. As seen in chapter 6, this is accomplished, and the results from this analysis are presented in figure 6.1.

In chapter 2, salmon welfare and the current state of the art within salmon re-identification are discussed. Relevant theory for the techniques used in this project is found in chapter 3. Datasets are constructed for each module and are used for model training and evaluation, as shown in chapter 4. In chapter 5, the developed re-identification pipeline and module evaluation is presented. Evaluation of the pipeline is presented in chapter 6, along with a discussion about the significance of the technology.

The results from this thesis are a step toward automatic welfare analysis of farmed Atlantic salmon. Knowing which parts of the salmon's body give information about the individual´s identity is valuable. This knowledge helps construct faster and more robust re-identification algorithms. Being able to re-identify each individual in a fish farm makes it easier to evaluate each individual´s welfare. When you know how the fish population perceives their life, you can take appropriate action to improve it if needed.

There are things that can be done to improve the approach and results in this thesis. For the short term, optimizing the re-identification models for each body is a way to gain higher accuracy for the re-identification. One can tune the model's hyperparameters by intuition or do a more systematic approach by using grid search, random search, or Bayesian optimization. Another thing is implementing a Siamese network to perform re-identification based on the distance between images. Such an approach can be generalized to open-set re-identification. This is good as one can compare the similarity between an individual in the training set and individuals not present in the set. There is, however, no guarantee that this approach would perform better than the one in this thesis on the closed-set task, but it would be interesting to compare the performance of these approaches. In the medium term, one can develop a semi-automated annotation tool that makes datasets from video recordings. Such a tool can be iteratively improved by manually inspecting, adjusting, and re-training accordingly. One can also add new body parts to the dataset so that more body parts are analyzed. A known problem within re-identification is perspective variation. One can lower the perspective variation between images by transforming the input data to a standardized perspective. This can make the re-identification models more robust towards perspective changes. In the long term, one can create other approaches for re-identifying salmon. One approach can be annotating the melanin spots on the salmon body using bounding boxes or segmentation masks. As mentioned above, one can semi-automate the annotation process, which can drastically improve the dataset creation. Then, one can re-identify salmon based on the pattern directly. One can use detection or segmentation models to extract the meaning pattern from the body, standardize the perspective, and compare patterns using metrics like, e.g., IoU. It would be interesting to compare this approach to the body part re-identification approach in this thesis.

Appendix A

Acronyms

AI Artificial intelligence

COCO Common objects in context

IoU Intersection over union

FPN Feature-pyramid network

R-CNN Region-based convolutional neural network

ResNet Residual network

RPN Region proposal network

XAI Explainable Artificial Intelligence

Appendix B

Training data and results from pipeline evaluation

B.1 Training datasets for re-identification models

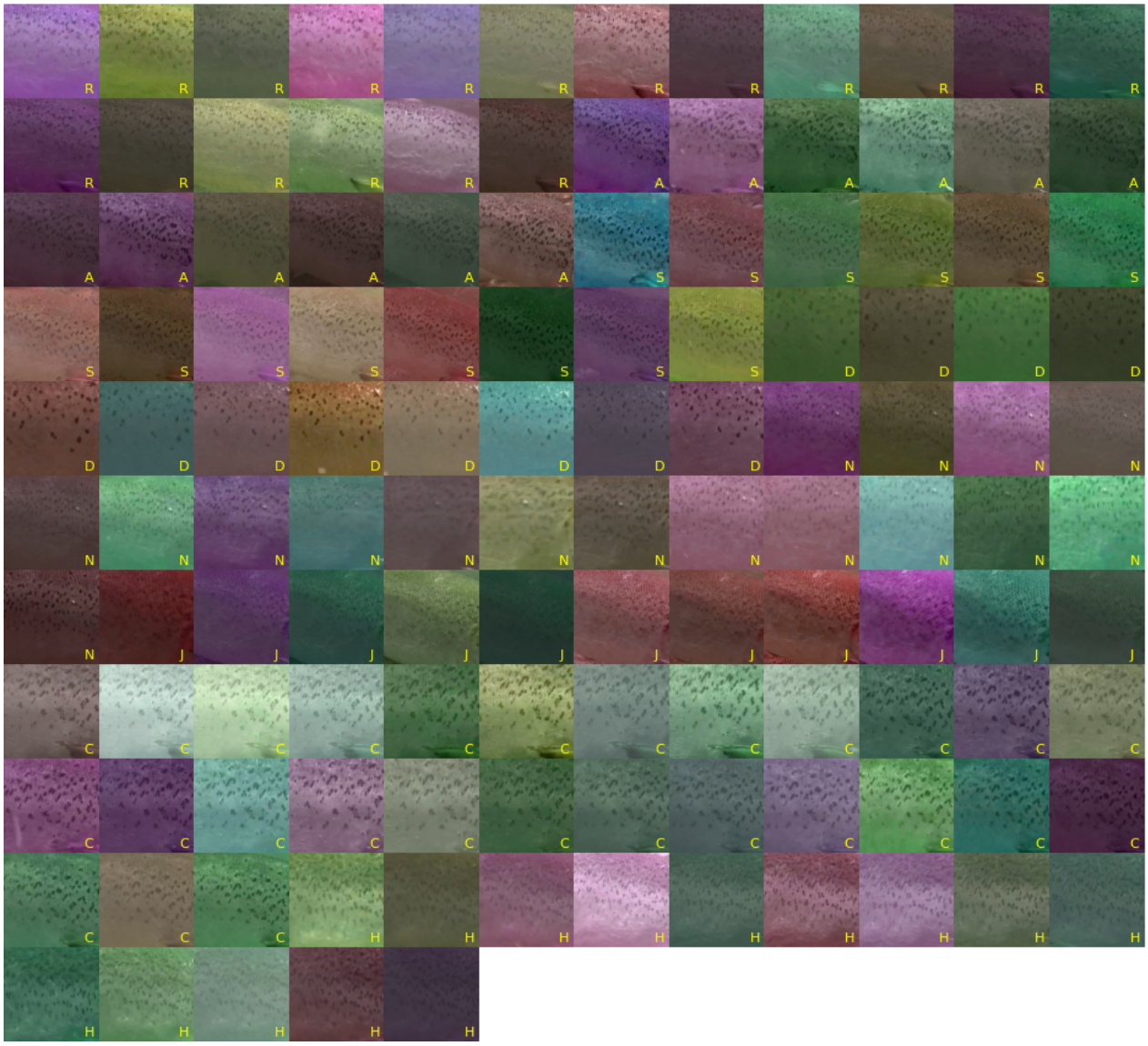


Figure B.1: Training dataset for thorax re-identification model with data augmentation applied. The images in the figure are sorted for ease of comparing images for different individuals. The letter in the bottom right of every image is the first letter of the name of the individual.

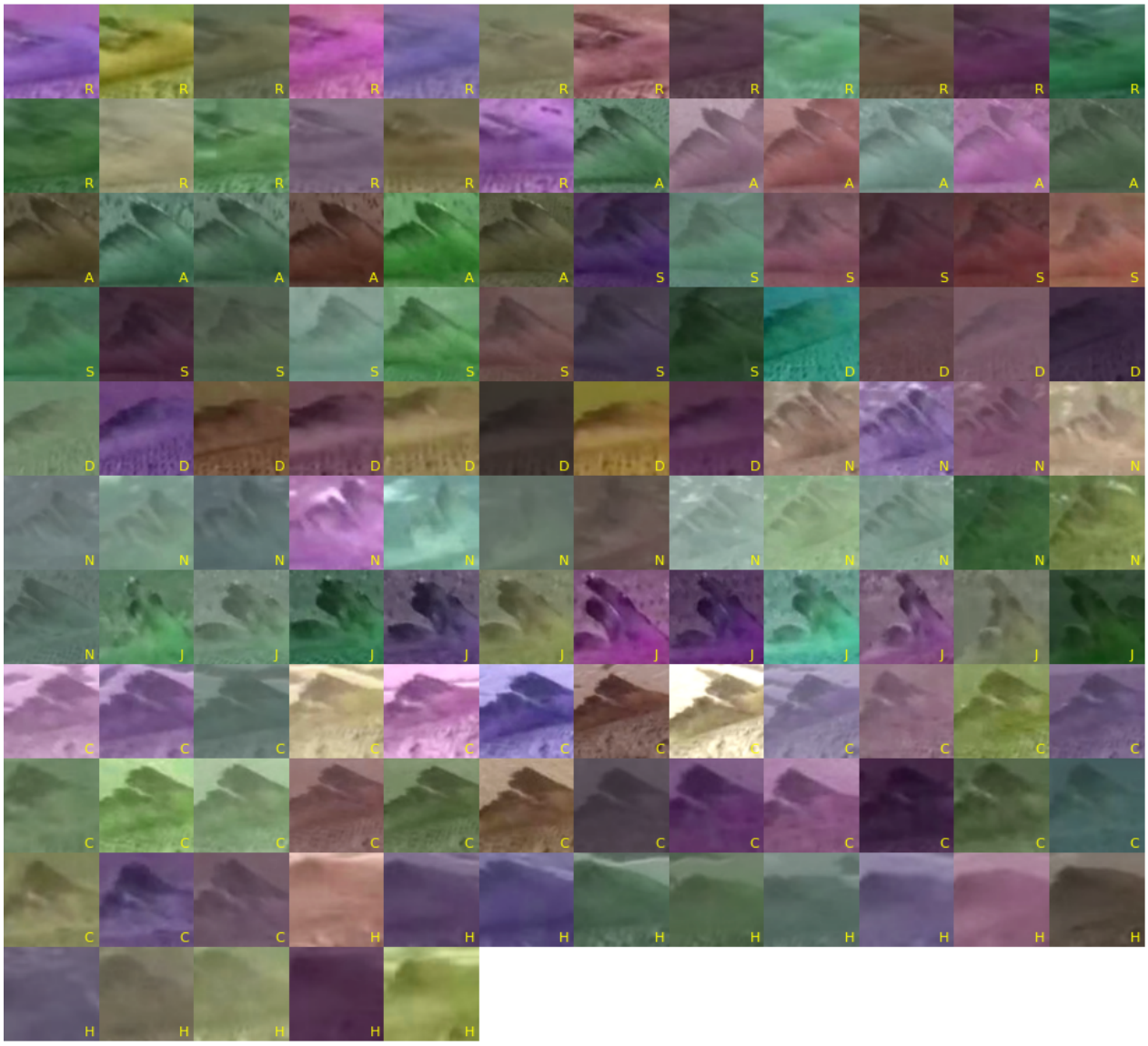


Figure B.2: Training dataset for dorsal fin re-identification model with data augmentation applied. The images in the figure are sorted for ease of comparing images for different individuals. The letter in the bottom right of every image is the first letter of the name of the individual.

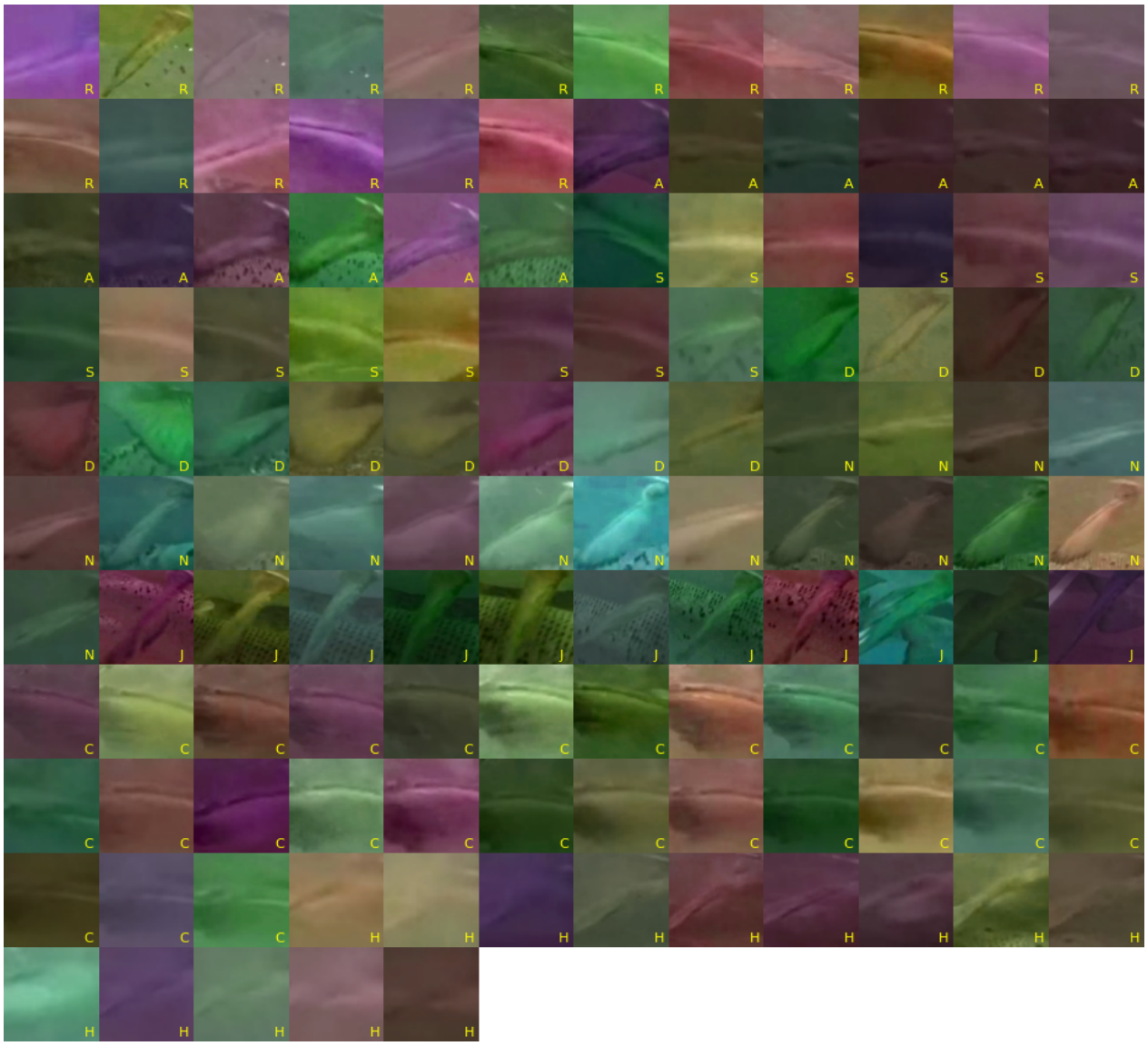


Figure B.3: Training dataset for pectoral fin re-identification model with data augmentation applied. The images in the figure are sorted for ease of comparing images for different individuals. The letter in the bottom right of every image is the first letter of the name of the individual.

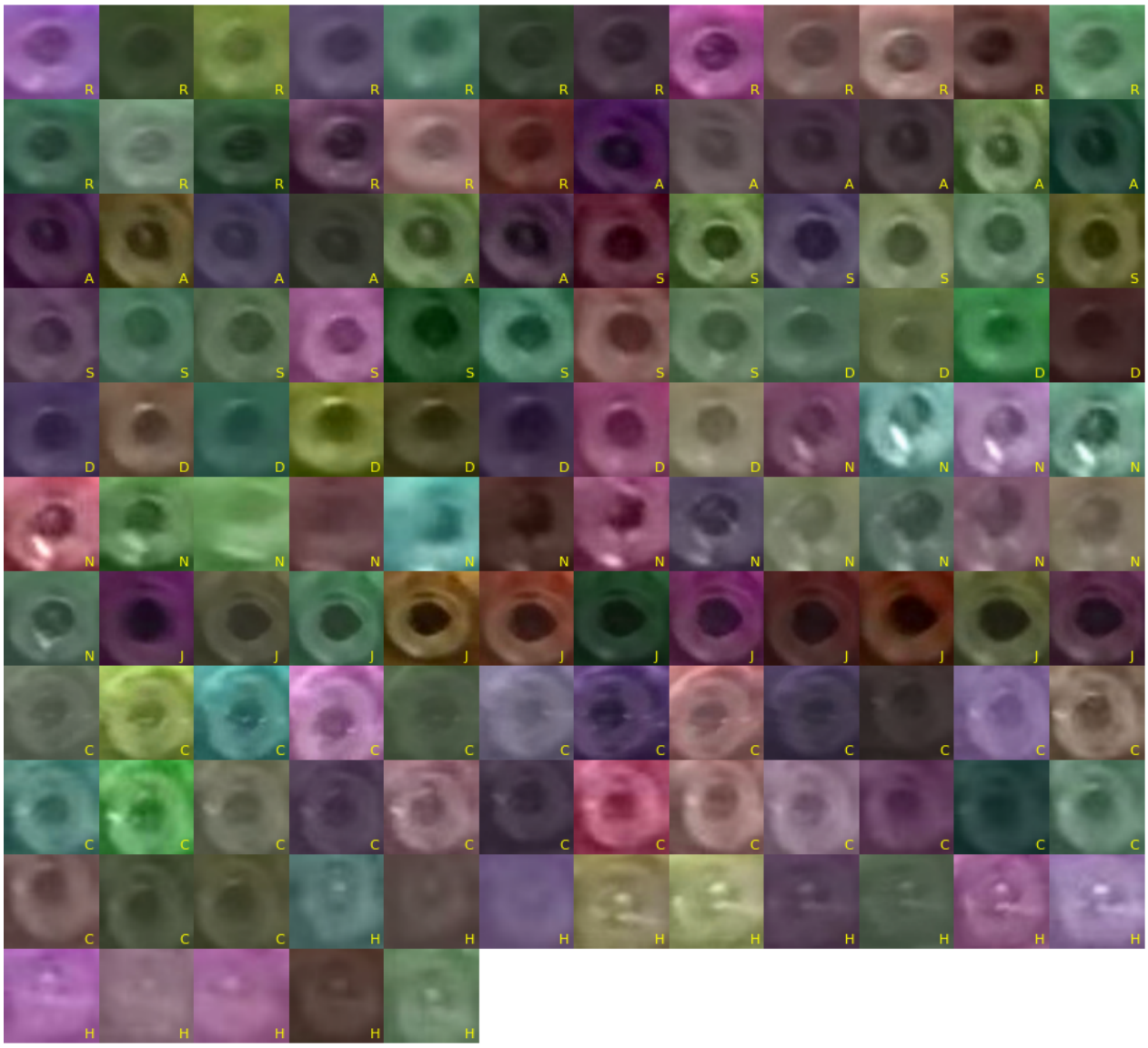


Figure B.4: Training dataset for eye re-identification model with data augmentation applied. The images in the figure are sorted for ease of comparing images for different individuals. The letter in the bottom right of every image is the first letter of the name of the individual.

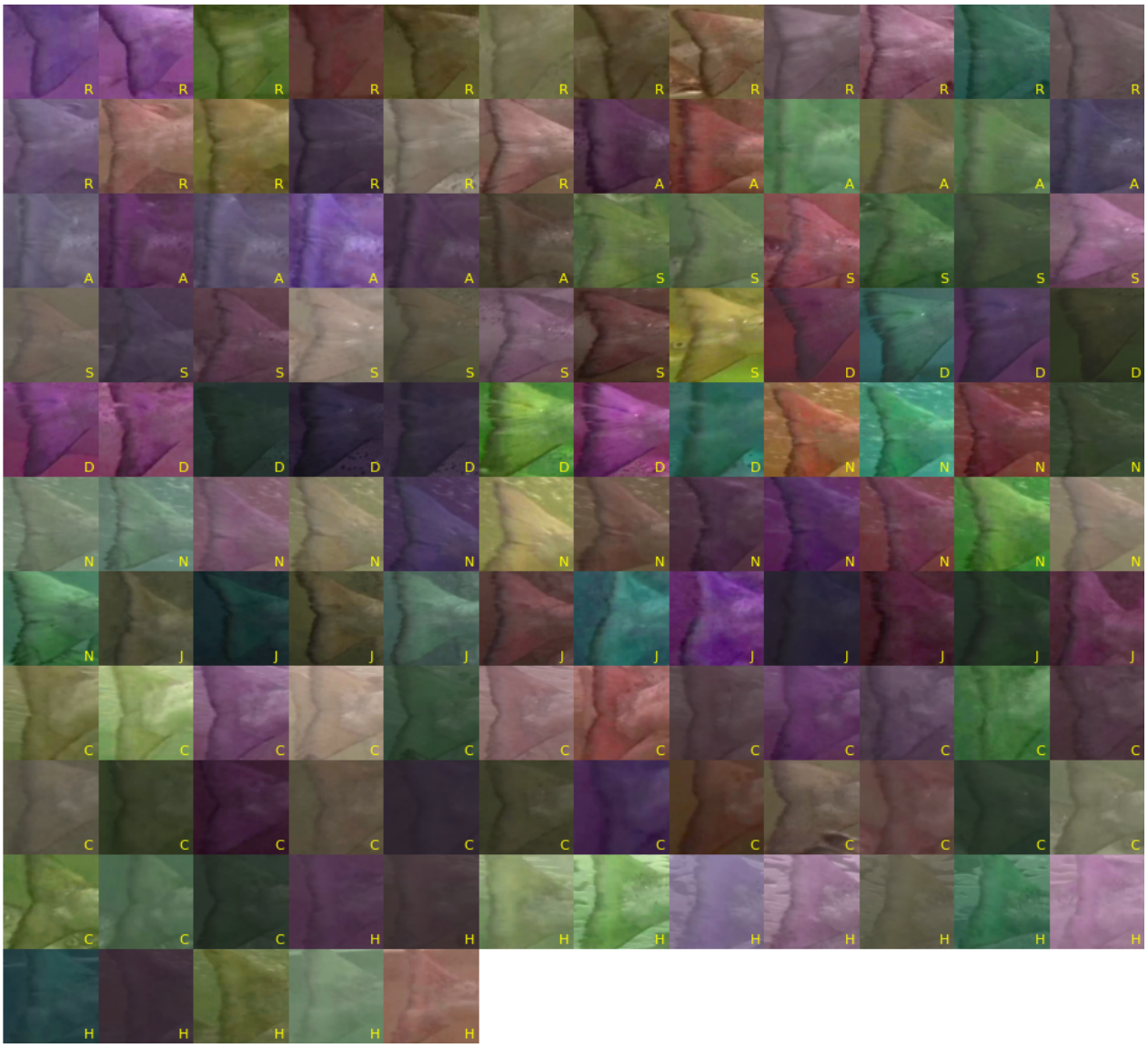


Figure B.5: Training dataset for tail fin re-identification model with data augmentation applied. The images in the figure are sorted for ease of comparing images for different individuals. The letter in the bottom right of every image is the first letter of the name of the individual.

B.2 Confusion matrices from pipeline evaluation

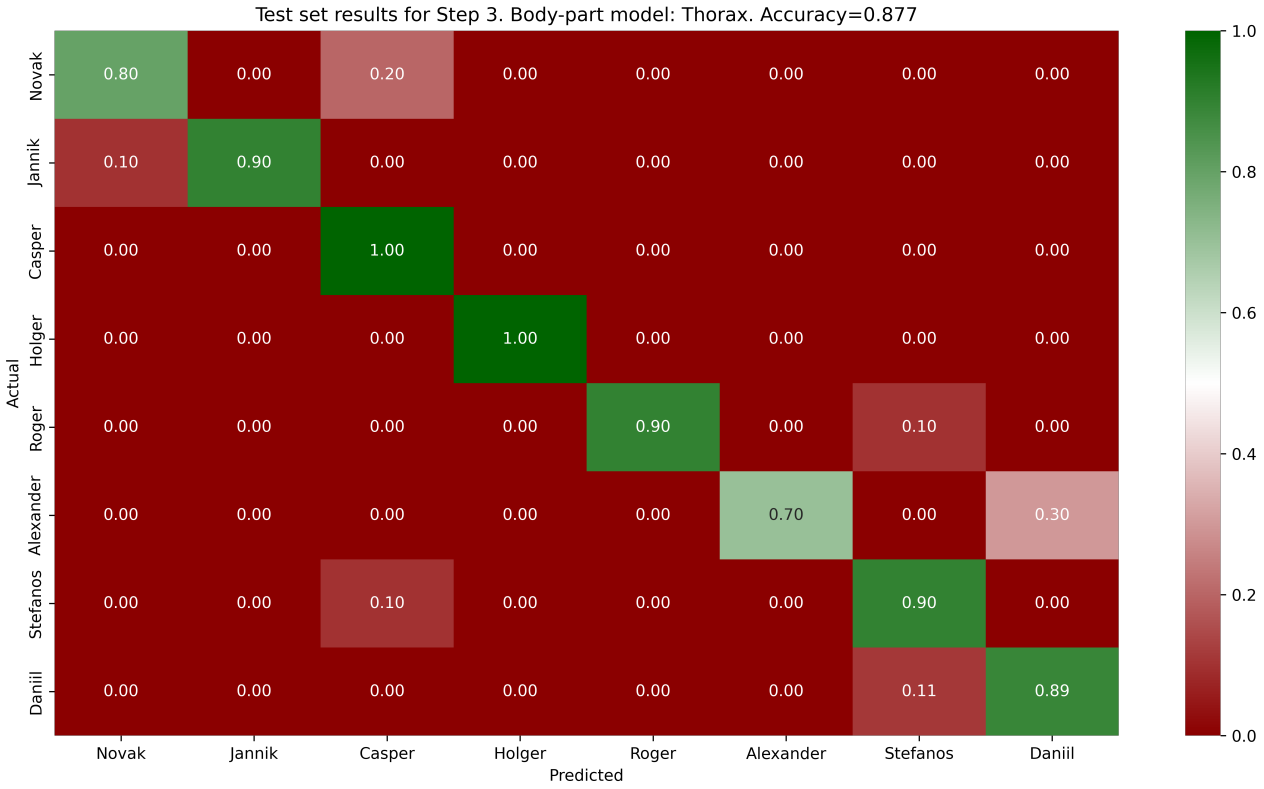


Figure B.6: Pipeline: Salmon re-identification test set results for thorax model.

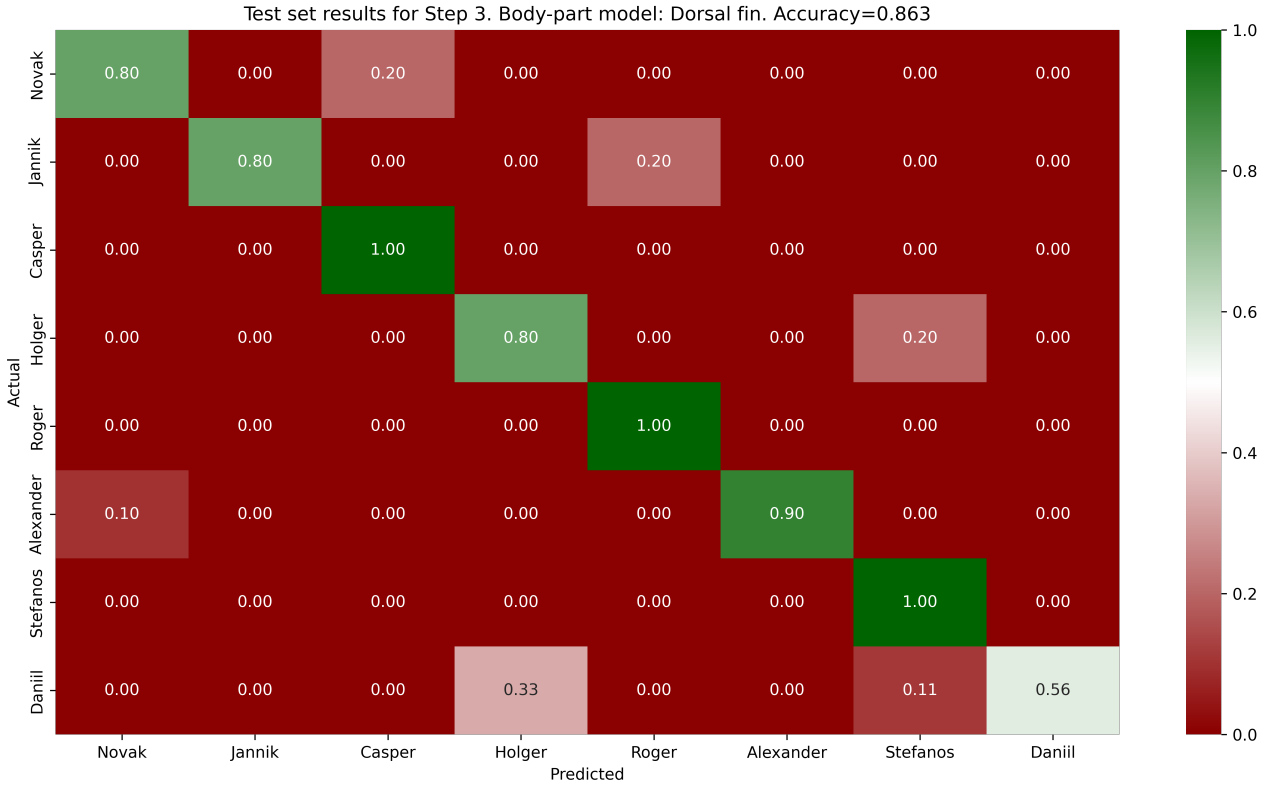


Figure B.7: Pipeline: Salmon re-identification test set results for dorsal fin model.

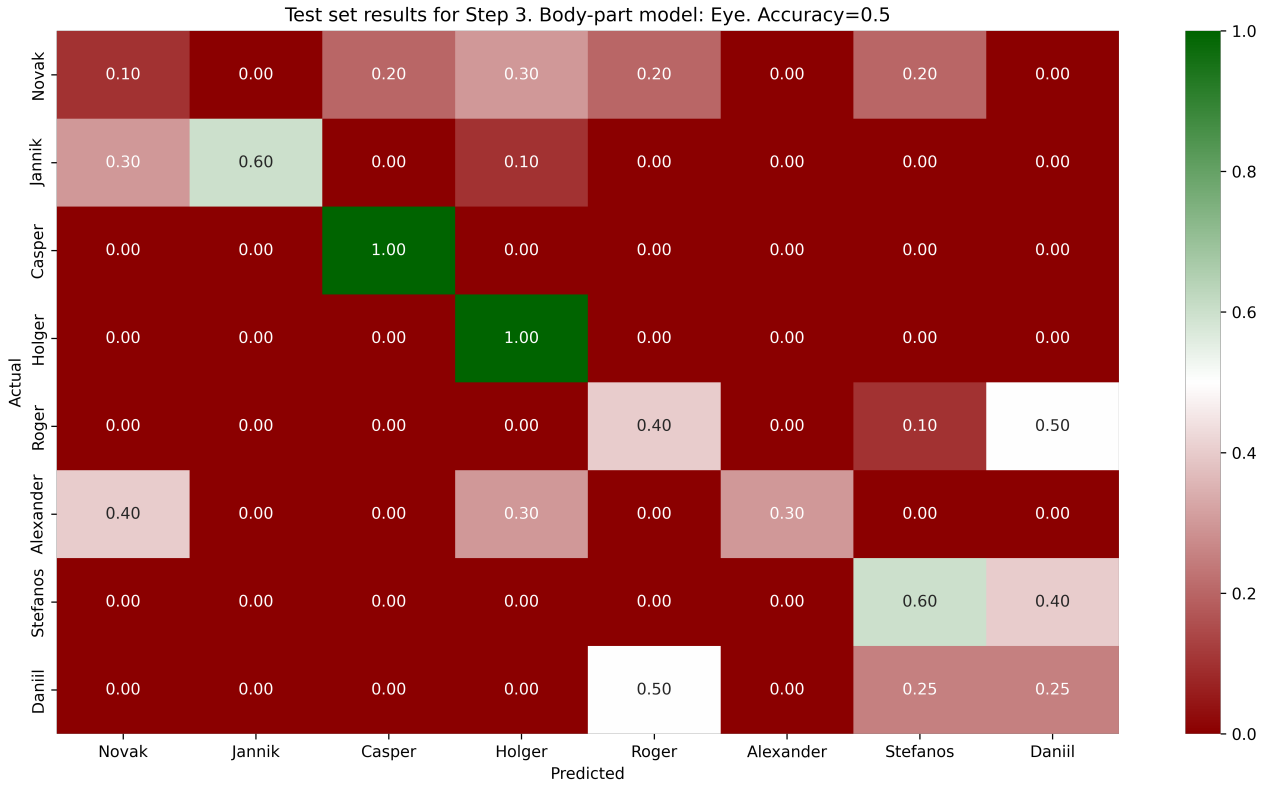


Figure B.8: Pipeline: Salmon re-identification test set results for eye model.

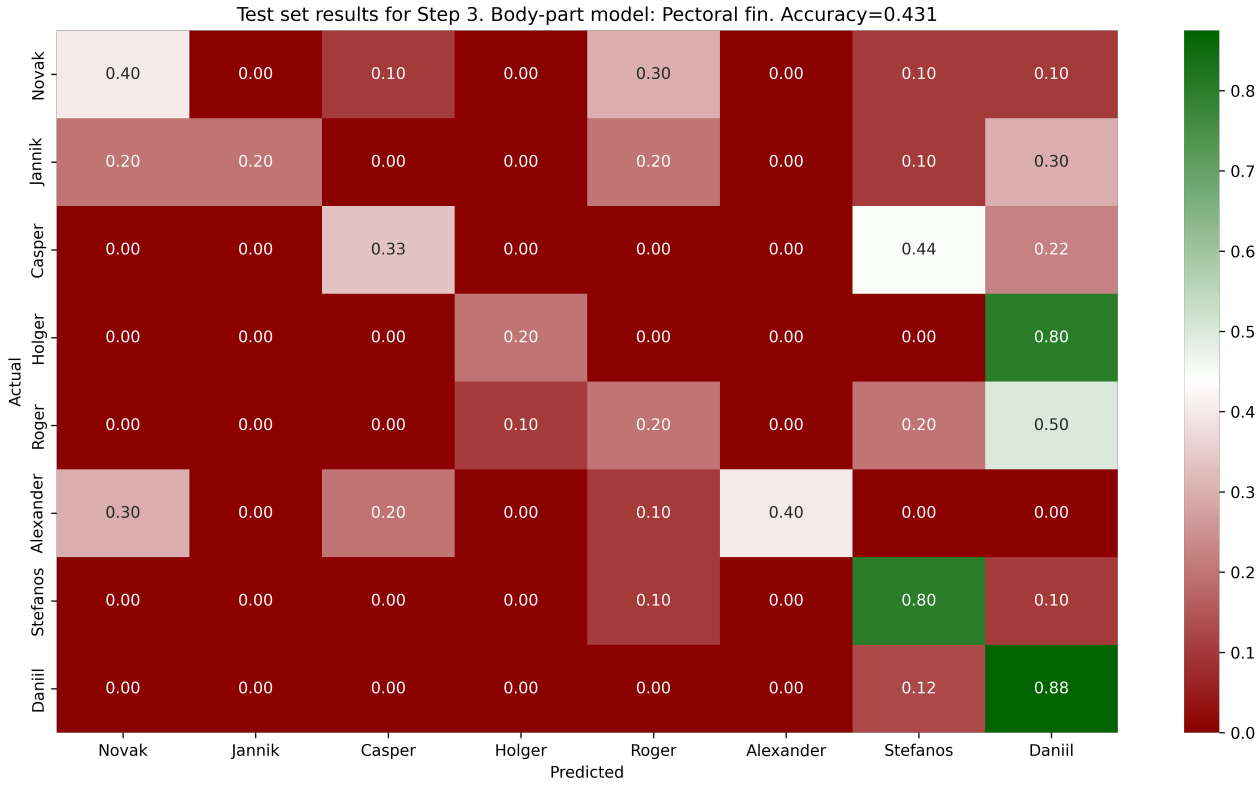


Figure B.9: Pipeline: Salmon re-identification test set results for pectoral fin model.

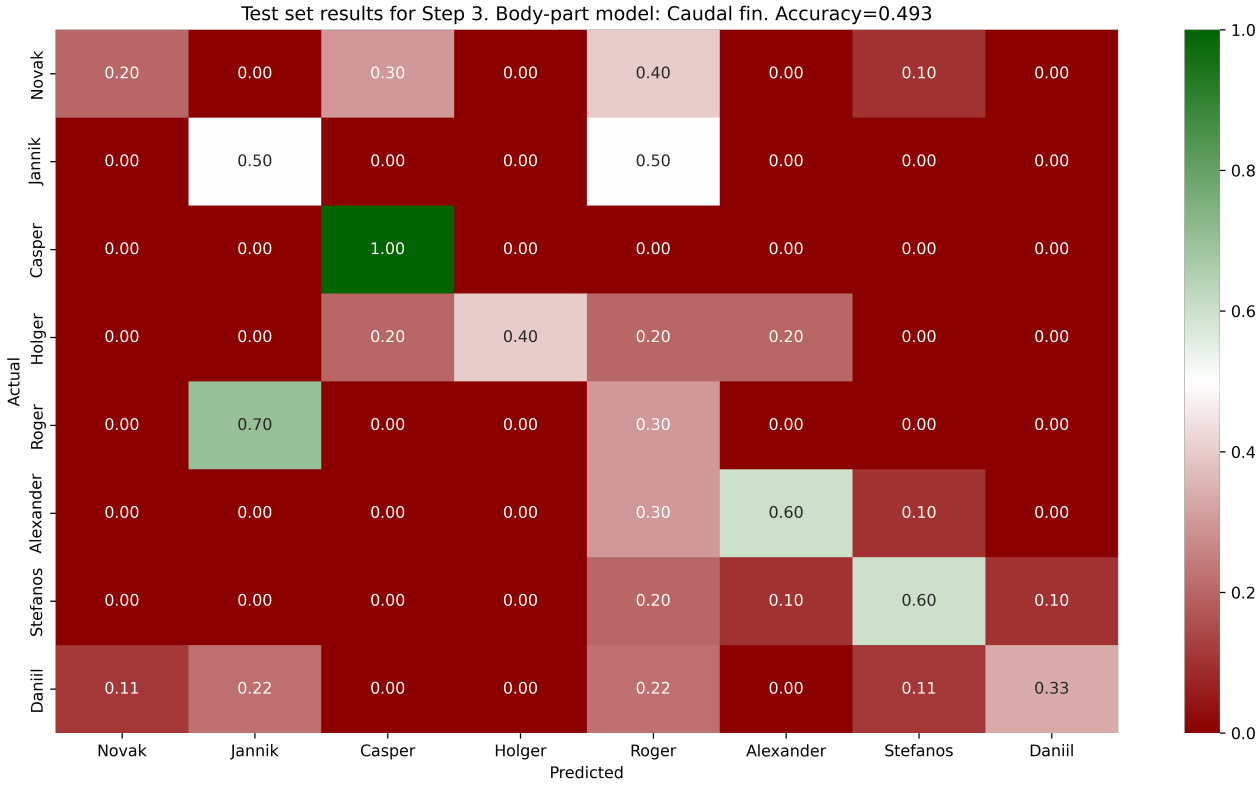


Figure B.10: Pipeline: Salmon re-identification test set results for caudal fin model.

Bibliography

[1] J. P. Fry, N. A. Mailloux, D. C. Love, M. C. Milli, and L. Cao, *Feed conversion efficiency in aquaculture: Do we measure it correctly?* Feb. 2018. [Online]. Available: https://iopscience.iop.org/article/10.1088/1748-9326/aaa273 (visited on 05/29/2024).

[2] O. Torrissen, R. E. Olsen, R. Toresen, *et al.*, "Atlantic Salmon (Salmo salar): The "Super-Chicken" of the Sea?" *Reviews in Fisheries Science*, vol. 19, no. 3, pp. 257–278, Jul. 2011. DOI: 10.1080/10641262.2011.597890. (visited on 05/29/2024).

[3] O. Torrissen, S. Jones, F. Asche, *et al.*, "Salmon lice – impact on wild salmonids and salmon aquaculture," *Journal of Fish Diseases*, vol. 36, no. 3, pp. 171–194, 2013, ISSN: 1365-2761. DOI: 10.1111/jfd.12061. (visited on 05/29/2024).

[4] K. Overton, T. Dempster, F. Oppedal, T. S. Kristiansen, K. Gismervik, and L. H. Stien, "Salmon lice treatments and salmon mortality in Norwegian aquaculture: A review," *Reviews in Aquaculture*, vol. 11, no. 4, pp. 1398–1417, 2019, ISSN: 1753-5131. DOI: 10.1111/raq.12299.

[5] A. Herrero, K. D. Thompson, A. Ashby, H. D. Rodger, and M. P. Dagleish, "Complex Gill Disease: An Emerging Syndrome in Farmed Atlantic Salmon (*Salmo salar* L.)," *Journal of Comparative Pathology*, vol. 163, pp. 23–28, Aug. 2018, ISSN: 0021-9975. DOI: 10.1016/j.jcpa.2018.07.004.

[6] M. Føre, K. Frank, T. Norton, *et al.*, "Precision fish farming: A new framework to improve production in aquaculture," *Biosystems Engineering*, Advances in the Engineering of Sensor-based Monitoring and Management Systems for Precision Livestock Farming, vol. 173, pp. 176–193, Sep. 2018, ISSN: 1537-5110. DOI: 10.1016/j.biosystemseng.2017.10.014.

[7] K. Banno, F. M. F. Gonçalves, C. Sauphar, *et al.*, "Identifying losers: Automatic identification of growth-stunted salmon in aquaculture using computer vision," *Machine Learning with Applications*, vol. 16, p. 100 562, Jun. 2024, ISSN: 2666-8270. DOI: 10.1016/j.mlwa.2024.100562.

[8] W.-C. Hu, L.-B. Chen, B.-K. Huang, and H.-M. Lin, "A Computer Vision-Based Intelligent Fish Feeding System Using Deep Learning Techniques for Aquaculture," *IEEE Sensors Journal*, vol. 22, no. 7, pp. 7185–7194, Apr. 2022, ISSN: 1558-1748. DOI: 10.1109/JSEN.2022.3151777.

[9] B. Zion, "The use of computer vision technologies in aquaculture – A review," *Computers and Electronics in Agriculture*, vol. 88, pp. 125–132, Oct. 2012, ISSN: 0168-1699. DOI: 10.1016/j.compag.2012.07.010.

[10] L. H. Stien, J. Nilsson, S. Bui, *et al.*, "Consistent melanophore spot patterns allow long-term individual recognition of Atlantic salmon," *Journal of Fish Biology*, vol. 91, no. 6, pp. 1699–1712, Dec. 2017, ISSN: 0022-1112, 1095-8649. DOI: 10.1111/jfb.13491.

[11] UN, *THE 17 GOALS | Sustainable Development*, Mar. 2024. [Online]. Available: https://sdgs.un.org/goals (visited on 06/03/2024).

[12] L. H. Stien, M. B. M. Bracke, O. Folkedal, *et al.*, "Salmon Welfare Index Model (SWIM 1.0): A semantic model for overall welfare assessment of caged Atlantic salmon: Review of the selected welfare indicators and model presentation," en, *Reviews in Aquaculture*, vol. 5, no. 1, pp. 33–57, 2013. DOI: 10.1111/j.1753-5131.2012.01083.x.

[13] M. S. Dawkins, "Battery hens name their price: Consumer demand theory and the measurement of ethological 'needs'," *Animal Behaviour*, vol. 31, no. 4, pp. 1195–1205, Nov. 1983. DOI: 10.1016/S0003-3472(83)80026-8.

[14] C. Noble, K. Gismervik, M. H. Iversen, *et al.*, *Welfare Indicators for farmed Atlantic salmon : tools for assessing fish welfare*, 1st ed. Nofima, 2018, ISBN: 978-82-8296-556-9.

[15] J. Panksepp, "Affective consciousness: Core emotional feelings in animals and humans," *Consciousness and Cognition*, Neurobiology of Animal Consciousness, vol. 14, no. 1, pp. 30–80, Mar. 2005. DOI: 10.1016/j.concog.2004.10.004.

[16] M. S. Dawkins, "From an animal's point of view: Motivation, fitness, and animal welfare | Behavioral and Brain Sciences | Cambridge Core," *Behavioral and Brain Sciences*, vol. 13, no. 1, pp. 1–9, 1990. DOI: https://doi.org/10.1017/S0140525X00077104.

[17] E. B. Høgstedt, "Developing a deep learning pipeline for automated salmon welfare analysis by respiration frequency," Master thesis, Engineering Cybernetics, NTNU, 2023. [Online]. Available: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3093227 (visited on 11/16/2023).

[18] P. Cisar, D. Bekkozhayeva, O. Movchan, M. Saberioon, and R. Schraml, "Computer vision based individual fish identification using skin dot pattern," *Scientific Reports*, vol. 11, no. 1, Aug. 2021. DOI: 10.1038/s41598-021-96476-4. (visited on 10/27/2023).

[19] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, Jun. 2015, pp. 815–823, ISBN: 978-1-4673-6964-0. DOI: 10.1109/CVPR.2015.7298682.

[20] B. M. Mathisen, K. Bach, E. Meidell, H. Måløy, and E. S. Sjøblom, *FishNet: A Unified Embedding for Salmon Recognition*, Oct. 2020. DOI: 10.48550/arXiv.2010.10475.

[21] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic Attribution for Deep Networks," en, in *Proceedings of the 34th International Conference on Machine Learning*, ISSN: 2640-3498, PMLR, Jul. 2017, pp. 3319–3328. [Online]. Available: https://proceedings.mlr.press/v70/sundararajan17a.html (visited on 05/10/2024).

[22] A. Foldvik, F. Jakobsen, and E. M. Ulvan, "Individual Recognition of Atlantic Salmon Using Iris Biometry," *Copeia*, vol. 108, no. 4, pp. 767–771, Nov. 2020. DOI: 10.1643/CI2020035.

[23] R. Schraml, G. Wimmer, H. Hofbauer, *et al.*, "CNN-based fish iris identification," in *2022 30th European Signal Processing Conference (EUSIPCO)*, Aug. 2022, pp. 628–632. DOI: 10.23919/EUSIPCO55093.2022.9909733.

[24] R. Schraml, H. Hofbauer, E. Jalilian, *et al.*, "Towards Fish Individuality-Based Aquaculture," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4356–4366, Jun. 2021, ISSN: 1941-0050. DOI: 10.1109/TII.2020.3006933.

[25] C. C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*. Cham: Springer International Publishing, 2018, vol. 1, ISBN: 978-3-319-94463-0.

[26] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, arXiv:1512.03385 [cs], Dec. 2015. [Online]. Available: http://arxiv.org/abs/1512.03385 (visited on 04/12/2024).

[27] PyTorch, *CrossEntropyLoss — PyTorch 2.3 documentation*, 2023. [Online]. Available: https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html (visited on 06/06/2024).

[28] S. Ren, K. He, R. Girshick, and J. Sun, *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, Jan. 2016. DOI: 10.48550/arXiv.1506.01497.

[29] R. Girshick, "Fast R-CNN," 2015, pp. 1440–1448. [Online]. Available: https://openaccess.thecvf.com/content_iccv_2015/html/Girshick_Fast_R-CNN_ICCV_2015_paper.html (visited on 04/03/2024).

[30] U. Erikson, C. Rosten, P. Klebert, S. Aspaas, and T. Rosten, "Live transport of Atlantic salmon in open and closed systems: Water quality, stress and recovery," *Aquaculture Research*, vol. 53, no. 11, pp. 3913–3926, 2022. DOI: 10.1111/are.15895. (visited on 11/14/2023).

[31] K. Wada, *Labelme: Image Polygonal Annotation with Python*, Feb. 2024. DOI: 10.5281/zenodo.5711226. [Online]. Available: https://github.com/wkentaro/labelme (visited on 02/06/2024).

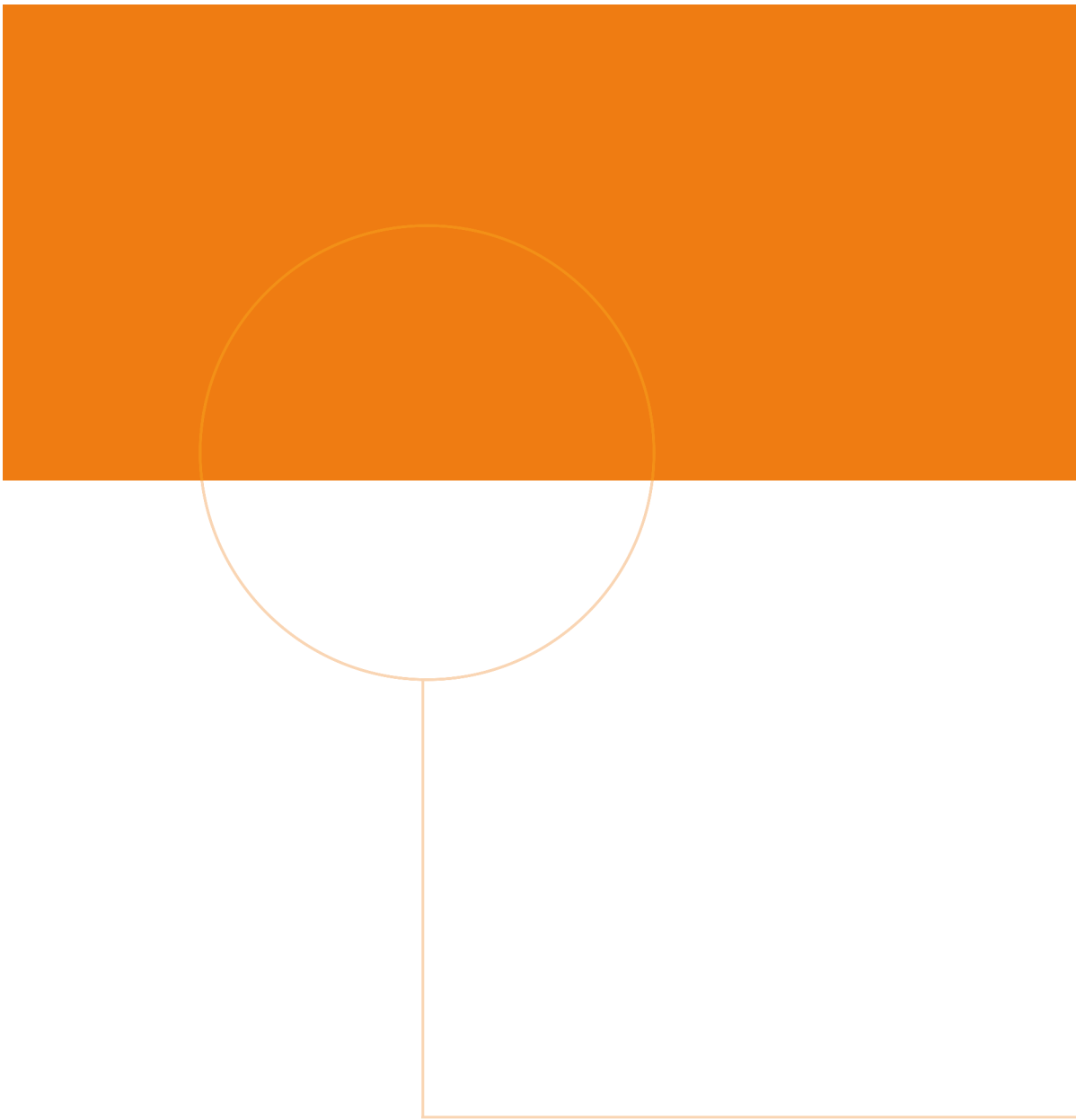
[32] *Models and pre-trained weights — Torchvision main documentation*, 2017. [Online]. Available: https://pytorch.org/vision/main/models.html (visited on 06/10/2024).

[33] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 936–944, ISBN: 978-1-5386-0457-1. DOI: 10.1109/CVPR.2017.106.

[34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848. (visited on 04/04/2024).

[35] S. Khirirat, H. R. Feyzmahdavian, and M. Johansson, "Mini-batch gradient descent: Faster convergence under data sparsity," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec. 2017, pp. 2880–2887. DOI: 10.1109/CDC.2017.8264077.

[36] M. Wiik, *AI and Computer Vision Technologies for Salmon Re-identification in Aquaculture (Unpublished)*, Dec. 2023.



NTNU

Norwegian University of
Science and Technology