Master's thesis

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

Espen Berntzen Høgstedt

Developing a deep learning pipeline for automated salmon welfare analysis by respiration frequency

Master's thesis in Cybernetics and Robotics
Supervisor: Annette Stahl
Co-supervisor: Christian Schellewald
June 2023

NTNU
Norwegian University of
Science and Technology

Espen Berntzen Høgstedt

Developing a deep learning pipeline for automated salmon welfare analysis by respiration frequency

Master's thesis in Cybernetics and Robotics
Supervisor: Annette Stahl
Co-supervisor: Christian Schellewald
June 2023

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

NTNU
Norwegian University of
Science and Technology

Abstract

Salmon farming is becoming increasingly important in the global food chain, which necessitates a deep understanding of how internal salmon states relevant for production yield and welfare is evinced externally on salmon, in order to facilitate real-time detection of inadequate conditions. In this thesis, one particular external trait of salmon, ventilation frequency, is elucidated in great detail to evaluate how it can inform on the condition of salmon in tanks and net pens.

To allow for easy and automatic extraction of this trait, a complete pipeline capable of estimating ventilation frequency of individual salmon from a video recording is developed. By the use of State Of The Art (SOTA) deep learning methods, the algorithm is capable of detecting and tracking salmon, estimating mouth poses in order to calculate ventilation frequency, and determining the unique salmon individual a fish belongs to.

Upon completion of the salmon ventilation frequency extraction method, it was applied to data from a salmon stress experiment, unveiling that salmon ventilation frequency increase in response to reduced Dissolved Oxygen (DO) content or disturbances, and that individual salmon show consistent respiration patterns across hours and weeks.

By observing the capability of the constructed salmon ventilation frequency pipeline to discern salmon welfare trends in a stress experiment, it can be concluded that the method is capable of automated evaluation of a salmon welfare indicator.

Sammendrag

Lakseoppdrett blir stadig viktigere i den globale næringskjeden, noe som nødvendiggjør en dyp forståelse av hvordan interne tilstander i laks som påvirker dennes velferd og økonomiske verdi manifesterer seg gjennom ytre trekk, slik at utilstrekkelige forhold kan avdekkes og utbedres. I denne oppgaven er et enkelt ytre trekk ved laks, ventilasjonsfrekvens, belyst i detalj for å undersøke hvordan denne fysiologiske parameteren kan informere om tilstanden til laks i tanker og merder.

For å kunne undersøke denne egenskapen enkelt og automatisk ble en algoritme utviklet for å estimere ventilasjonsfrekvensen til individuelle laks fra et videoopptak. Ved hjelp av moderne metoder for dyp læring ble denne algoritmen i stand til å lokalisere og spore laks, estimerer disses munnåpning for å beregne ventilasjonsfrekvensene deres, og identifiserer hvilket unike lakseindivid den lokaliserte fisken tilhørte.

Denne ventilasjonsestimeringsalgoritmen for laks ble brukt til å analysere data fra et stresseksperiment da den var ferdig utviklet, noe som avdekket at laksens ventilasjonsfrekvens øker som respons på forstyrrelser og redusert oksygenmetning i vann, og at individuelle laks viser konsistente respirasjonsmønstre over timer og uker.

Ved å observere ventilasjonsestimeringsalgoritmens evne til å bestemme laksens velferd i et stresseksperiment, kan det konkluderes med at metoden er i stand til automatisert evaluering av velferd hos laks gjennom analyse av respirasjonsfrekvens.

Preface

This document represents my master thesis written at the Engineering Cybernetics faculty at the Norwegian University of Science and Technology (NTNU). It is a continuation of my preproject[1], and as such some paragraphs share similarities between the two documents. This is the case for the chapters that have seen little adjustments during the thesis work, in particular the theory, experiment and biology literature review sections.

This thesis is associated with a Norwegian Seafood Research Fund (FHF) project (project number 901736) with title *Kunnskapsgrunnlag for biologisk relevante velferdsindikatorer for laks i akvakultur (BIORELEVANS)*. The work was conducted between 9^{th} of September 2022 and 5^{th} of June 2023. Supervision was performed by Christan Schellewald from Sintef and Annette Stahl from NTNU. Fish experiments were performed as part of the BIORELEVANS project together with the Norwegian Institute for Water Research (NIVA) and the Norwegian University of Life Sciences (NMBU) at Solbergstrand research facility, and the software employed in the subsequent work was constructed in Python. All code employed to create the results of this thesis is available in my GitHub repository[2].

Acknowledgement

I would like to thank Christian Schellewald at SINTEF for guidance during the writing process.

Contents

List of Figures

List of Tables

Acronyms

2D 2 Dimensional. 15, 37, 39

3D 3 Dimensional. 15

A/D Analog to Digital. 14

ACTH AdrenoCorticoTropic Hormone. 6

CNN Convolutional Neural Network. 20, 22, 25–27

CRH Corticotropin Releasing Hormone. 6

DLT Direct Linear Transform. 39

DO Dissolved Oxygen. 2, 5, 9–11, 47, 48, 88, 89, 110–112, 115–121, 167–172, I, VI, VIII, IX, XIII

e.g. *Exempli Gratia*. 9, 16, 19, 36, 43–45, 49, 52, 66, 109, 114, 117–119

FHF Norwegian Seafood Research Fund. III

GPU Graphics Processing Unit. 22, 57, 64, 119

HBI Head-Body Intercept. 54, 60, 65

HPI Hypothalamic-Pituitary-Interrenal. 5, 6

HSC Hypothalamic-Sympathetic-Chromaffin cell. 5, 6

i.e. *Id Est*. 26, 69, 78, 109, 118, 120

ID IDentifier. 60

IoU Intersection over Union. 17, 31, 32, 70, V

KKT Karush-Kuhn-Tucker. 34

KNN K Nearest Neighbors. 21

LM Levenberg-Marquardt. 19, 44, 45, 65–67, 108

LSTM Long Short-Term Memory. 15

NIVA Norwegian Institute for Water Research. 3, 47, III

NMBU Norwegian University of Life Sciences. III

NMS Non Maximum Suppression. 32, 57, V

NTNU Norwegian University of Science and Technology. III

RANSAC RAndom SAmple Consensus. 19, 44–46, 65–68, 108, 109, V

RCNN Region-based Convolutional Neural Network. 16, 19–21, 23, 27–30, 32, 52, 54–57, 60, 64, 65, 70, 71, 78, 100–105, 122, 146, 147, V–VIII

RGB Red Green Blue. 63

ROI Region Of Interest. 27, 28, 31

SOTA State Of The Art. 2, 21, 27, I

SWIM Salmon Welfare Index Model. 11

Terminology

1. **Salmon**: Refers to atlantic salmon (*Salmo salar*).
2. **Tracker**: Refers to an object that follows a salmon individual.
3. **Downbreathing and upbreathing**: Downbreathing refers to restricting oxygen to a salmon tank and letting the metabolism of salmon gradually reduce the DO content in the water. Upbreathing is the reverse mechanism, and involves introducing oxygenated water to a hypoxic environment, causing a gradual increase in DO content. These terms are used as both nouns and verbs in the document.
4. **School and shoal**: These terms will be used interchangeably about a collection of fish, even though the terms technically have slightly different meanings[3].
5. **Breathing, respiration and ventilation**: These terms will be used as synonyms for salmon oxygen uptake.
6. **Salmon classification**: When the two terms salmon and class/classification are used together, they are referring to classifying salmon based on which individual the salmon belongs to. Species classification, which these terms also could refer to, is not discussed in this document.
7. **Statistical significance**: This refers to a p value below 0.05.
8. **Mean and average**: These terms are used interchangeably and refers to the arithmetic mean ($\sum_{i=1}^{N} \frac{x_i}{N}$).

Chapter 1

Introduction

Aquaculture is the fastest growing food animal sector in the world[4], with Norway establishing itself as a major player in the field. Due to the high feed conversion ratio, low CO_2 footprint and large stock densities, salmon farming has been praised as the solution to the food requirements of a growing population[5]. Like all success stories, however, the industry is not without its problems. Lice infestation[6], escaping fish[7], sick fish[8], fish mortality[9], retarded growth[10] and pollution[11] are some of the issues researchers and farmers are faced with, causing reduced salmon welfare and economic losses.

To effectively combat these problems, the underlying causes of the adversary effects must be elucidated. It is firmly established that stress has some role in this discussion[12], causing both acute and chronic changes in the physiology of the fish. Stress responses can be examined by chemical screening of blood and organs, external observation by humans, or even analysis of the environment of the fish[13]. All of these methods have their downsides ([14], chapter 11), which motivates a new, automatic framework for stress evaluation in fish.

Among the possible candidates, computer vision is standing out as a cheap, effective and non-invasive way of performing this automation. It has been successfully applied to a number of related tasks, such as measuring feeding activity in salmon[15], analysing hypoxia response in goldfish[16], measuring salmon welfare from behaviour indicators[17], counting fish[18], estimation of fish size[18], quality assessment[18] and species identification[18]. All of these projects related to welfare and stress used behaviour indicators like speed, direction and location of the fish as features. By doing so, they are neglecting an easily discernible and possibly important parameter; respiration. This metric requires no three dimensional reconstruction, is always present and shows a quick dynamic. Furthermore, as will be elucidated later in the document, stress is closely related to ventilation frequency in teleosts (fish belonging to the infraclass Teleostei[19], which includes salmon). These considerations motivate the construction of a deep learning pipeline extracting salmon breathing frequency from video recordings.

In the preproject associated to this thesis, the feasibility of extracting respi-

ration frequency of a salmon shoal by applying a deep learning pipeline to videos was demonstrated[1]. The method demonstrated there is further investigated and improved in this report, before applied to data from a salmon experiment in order to hypothesize how ventilation frequency relates to stress induced by temperature, hypoxia and disturbance.

As this is a technical thesis, the main focus of the report will be on the computer vision frequency extraction pipeline. Even so, some elucidation on the biological context around the experiment data is necessary to understand how the fish breathing frequency can inform about important parameters and relationships. Therefore, the report will start with a literature review of relevant salmon biology. This will consist partly of a distilled version of earlier work of the author, and the reader is referenced to[1] for more in-depth covering of the topics.

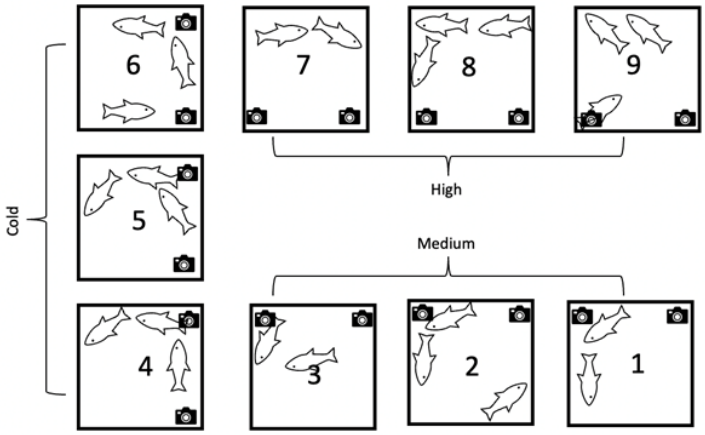
Succeeding this, literature relevant for the computer vision pipeline will be presented. This will cover precise problem formulations, together with discussion of SOTA methods that can contribute to algorithms solving the stated problems. Afterwards, the mathematics behind the most intriguing methods will be described in the theory section.

Next, the experiment will be described. It was performed at NIVAs research facility at Solbergstrand (figure 1.1a[20]), and involved seven salmon in each of nine different tanks, exposed to three different temperatures (figure 1.1b[21]). At two different times over a two week period, downbreathings were performed in all tanks, involving an initial decline in dissolved oxygen concentration, and a subsequent increase. All salmon were recorded by cameras during the downbreathings (see figure 1.1c for an example frame), providing the data used in this report.

Armed with theory from both biology and engineering, as well as experimental data, the salmon ventilation analysis pipeline is described in detail. This pipeline is then used for a thorough analysis of the experiment data, on an individual salmon level. All results from both the salmon ventilation frequency extraction pipeline and the experiment will be presented and discussed in detail, in order to both display the capabilities of the constructed method to extract ventilation frequencies from video recordings, and to explore the link between salmon ventilation and welfare.

This discussion will make it clear that ventilation frequency can be automatically detected by computer vision and deep learning methods, and that the extracted ventilation frequencies can inform about welfare reduction induced by acute disturbances or reduced DO content, as well as determining the personality type of individual fish, which correlates to their welfare.

(a) NIVA Research Facility at Solbergstrand[20]. This was where the salmon stress experiment was performed.



(b) Layout of fish tanks[21]. Each tank contains seven fish, exposed to two downbreathings over a two week period. Tank four to six are cold (12.3 $^{\circ}C$ to 13.4 $^{\circ}C$), tank one to three hold intermediate temperatures (14.0 $^{\circ}C$ to 15.6 $^{\circ}C$), and tank seven to nine are warm (15.9 $^{\circ}C$ to 17.7 $^{\circ}C$).

(c) Example of a video frame from the tank recordings. Each tank was recorded at 60 Hz with a consumer camera regularly over a two week period, and continuously with two cameras during the two downbreathings.

Chapter 2

Literature survey biology

2.1 Chapter introduction

This literature review will start with a presentation of the main physiological mechanisms involved in teleost ventilation, before the internal mechanisms and evaluation of fish stress are presented. Subsequently two stressors are examined in light of the discussed concepts, and the notion of fish welfare is briefly elucidated. The primary search engines used during the survey were google, Google Scholar[22] and Science Direct[23].

2.2 Fish ventilation

Teleosts absorb oxygen for metabolism by gas exchange between erythrocytes and water in lamellae attached to the gill arch[24]. Two modes of breathing have been observed; buccal pumping and ram-ventilation[25]. In the former, the operculum closes and the mandible is lowered to draw water into the mouth, followed by closing of the jaws and opening of the operculum to thrust water out over the gills[26]. In the latter, high relative water velocity drives water over the gills without energy expenditure of the opercular and buccal pumps.

Buccal pumping is effectuated by muscles in the jaw, operculum and gills, driven by the Vth, VIIth and IXth/Xth cranial nerve, respectively[27]. Their nuclei are located in the medulla oblangata, where they are interacting with the reticular formation. This loosely defined network contains pacemaker cells that generate the respiratory rythm of the fish. The medulla alone is capable of sustaining ventilation, however the midbrain synapse with both the reticular formation and the cranial nerves, and is through this capable of altering respiration. In addition to efferent pathways, the cranial nerves possess afferent branches, projecting to different locations in the brainstem. The sensory pathways are receiving stimuli from mechanoreceptors and chemoreceptors, modulating the breathing to adapt to the millieux.

The main modulatory sensory input for fish ventilation stems from oxygen

Author	Weight (g)	resp. freq. (num/min)	Stressor
Hosfeld[32]	29.2	73	control
Hosfeld[32]	29.2	65	125% DO
Hosfeld[32]	29.2	57	145% DO
Hosfeld[32]	29.2	53	178% DO
Knoph[33]	430	56	control
Knoph[33]	430	61	30 mg/l ammonia
Knoph[33]	430	60	56.2 mg/l ammonia
Millidine[34]	1.8-12.6	61-140	Diverse
Erikson[35]	1641	161	Commercial transport
Erikson[35]	1641	137	Open transport
Erikson[35]	1641	153	Closed transport
Erikson[36]	4410	55 and 64	Control
Erikson[36]	4410	80 and 81	Crowding

Table 2.1: Overview over salmon ventilation frequency

chemoreceptors[28]. Branchial receptors of this type, oriented either towards the blood or the water, have a pronounced effect on both ventilation frequency and amplitude. Extrabranchial externally oriented sensors, such as those located in the orobranchial cavity, are mainly altering ventilation amplitude[29]. In addition to oxygen receptors, fish possess externally facing CO_2 sensitive chemoreceptors at their gill arch[29], and internally oriented H^+ sensors close to the fish arteries[30]. In salmon, CO_2 sensors are far more prominent than H^+ receptors[31].

Table 2.1 presents an overview of salmon ventilation frequency at different stressors, environments and sizes. Most salmon seems to fall in the ventilation range [1, 2.5] Hz, with 1 Hz being the resting respiratory frequency.

2.3 Stress in fish

Fish get stressed when subjected to real or perceived stressors such as high temperature[37] or hypoxia[38]. The physiological responses to these irritants are initiated and controlled by two endocrine axes; the Hypothalamic-Sympathetic-Chromaffin cell (HSC) axis and the Hypothalamic-Pituitary-Interrenal (HPI) axis. The functioning and effects of these will be expanded on in the following, structured according to the three layers of physiological responses as described by Barton[39].

2.3.1 Primary responses

The primary responses are the endocrine changes that occur in the fish immediately after being exposed to a stressor. The main agents are cathecolamines

and corticosteroids, which are effectuated by the HSC axis and the HPI axis, respectively[12].

Cathecolamines, of which ephinephrine and norephinephrine are the main actors, are secreted into circulation by direct action of the hypothalamus on the chromaffin cells of the head kidneys by the sympathetic branch of the autonomous nervous system[12]. These stress hormones can also be excreted from chromaffin cells in direct response to reduced oxygen levels, without being stimulated by the hypothalamus[40].

The HPI axis is initiated by the release of Corticotropin Releasing Hormone (CRH) by the hypothalamus, causing the anterior pituitary to secrete AdrenoCorticoTropic Hormone (ACTH). This in turn causes interrenal cells to synthesize and release cortisol. In addition to this main pathway, several other endocrines have been shown to support, inhibit or replace CRH and ATCH in the cortisol pathway of fish[12]. As this endocrine axis involves hormonal signal transfer, it works slower than the nerval driven HSC axis ([14], chapter 11).

Some fish circumvent these main hormonal axes, and rely on sympathetic nerve activity as a response to stress instead[41]. Salmon is not in this group, so this peculiarity will not be pursued further.

2.3.2 Secondary responses

The secondary responses are the direct effects of released endocrines on physiological systems in the fish, such as the cardiovascular system, the defense system and the respiratory system, as well as changes in metabolism and hydromineral balance[39].

Starting with the cardiovascular system, stress modulates the teleost heart by cholinergic and adrenergic neural pathways ([14], chapter 7) in order to raise[42][43] or lower[24][42] the beating frequency. Also direct endocrine actions modulate the cardiovascular system of fish. Catecholamines[44] and cortisol[45] increase cardiac contractility, catecholamines induce vasoconstriction in the systemic circuit, and vasodilation in the branchial capillaries[12], adrenaline increase the permeability of gill epithelium[46] and catecholamines increase the blood transport capacity[47] and hematocrit[12].

The increased diffusion through the gills driven by the mechanisms above disturbs the hydromineral equilibrium of teleosts. Cortisol have some regulating effects on this balance, causing proliferation of the osmolarity regulating[48] chloride cells[49], and might[50][51] have an effect on sodium-potassium ATPase activity.

With regards to metabolism, catecholamines are the hormones that exert the main glycemic control during stress. They cause elevated glycogenolysis, slightly elevated gluconeogenesis, and might cause an increase of free fatty acids[12]. Cortisol promote some gluconeogenesis and lipolysis.

Lastly, stress hormones alters ventilation, however the relationship is controversial[12]. Evidence have been presented for both a positive[52][53] and a negative[54] correlation between catecholamines and ventilation response. In

Author	Cortisol	Stress	Growth rate
Mcckormick[55]	Reduced	Handling	Reduced
Pickering[59]	No effect	Handling	Reduced
Barton[60]	Injected	No	Reduced
Barton[60]	Reduced	Handling	No effect
Fast[61]	No effect	Handling	No effect

Table 2.2: Overview over papers examining the relationship between cortisol, stress and growth rate

the case of severe hypoxia, where feeding muscles are recruited in an attempt to stave off further oxygen deficit, catecholamines seem to have an effect[53].

2.3.3 Tertiary responses

This kind of responses is the main reason farmers are concerned about stress in fish. They describe how the organism is affected by stress, causing reduced growth, weaker defense system, accelerated ageing, and reproduction problems, culminating in reduced revenue of the farm.

Fish subjected to stressors show reduced nutrient assimilation by actions on the appetite[55][38] and gut[56], and stress induced release of cortisol promote proteolysis[57] and reduce myogenesis[58]. Some of the research on the long term effects of cortisol and stress on growth is displayed in table 2.2, and show that chronic stress lower basal plasma cortisol levels, while the effects of handling and cortisol on growth differ between experiment setups and salmon strains.

Furthermore, both immune enhancing and immune suppressive effects have been observed from stress, with the acute responses tending towards increased immunocompetence, while the chronic effects are immunodepressive. The depressive effects far outweigh the enhancing ones ([14], chapter 10). In salmon, acute stress cause enhanced expression of inflammatory genes, while chronic stress leads to decreased stimulation and survival of leukocytes[61]. Furthermore, cortisol seems to prevent skin growth[62].

Accelerated aging has also been observed in teleosts during stress, where cortisol has been shown to promote apoptosis in pavement, mucous and chloride cells[63][64].

Lastly, stress affects all levels of the reproductive system of fish, with a clear detrimental effect ([14], chapter 8). In salmon, elevated maternal cortisol levels lead to increased mortality, reduced size and increased morphological malfunction in offsprings[65].

2.3.4 Fish personality

Having discussed the general response of teleosts to stress, it is necessary to state that fish responds in different manners from individual to individual. In the literature, it is common to separate fish into two personality categories

based on their stress response, proactive and reactive. They are diverging by the former category tending towards low cortisol response after acute stress[66], high sympathetic activity[67], social dominance[68], low locomotor activity[69], quick feeding in novel environments[69] and lower ventilation rate[70].

2.4 Stress evaluation

To properly research and draw inference on stress, some method of observation and quantification of indicators are necessary. Catecholamines and cortisol are the initiators of later modulation of physiological systems, and measuring these should provide a perspicuous perception of the fish stress state.

Catecholamine concentration in blood plasma rise quickly after stressor exposure ([14], chapter 11), making it difficult to use as a stress indicator, since capture and handling prior to blood sampling have a great effect on measured levels. Furthermore, the measurement of catecholamines requires highly specialized equipment[71].

Cortisol is better fit as a stress metric than catecholamines due to its slower response and easier means of analysis. It can even be measured in water around the fish[13]. Furthermore, a correlation between crowding stress and immediate blood cortisol elevation has been firmly established[72][73][13][74], although, as previously seen, some accommodation effect might occur after long term exposure. This long term effect will influence whether cortisol can predict chronic stress as well as acute stress[75]. Considering the above, cortisol can be considered the current gold standard of fish stress evaluation, at least for acute stress ([14], chapter 11).

Even if cortisol measurements are precise, their invasiveness limit their practical applicability. Furthermore, solutions for reducing invasiveness, such as measuring water bound cortisol, cannot differentiate between individual salmon. This motivates exploiting an easily accessible secondary response for stress evaluation instead. One such marker could be respiration frequency, which is both non-invasive and individual specific.

The use of ventilation as a stress indicator, however, is controversial. Some researchers argue against its use, claiming it is insensitive to stress[76], that it has too many confounding variables[73], and that it requires the fish to ventilate by pure buccal pumping. Others hold that the variable is among the most informative for stress elucidation[36][77]. A bonus of using this metric is that ventilation frequency can inform about more than stress state, such as type of stressor[78] and stress coping strategies[70].

Traditionally, ventilation has been measured by counting opercular[79] or jaw oscillations, or measuring electrical activity of buccal muscles by external[80] or internal[81] electrodes. The measurement methods related to counting is possible to automate by computer vision methods.

2.5 Stressors

Temperature[82][83][84] and oxygen concentration[84][85] in the epipelagic zone shows diurnal and yearly cycles, as well as spatial variations, often to an extent that moves the environment in and out of salmon stress and welfare thresholds. This affects farming sites as well, motivating a further exploration of these stressors.

2.5.1 Oxygen

Salmon consume oxygen for metabolic activity, and stress, size[86], temperature[87] and exercise ([14], chapter 7) change the oxygen requirements of the fish. When the fish is unfed, resting, and only maintaining normal physiological functions, its metabolism is in a state known as standard or resting metabolic rate. This distinguishes itself from the maximum metabolic rate, in which the fish cannot increase its aerobe metabolism further. The difference between these two are known as the metabolic scope[88], which, colloquially speaking, is the capacity the fish has for activity. When the oxygen requirements of the cells surpass the current availability of blood oxygen, salmon normally cope by increasing the water flow over their gills in order to raise oxygen availability around its lamellae, ensuring sufficient oxygen delivery to the cells[89]. This strategy is known as oxyregulation[90], and is only a feasible approach as long as the fish metabolism is inside the metabolic scope. At excessive metabolic challenges, increased water flow over the gills is not a viable compensation strategy, and the salmon start conforming to the environment by decreasing the metabolic activity, increasing anaerobic metabolism and focus energy expenditure on immediate survival[85]. The oxygen saturation when salmon switch from regulators to conformers is called p_{crit}[91].

Low DO saturation cause three main effects on salmon ventilation frequency.

1. The salmon ventilation frequency increase to compensate for the low oxygen gradient at the lamellae-water interface in hypoxic environments.

2. The salmon ventilation frequency is influenced by the transient cortisol release in fish induced by reduced water DO content[92][93] (see section 2.3.2). Since the stress response of salmon is somewhat adaptive, as seen from e.g. the effects of stress on the teleost heart (section 2.3.2), it is likely that stress responses caused by hypoxic conditions cause an increase in the ventilation frequency.

3. The metabolic rate of fish change together with the DO content according to the oxyconformance effects discussed above. For high oxygen saturations, the oxygen requirements of fish are constant or slowly decaying, while the oxygen requirements decrease linearly (with a faster rate than before) after a breaking point[94][1][95].

[1] This article discusses Sturgeon, which is part of the same subclass as Salmon (Actinoptery-

By assuming the oxygen gill uptake is proportional to the total amount of oxygen flowing over them, such that two litre of 50% oxygen saturated water provide the same amount of oxygen to the fish as one litre of 100% oxygen saturated water, the oxygen gradient alone would yield a linear increase in ventilation frequency in response to decreasing DO content. This linear ventilation frequency increase is subjected to two opposite curvature influences from the two remaining ventilation frequency effects; the reduced metabolism under hypoxic conditions (oxyconformance) pushing it towards concavity, and the stress effects pushing it towards convexity. The reason that these effects influence the curvature, is that they are only active after a breaking point, hence the rate of increase of the ventilation frequency will change as the DO saturation decreases.

In [85][2], empirical ventilation frequency to DO curves are presented, displaying a slightly sigmoid (almost linear) ventilation frequency shape as the DO saturation decreases. This is consistent with the above discussion. The oxygen gradient is the main driver of the ventilation frequency increase, with convex stress effects being most significant at the beginning of the DO decrease, and concave oxyconformity effects being more pronounced as the DO content approaches p_{crit}.

If the DO content decreases past p_{crit}, oxyconformity will be the main influence on the fish ventilation, causing the ventilation frequency to decrease.

2.5.2 Temperature

Fish, with some exceptions[96], are unable to regulate their body temperature[97]. They deal with varying ambient temperatures by locating themselves at water areas holding optimal[3] temperatures[99]. Salmon have a higher temperature tolerance than other salmonids[100], with farming industry theory suggesting survival between zero and 23 degrees Celsius, reasonable growth between six and 16 degrees Celsius, and optimal temperature between 12 and 15 degrees Celsius[101]. This is challenged by later work, suggesting optimal temperature between 16.5 and 20 degrees Celsius[102][85]. These ranges are simplifications, and are not true for all cases. For example, acclimation has an effect on lethal limits[103][104], and viral infections[105] and stress[106] have an effect on optimal temperature.

The relationship between resting metabolic rate and temperature is still an area of active research. Several curves have been suggested to explain the correlation between the two variables, such as exponential[107][101], sigmoid[94] and Gaussian[108]. The unexplained variance of the fit of these curves are large, and some papers have findings that disagree with all the suggested curves, such as seeing a ventilation decrease with increasing temperatures at water

gii), but not the same infraclass (Teleostei). Sturgeon is instead part of the infraclass Chondrostei[19]. This might limit the relevance of this paper to salmon physiology.

[2]The version of [85] that the URL in the bibliography of this document link to does not contain the DO to ventilation frequency curves.

[3]Optimal temperature for fish can be defined as the point where the difference between standard and maximum metabolic rate is highest[98].

temperature below ten degrees[109]. The general trend seems to be an increase of the resting metabolic rate as the temperature increases.

The complex biological interplay in salmon makes the extent to which stress is responsible for the observed temperature induced ventilatory changes difficult to delineate. It seems clear that deviation from the optimal temperature increasingly brings the fish into a range of stress ([14], chapter 9), and that increased temperature increases cortisol in fish[110][111]. Some papers, however, find that the cortisol increase is only significant at extreme temperatures[112].

As another point when it comes to the relationship between temperature, ventilation and stress, elevated temperature boosts cortisol increase after exposure to other stressors[87].

2.5.3 Interplay of temperature and oxygen

The salmon oxygen uptake and temperature are closely related. As the temperature increase, the percentage of oxyconformers increase, the critical oxygen saturation (P_{crit}) increase[91][85] and the metabolic scope changes (it decreases as the temperature deviates from the optimal)[101].

In addition to the difference in salmon hypoxia response at different temperatures, increased temperature changes the environment by decreasing oxygen solubility[85]. Furthermore, the discussed stressors affect other internal mechanisms, as an example, higher temperatures and higher DO content leads to faster growth[113][114].

2.6 Fish welfare

When it comes to fish welfare, three different definitions have been proposed. Feelings-based welfare evaluation looks at the subjective experience of the fish, function-based welfare evaluation looks at the extent to which the animal is coping with its environment, while the comparison with natural lives is evaluating in what manner the fish is exhibiting behaviour similar to that of a wild fish ([14], chapter 12). Some work has been done in moving from these rather vague definitions to more tangible, quantitative analysis. The Salmon Welfare Index Model (SWIM) index is one such approach, where several indicators of welfare in fish farms, such as temperature, salinity, mortality and condition factor, are evaluated and summed to acquire a single welfare score[9][10].

By comparing the indicators of the SWIM model with the secondary and tertiary responses covered in this study, it is clear that stress and welfare of fish is intimately linked. Some welfare indicators are directly affected by chronic stress response, such as growth, reproduction and appetite. In the rest of the indicators, the causal direction is turned, and reduced welfare indicators cause stress. This is the case for temperature, stocking density and disturbances.

One aspect of welfare not addressed in the SWIM index is that fish subjected to reduced well-being alters behaviour. In salmon, fixed feeding and underfeeding cause aggression, hyperoxia reduce swimming speed, parasites reduce max

swimming speed, cage submergence increase swimming speed, environmental gradients cause changed space use, and scheduled feeding cause higher swimming speeds[115]. As seen above (section 2.3), cortisol is linked to both feeding behaviour and oxygen requirements, hence some of these behaviour responses might be a tertiary stress response. Furthermore, a correlation is observed between cortisol levels and hierarchical position, suggesting a stress mediated pathway in displacement aggression[116]. Swimming behaviour have been considered in other welfare indexes[117], as well as in earlier computer vision based works for welfare estimation[17].

considering the findings in this section, we establish that stress and welfare are equal for our purposes. To be more specific, a stressed salmon, at least if this stress is chronic, has poor welfare, while a salmon with low welfare is frequently stressed. Due to this, we will focus on stress in the rest of this report, as this property is more tangible and has a stronger link to physiological effects than welfare.

Chapter 3

Literature survey Computer vision

3.1 Chapter introduction

This chapter discusses different ways of solving the main computer vision related challenges of this report by first presenting the technical description of the problem, moving on to discussing the input, the output and the core obstacles of the problem, and finally how the input-output mapping should be constructed. Two approaches are suggested for the mapping; an end-to-end and a composite pipeline. The composite pipeline disquisition is further split into subsections covering possibilities at each pipeline segment.

3.2 Problem statement

The problem that this report endeavours to solve can be framed as follows.

> *P: Find the transformation T that maps an ordered set of tensors V onto a set of salmon frequencies F.*

This is a very general description of the task of measuring the breathing frequencies of salmon from underwater video recordings taken in fish tanks, however it allows an initial separation of the problem into three parts; input, transformation and output.

3.3 Problem input

The input V is received from cameras, which all have a similar mode of operation. Reflected light hits a lens, is steered through an aperture, before it hits color-sensitive photodiodes. When a shutter, placed behind the aperture, is

closed, the intensity values of the photodiodes are read by an Analog to Digital (A/D) converter, and a tensor of integer values is passed on for further processing ([118], chapter 2). Some cameras, termed event type, omit the shutter by replacing the A/D converters with value triggered integrators, facilitating continuous time registration at the cost of spatial resolution and color[119].

Most consumer cameras today have a frame rate of at least 30 Hz[120], which is around ten times the Nyquist frequency[121] of salmon ventilation, making this frame rate a well suited sampling rate. Considering this, the time discretization induced by the shutter does not pose an obstacle for respiration frequency elucidation, and a reasonable high resolution water proof camera is a good choice for capturing the salmon scenery.

3.4 Problem output

The value of the salmon frequency data at the pipeline output depends on the extent to which it describes salmon stress state. As such, the algorithm should return the breathing frequency itself, as well as variables that correlates with both ventilation rate and stress. To formalize, the output should be points on an underlying function $g() : M \to \mathbb{R}$ in some euclidean topology M, where the axes of M align with those of frequency variability, and the function output of g is a ventilation frequency.

The review in the previous chapter (chapter 2) suggests that the breathing frequency of the shoal depends on two main factors; time and individuals. The former is a proxy for other variables, like temperature and dissolved oxygen, that can be ultimately linked to stress. Considering this, the output set should consist of four-tuples such that $F = \{\tilde{f} | \tilde{f} = (f, t, c, l)\}$ where f is a single frequency, t is a short time range c is an identifier for salmon individuals and l is the location of a fish. The l variable allows linking $\tilde{f}$ instances with close time proximity, ensuring redundancy in case of occasional errors in c. From $\tilde{f}$, then, it is possible to estimate g(t, c) = f. As g() is a function dealing with biological relationships, it will be highly stochastic, and as such primarily has descriptive power over general trends, and not necessarily specific details.

3.5 Core obstacles

Before delving into the construction of the pipeline T, it is useful to state some of the obstacles that must be overcome.

1. The size of the input data is huge. A conservative estimate is one million pixels per image channel and a frame rate of 60 Hz[120], summing to over ten billion pixels for each minute of video.

2. The output is far smaller than the input, and has a completely different form.

3. Salmon rotate, translate, deform and are exposed to different lighting conditions, complicating detection, individual identification and keypoint localization.

4. Lense and image-water interface cause deviation from an ideal pinhole camera model, resulting in deformation in the 2 Dimensional (2D) mapping of the salmon scenery unless compensated by an elaborate camera model. This comes in addition to the information loss intrinsic to any 3 Dimensional (3D) to 2D mapping.

5. Individual salmon look very similar, which complicate salmon identification.

Considering the size of V, deep learning should, and does, play a critical role in the pipeline. This machine learning method is able to learn very accurate transformations from input-output relationships, and shows better results than manually engineered features as long as the dataset is big enough[122]. Below, two different approaches for incorporating deep learning into T (the transformation from videos to salmon breathing frequency information) are discussed.

3.6 End-to-end training

The output of the breathing frequency extraction pipeline F could be annotated directly on a small video sequence. Assuming a minimum salmon frequency of 0.8 Hz (see table 2.1), a 75 frame video with a frame rate of 60 Hz is sufficient to capture at least one period of salmon breathing. As several salmon can be present in one frame, each salmon in each frame must be labeled with the identifier c. Then, one datapoint consists of a $\mathbb{R}^{75\times W\times H\times 3}$ tensor as input, and a tuple $(f \in \mathbb{R}^N, t \in \mathbb{R}^1, c \in \mathbb{N}^{75\times N}, l \in \mathbb{N}^{75\times N\times 4})$ as output, where N is the number of salmon in the sequence, l is the locations of all salmon and (W, H) is the dimensions of the image. Assuming seven fish and fourteen[1] dot structure classes, 52500 classes and locations must be annotated for a dataset with 50 instances from each class.

Here we consider two types of deep learning frameworks that can be employed for video data, whereof the first is sequentially passing frames into the network, and letting the network remember previous images by the use of memory units. These memory units can be backward connections in the network([123], chapter 15), Long Short-Term Memory (LSTM) cells[124] or image LSTM cells[125]. Object detection and classification have been successfully solved by exploiting image LSTM cells[119], and extending such a network with a frequency estimation head should be trivial (at least conceptually).

In the second framework images are stacked in a new dimension, and this large tensor is treated as a single unit. This has been successfully applied to

[1] The network must learn each side of the salmon independently.

action recognition[126], and one could frame the problem P as an action classification task by creating classes for all permutations of individual fish and frequencies. To be more specific, it is possible to define c as $c \in \{a10, a12, ..., n22, n24\}$, where the letter corresponds to the fish individual and the number corresponds to frequencies (in decihertz). These kind of networks have a high number of tunable parameters, and demand more data and training time than their frame-by-frame counterparts[126]. Some complexity reduction can be achieved by different methods, one of these is to incorporate Spatio-Temporal short term Fourier Transform (STFT) blocks into the neural network. These blocks use fixed convolutional layers to extract local Fourier information, together with learnable linear weights determining how the fixed Fourier channels correlate[127].

3.7 Composite pipeline

3.7.1 Deep learning

The end-to-end networks described above are intriguing, however their large data requirements make them impractical for a thesis project. One alternative approach is to use deep networks as part of a composite pipeline that eventually outputs F. The primary external manifestation of salmon breathing is its jaw movements, which can be inferred by the evolution of framewise jaw pose. The problem of jaw pose estimation is a special case of the more general keypoint detection problem, satisfactory solved by networks such as Keypoint Region-based Convolutional Neural Network (RCNN)[128].

Keypoint RCNN has as input the same V as the problem statement P, and can learn to predict keypoints, bounding boxes and classes through supervised learning techniques. Using the network to predict classes would cause the same large training data predicament as faced earlier (section 3.6), but by disregarding classification in the first deep learning phase, reasonable salmon detection and keypoint localization results can be obtained by annotating only 50 frames[2].

Jaw movement detection could be enhanced by some classical machine learning methods, such as incorporating optical flow[129][130] grids into the network, *Exempli Gratia* (e.g.) by warping the color image grid[131] in a way that enhances (zooms in on) locations with high velocity. Due to the deformation, distortion and complex motions of the salmon, however, creating a beneficial warping is far from trivial. Some of the problems that pose themselves are inconsistent warping between frames and warping of non-jaw areas.

Even though jaw pose estimation seems like the most straight forward way to evaluate respiration frequency, other biological mechanism could be elucidated instead. One such mechanism is the hydromineral balance, which can be measured by change in colorimetric properties of water based on calcium content[132]. As respiration is correlated with hydromineral balance (section 2.3.2), detection of hydromineral disturbance would indirectly provide information on salmon respiratory state. It is unlikely that this method would be able to extract

[2]This number is based on feasibility examination performed by the author.

ventilation information about individual salmon, and the extracted respiration information would probably be imprecise (due to the very small color changes the hydromineral disturbance induce). As such, measuring jaw pose evolution come across as a better approach for solving the problem P.

3.7.2 Linear assignment

Assuming two consecutive frames with object and keypoint detections, any temporal inference of the objects require a method that matches items between frames. A simplistic way to accomplish this is to match instances by bounding box Intersection over Union (IoU) score (see section 4.2.6) across frames. This requires a tight thresholding parameter and is prone to wrong matches when detections are overlapping. More sophisticated methods can be explored by formalizing the problem as a linear program, as shown in equation 4.7[133]. Here, a cost matrix $\mathbf{c}$ is introduced to define the proximity between all salmon detections in the two frames, as well as an index matrix $\mathbf{x}$, representing the matches of the salmon between frames.

Dantzig showed that the linear assignment problem (eq: 4.7) can be solved by classical linear programming algorithms such as the simplex method [134][3]. This approach is not very efficient, with exponential time complexity for some unfortunate configurations[135], motivating the development of better algorithms specifically tailored for item matching.

A well performing group of algorithms focuses on the idea of applying the primal-dual relationship of optimization theory to bipartite graphs. Kuhn[136] and Munkres[137] started this line of research when they framed the linear assignment problem as a maximum flow problem, which they solved by iteratively bringing the dual and primal solutions closer to each other, until convergence at the optimum.

Although still popular, this algorithm is outperformed by newer methods[138]. A popular algorithm, used in modern python libraries[139], is presented by Crouse[133] as the culmination of a long line of contributions originating from Kuhn and Munkres. The first step was taken by Tomizawa[140], which improved on Munkres method by framing the linear assignment method as a shortest path problem, solved by repetitive uses of Dijkstras[141] shortest path algorithm. Jonker and Volgenant[142] further improved on this idea, before Crouse finally demonstrated how to deal with the case of different number of objects to match ($N_r \neq N_c$).

All the methods discussed above converge at the global optimum, and considering the small problem size when matching salmon across frames, any of the mentioned algorithms will perform satisfactory. The main engineer decision in this part of the pipeline is what kind of cost the $\mathbf{c}$ matrix should be filled with. The euclidean distance is often a standard choice when measuring distances, however employing the Mahalanobis distance[143] instead is intriguing. This

[3]Strictly speaking, he demonstrated it for a closely related problem known as the transportation problem.

measurement weights the euclidean distance by the inverse covariance matrix, so uncertain detections have reduced influence when the distance is short, and increased influence when the distance is long. As will be discussed later, the current salmon detection could be matched with not only the previous salmon, but with some composition of several previous measurements, contained in an item known as a tracker. Some of these trackers, like the Kalman filter, keeps an updated covariance matrix describing the uncertainty of the composite state estimate ($\mathbf{P}$ in the Kalman filter). This covariance matrix can be used for calculating the Mahalanobis distance in the linear assignment problem.

3.7.3 Object tracking

Matching detections between frames can provide good trajectories, however any one frame without detection of a fish will end that track. Several methods have been proposed to accurately track multiple objects in the presence of lost detections, false positives, and noise, with the Kalman filter[144] based algorithms being popular for computer vision applications[145]. These types of trackers require a model of the swimming behaviour of the fish, which could be as easy as assuming some coherence of swimming direction and velocity, or something more advanced, like what was developed in Føres fish modelling experiments[146][147][148].

A possible extension of a purely linear velocity based Kalman filter is to include salmon rotation and camera projection into the mathematical model. The nonlinearity of the new model requires some adjustment to the Kalman filter, such as seen in the Extended Kalman filter or the Unscented Kalman filter[149].

Another possible improvement of the tracking method is to replace the simple frame-by-frame linear assignment association method with a more sophisticated probabilistic algorithm, such as Cox and Reids multiple hypothesis tracking algorithm[150][151].

3.7.4 Jaw poses

Tracking keypoints attached to the same salmon over a time period allows estimating jaw pose evolution, which can be used for respiration frequency estimation, as long as the salmon is breathing by buccal pumping. Several jaw poses can be envisioned, such as euclidean pixel distance between the upper and lower jaw, angle of mouth opening, or jaw distance after normalizing the salmon size by warping it onto a fixed configuration by the application of a homogeneous transformation matrix. In table 3.1, the strengths and weaknesses of the different considered methods are compared. Generally, as the jaw pose estimation method become more complex, the robustness decrease, as more keypoints must be correctly placed for the jaw pose to be correctly estimated. On the other hand, more keypoints makes it is easier to determine whether the jaw pose measurements are correct, as fewer relative keypoint placements are valid.

Features	Euclidean	Angular	Normalized euclidean
Number of keypoints	2	3	6
Rotation invariance	Dorsoventral and lateral axes	Lateral axis	All axes
Scale invariance	No	Yes	Yes
Ease of error detection	Difficult	Medium	Easy
Robustness	Good	Medium	Poor
Amplitude measurement	No	Yes	Yes

Table 3.1: Overview over methods of jaw pose estimation

3.7.5 Frequency extraction

The output after Keypoint RCNN detections, linear assignment and tracking is a set of jaw pose time series of unequal length. These contain oscillations corresponding to the mouth opening frequency of salmon jaws, together with offset drift, varying amplitude and noise. As the jaw pose errors come from wrong neural network detections, they will not necessarily be described by a well-behaving probability distribution, and their probability distribution could be both multimodal and skewed. Both the complex dynamics of the true jaw gape evolution and the difficult error distribution makes direct application of frequency extraction algorithms such as autocorrelation or Fourier transforms, as well as classical outlier rejection algorithms, unsuited.

RAndom SAmple Consensus (RANSAC) type algorithms can to some extent overcome difficult noise distributions by sampling a subset of (time series) points, estimate a model from the sampled points (by e.g. the Levenberg-Marquardt (LM) method, see section 4.9.1), and evaluate the model fit by counting inliers[152]. This moves the problem of discerning outliers from before to after an underlying model exists. The model to be fitted could be a binary switch explaining whether the jaw gape is larger than the mean, or a sine wave. The original RANSAC method[153] can be extended with a series of improvements such as smart sampling[154][155], effective model validations[156], refining the best model[157] or more elaborate loss functions[158].

Other options for detecting the time series frequencies are the use of genetic algorithms[159], or exploring clever outlier pruning algorithms suited for difficult outlier distributions.

3.7.6 Salmon identificators

The earlier parts of this chapter completes the pipeline for three of the four elements in the output tuple $\tilde{f}$. In the final parts of this chapter, the last element of $\tilde{f}$, c (individual salmon class), is dealt with. Salmon classification is not strictly necessary to elucidate frequency change over time, however it reduces estimation variance if there is any correlation between salmon individuals and frequency inside a school. Earlier work has demonstrated that computer vision methods can differentiate individual fish both for salmon [160] and other species[161].

Cisar[160] used the salmon dot structure of the dorsoanterior quadrant of the fish body to distinguish between fish instances. His out of water photography setup made it possible to remove almost all pose dependent variability, which facilitated a dot by dot distance calculation between fish images that directly explained the dissimilarity between fish. Although the problem Cisar solved is an easier problem than the one discussed here, it demonstrate the feasibility of salmon recognition by dot structure.

The amount of literature on human re-identification is far greater than that of salmon. Typical approaches in this field is to match equal individuals by comparing distances between body landmarks[162], gait[163], height[164] or body area. These strategies could also *mutatis mutandis* be applied to salmon.

3.7.7 Salmon identification

Assuming the use of Keypoint RCNN in the pipeline, clippings of salmons displaying part of its dot structure are readily available. This is a natural input to the salmon identification algorithm, and the output is given from F as a single class represented by an integer. The depicted fish exhibits texture, deformation and 3 dimensional rotation and translation, complicating the problem. Furthermore, fish eye cameras add radial distortion when mapping a fish scene onto camera sensors. If the fish is sufficiently far away from the camera, the texture becomes inconsequential and the radial distortion approaches constant over the span of the fish. Deformation is significant over the entire fish, however the head is almost entirely acting like a rigid object, except for the lower jaw. If the region of interest is extended further posterior on the fish, more distortion occurs, however the extent of the distortion can be considered small as long as the region of interest does not extend past (more posterior than) the dorsal fin. By disregarding the minor effects of uneven scaling, deformation and texture, the fish-camera system can be modelled as a rigid plane exposed to a pinhole camera.

One way to classify individual fish is to train a convolutional neural network[123] classifier. The filters of deep CNNs (Convolutional Neural Networks) are powerful when it comes to pattern recognition under spatial variance, however they cannot, without certain tricks, deal with rotation, scaling or deformation of objects. A common approach to overcome this is to augment the input data in plausible ways, and let the network learn distortion invariance by generating a set of parameters for each configuration[165]. For rotation invariance specifically, an alternative strategy is to learn only one canonical filter representation, apply this to the input in a set of rotated configurations, and max pool the feature maps at the network output[166]. Also restricting convolutional weights to adhere to isotropic constraints are popular[167]. A final possibility in the search for pose invariance is to reduce the need for distortion robustness of the network by mapping the input objects to a predefined configuration before passing them through the network. By exploiting the keypoints found by the Keypoint RCNN network, a homogeneous transformation can transform the salmon clippings to a predefined pose (see figure 6.3). This

preserves the salmon dot structure as long as the simplifications of the previous paragraph are valid. The mapping could be made even more precise by rectifying the video[168] before applying the Keypoint RCNN model to the salmon recordings, or introducing more keypoints when calculating the homogeneous transformation matrix.

A main consideration in salmon reidentification problems is that the amount of data for each class can be very low, or even missing, due to the infeasability of building a database over every fish in a farm. First consider the case where the number of fish are limited, and it is possible to build a small individual fish dataset. This problem is known as few-shot classification.

Instance recognition is usually performed in four steps; instance detection, instance alignment, feature extraction and classification[169], and the case of few-shot learning is no exception[170]. The first step is handled earlier in the salmon breathing frequency extraction pipeline (section 3.7.1 about Keypoint RCNN), the second step was discussed in the first paragraphs of this section (homogeneous transformation), and the last two steps will be dealt with below.

In SOTA methods, feature extraction is mostly performed by some sort of neural network, while the classification itself tests a distance measure[171]. This distance could be euclidean, cosine, Mahalanobis[170], L1 (as seen in siameese networks) or neural networks[172]. Below, some popular feature extraction methods are presented.

1. K Nearest Neighbors (KNN) networks consider the image to be a point in a $\mathbb{R}^{W \times H \times C}$ space, and classify novel datapoints by a max vote of the k nearest neighbors ([173] for an introduction to the k nearest neighbor algorithm).

2. Prototypical networks learn an informative embedding space, and represent each class as the mean of a set of examples in this space. At inference time, the test point is assigned to its closest mean[171].

3. Siameese networks accept two images, and outputs a value describing the similarity of these images[174]. Traditionally, examples of positive and negative pairs are needed during training, however a late approach is able to generate a descriptive feature map without negative pairs, and without convergence to the trivial solution[175].

4. The finetuning approach consists of first training a network on a huge, general dataset, and then continue to train this network with a small, few-shot dataset. Finetuning can be done with or without frozen layers, generally performance increase by not freezing layers[176].

5. Model Agnostic Meta Learning specifically trains on quick adaption to new datasets[177].

6. CNAPS use a novel layer, FiLM, which translate and scale fixed feature maps. When exposed to novel datasets, only the FiLM parameters and

the final classifier are adjusted, enabling quick adaption on a small number of parameters to tune the network according to the few-shot dataset[178].

7. TADAM is relatively similar to CNAPS, but it has a learnable scale factor before the softmax output to reduce the network sensitivity to extreme values and choice of distance metric[179].

Several of these few-shot classification approaches require a neural network backbone, i.e. a convolution neural network that the more specific method is built around[180]. These backbones also work as stand-alone classification networks, with color images as inputs, and sets of class probabilities as outputs. Below, the main peculiarities of three possible neural network backbones are described.

1. AlexNet[181] is one of the first, and perhaps the most influential, paper on deep CNN networks with Graphics Processing Unit (GPU) accelerated training[182][183]. It has a classical pyramidal shape, and can be considered a baseline for later CNN networks.

2. Efficientnet[184] is able to improve accuracy and decrease the number of parameters (compared to AlexNet) by a balanced scaling of depth, width and resolution of the network.

3. Resnet[185] uses skip connections between blocks (collections of a few neural network layers) to combat the problem of vanishing gradients, and are through this capable of greatly increasing the network depth. Benali suggests that this network is a good neural network backbone[180].

The ideal salmon re-identification network is one that is able to correctly identify already seen salmon, detect images that belong to a novel salmon, and facilitate later re-identification of the novel salmon at operation time. This can be considered a zero-shot classification task[186], as some instances are unseen at deployment time.

Several of the few-shot models can trivially be generalized to zero-shot models. Siameese networks can be trained on a set of fish instances, and at deployment the similarity between a novel fish and each of the seen classes can be calculated. If the most similar class has too low score, the fish is a novel instance, while a high score for the most similar class allows the fish to be assigned to that class. Prototypical networks can be adapted to the zero-shot case by clustering seen fish in an embedding space by methods such as k-means[187], DBSCAN[188] or Gaussian mixture models[189]. In recent years, some clustering techniques have been specifically designed for unsupervised re-identification tasks[190].

For the case of salmon re-identification, leveraging knowledge of the problem domain can provide better results than general re-identification solutions. A discriminative feature on salmon is its dot structure[160], so by generating a salmon dot mask, and then use this to classify salmon instances, better results

might be achieved. Several ways of detecting dot masks can be imagined, three of them are suggested below.

1. Use a segmentation neural network ([191] for an extensive review) to generate a dot mask directly.

2. Split the image into small windows, and classify each window as "dot" or "no dot", as seen in[160]. Afterwards, a dot mask is constructed by concatenating the window classification results.

3. Use classical machine vision methods such as adaptive thresholding (see section 4.6) to detect dot areas.

Consider the case where masks of prototypes of fish individuals, together with the mask of a novel fish, are available. To classify the novel fish, one can compare its dot mask directly with that of the prototypes, by assigning the dots on the two masks to each other by Jonker-Volgenant[142] and sum up the euclidean distance over all the dot matches. The sum of the distances can then be used as a similarity metric between the protoype and the novel instance. Another option is to transform the dot mask to an embedding space, which could be as small as the $\mathbb{R}^1$ space *Number of dots*, and use distances in this space to evaluate proximity to prototypes.

A clear issue with this approach is that it requires a very accurate instance alignment (pose match) between the novel fish and the prototypes. It is doubtful that a homogeneous transformation with four points detected by the Keypoint RCNN is accurate enough, however using more keypoints, or a more complex transformation, might improve the accuracy of the instance alignment.

Chapter 4

Theory

4.1 Chapter introduction

This chapter presents the mathematics behind the algorithms that are employed in the final breathing frequency analysis approach. It is structured according to the order the different methods appear in the final processing pipeline, and as such will begin with discussing neural networks, before continuing with assignment methods, the Kalman filter, homogenous transformations and pixelwise image transformations. After this, a probabilistic method for removing uncertain classification sets will be presented, some statistical methods are displayed, and algorithms for extracting frequency information from time series data are elaborated on.

4.2 Neural networks

The cerebral cortex is the outer layer of our brain, and carries out essential human functions including emotional response, execution of cognitive functions, spatial awareness, motion planning, voluntary movement and sensory processing ([192], table 13.1). It spends 30 percent of its capacity on processing visual information[193], demonstrating the importance and difficulty of understanding complex scenery. Tackling this problem by artificial means has seen several proposed solutions, with deep neural networks establishing itself as the top contender in recent publications[194][122]. Deep neural networks mimic the way our cortex processes visual information. It does so by organizing huge amounts of nodes (neurons) and connections (synapses) in a layered fashion, in order to create a generalized function with impressive descriptiveness. Up to 135 billion parameters have been used in a single function approximator[195].

4.2.1 Fully connected layers

The conceptually easiest, and mathematically most complex, layer is known as a fully connected layer. Such a module calculates a weighted sum of all nodes in the previous layer, passed through a nonlinear function. Let σ be an elementwise nonlinear function, $\mathbf{X}$ be a row vector containing the inputs to the layer, $\mathbf{W}$ be a square matrix containing the weights between all nodes, $\mathbf{b}$ be a row bias vector and $\mathbf{Y}$ be the output of the layer. Then, a fully connected layer can be represented as shown in equation 4.1 ([123], chapter 10).

$$\mathbf{Y} = \sigma(\mathbf{X}\mathbf{W} + \mathbf{b}) \tag{4.1}$$

4.2.2 Convolutional layers

Fully connected layers are not well suited for extracting features from images. They can not learn spatial invariance, and connection between all pixels in an image is excessive. Convolutional layers improve on both of these issues by letting filters scan previous layers for learned items, facilitating both detection and localization of objects with a single weight set ([123], chapter 14). If the location of an object is irrelevant for the task, pooling layers can reduce complexity and leave the network almost[196] invariant to spatial location. A common network structure is to begin with convolutional and pooling layers to extract features, and then end the network with some fully connected layers to combine all the local feature information into a global output. To move from convolutional to fully connected layers, a flattening of the feature tensor is necessary.

Keeping to the notation introduced above, an arbitrary layer $\mathbf{X}$ of a convolutional neural network will be a depthwise stack of feature maps. Let the subscripts w, h and d describe width, height and depth of the input map, and f and s describe receptive field and stride. The receptive field of a kernel is the dimensions in the input layer $\mathbf{X}$ that is captured in one node in the output layer $\mathbf{Y}$, and the stride of a layer is the size of pixel shift between each filter computation. Then, equation 4.2 ([123], chapter 14) describes how a node in the output layer is related to the previous feature tensor in a convolutional neural network. Note that the dimensions of $\mathbf{X}$ are extended compared to the fully connected layer representation, from an $\mathbf{X}$ spanning only the two dimensions x and y (a feature map plane), to an $\mathbf{X}$ spanning a three dimensional (x, y, z) feature map space. The CNN filters vary along the added depth axis in the feature maps. No stride is applied in the depth direction.

$$Y_{i,j,k} = b_k + \sum_{u=0}^{f_h-1} \sum_{v=0}^{f_w-1} \sum_{z=0}^{f_d-1} \mathbf{X}_{i \times s_h+u, j \times s_w+v, z} \times \mathbf{W}_{u,v,z,k} \tag{4.2}$$

4.2.3 Pooling layers

Normally, pooling layers are used with more or less regularity between convolutional layers. They map a cube in the feature tensor to a single value, often by using the max operator. Keeping to the notation of equation 4.2, this cube is of size $f_h \times f_w$ along the spatial axes of the feature map, and includes f_d filters along the depth direction. By letting a function p define a pooling operation (such as the max() operation) acting on a cube of pixel values, one element in the output of a pooling layer is defined as shown in equation 4.3.

$$Y_{i,j,k} = p(\mathbf{X}_{[i\times s_h, i\times s_h+f_h],[j\times s_w, j\times s_w+f_w],[k\times s_d, k\times s_d+f_d]}) \tag{4.3}$$

4.2.4 Training of neural networks

Having described the flow of data through different layers, it is now necessary to elucidate on how the parameters are adjusted during the training phase. The training is done by applying the generalized delta rule (backpropagation)[197], which works in four steps:

1. Passing a feature vector through the network.
2. Compare the generated output vector to a manually annotated output vector by computing the value of a loss function.
3. Calculate the gradient of the loss function with regards to all weights.
4. Adjust all weights by some form of gradient descent.

To understand the power of this method, note that the gradients in the l_{th} last layer can be decomposed. Assume we are in the middle of a fully connected neural network, and wish to adjust a weight $w^l_{2,3}$ that goes from node two in layer $l-1$ to node three in layer l. Let superscript describe layers (indexed by $l-1$, l or $l+1$), subscript describe node connections (nodes are indexed by $\mathbb{N} \leq n$), w describe weights, v describe node outputs (node values) and n describe the size of hidden layer l+1. Then, the gradient of the loss with regards to node $w^l_{2,3}$ is given by equation 4.4. All gradients can be unwrapped based on deeper nodes this way, until a full loss gradient is acquired. *Mutatis mutandis*, the CNN gradients are calculated in the same manner.

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial w^l_{2,3}} = (&\frac{\partial \mathcal{L}}{\partial v^{l+1}_1}\frac{\partial v^{l+1}_1}{\partial v^l_3} + \frac{\partial \mathcal{L}}{\partial v^{l+1}_2}\frac{\partial v^{l+1}_2}{\partial v^l_3} + ... \\ &+ \frac{\partial \mathcal{L}}{\partial v^{l+1}_n}\frac{\partial v^{l+1}_n}{\partial v^l_3})\frac{\partial v^l_3}{\partial w^l_{2,3}}\end{aligned} \tag{4.4}$$

Any gradient descent method can iteratively improve the neural network parameter set ($\mathbf{W}$) after backpropagation has provided $\nabla_W \mathcal{L}(\mathbf{W})$ (the partial derivative of the loss function with regards to each parameter, i.e. $\frac{\partial \mathcal{L}}{\partial w^l_{i,j}} \forall i \in$

$n_i, j \in n_j, l \in n_l$). A good option is a stochastic version with momentum and weight decay, allowing batch training, improving the likelihood of avoiding local minima and reducing the erroneous effects of noisy gradients ([123], chapter 11). Let $\mathbf{b}$ with parameter μ be the momentum, $\mathbf{g}$ be the generalized gradient, λ be the regularization parameter, γ be the learning rate and $\mathcal{L}$ the loss function. Then, the gradient is calculated using a random subset of the training examples of the dataset (batch), and parameters updated, as displayed in algorithm 1[198].

Algorithm 1 Stochastic gradient descent

$\mathbf{g}_t \leftarrow \nabla_W \mathcal{L}(\mathbf{W}_{t-1}) + \lambda \mathbf{W}_{t-1}$
$\mathbf{b}_t \leftarrow \mu b_{t-1} + \mathbf{g}_t$
$\mathbf{g}_t \leftarrow \mathbf{b}_t$
$\mathbf{W}_t \leftarrow \mathbf{W}_{t-1} - \gamma \mathbf{g}_t$

4.2.5 Keypoint RCNN

Architecture

The layers and training methods described above (in the previous subsections of this section) have been incorporated into a SOTA network called Keypoint RCNN, which is described in the original Mask RCNN paper[128][1]. This network is based upon Faster RCNN[200], which again draws inspiration from Fast RCNN[201].

Fast RCNN uses a pipeline with two components; a Region Of Interest (ROI) extractor algorithm, and a neural network to predict classes and bounding boxes of the ROIs. The pipeline uses a conventional CNN to extract a feature map of the complete input image, crops the feature map into several smaller tensors determined by the ROIs, and then pools all the feature map crops to a fixed size. Then, the equally sized feature maps are passed through fully connected layers individually to estimate a softmax class vector and a per-class bounding box (within the cropped area) for each feature map crop. An image from the original paper visualizing the network structure is displayed in figure 4.1a to make it easier to follow the description of the algorithm[201].

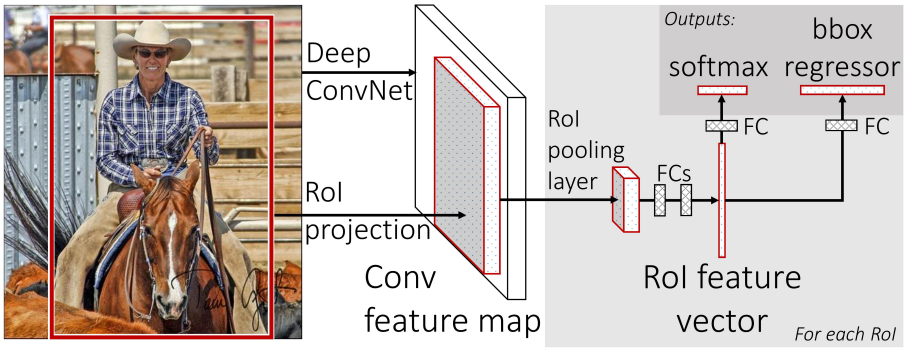
The main drawback with Fast RCNN is the two-step procedure, where the ROI algorithm and the classifier must be developed separately. This is fixed in Faster RCNN, where a common feature map is generated, a separate branch is using these features to generate region proposals, the common feature map is cropped into smaller tensors determined by the extracted region proposals, before these crops are fed into into the ROI pooling layer of the Fast RCNN. The network structure is visualized in figure 4.1b[200].

Finally, Mask RCNN[128] adds segmentation to the network. This is achieved by introducing a third branch after the ROI pooling layer, which uses several

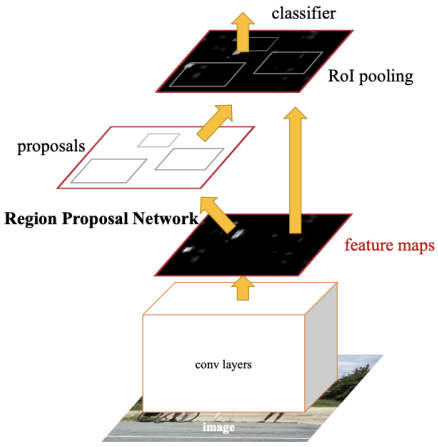
[1]The explanations related to the Mask RCNN construction is partly based on unpublished work of the author[199].

fully convolutional layers[202] to generate a pixel mask for each ROI. As an additional improvement, Mask RCNN uses floating numbers and bilinear interpolation when mapping ROIs to the common feature map, in order to properly align the two domains, which removes the quantization noise present in Fast RCNN. The Mask RCNN architecture is visualized in figure 4.1c[128].

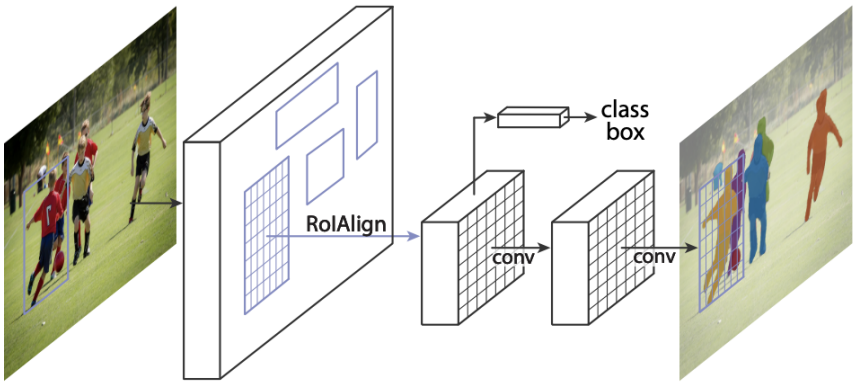
Keypoint prediction is achieved by introducing k one-hot masks to the Mask RCNN model, where each mask represents one keypoint. One-hot masks are binary pixel grids filled with zeros, except for a single 1 value at the location of the keypoint. The final network is now a complete end-to-end instance detection and keypoint prediction net, capable of extracting the anterior body area of salmon, together with informative landmarks.



(a) Fast RCNN[201]



(b) Faster RCNN[200]



(c) Mask RCNN[128]

Figure 4.1: Three architectures based on residual neural networks.

Training

To train the network, a regular back propagation approach is employed, which requires both manually annotated training data, and a loss function.

The training data consists of images annotated with bounding boxes and bounding box specific keypoints, which it learns to predict on images during training. As discussed in the literature review chapter (section 3.7), the detected keypoints must be usable for two tasks.

1. Retrieving states relevant for salmon respiration, which can be concatenated in order to create a ventilation signal.

2. Creating a transformation matrix that maps salmon to a predefined configuration, so that the dot structure can be analysed in order to re-identify fish. A reasonable fixed configuration is one where the plane spanned by the anterioposterior and dorsoventral axes of the fish is parallel to the image plane (see figure 6.3).

To accommodate the first point, the upper and lower jaws must be part of the keypoint set, hence the bounding box needs to cover the fish snout. The second point requires the inclusion of a significant part of the fish body to reduce the risk of keypoint colinearity by increasing the distance between points. Additionally, including parts of the salmon trunk in the bounding box ensures sufficient amount of dots to discern a large number of fish. On the other hand, parts of the most posterior sections (such as the caudal fin) of the fish should be removed to avoid excessive deformations in the fish clip, and to reduce the risk of overlapping salmon detections. A reasonable trade off is to include all aspects of the salmon anterior to the posterior root of the dorsal fin in the bounding box (see figure 6.2a).

By annotating all salmon with the coordinates of the upper jaw, lower jaw and jaw root for jaw gape estimation, and four points for the homogeneous transformation, all bounding boxes contain $K = 7$ keypoints, each represented by two numbers (x, y). See figure 4.2 for a visualization of how a single salmon is annotated. Furthermore, the number of bounding boxes will be equal to the number of fish in the frame, and each box will have dimension $\mathbb{R}^4$, two numbers (x, y) for two of the box corners. Having described the network labels, the next step is to explicate the loss function.

For simplicity, assume batches of single frames. The ground truth targets consist of R bounding boxes contained in a $v \in \mathbb{R}^{Rx4}$ tensor and a set of K different keypoints for each bounding box, such that all keypoints are contained in a $s \in \mathbb{R}^{RxKx2}$ tensor. Additionally, the tensor $\mathbf{p} \in \mathbb{R}^R$ contains one integer representing salmon class for each bounding box. For the network output, let the number of classes be C, the number of region proposals be R_{ROI}, and the dimension of the keypoint masks be $w_k \times h_k$ pixels. Then, the Keypoint RCNN estimates consist of a class probability tensor ($\hat{p} \in \mathbb{R}^{R_{ROI},C}$), a bounding box regression tensor ($\hat{v} \in \mathbb{R}^{R_{ROI},4\cdot C}$) and a keypoint mask tensor ($\hat{s} \in \mathbb{R}^{R_{ROI},K,C,h_k,w_k}$).

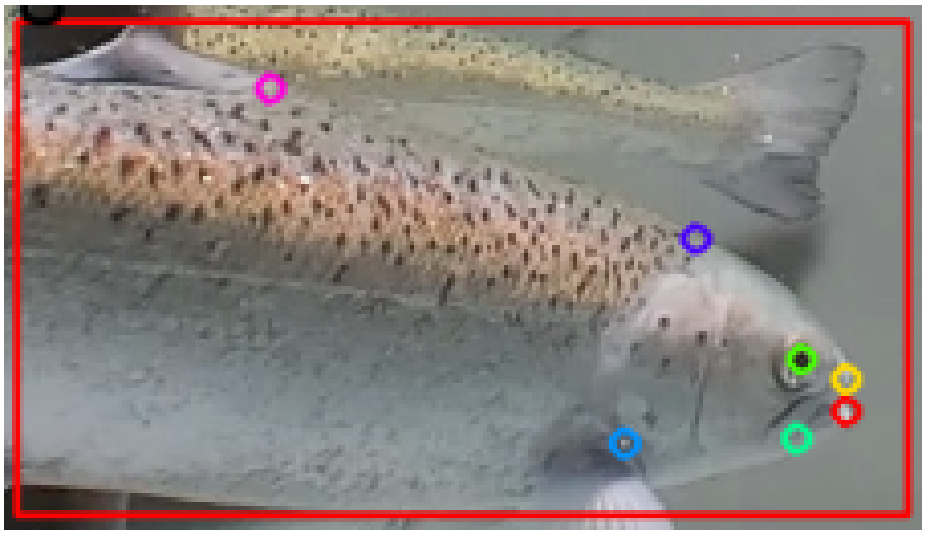


Figure 4.2: Annotation of a salmon instance

The total loss function (equation 4.5a) for a frame is the sum of the losses of each ROI, and each ROI has a loss equalling a sum of classification loss, localization loss and keypoint loss. An optimal mapping between ROI and ground truth is performed by IoU thresholding[200], facilitating a synchronized ordering of the network estimates and ground truth targets during training, such that equal subscripts refer to the same salmon (i.e. p_r and $\hat{p}_r$ refers to the ground truth class and the estimated class of the same salmon, respectively). Localization loss and keypoint loss are only calculated for ROIs that are associated with a positive (salmon) class, hence the encoding function λ is introduced, defined as one when a ROI is associated with a positive ground truth class, and zero otherwise. In the complete loss function (equation 4.5a), N_{cls} is equal to R_{ROI}, and N_{loc} and N_{key} is equal to the number of ground truth bounding boxes in the frame. The individual components of the loss function are expanded on below.

1. The classification loss is estimated as a log loss over the two classes *object* and *not object* (equation 4.5b), where $\hat{p}_r$ refers to the estimated class, p_r refers to the ground truth class, and $sm(\hat{p}_r)$ refers to the softmax value at the neural network output of the estimated class. Furthermore, the binary variable $h \in \{0, 1\}$ is one if $p_r = \hat{p}_r$, and zero if $p_r \neq \hat{p}_r$.

2. The localization loss is calculated as a smooth L1 loss, (equation 4.5c, see the documentation of [198]), where $\hat{v}_r$ refers to the network bounding box estimation values of the ground truth class (p_r).

3. The keypoint loss is calculated as the cross-entropy over an $K \cdot h_k \cdot w_k$ way softmax output (equation 4.5d), where $\hat{s}_r$ refers to the one-hot mask estimates of the ground truth class (p_r).

$$\mathcal{L}(p, v, s, \hat{p}, \hat{v}, \hat{s}) = \sum_{r=0}^{R_{ROI}} \frac{1}{N_{cls}} \mathcal{L}_{cls}(p_r, \hat{p}_r) + \frac{1}{N_{loc}} \lambda_r \mathcal{L}_{loc}(v_r, \hat{v_r}) + \frac{1}{N_{key}} \lambda_r \mathcal{L}_{key}(s_r, \hat{s}_r) \tag{4.5a}$$

$$\mathcal{L}_{cls}(p, \hat{p}_r) = -(h \cdot log(sm(\hat{p})) + (1 - h) \cdot log(1 - sm(\hat{p}))) \tag{4.5b}$$

$$\mathcal{L}_{loc}(v, \hat{v}) = \begin{cases} 0.5 \cdot (v - \hat{v})^2, & if \|v - \hat{v}\| < 1 \\ \|v - \hat{v}\| - 0.5, & otherwise \end{cases} \tag{4.5c}$$

$$\mathcal{L}_{key}(s, \hat{s}) = -\frac{1}{K \cdot h_k \cdot w_k} \sum_{k=0}^{K} \sum_{i=0}^{h_k} \sum_{j=0}^{w_k} (s_{k,i,j} \cdot log(\hat{s}_{k,i,j})) + (1 - s_{k,i,j}) \cdot log(1 - \hat{s}_{k,i,j})) \tag{4.5d}$$

4.2.6 NMS and IoU

After training Keypoint RCNN for salmon detection, every true salmon instance will be detected by multiple overlapping bounding boxes. One way to deal with this situation is by using the Non Maximum Suppression (NMS) method, built on the IoU metric.

The IoU of two bounding boxes A and B is the correct overlap of the boxes, divided by the total proposed area of both of them, defined in equation 4.6.

$$IoU = \frac{A \cap B}{A \cup B} = \frac{true\ positives}{true\ positives + false\ negatives + false\ positives} \tag{4.6}$$

The NMS algorithm is build on this metric, and works by removing all bounding boxes with too high IoU overlap, keeping the most confident bounding boxes. Let B be the initial set of bounding boxes in a frame, C a confidence function ($C : b \rightarrow [0, 1]$, with b being a bounding box in B), T a threshold and U the set of non-overlapping bounding boxes. NMS is then described in algorithm 2.

Algorithm 2 NMS

Require: B, T
  $U \leftarrow \emptyset$
  while B *not* $\emptyset$ **do**
    $\tilde{b} \leftarrow \{\tilde{b} | C(\tilde{b}) \geq C(b) \forall b \in B, \tilde{b} \in B\}$
    for $b \in B$ **do**
      if $IoU(\tilde{b}, b) > T$ **then**
        $B \leftarrow B \setminus \{b\}$
      end if
    end for
    $B \leftarrow B \setminus \tilde{b}$
    $U \leftarrow U \cup \tilde{b}$
  end while

4.3 Assignment methods

The problem of linear assignment can be formalized as the linear optimization problem below (equation 4.7), with the cost matrix **c** and index matrix **x** describing the proximity and matches between the two assignment sets, respectively[133]. For the salmon tracking problem considered in this report, the two assignment sets are the salmon detections in the current frame, and the set of active salmon trackers.

$$\begin{aligned} \mathbf{x}^* &= \underset{x}{argmin} \sum_{i=1}^{N_R} \sum_{j=1}^{N_C} c_{ij} \cdot x_{ij} \\ s.t. &\begin{cases} \sum_{j=1}^{N_C} x_{ij} = 1 \forall i \\ \sum_{i=1}^{N_R} x_{ij} \leq 1 \forall j \\ x_{ij} \geq 0 \forall x_{ij} \end{cases} \end{aligned} \tag{4.7}$$

This formulation assumes that the number of salmon in the current frame (N_C) is equal or larger than the number of salmon trackers (N_R), however the assumption of $N_C \geq N_R$ implicit in the above formulation (equation 4.7) does not reduce generalizability, as the cost matrix can be transposed without any effects on the original problem[2]. Furthermore, the last constraint is a relaxation of the more intuitive $x_{ij} \in \{0, 1\} \forall x_{ij}$. The reason for performing this relaxation is to ensure linearity of constraints, and the change of constraint does not affect the final solution[133].

In this section a popular solution developed by Crouse[133], leveraging the primal-dual dichotomy of optimization theory, is described. To begin, the dual objective function of the assignment problem[133] is presented in equation 4.8 (see [204] for a thorough description of dual problems).

[2]This is because the symmetry property of distances state d(a, b) = d(b, a)[203].

$$
\begin{aligned}
g(\mathbf{u},\mathbf{v}) = \min_{x,x\geq 0} & \sum_{i=1}^{N_R}\sum_{j=1}^{N_C} c_{ij}\cdot x_{ij} + \\
& \sum_{i=1}^{N_R} u_i \cdot (1 - \sum_{j=1}^{N_C} x_{ij}) + \\
& \sum_{j=1}^{N_C} v_j \cdot (1 - \sum_{i=1}^{N_R} x_{ij})
\end{aligned}
\tag{4.8}
$$

By rearranging, equation 4.8 can be written as shown in equation 4.9[133].

$$
\begin{aligned}
g(\mathbf{u},\mathbf{v}) = \min_{x,x\geq 0} & \sum_{i=1}^{N_R}\sum_{i=j}^{N_C} (c_{ij} - u_i - v_j)\cdot x_{ij} \\
& + \sum_{i=1}^{N_R} u_i + \sum_{j=1}^{N_C} v_j
\end{aligned}
\tag{4.9}
$$

Note that a maximization of $g(\mathbf{u},\mathbf{v})$ has no (finite) solution if $c_{ij} - u_i - v_j < 0 \,\forall\, i,j$ (as x_{ij} has no upper constraint on individual matrix entries), hence the constraint $c_{ij} - u_i - v_j \geq 0$ is introduced (see [133] for a further motivation). Then, the dual objective is minimized with regards to $\mathbf{x}$ at $\mathbf{x} = 0$, which reduces the dual objective to equation 4.10.

$$
g(\mathbf{u},\mathbf{v}) = \sum_{i=1}^{N_R} u_i + \sum_{j=1}^{N_C} v_j \tag{4.10}
$$

Now, since strong duality is guaranteed for linear programs ([204], ch. 13.1), the following dual problem (equation 4.11) will converge to the same optimum as the primal representation (equation 4.7).

$$
\begin{aligned}
& \mathbf{u}^*,\mathbf{v}^* = \underset{u,v}{argmax}\; g(\mathbf{u},\mathbf{v}) \\
& s.t. \begin{cases} v_j \leq 0 \\ \sum_{i=1}^{N_R}\sum_{j=1}^{N_C}(c_{i,j} - u_i - v_j) \geq 0 \end{cases}
\end{aligned}
\tag{4.11}
$$

The solution to problem 4.11 provides the optimal cost of the assignment problem, however it does not directly provide the value of the primal optimization variable at the optimum (it provides $\sum_{i=1}^{N_R}\sum_{j=1}^{N_C} c_{ij}\cdot x_{ij} = g(\mathbf{u}^*,\mathbf{v}^*)$, not $\mathbf{x}$).

Since both the objective and constraints in the original formulation (equation 4.7) are linear, the Karush-Kuhn-Tucker (KKT)[205] requirements are applicable to the problem, and provide necessary conditions for the dual variables at

the optimum ([204], ch. 13.1). This ensures that the complementary slackness condition holds, which states that if the dual inequality variable (v_j) is slack (non-zero), the primal optimization variable must be tight (in the case of this problem, $\sum_{i=1}^{N_R} x_{ij}$ must be one). For problem 4.8, the complementary slackness condition takes the form displayed in equation 4.12. This condition allows the determination of the exact values of $\mathbf{x}$[133].

$$v_j \cdot (1 - \sum_{i=1}^{N_R} x_{ij}) = 0 \forall j \in \{\mathbb{N}_1 \leq N_C\} \tag{4.12}$$

The primal-dual relationship discussed hitherto can, as elucidated by Egervary and König, be applied to edges and vertices in a bipartite graph. Consider the columns and rows of the cost matrix to be represented by vertices of a graph, the dual variables $\mathbf{u}$ and $\mathbf{v}$ to represent vertex budgets (b_v), and the costs in the cost matrix to represent edge weights (w_e). Then, the general form of Egervarys theorem is presented below[206].

> *Let G be a bipartite graph. Let a non-negative integral weight w_e be assigned to every line e, and also let a non-negative integral weight b_v be assigned to every point v. Then, the maximum weight of a t-matching in G is equal to the minimum b-weight of a collection of points from which every line e contains at least w_e elements.*

Crouse[133] leverage this theorem to alternately reduce edge weights and increase vertex weights until the primal and dual optimum have reached equality. A central concept in his algorithm is the notion of a shortest augmenting path, which is explicated below through a list of definitions from the field of graph theory.

1. A **matching** is a set of edges such that no edges share vertices. A **maximum matching** is the highest cardinality matching in a given graph. An **assignment** is one of the edges in the matching.

2. Consider graph edges E and vertices V, and let a matching $M \subset E$ be given. An **alternating path** is then defined as a path alternating between M and $\{E \setminus M\}$.

3. An **augmenting path** is an alternating path starting in a row (a left vertex in the bipartite graph) not connected to M, and ending in a column (a right vertex in the bipartite graph) not connected to M. By reassigning edges along an augmenting path, the cardinality of the match, $|M|$, increase by one.

4. A **shortest augmenting path** is an augmenting path that solves $min \sum E$.

Having specified the primal and dual optimization problem that is to be solved, as well as how this can be framed as a graph problem, the full linear assignment algorithm is presented below. Crouse[133] describe five steps to follow in order to find the optimal matching from a given cost matrix.

1. Initialize the data structures: When the algorithm is implemented into code, this step must be considered in detail. For the high-level theoretical description provided here, it suffices to say that the dual variables **u** and **v** is initialized to zero vectors.

2. Find the shortest augmenting path The algorithm begins by assigning the first row vertex (u_1) in the graph representation of the cost matrix to the column (v_j) that has the lowest edge weight connecting it to u_1. This row can never be connected to a column by a smaller edge weight, hence this assignment provides a lower bound on the price (vertex budget) of the vertices (u_1 and v_j) it connects. As the algorithm propagates to subsequent rows (u_i) in later iterations, the event might occur that the edge with the lowest weight (connected to that row) leads to an already assigned column, hence the simple minimal assignment (using the edge with the lowest weight) is not applicable. Instead, a slightly modified version of Dijkstras shortest path algorithm[141] is used, that search for a shortest augmenting path starting at the current row (u_i) and ending at an unassigned column (v_j). Reassignment along this path increase the matching cardinality by one, while resolving edge conflicts (e.g. the smallest edges of two rows connect to the same column) in an optimal manner. This shortest path is then saved for later steps in the algorithm.

The search for a shortest augmenting path is performed on a reduced cost matrix, in which each element is calculated as shown in equation 4.13. The entries of the reduced cost matrix will decrease as the algorithm propagates, and when the algorithm terminates, all entries in the matrix corresponding to an assignment will be zero. Note that the weights of the reduced matrix will never be negative, hence the requirements for Dijkstras shortest path algorithm is fulfilled.

$$\bar{c}_{ij} = c_{ij} - u_i - v_j \tag{4.13}$$

3. Update the dual variables to assure complementary slackness: Equation 4.11 states that the dual variables should be maximized under the constraint in equation 4.14, which means that the sum of the two vertices in an assignment (u_i and v_j) should equal the value of the edge connecting them (c_{ij}). After a shortest augmenting path is found, all assignments weights (edge weights of assignments) will stay the same or increase when this shortest augmenting path is used to increase the cardinality of the matching. Hence, the constraint given in 4.14 will no longer be tight, and the variables u_j and v_j can be increased until the sum of the dual variables equal that of the new edge weight.

$$c_{ij} - u_i - v_j \geq 0 \tag{4.14}$$

4. Augment the previous solution with the shortest path: After updating the dual variables, the graph matching is changed according to the shortest augmenting path found in step two of the algorithm. This new matching

will consist of the right assignments (assignments from row to column) in the shortest augmenting path.

5. Loop: If all rows have been assigned, then the problem is solved. Otherwise, go to step 2, and consider the next row in the graph. As discussed in step two and three, the propagation of the algorithm leads to decreased values in the reduced cost matrix, and increased dual variable values, which in effect is the dual and primal problems converging to the same optimum.

When the algorithm exits, the optimal value ($\sum_{i=1}^{N_R}\sum_{j=1}^{N_C} c_{ij}{\cdot}x_{ij}^* = g(\mathbf{u}^*, \mathbf{v}^*)$) is found as the sum of the elements of the dual variables $\mathbf{u}$ and $\mathbf{v}$, and the value of the primal optimization variable $\mathbf{x}$ is found by considering the elements of the reduced cost matrix (a zero value in the reduced cost matrix is a necessary condition for being an optimal assignment), together with the complementary slackness constraint from equation 4.12. The sum of the entries of the (unreduced) c_{ij} at the locations specified by $\mathbf{x}$ will be equal to the optimal dual value ($g(\mathbf{u}^*, \mathbf{v}^*)$).

4.4 Kalman filter

Assume a set of detections at discrete timesteps, and that a matching exists between the detections at time k-1 and time k. Let the detections be defined by their euclidean coordinates $\mathbf{p} \in \mathbb{R}^2$. The problem of finding the optimal estimate of the real $\mathbf{p}$ from a number of measurements subjected to uncertainty is a reoccurring problem in several technical fields, and was solved by Kalman in his famous paper from 1960[144] in the case of linear, discrete models exposed to Gaussian noise.

Estimating the state by a Kalman filter requires a known system model. A salmon in a video stream moves in three dimensions, and has an erratic dynamic that is difficult to model, however the 2D projection of a salmon trajectory onto camera sensors tend to be approximately linear with fixed velocity. Hence, the model displayed in equation 4.16 can predict salmon movement in the immediate future with reasonable accuracy. Here, Δt is the time between frames, $\mathbf{x} = [\mathbf{p}^T\ \dot{\mathbf{p}}^T]^T$ is the true state vector, $\mathbf{w}$ is the model uncertainty and $\mathbf{v}$ is the measurement uncertainty.

$$\mathbf{x}_k = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x_{k-1}} + \mathbf{w}_k = \mathbf{A}\mathbf{x_{k-1}} + \mathbf{w}_k \tag{4.15}$$

$$\mathbf{z}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \mathbf{x}_k + \mathbf{v}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k \tag{4.16}$$

At any timestep k, the best linear estimate of $\mathbf{x}_k$, $\hat{\mathbf{x}}_k$, will be a weighted sum of the two sources of information available; the model estimate $\hat{\mathbf{x}}_k^-$ and the

measurement $\mathbf{z}_k$. This is displayed in equation 4.17, with the weight $\mathbf{K}$ called the Kalman gain.

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^-) \tag{4.17}$$

Now, call the covariance matrix of the expected error between the true and estimated state $\mathbf{P}_k$, and observe that the optimal state estimate is at the point where this is minimized. By simple gradient considerations, a recursive formula calculating the Kalman gain $\mathbf{K}$ at this minima can be found, as demonstrated in[207]. This iterative method is presented in algorithm 3, with superscripts $-$ describing estimates before received measurements, subscripts describe time step, hat describe estimates, and $\mathbf{R}$ and $\mathbf{Q}$ are describing the magnitude of the measurement and model uncertainty, respectively.

Algorithm 3 Kalman filter

Require: $\hat{\mathbf{x}}_0, \hat{\mathbf{P}}_0$
  $k \leftarrow 1$
  while $True$ **do**
    $\hat{\mathbf{x}}_k^- \leftarrow \mathbf{A}\hat{\mathbf{x}}_{k-1}$
    $\mathbf{P}_k^- \leftarrow \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q}$
    if $\mathbf{z_k} \neq \emptyset$ **then**
      $\mathbf{K}_k \leftarrow \mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R})^{-1}$
      $\hat{\mathbf{x}}_k \leftarrow \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-)$
      $\mathbf{P}_k \leftarrow (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^-$
    end if
    $k \leftarrow k + 1$
  end while

Several ways of modelling the noise, in order to determine $\mathbf{Q}$, exists. Here a piecewise white noise model will be employed[208]. Let σ_i be a random acceleration disturbance, constant over each Δt, and with $i \in \{x, y\}$ describing the direction of acceleration. This noise will act as an additive disturbance on the model dynamics, with magnitude $\sigma_i \Delta t$ for $\dot{\mathbf{p}}$, and $\sigma_i \frac{\Delta t^2}{2}$ for $\mathbf{p}$. These effects can be incorporated into the initial model equation (eq: 4.16) by defining $\mathbf{w}$ as follows.

$$\mathbf{w}_k = \begin{bmatrix} \sigma_x \frac{\Delta t^2}{2} \\ \sigma_y \frac{\Delta t^2}{2} \\ \sigma_x \Delta t \\ \sigma_y \Delta t \end{bmatrix} \tag{4.18}$$

The covariance of the process noise is then $\mathbf{Q} = \mathbb{E}(\mathbf{w}_k \mathbf{w}_k^T)$, showed in equation 4.19. The final form (of equation 4.19) has been further simplified by assuming no covariance between the acceleration in the two directions ($\sigma_x \sigma_y = 0$) and that the magnitude of both standard deviations are equal ($\sigma_x = \sigma_y = \sigma_a$).

The standard deviation of the acceleration, σ_a, acts as a tunable parameter for algorithm 3.

$$
\begin{aligned}
Q &= \begin{bmatrix}
\sigma_x^2 \frac{\Delta t^4}{4} & \sigma_x \sigma_y \frac{\Delta t^4}{4} & \sigma_x^2 \frac{\Delta t^3}{2} & \sigma_x \sigma_y \frac{\Delta t^3}{2} \\
\sigma_x \sigma_y \frac{\Delta t^4}{4} & \sigma_y^2 \frac{\Delta t^4}{4} & \sigma_x \sigma_y \frac{\Delta t^3}{2} & \sigma_y^2 \frac{\Delta t^3}{2} \\
\sigma_x^2 \frac{\Delta t^3}{2} & \sigma_x \sigma_y \frac{\Delta t^3}{2} & \sigma_x^2 \Delta_t^2 & \sigma_x \sigma_y \Delta_t^2 \\
\sigma_x \sigma_y \frac{\Delta t^3}{2} & \sigma_y^2 \frac{\Delta t^3}{2} & \sigma_x \sigma_y \Delta_t^2 & \sigma_y^2 \Delta_t^2
\end{bmatrix} \\
&= \begin{bmatrix}
\frac{\Delta t^4}{4} & 0 & \frac{\Delta t^3}{2} & 0 \\
0 & \frac{\Delta t^4}{4} & 0 & \frac{\Delta t^3}{2} \\
\frac{\Delta t^3}{2} & 0 & \Delta_t^2 & 0 \\
0 & \frac{\Delta t^3}{2} & 0 & \Delta_t^2
\end{bmatrix} \sigma_a^2
\end{aligned}
\tag{4.19}
$$

For the measurement covariance $\mathbf{R}$, similar simplifications can be made as for $\mathbf{Q}$. Assuming no covariance between the measurement errors in the x and y directions, and equal uncertainty in both directions (σ_m), the measurement covariance matrix takes the following form (equation 4.20).

$$
\mathbf{R} = \begin{bmatrix} \sigma_m^2 & 0 \\ 0 & \sigma_m^2 \end{bmatrix} \tag{4.20}
$$

Some of the noise modelling simplifications might be unreasonable. In particular, the model noise covariance ($\mathbf{Q}$) in the x and y directions are different, as the direction of fish velocity is mainly in the x direction. However, due to the already large simplifications made when modelling the salmon movement, this should not cause too much performance drop of the Kalman filter. Furthermore, keeping only one parameter greatly simplifies the filter tuning.

4.5 Homogenous transformation

A homography is a mapping between two planes in $\mathbb{R}^3$ defined by a 3x3 matrix $\mathbf{H}$ ([209], s. 33). Let any salmon in the tank be simplified to an oriented plane segment, and coordinates in a local frame attached to this bivector (plane segment) be given by a 2D vector $\mathbf{x} = [x, y]^T$ ($\mathbf{x}$ is describing displacement in the 2D plane segment). Furthermore, let the image frame coordinates be defined as $\mathbf{u} = [u, v]^T$ and the homogeneous transformation matrix as $\mathbf{H}$. Then the relationship between the two planes can be written as $c\tilde{\mathbf{u}} = \mathbf{H}\tilde{\mathbf{x}}$, with $\tilde{\mathbf{x}} = [\mathbf{x}^T 1]^T$ and $\tilde{\mathbf{u}} = [\mathbf{u}^T 1]^T$. As seen from the scalar c, the homogeneous transform is only unique up to scale[210], giving it eight degrees of freedom.

If four point correspondences between the two planes are known, the problem of estimating $\mathbf{H}$ is straight forward, by an approach known as the Direct Linear Transform (DLT). To explain this method, first notice how $\tilde{\mathbf{u}} = \mathbf{H}\tilde{\mathbf{x}}$ can be written as shown in equation 4.21 after some algebraic manipulations[210][3].

[3]In this article the last term in the second equation is multiplied by u. This is likely a typo, as subsequent matrix forms in that article, as well as my own derivation, suggests a multiplication by v instead.

$$-h_{11}x - h_{12}y - h_{13} + (h_{31}x + h_{32}y + h_{33})u = 0 \tag{4.21a}$$
$$-h_{21}x - h_{22}y - h_{23} + (h_{31}x + h_{32}y + h_{33})v = 0 \tag{4.21b}$$

This form is linear in the matrix ($\mathbf{H}$) elements, and can be written as $\mathbf{Ah} = \mathbf{0}$ if $\mathbf{A}$ and $\mathbf{h}$ are defined as below.

$$\mathbf{A} = \begin{bmatrix} -x & -y & -1 & 0 & 0 & 0 & ux & uy & u \\ 0 & 0 & 0 & -x & -y & -1 & vx & vy & v \end{bmatrix} \tag{4.22a}$$
$$\mathbf{h} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{21} & h_{22} & h_{23} & h_{31} & h_{32} & h_{33} \end{bmatrix}^T \tag{4.22b}$$

By stacking the $\mathbf{A}$ matrices of four points vertically, a linear matrix equation with eight rows result. This allows the inference of the eight degrees of freedom in the homography, and hence a determination of $\mathbf{H}$.

It is important to note that if three of the points in one plane segment are collinear, while the corresponding points in the other plane segment are noncollinear, no transformation can be found, as a projective transformation (synonym to homogeneous transformation) preserves colinearity ([209], s. 91).

In the case of more than four points, the redundancy can be handled by introducing a cost function that is minimized by an appropriate transformation matrix $\mathbf{H}$. By assuming the true correspondences have zero error, and that the measured point correspondences are Gaussian distributed around these true matches, the maximum likelihood estimate of the transformation matrix is given by the minimizer of the reprojection error ([209], s. 103). This error function is shown below, with N representing the number of points.

$$\sum_{i}^{N}(u_i - \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}})^2 + (v_i - \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}})^2 \tag{4.23}$$

Other types of transformations could be applied to the salmon plane segments, such as affine transformations. These kind of transforms can compensate two dimensional rotation, translation, shear, and scale, however it is not able to correct for perspective, like the homogeneous transformation.

4.6 Pixelwise image transformations

Some classical computer vision methods are used in the pipeline, they are briefly covered below. The theory is based on the OpenCV documentation[211], and can also be found in classical computer vision books such as [118] and [212].

4.6.1 Grayscale

To convert an image to grayscale, a pixelwise weighted sum of the color channels is performed. In openCV, the following calculus is used (equation 4.24).

$$gray = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \tag{4.24}$$

4.6.2 Adaptive threshold

Thresholding transforms an image to a binary grid, where each pixel is either on (1) or off (0). As the interesting image information when capturing salmon pose and dot structure is not color dependent, the thresholding can be performed on an image with a single channel, such as a gray scale one. Let A be the set of pixels where some requirement is met, then the threshold image $\mathbf{1}_A(x, y)$, mathematically an indicator function, is presented below.

$$\mathbf{1}_A(x, y) = \begin{cases} 1 \text{ if gray(x, y)} > \text{T(x, y)} \\ 0 \text{ else} \end{cases} \tag{4.25}$$

T can be defined a couple of different ways. OpenCV has two main formats, differing from each other by whether an identity[4] or Gaussian kernel is applied to the receptive field of the thresholding function. Let g() describe a Gaussian function with mean at (x,y), N the field of view of the threshold function, and C an offset constant. Then, the two thresholding functions are defined as shown below (equation 4.26a describes the linear kernel thresholding function, while equation 4.26b describes the Gaussian kernel thresholding function).

$$T(x, y) = \frac{1}{N \times N} \sum_{i=-\frac{N}{2}}^{\frac{N}{2}} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}} gray(x + i, y + j) - C \tag{4.26a}$$

$$T(x, y) = \frac{1}{N \times N} \sum_{i=-\frac{N}{2}}^{\frac{N}{2}} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}} gray(x + i, y + j) \times g(x + i, y + j) - C \tag{4.26b}$$

4.6.3 Laplacian

The Laplacian of an image is a way of describing how much the color intensity changes around pixels. It is based on the Sobel operator[213], referred to by $\tilde{\partial}$ in this document, which is a filter that approximates the image intensity gradient when convolved with the image. The sobel operator is isotropic, so to elucidate the full intensity change at a pixel, the gradient along each axis must be found independently, and then composed afterwards. Since convolution is associative,

[4]Identity kernel is referring to treating every pixel value in the receptive field with equal weight.

higher order derivatives can be computed by first convolving gradient filters with itself, and then convolving the larger filter with the image.

Let the notation for the Sobel operator be similar to that of partial derivatives, and *gray* refer to a grey scale image. Then the Laplacian of an image is shown in equation 4.27.

$$lap = \frac{\tilde{\partial}^2 gray}{\tilde{\partial} x^2} + \frac{\tilde{\partial}^2 gray}{\tilde{\partial} y^2} \tag{4.27}$$

4.7 Probability

Assume a time series of length n has a predicted class for each sample. The class with most predictions is called $bestClass$, and this class has $bestClassNum$ number of predictions. An important question that poses itself is:

> P_{prob}: *What is the probability that* $bestClass$ *would have more than* $bestClassNum$ *predictions under a random classifier.*

If this probability is large, the certainty of the estimated class over the entire time series is low, and the whole data point set (time series) should be discarded.

To answer P_{prob}, the first step is to model the random classifier. Let the number of classes in a tank be N, and assume each class prediction is an independent experiment with constant probability $p = \frac{1}{N}$ of estimating $bestClass$, and $q = \frac{N-1}{N}$ of predicting another class. This is a common setup in probability theory, with each experiment called a Bernoulli trial, and the probability of drawing X instances of $bestClass$ in n number of frames is given by the probability mass function of the binomial distribution[214] (equation 4.28).

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \tag{4.28}$$

By summing up the probability mass function over the range that concerns the initial problem, P_{prob} is answered in equation 4.29.

$$P(X \geq k) = \sum_{i=k}^{n} \binom{n}{i} p^i q^{n-i} = 1 - \sum_{i=0}^{k-1} \binom{n}{i} p^i q^{n-i} \tag{4.29}$$

Equation 4.29 gives a quantitative measure for the classification certainty of a set of salmon identifications, which facilitates an automatic rejection of time series if $P(X \geq bestClassNum | random\, classifier) > p$.

The assumption of independent draws is a major simplification, as frames close together in time share a large resemblance, making consecutive frames likely to share the same misclassifications. A practical solution to this issue is to lower the class number N to two, making all misclassifications the same class. This solution leads to a higher $bestClassNum$ threshold for a time series to be accepted, however if the classifier is reasonable well performing, this requirement is not unreasonable. As an example, if $n = 100$, $P(X \geq 63 | random\, classifier, N = 2) = 0.006$.

4.8 Statistics

4.8.1 t-test

Statistics will play a role in this report when analysing the results of the salmon stress experiment. Apart from calculation of average ($\mu = \sum_{i=1}^{N} \frac{x_i}{N}$), mode (the most common value of a sample set) and standard deviation ($std(x) = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N-1}}$) of samples (see e.g. [214]), the main use of statistic will revolve around t-tests. These t-tests are used to evaluate the probability of two sample sets having different sample means, or the probability of a single sample set having mean above zero. First, consider the problem of deciding whether two samples have different means.

Due to the complex dynamics of fish ventilation, it is not reasonable to assume equal variance between the populations examined in this report, and as such a version of the t-test known as Welch's t-test is employed, which allows for both unequal variance and unequal sample size of the populations[215]. Consider two sample sets containing samples X_i and Y_j from random variables X and Y with distributions $N(\mu_X, \sigma_X)$ and $N(\mu_Y, \sigma_Y)$, with N being the normal distribution. Furthermore, let $\bar{X}$ and $\bar{Y}$ be the sample means, s_X and s_Y be the sample standard deviations, and N_X and N_Y be the sample sizes of the two populations. Then the statistic T, given by equation 4.30[216], can be approximated as Student's t-distributed with degree of freedom ν (see equation 4.31).

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{N_X} + \frac{s_Y^2}{N_Y}}} \tag{4.30}$$

$$\nu = \frac{\frac{s_Y^2}{N_X} + \frac{s_Y^2}{N_Y}}{\frac{1}{N_X - 1}(\frac{s_X^2}{N_X})^2 + \frac{1}{N_Y - 1}(\frac{s_Y^2}{N_Y})^2} \tag{4.31}$$

Considering this, the following steps can be followed to find the probability that the random variables X and Y have different means.

1. Insert the values of the population samples (X_i and Y_j) into the formula for T (equation 4.30) to retrieve a sample t value (t').

2. Use a lookup table to find the critical values of Student's t-distribution with ν degrees of freedom. If the p value is set to 0.05, the critical values are $t_{crit} = ^{+}_{-} t_{0.025,\nu}$.

3. If $|t'| > t_{0.025,\nu}$ (the distribution is symmetric) it is under 5 percent probability that random variable X and Y (of which only a set of samples are known) has the same mean.

When evaluating the probability of a single sample set having mean above zero, the approach is very similar. The T statistic is now given by equation 4.32, with $\nu = N_X - 1$[214]. Furthermore, since the hypothesis to be tested is

whether the sample mean is above zero, a one-tailed t-test can be used, giving a critical value of $t_{crit} = t_{0.05,\nu}$.

$$T = \frac{\bar{X} \cdot \sqrt{N_X}}{s_X} \tag{4.32}$$

4.8.2 Pearson correlation

The Pearson correlation is a quantitiave measure for the linear correlation between two sample sets. This metric is in practice a normalized covariance, as displayed in equation 4.33 (see e.g. [214], where the Pearson correlation is called the *Correlation coefficient*). As this coefficient compare datapoints between the two sets, it requires equal sample size ($N_X = N_Y = N$).

$$r = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2 \sum_{i=1}^{N}(Y_i - \bar{Y})^2}} \tag{4.33}$$

To determine whether the calculated correlation coefficient (r) is significantly different from zero, the approach discussed in the previous subsection (section 4.8.1) can be employed. The T statistic for the Pearson coefficient in the case of normally distributed random variables (see equation 4.34[214]) is Student's t-distributed with degree of freedom $\nu = N - 2$.

$$T = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} \tag{4.34}$$

4.9 Frequency estimation

Frequency estimation of a one dimensional signal is an important part of the final salmon breathing frequency pipeline. Due to this, the section below is dedicated to explaining methods of non-linear function fitting and RANSAC.

4.9.1 Levenberg-Marquardt

The problem of fitting a sinusoid to a set of datapoints (x, y) can be framed as a non-linear least square problem. Let the parameter vector $\mathbf{p}$ (4.35a) determine the function $f(x, p)$ (4.35b), and let $\hat{\mathbf{p}}$ be the optimal $\mathbf{p}$, defined as the parameter vector that minimizes the sum of squared errors (4.35d). An effective and robust algorithm to find $\hat{\mathbf{p}}$ is the Levenberg-Marquardt (LM) method, which effectively combines the advantages of gradient descent and Gauss-Newton[217] by finding the $\mathbf{s}_k$ that solves equation 4.35e, and then updates the current $\mathbf{p}$ estimate in direction of $\mathbf{s}_k$. This algorithm will converge to a local minima, and under some conditions achieve locally quadratic convergence[218].

$$\mathbf{p} = [\ A\ f\ \phi\ O\ a\ ] \tag{4.35a}$$

$$f(x, \mathbf{p}) = A \cdot sin(2\pi \cdot f \cdot x + \phi) + o + a \cdot x \tag{4.35b}$$

$$\mathcal{L}(x, y, \mathbf{p}) = \frac{1}{2}\sum_{i=1}^{N}(f(x_i, p) - y_i)^2 = \frac{1}{2}\sum_{i=1}^{N} r_i^2 \tag{4.35c}$$

$$\hat{\mathbf{p}} = min_p \mathcal{L}(x, y, \mathbf{p}) \tag{4.35d}$$

$$(\mathbf{J}_k^T \mathbf{J}_k + \mu_k \mathbf{I})\mathbf{s}_k = -\mathbf{J}_k^T r_i(\mathbf{p}_k) \tag{4.35e}$$

The method needs an initial parameter vector, and the better this guess is, the larger the possibility that the algorithm converges to the wanted optimum. Below is an overview of possible heuristics to perform this initialization.

1. **Offset (O)**: This describes the vertical bias of the data, and can be initialized to the average of the datapoints. $O_0 = \mu = \sum_{i=1}^{N} \frac{y_i}{N}$.
2. **Amplitude (A)**: The standard deviation of a perfectly sampled sinusoid is $\frac{A}{\sqrt{2}}$[219], hence a reasonable initialization of the amplitude is $A_0 = \sqrt{2} \cdot std(y)$, where $std(y) = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N-1}}$.
3. **Frequency (f)**: This should be initialized to an expected respiration frequency.
4. **Phase**(ϕ): This could be initialized by looking at the numerical gradient, or the absolute gape size, at the beginning of the signal.
5. **Offset inclination** (**a**): This could be initialized by fitting a line to the data, and extract the rate of increase. As it is reasonable to assume a low offset inclination, zero is also a good parameter guess.

4.9.2 RANSAC

The RANSAC algorithm is able to simultaneously do outlier rejection and function fitting. It does so by randomly sampling minimal subsets S of size n from from the full set of datapoints $P \in \mathbb{R}^N$, before calculating a set of inliers I containing points close to a model $M := f(S, M_0)$ fitted on the minimal sample. This model fitting could be performed by e.g. the LM method. The sample extraction and model fitting is repeated k_{max} times, and the model with the largest set of inliers is returned as the best model fit when the algorithm finishes. Close is in the RANSAC setting defined as any point with loss function $\mathcal{L}(M|_x$, y) smaller than a threshold T. Setting a fixed threshold T is problematic if the scale of the data is changing between datapoint sets (P), however this issue can be tackled by normalization methods such as multiplying T by the standard deviation between the full datapoint set (P) and a straight line.

The random sampling is done k_{max} times, where k_{max} can be found by the relationship shown in equation 4.36, giving the minimum number of RANSAC

iterations required to find one outlier free set of size n with confidence η_0 from a dataset with ϵ fraction of inliers[152]. Since we do not know the inlier ratio *a priori*, k_{max} can be dynamically updated every time a new maximum inlier set is found. Composing all the ideas above, the full RANSAC method is summarized in algorithm 4[152]. Upon completion, this algorithm returns the best inlier set (I^*) and model parameter set (M^*) it has encountered during its execution, which most likely are close to the optimal sets.

Algorithm 4 RANSAC

Require: $P, N > n, M_0, \eta_0, T$
  $I_{max}, k \leftarrow 0$
  $k_{max} \leftarrow g(\frac{1}{N}, \eta_0, n)$
  while $k < k_{max}$ **do**
    $S_k \leftarrow \{A | dim(S_k) = n, A \in P\}$
    $M_k \leftarrow f(S_k, M_0)$
    $I_k \leftarrow \{(x, y) | \mathcal{L}(M|_x, y) < T, (x, y) \in P\}$
    if $dim(I_k) > I_{max}$ **then**
      $M^*, I^* \leftarrow M_k, I_k$
      $I_{max} \leftarrow dim(I_k)$
      $k_{max} \leftarrow g(\frac{I_{max}}{N}, \eta_0, n)$
    end if
    $k \leftarrow k + 1$
  end while

$$g(\epsilon, \eta_0, n) = \frac{log(1 - \eta_0)}{log(1 - \epsilon^n)} \tag{4.36}$$

Chapter 5

Experiment

5.1 Chapter introduction

This chapter describes the salmon stress experiment performed in the autumn of 2022 at NIVA Research Facility Solbergstrand. The data from this experiment was later used to develop and test the salmon breathing frequency pipeline discussed in this report.

5.2 Experiment

Nine tanks, each with seven fish, were subjected to changing oxygen and temperature millieux (see figure 1.1b). The experiment lasted for fifteen days, 22.09.22 to 06.10.22, and contained four phases.

1. **Phase 1**: This phase lasted six days, and were used to raise the temperature from twelve degrees to fifteen degrees in tanks one to three, and from twelve degrees to eighteen degrees in tanks seven to nine. Tanks four to six were kept at twelve degrees.

2. **Phase 2**: This phase lasted two days, and contained a downbreathing of the tanks. During downbreathings, the oxygen supply to the tanks was restricted, causing a gradual decline in DO content. This was enabled by removal of air stones and water flow to the tanks. When the DO levels reached a lower limit (see table 5.1), oxygen was let back into the tanks, allowing the fish to upbreathe the water to previous DO levels. Tanks four to six were downbreathed the second day, the rest of the tanks were downbreathed the first day.

3. **Phase 3**: This phase lasted for five days, in which the tanks held constant temperature.

4. **Phase 4**: The last phase lasted two days, and was a new downbreathing similar to the first one.

Tank	Date	Temperature ($\pm 0.1°C$)	Min DO ($\pm 5\%$)	Max DO ($\pm 5\%$)	DB time
1	28.09	14.0 $°C$	37%	90%	300 min
1	05.10	15.6 $°C$	40%	94%	333 min
2	28.09	14.0 $°C$	40%	91%	301 min
2	05.10	15.6 $°C$	44%	91%	344 min
3	28.09	14.0 $°C$	18%[1]	76%	311 min
3	05.10	15.6 $°C$	38%	87%	319 min
4	29.09	12.3 $°C$	46%	91%	363 min
4	06.10	13.4 $°C$	39%	89%	365 min
5	29.09	12.3 $°C$	38%	90%	371 min
5	06.10	13.4 $°C$	39%	89%	334 min
6	29.09	12.3 $°C$	34%	89%	335 min
6	06.10	13.5 $°C$	39%	90%	329 min
7	28.09	15.9 $°C$	45%	90%	218 min
7	05.10	17.7 $°C$	49%	89%	215 min
8	28.09	15.9 $°C$	43%	88%	204 min
8	05.10	17.8 $°C$	47%	88%	231 min
9	28.09	15.9 $°C$	45%	94%	213 min
9	05.10	17.7 $°C$	49%	94%	213 min

Table 5.1: Overview over downbreathing data from the salmon stress experiment

In table 5.1, an overview of temperature, DO content and length of oxygen cutoff (DB time) during the downbreathings are displayed. The warm tanks had less oxygen reduction as higher temperature reduce hypoxia tolerance in salmon[85]. As can be seen from the table, the ideal temperatures described in the phase list above were not reached precisely, and the minimum DO content in the tanks holding equal temperature was not exactly the same. The temperature was, however, equal for equal days at equal temperature steps (cold, medium and hot). Under the column DB time, the length between the start of the oxygen decline and the time of minimum DO content is presented.

In the phases with constant oxygen supply, salmon were filmed for 1 hour two times each day, centered around the feedings at 10:00 and 14:00 o'clock. Spotlights were placed above the tanks to ensure sufficient light, and GoPro cameras were lowered down to recording positions on a stick. This position was horizontal and 21 cm from the bottom of the cage, alternating between two corners of the tank. Every fourth recording was done close to the surface and with an angle, facilitating behaviour analysis. During downbreathings, the fish were filmed constantly from two bottom corners, with minor breaks during battery and memory card switching. This switching was performed approximately once every hour.

Due to camera restrictions, each video clip had a length of 42420 frames

[1] Likely erroneous oxygen measurement

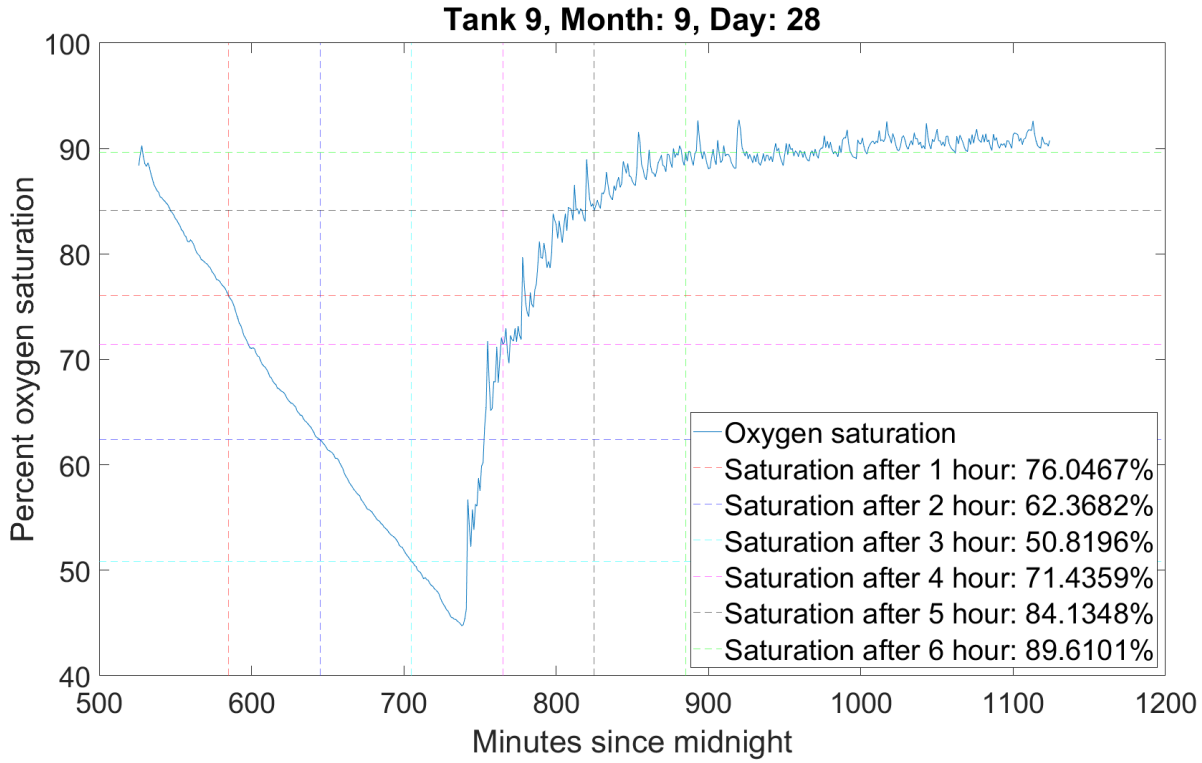
(707 seconds)[2], unless being stopped prematurely by e.g. battery change. These video clips were organized into *hours* by grouping all clips produced between a battery change into the same *hour*. The exact length of this *hour* varied according to the exact time of battery change, but usually one *hour* contained five to seven clips. As such, the temporal length of a one *hour* video clip group was usually a bit longer than an actual hour.

In figures 5.1 and 5.2, representative oxygen trajectories during both downbreathings in two tanks are displayed.

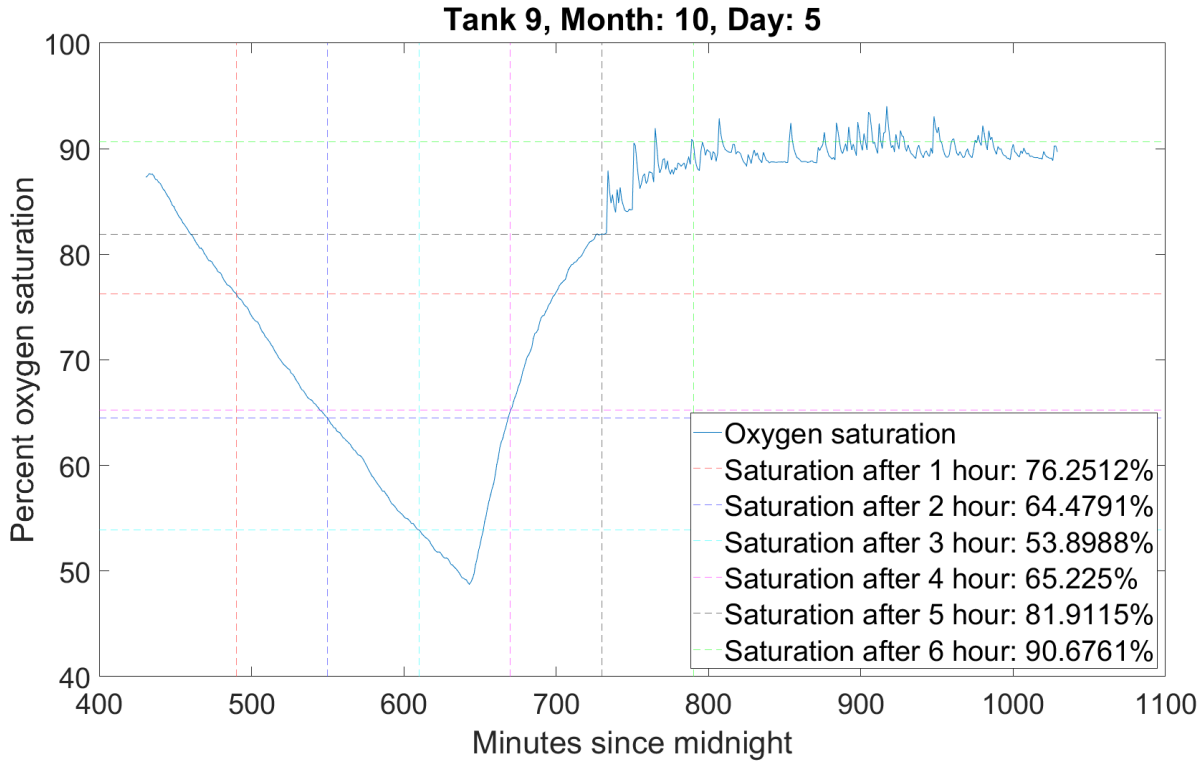
Dissolved oxygen and temperature were measured continuously in all tanks with the $miniDO_2T\,Logger$[220]. During the downbreathings, oxygen and temperature were measured each hour with a handheld device additionally[3].

[2]In a couple of downbreathings, the settings were changed (by accident), resulting in a clip length of 531 seconds instead.

[3]For this project, the data extracted from the handheld device was not used.

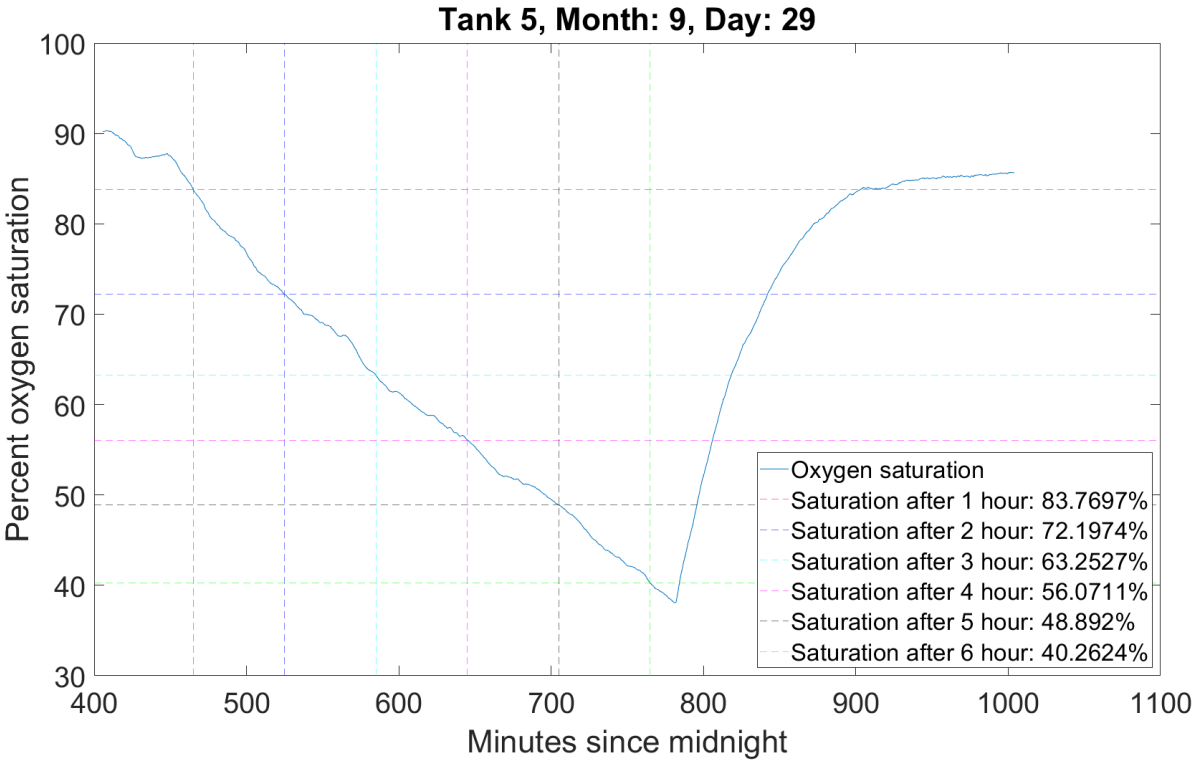


(a) Oxygen evolution in tank nine at 28.09

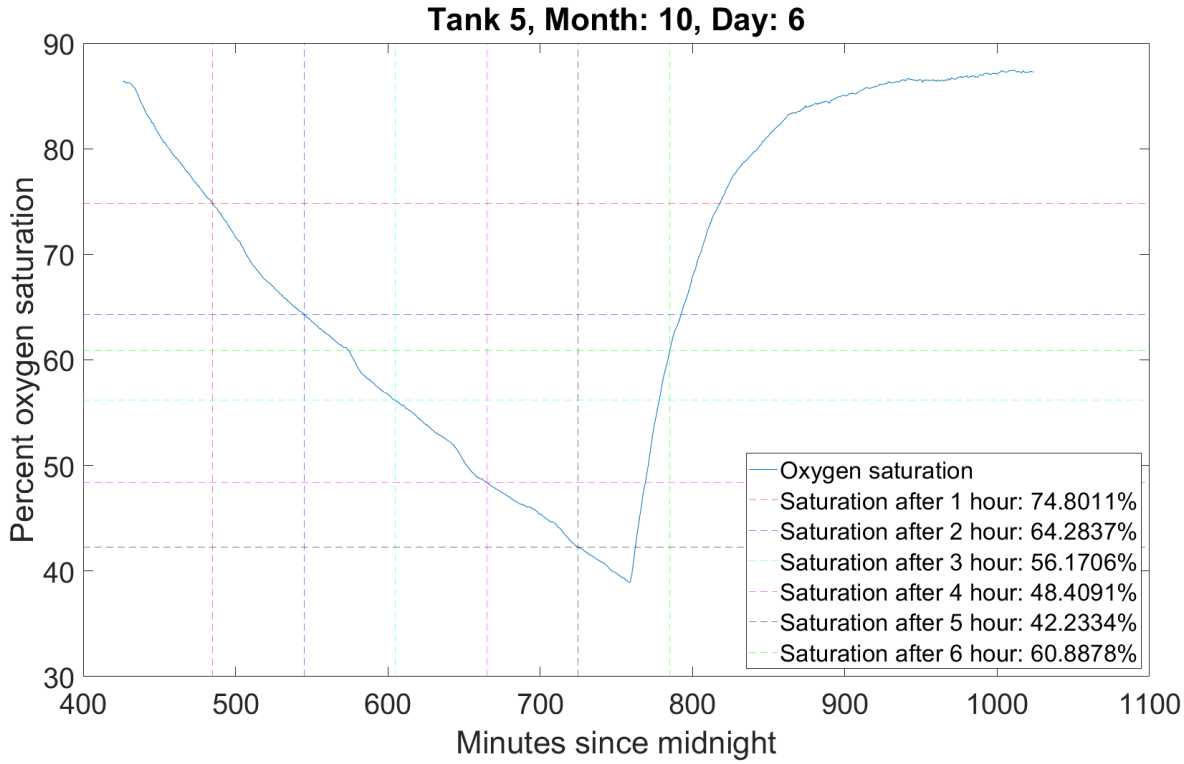


(b) Oxygen evolution in tank nine at 05.10

Figure 5.1: Oxygen evolution of tank nine during two downbreathings



(a) Oxygen evolution in tank five at 29.09



(b) Oxygen evolution in tank five at 06.10

Figure 5.2: Oxygen evolution of tank five during two downbreathings

Chapter 6

Method

6.1 Chapter introduction

In this chapter, the complete pipeline for extracting breathing frequencies from underwater salmon recordings is presented. To ensure a firm grasp on the overarching structure of the method, this section will be dedicated to a high level description of the developed algorithm.

In figure 6.1, the entire salmon breathing frequency extraction method is displayed. The basis of the pipeline is a Keypoint RCNN model that detects salmon with keypoints for all frames in a small video sequence. From there, the method splits into two branches. The first one tracks salmon individuals between frames, and the second classify each salmon detection into one of the seven fish in the tank. From the trackers of branch one, a set of one dimensional salmon jaw pose signals is extracted and fitted to sinusoids, before these time series are associated to a salmon class as predicted by branch two. Some frequency central tendency is found for each salmon class over the small video sequence, which is considered the salmon frequency at a given time. This salmon frequency is finally evaluated with a certain time interval, e.g. 10 minutes, informing on fish ventilation frequency evolution for all salmon individuals over a longer time period.

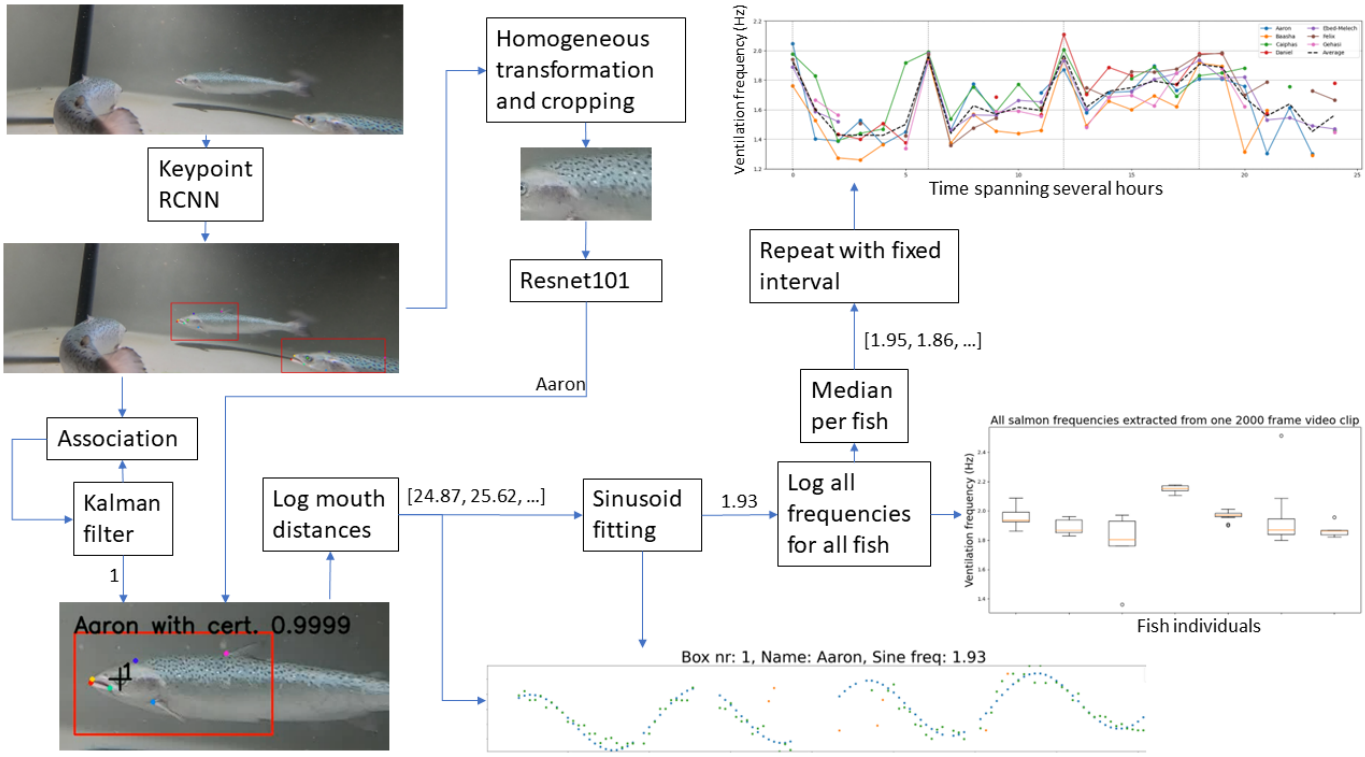


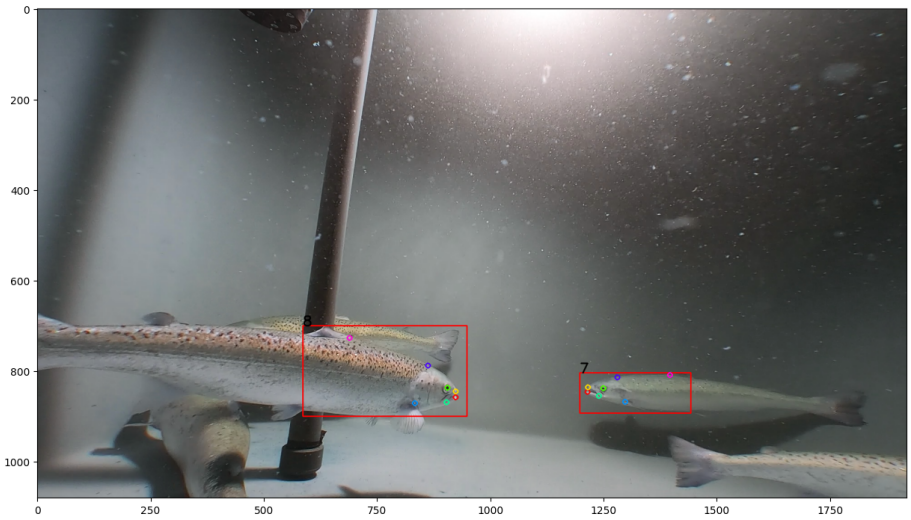
Figure 6.1: Overview of the pipeline performing breathing frequency extraction of all individual salmon in a tank from an underwater video.

6.2 Keypoint RCNN

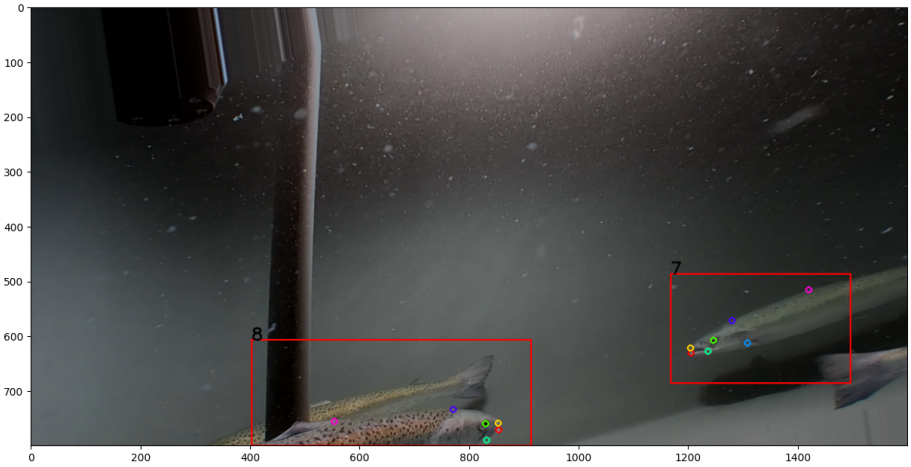
A 10 minute video from the second clip of the second hour into downbreathing one in tank nine was split into frames, and each tenth frame was saved to disk. The first 200 of these frames were annotated with the items presented below. When annotating keypoints for fish partly outside the frame, points were drawn at the frame edge closest to where the points would be. The software used for the annotation was labelme[221], and the keypoint positions are visualized in figures 6.2a and 4.2.

1. **Bounding box**: Placed such that every part of the fish anterior to the posterior root of the dorsal fin were inside the box. In addition, a margin equal to half of the length between the premaxilla[222] and the point where the maxilla attaches to the angular bone[222] were added at the snout of the fish.
2. **Keypoint eye**: Placed in the middle of the eye.
3. **Keypoint lower jaw**: Placed at the kype (dentary prominence).
4. **Keypoint upper jaw**: Placed at tip of the premaxilla.
5. **Keypoint root jaw**: Placed at the intersection betwen the maxilla and the mandible, where the gap between the two bones disappears.
6. **Keypoint Head-Body Intercept (HBI)**: Placed at the anterior extreme of the nape([223], s. 18). This is at the intersection between the scaly salmon body and the smooth, scaleless salmon head.
7. **Keypoint pectoral root**: Placed at the point of the pectoral fin attachement most proximal to the eye of the fish.
8. **Keypoint dorsal fin**: Placed at the anterior root of the dorsal fin.
9. **Class**: Two classes were defined for each fish, one for swimming left, and one for swimming right. In addition, an *occluded* class was defined for when both jaw keypoints were visible, but one or more other keypoints were occluded, and a *front* class for when the pattern on the fish operculum were invisible, or the angle that the eye made with the HBI and the pectoral root exceeded 170 degrees. Finally, adding 0 as a *background* class gave a total of 17 classes. In figure 6.3, all seven fish in tank nine are displayed, with both numeric and letter classes. The postfixes *sl* and *sr* describes the fish swimming in the left or right direction, respectively.

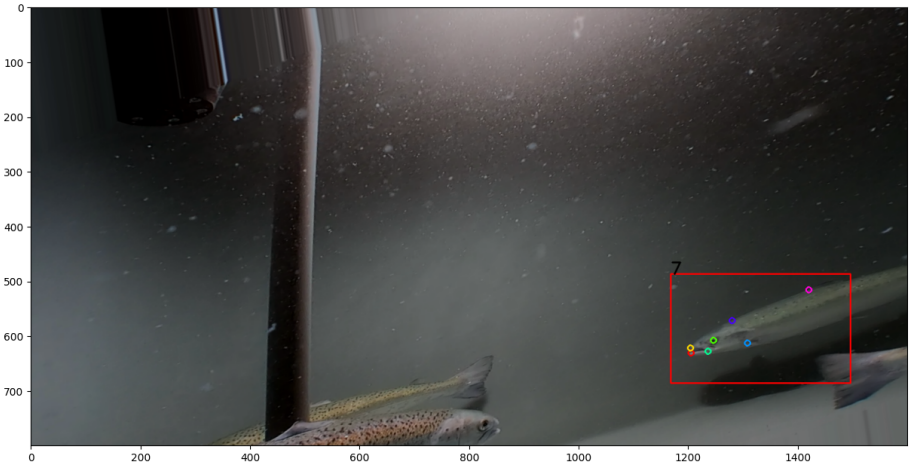
To increase the size of the dataset, plausible transformations were applied to each of the annotated images according to algorithm 5. The algorithm uses the following variables and functions.



(a) Manually annotated frame (1080x1920 pixels)



(b) Augmented frame before label pruning (800x1600 pixels)



(c) Augmented frame after label pruning (800x1600 pixels)

Figure 6.2: Annotation pipeline for Keypoint RCNN

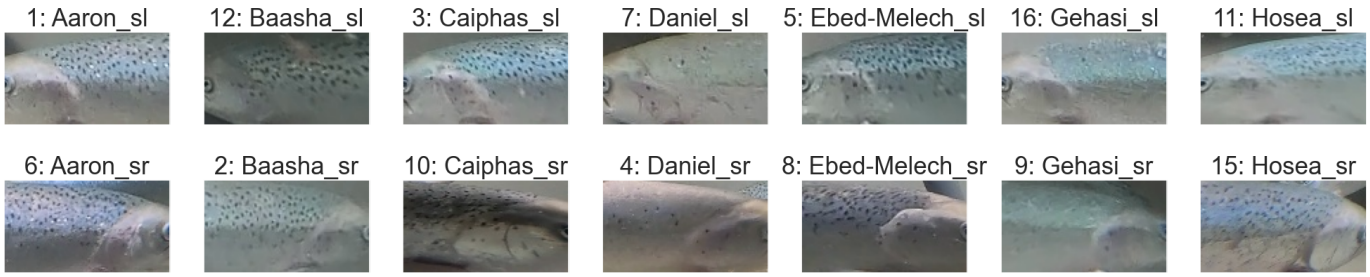


Figure 6.3: All fish in tank nine with class labels

1. V is an ordered set containing annotated image frames, such that $V = \{v|v = (I, L)\}$ where I is an image tensor, L is a set of salmon labels and v is an annotated frame.

2. $inflation\_ratio$ is the number of augmented images per manually annotated frame.

3. $apply\_augmentation()$ is a function performing rotation, four-point random perspective transformation, scaling, shearing, random cropping, brightness adjustment and contrast adjustment in a random fashion by the use of the Python Albumentations[224] library. The four first transformations are all subsets of the projective transformation, described in section 4.5.

4. $remove\_incomplete\_labels()$ is a function that removes all labels with one or more keypoints placed outside of the frame after all transformations.

By setting the $inflation\_ratio$ to 30, a dataset of 6000 images were obtained (the S set of algorithm 5). In figure 6.2, three different intermediate outputs in the augmentation pipeline are visualized.

Algorithm 5 Augmentation pipeline for Keypoint RCNN data

Require: V, $inflation\_ratio$
  $S \leftarrow \emptyset$
  for all $(I, L) \in V$ **do**
    k = 0
    while $k < inflation\_ratio$ **do**
      $(I, L) \leftarrow apply\_augmentation(I, L)$
      $L \leftarrow remove\_incomplete\_labels(L)$
      if $L \neq \emptyset$ **then**
        $k \leftarrow k + 1$
        $S \leftarrow \{(I, L)\} \cup S$
      end if
    end while
  end for

The Keypoint RCNN implementation from PyTorch[198], constructed as explained in the theory chapter, was used for salmon detection and keypoint prediction. The backbone was pretrained on imagenet[225], while the keypoint and bounding box heads were randomly initialized. All parameters in the model were adjustable, and the model was trained for 13 epochs on 2 T4 GPUs. The initial learning rate was set to 0.005, and divided by 10 every 3 epochs. Furthermore, the momentum was set to 0.9 and the regularization to 0.0005. Total training time was between eight and nine hours.

At the output of the model, all bounding boxes were pruned by the PyTorch NMS method when making predictions, with the overlap threshold set to 0.1.

6.3 Tracking

For tracking salmon instances, a Kalman filter based method was employed. Consider an indexed set of trackers T with $|T| = M$, where each tracker t in the set contain information pertaining to a single fish instance. For the high level description of the algorithm, it is sufficient to consider t as a Kalman filter, with domain $\mathbb{R}^2$ describing the spatial location of the filter. Let the matrix $\mathbf{eye} \in \mathbb{R}^{N\times 2}$ contain the N eye detections of the Keypoint RCNN model in a frame and the cost matrix $\mathbf{c} \in \mathbb{R}^{M\times N}$ contain the euclidean distances between all tracker positions and eye detections. Furthermore, let W and H be the width and height of a frame, T_{log} be the set of kalman filters which are no longer active, and S be the number of matches between trackers and detections in a given frame. Before presenting the algorithm to track the fish, a number of functions are defined below, with the non-trivial ones expanded upon in the theory chapter (chapter 4). The Kalman filter implementation was a modified version of [226].

1. $predict() : t \to t$ propagates the Kalman filter one timestep by updating the pre measurement model estimate (see section 4.4).

2. $eyeDetect() : \mathbb{R}^{H\times W\times C} \to \mathbb{R}^{N\times 2}$ finds salmon eyes in a frame by applying the Keypoint RCNN model.

3. $position() : T \to \mathbb{R}^{M\times 2}$ returns the position of the Kalman filters.

4. $dist() : \mathbb{R}^{N\times 2} \times \mathbb{R}^{M\times 2} \to \mathbb{R}^{MxN}$ finds the cost between detections and trackers, and store the results in a cost matrix $\mathbf{c}$. In this implementation, euclidean distance was used as cost.

5. $JonkerVolgenant() : \mathbb{R}^{M\times N} \to \mathbb{R}^{2\times S}$ finds the best matches between the trackers and the eye detections, given a cost matrix. The first row of the function output contain eye detection indices, and the second row contain tracker indices. Included in this function is a distance threshold ($max\_tracker\_detection\_dist$), that discards the match if the tracker and the eye detection is sufficiently far away from each other.

6. $update() : t \times \mathbb{R}^2 \rightarrow t$ performs the measurement update step in the Kalman filter (see section 4.4).

7. $incrUnmatched() : t \rightarrow t$ increments a flag in the Kalman filter, which enables the filter to remember the time span it has not received a measurement.

8. $unmatched() : t \rightarrow \mathbb{R}^1$ returns the amount of time steps which the tracker has not received a measurement.

9. $newTracker() : \mathbb{R}^2 \rightarrow t$ initialize a new kalman filter with position at the argument, and zero velocity.

The salmon tracking method is presented in its entirety in algorithm 6. In the actual implementation, the trackers contain more information than the Kalman filters, such as mouth pose descriptors, however they do not contribute to the logic of the tracking and are hence omitted here for clarity. Subscript : is indexing a full matrix axis, and k is a tunable parameter defining the number of frames a tracker is allowed to have no matches.

When algorithm 6 finish, the two tracker sets T and T_{log} contains objects (t) that hold information pertaining to a fish individual for the time span that this fish occupy the camera field of view. Included in these objects are information about salmon class (discussed in subsequent sections 6.4 and 6.5), and salmon mouth pose (which is given directly from the keypoint placement on the salmon detections) for each frame.

The parameters were assigned the following values in the final implementation of the salmon tracking algorithm:

1. **Kalman filter model uncertainty (σ_a^2 in equation 4.19)**: $\sigma_a^2 = 0.5$

2. **Kalman filter measurement uncertainty (σ_m^2 in equation 4.20)**: $= \sigma_m^2 = 0.01$

3. **Survival time of tracker without detections (k)**: 50 frames

4. **Maximum distance between tracker and eye detection (**$max\_tracker\_detection\_dist$**)**: 100 pixels

Algorithm 6 Salmon tracking

Require: V, k
  $T \leftarrow \emptyset$
  $T_{log} \leftarrow \emptyset$
  for all $v \in V$ **do**
    $T \leftarrow predict(T)$
    eye $\leftarrow eyeDetect(v)$
    c $\leftarrow dist(\textbf{eye}, position(T))$
    $m \leftarrow JonkerVolgenant(\textbf{c})$
    for all $i \in m_{2,:}$ **do**
      $T_{m_{2,i}} \leftarrow update(T_{m_{2,i}}, \textbf{eye}_{m_{1,i}})$
    end for
    $I_{ut} \leftarrow \{j \in \mathbb{N} | j \leq M, j \notin m_{2,:}\}$
    for all $i \in I_{ut}$ **do**
      $T_i \leftarrow incrUnmatched(T_i)$
      if $unmatched(T_i) > k$ **then**
        $T_{log} \leftarrow T_{log} \cup T_i$
        $T \leftarrow T \setminus \{T_i\}$
      end if
    end for
    $I_{ud} \leftarrow \{j \in \mathbb{N} | j \leq N, j \notin m_{1,:}\}$
    for all $i \in I_{ud}$ **do**
      $T \leftarrow newTracker(\textbf{eye}_i) \cup T$
    end for
  end for

6.4 Semi-automated labelling

The dataset used for training salmon identification networks was constructed by the help of the previously developed detection (section 6.2) and tracking (section 6.3) algorithms.

In this dataset, the input features consisted of salmon images warped onto a fixed configuration, and cropped to display only part of the fish. The image warping was performed by a homogeneous mapping, using the OpenCV Python library[211] and four of the keypoint predictions of the Keypoint RCNN model. As the matching points should be noncollinear for a large range of poses, as well as attached to anatomical landmarks that does not move independent of the fish[1], the points used for the homogeneous transformation were the four keypoints most posterior on the fish (the eye, HBI, dorsal fin and pectoral fin keypoints). After the warping, the fish were cropped to display the area between the dorsal fin and eye keypoints along the anterioposterior axis and the area between the dorsal fin and pectoral root keypoints along the dorsoventral axis. An image of this form will be called the torax of the fish (see figure 6.3). The variability between torax images were almost exclusively due to the dot structure of the fish, which a network can use to identify individual salmon.

Having specified the data input, it is time to describe the method for labelling the output y. The output consists of one integer c for each input torax image, describing which salmon individual the fish belongs to. The semi-automated labelling consisted of four steps.

1. Pass all frames in a video sequence of 500 to 3000 images through the Keypoint RCNN model.

2. For each frame, generate torax images and annotated complete frames by the help of the Keypoint RCNN detections. Save these to disk.

3. Review the full frame images, and create an annotation containing the tracker IDentifier (ID), start frame, end frame and c for all good detection trajectories. A good detection trajectory is a consecutive set of detections of the same salmon, where all keypoints used for the homogeneous transformation of the salmon is placed relatively well.

4. Iterate through all trackers, and use the manual annotations to extract all salmon torax images in the range [start frame, end frame]. These are then linked to the specified c.

This method allows a single manual annotation to generate several hundreds of datapoints, as all torax images between the start and end frame are included in the dataset. After the labelling, all torax images were manually reviewed to check for degenerate images.

Using this method of semi-automatic annotation, datasets of at least 500 images per fish class were created for each tank. It was ensured that each fish

[1] Points such as the lower jaw moves independent of the rest of the salmon.

	T1	T2	T3	T4	T5	T6	T7	T8	T9
A_{sl}	633	806	850	595	1566	786	819	650	1065
A_{sr}	744	1256	630	1587	1728	608	673	845	2152
B_{sl}	551	831	887	575	980	863	748	772	2864
B_{sr}	1213	583	525	1320	1207	569	570	1753	1196
C_{sl}	628	577	624	755	1051	669	648	822	2954
C_{sr}	1623	633	536	639	1546	708	1047	610	1070
D_{sl}	764	694	1210	533	1020	606	623	1050	1049
D_{sr}	1305	1079	661	1008	1463	651	1006	1002	1157
E_{sl}	730	583	1006	684	1109	819	690	1044	1542
E_{sr}	658	627	500	1350	1152	734	1243	635	1129
G_{sl}	710	828	576	523	1006	930	573	744	1014
G_{sr}	818	549	935	1271	1224	644	584	1307	1138
H_{sl}	1084	842	775	820	1185	778	1087	1436	1471
H_{sr}	687	874	545	870	1248	537	953	653	1081

Table 6.1: Overview over dataset size. Subscripts in the left column (sr and sl) refers to swimming direction, and the letters correspond to the fish individual.

class contained at least 100 datapoints from each downbreathing. Due to the nature of the fish data, with some fish classes being frequently filmed and some staying mostly out of the camera field of view, there was a large difference in the number of datapoints in the different fish classes. To avoid imbalanced datasets, the size of all fish classes were reduced to the size of the smallest (least number of datapoints) class in each tank by the use of random sampling. The number of datapoints of all fish classes are displayed in table 6.1, with the tanks changing along the horizontal axis, and the fish individual differing along the vertical axis.

For one tank, tank nine, there was also created a small test dataset to evaluate network performance. The test dataset was created from different video clips than the training data, to ensure any results would be descriptive of network performance on unseen trajectories. The test data had 105 images per fish class, or 1470 images in total. Note that the *front* and *occluded* classes were excluded from all torax datasets as their keypoint placement did not allow homogeneous transformations of salmon into the desired torax images, and classification of these classes are of less interest to the salmon ventilation frequency pipeline.

Finally, the datasets of all tanks were inflated thirtyfold by an Albumentations[224] pipeline. The augmentations of the Albumentations pipeline consisted of blurring, randomizing the tone curve, distorting the grid, square dropouts and altering the brightness and contrast. An example of an augmented image is shown in figure 6.4.

For tank nine, an additional modified dataset was created, in which the color images were replaced with three feature engineered channels. These channels were gray scale images (eq: 4.25), adaptive thresholding images with identity kernel (eq: 4.26a) and image Laplacians (eq: 4.27). The dataset with the

Figure 6.4: A non-augmented (top) and an augmented (bottom) torax image

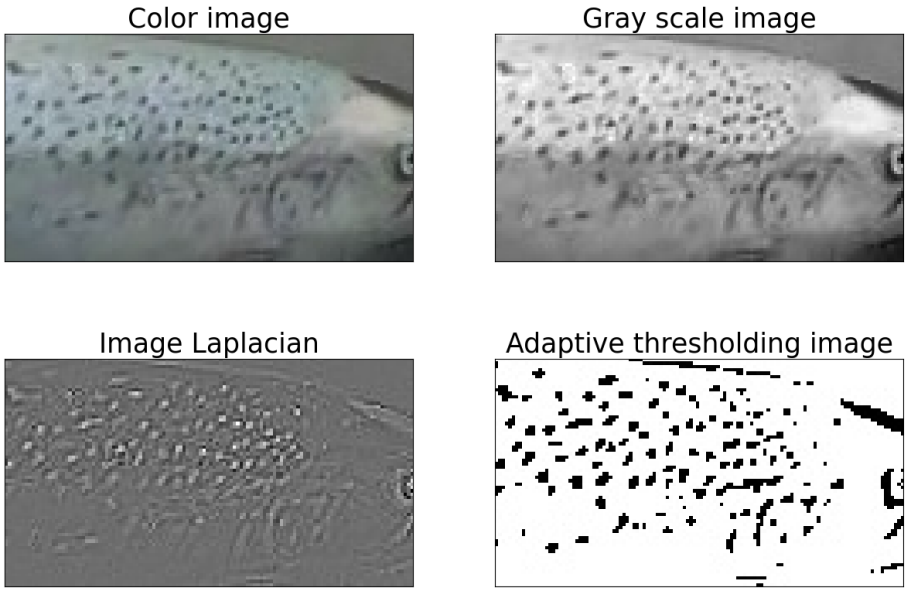


Figure 6.5: Visualization of the three different channels in a dot image, as well as the color image they are created from.

feature engineered channels is called the *dot dataset*, in contrast to the *color dataset* which are composed of Red Green Blue (RGB) images. In figure 6.5, the three different channels in a dot image, as well as the color image they are created from, is visualized.

The tunable parameters in the adaptive thresholding function were tuned as follows.

1. **Offset constant (C)**: 21
2. **Thresholding function field of view (N)**: 9

6.5 Salmon classification

A set of deep neural networks from the Python Pytorch[198] library were trained on the semiautomaticly generated dataset from tank nine. For the color dataset, Alexnet[181], Efficientnet[184], Resnet 50[185] and Resnet 101[185] were used, each trained for two epochs. The dot dataset was trained only on Resnet 50 and Efficientnet, also for two epochs. The two best networks from the color dataset, and one network from the dot dataset, were then trained for a longer time. All models were evaluated on the test dataset from tank nine, using accuracy as optimality metric and a confusion matrix for enhanced understanding of the error distribution. The network performing best on the test dataset of tank nine was saved for later usage in the full salmon breathing frequency pipeline.

This turned out to be the Resnet 101 architecture (see section 7.4 later in the report).

For each new tank analysed, the same Resnet 101 model that was used on the previously analysed tank was fine-tuned on the dataset of the novel tank. This way, the Resnet 101 backbone accumulated information from previous tanks, while the network head, which was randomly initialized for each new tank, only contained knowledge pertaining to the set of fish classes in the current tank. The networks were trained on two T4 GPUs, with a learning rate of 0.005, a weight decay of 0.0005 every fourth epoch, a momentum of 0.9, and a total training time of eight epochs. With regards to resource requirements, the finetuning took between four and five hours per tank.

6.6 Frequency analysis

As discussed in section 6.1, the pipeline initially splits into two branches; salmon classification, and mouth pose tracking. At the current step in the pipeline, these two branches merge to create time series over the mouth pose of the salmon, and assign these time series to a fish class. The process can be described in four steps.

1. The Keypoint RCNN model predicts salmon and corresponding keypoints for each frame.
2. Each fish is assigned to a tracker.
3. The Keypoint RCNN detections are used to classify fish and calculate mouth pose. This information is saved together with the tracker.
4. When all frames are finished, the trackers are used to generate time series, where the time index (x) is the frame number, and the time series value (y) is a tuple containing fish class and mouth pose.

Three different metrics were tried for salmon jaw pose; the euclidean distance between the top and bottom jaw keypoints, the same distance, but after the homogeneous transformation, and the angle the root jaw makes with the upper and lower jaw (see section 3.7.4). The best of these metrics, the non-normalized (raw) euclidean distance, was chosen.

To reduce the number of outliers and erroneous classifications, some requirements were put on the relative keypoint placement. Jaw pose measurements were removed if the upper jaw keypoint was ventral to the lower jaw keypoint, if the lower jaw keypoint with an added margin ($eyeLjawMargin$) was dorsal to the eye keypoint, or if the upper jaw keypoint was further away from the eye keypoint in the vertical direction than in the horizontal direction[2]. Similarly, the classification results were disregarded if the four points used for the

[2]Even though anatomical terminology is used to described relative keypoint placement in this paragraph, the implemented checks consider image axes, not the anteroposterior and dorsoventral axes of the fish. These sets of axes will normally be relatively aligned, and anatomical terms were used to make the mental picture clearer for the reader.

homogeneous transform were placed fallaciously relative to each other. This was checked by ensuring the eye to be the most anterior keypoint, the dorsal fin to be the most dorsal keypoint, and the HBI keypoint to be more dorsal than the keypoint at the root of the pectoral fin. Aditionally, it was ensured that the upper and lower jaw keypoints were placed anterior to the four homogeneous transformation keypoints. The function performing removal of erroneous keypoints, as explained in this paragraph, is called $remove\_err\_keypoints()$.

After this pruning, the available data consisted of a set of time series of varying length. A moving window approach was used to extract time series of fixed length L and stride s from these. By letting TS_f be a full time series, the moving window approach is displayed in algorithm 7. This algorithm was repeated for all the original time series (TS_f) in the trackers.

Algorithm 7 Moving window approach for time series splitting

Require: $L, s, TS_f, \|TS_f\| > L$
  $TS \leftarrow \emptyset$
  $N \leftarrow \|TS_f\|$
  $sf = \{j \in \mathbb{N} | j \leq N - L, j\%s == 0\}$
  for all $i \in sf$ **do**
    $TS \leftarrow TS \cup TS_f[i \cdot s, i \cdot s + L]$
  end for

Each of the L-length jaw gape time series were fitted to a sinusoid by the LM method wrapped inside a RANSAC loop for outlier rejection (see section 4.9). This RANSAC loop was again wrapped in a line search over a small number of frequencies ($freqVal$, used for RANSAC initialization), to avoid frequency bias in the results. Amplitude (A), offset (O) and offset inclination (a) were initialized as described in the theory section, with an added perturbation ($ransacInitPerm$) each model fit (LM run) to avoid local minima. Furthermore, the phase (ϕ) was randomly initialized in the range $[-\pi, \pi]$. Both the RANSAC loop and the frequency line search used number of inliers below a threshold as the optimality criterion, with the thresholds ($ransacThresh/lineSearchThresh$) normalized by the standard deviation between the datapoint sample (time series data) and a straight line fitted to the sample ($std_{sl}(y)$). Additionally, the RANSAC algorithm output was discarded if the model it returned had a frequency standard deviation above $maxFreqStd$ on the LM fit of this model.

After the time series sine wave fitting, the data (time series) was discarded if any of the following were fulfilled. N is describing the length of the time series, and the function checking these requirements is called $acceptable\_time\_series()$.

1. The number of valid classification points of the time series was less than $minValClsf$. This happened if a large amount of Keypoint RCNN detections had poor keypoint configurations.

2. A random classifier had probability above p of beating the results achieved

by the neural network model used to generate the time series classifications.

3. The time span (frame number) between the first and last frame was longer than $maxFrameDiff$.

4. The inlier number of the best model was less than $minInlCnt$.

By putting all the above together, the complete frequency extraction method is displayed in algorithm 8. The notation used in this method is explained in the list below.

1. T is the set of trackers from the video clip V ($T \cup T_{log}$ from algorithm 6).

2. F is a set of tuples such that $F = \{\tilde{f} | \tilde{f} = (c, f)\}$, where c is an integer representing a salmon individual, and f is a respiration frequency. This set will be the output of algorithm 8.

3. G is a set of time series (TS) of fixed length L. Furthermore, $TS_g \in \mathbb{R}^{2\times L}$ and $TS_c \in \mathbb{R}^L$ are extracting the mouth gape time series and a list of classes, respectively, from TS.

4. $freq\_val$ is a list of frequencies used for RANSAC initialization. Each instance in this list is used as the f value of $\mathbf{p}_0$ (see section 4.9.1).

5. **params** are tunable parameters included in the method. For simplicity, they are not stated independently in algorithm 8, however all of them can be found in the list of parameter values later in this section.

6. The $moving\_window()$ function are an implementation of algorithm 7.

7. The $ransac()$ function is an implementation of algorithm 4. To fit the sine wave model with the LM approach (line $M_k \leftarrow f(S_k, M_0)$), the function $curve\_fit$ from the scipy.optimize[139] Python library was employed.

8. $\mathbf{p}$ is describing the parameter list of equation 4.35b. Superscript star refers to some optimal $\mathbf{p}$, and subscripts are extracting a particular parameter (e.g. p_f^* are referring to the optimal frequency parameter). Furthermore, $\mathbf{p}_0$ is an initial parameter vector, and f($\mathbf{p}_0$) is an initial model M_0 (see section 4.9.1).

9. mode() are returning the most common fish class in the time series.

10. The function $acceptable\_time\_series()$ returns True or False depending on whether the requirements stated in the list above are fulfilled.

11. $\mathbf{e}$ is the estimated covariance of the best model fit, extracted from the LM method. $\mathbf{e}_f$ is the estimated variance of the frequency parameter.

12. The $remove\_err\_keypoints()$ function removes keypoints with erroneous configurations, as explained earlier in this section.

Algorithm 8 Salmon frequency estimation from time series

Require: $T, freq\_val,$ **params**
  $F \leftarrow \emptyset$
  for all $t \in T$ **do**
    $t \leftarrow remove\_err\_keypoints(t)$
    $G \leftarrow moving\_window(t)$
    for all $TS \in G$ **do**
      $I_{max} \leftarrow 0$
      for all $f_0 \in freqVal$ **do**
        $\mathbf{p_0} \leftarrow [A_0, f_0, \phi_0, O_0, a_0]$
        $\mathbf{p}, \mathbf{I}, \mathbf{e} = ransac(TS_g, f(\mathbf{p_0}), \eta_0, ransacThresh)$
        if $dim(\mathbf{I}) > I_{max}$ and $e_f < maxFreqStd$ **then**
          $\mathbf{p}^* \leftarrow \mathbf{p}$
          $I_{max} \leftarrow dim(\mathbf{I})$
        end if
      end for
      $c \leftarrow mode(TS_c)$
      if $acceptable\_time\_series(TS)$ **then**:
        $F \leftarrow F \cup \{(c, p_f^*)\}$
      end if
    end for
  end for

After receiving a single frequency and a single class for all L-length time series (the F set of algorithm 8), the data was reduced to one number per fish by taking the median over the frequencies. This way, the entire video sequence V provided a single frequency estimate for each fish individual in the tank.

The parameters were tuned as follows.

1. **Mouth pose metric**: Raw euclidean distance.
2. **Length of each time series (L)**: 100
3. **Length of time series stride (s)**: 20
4. **Minimum number of valid classifications (minValClsf)**: $\frac{N}{3} = 34$
5. **Probability of more than bestClassNum number of predictions under a random classifier (p)**: 1%
6. **Max frequency standard deviation of LM fit (maxFreqStd)**: 0.05
7. **Maximum time span between first and last time series frame (maxFrameDiff)**: 200
8. **Minimum inlier count (minInlCnt)**: $\frac{11 \cdot N}{20} = 55$
9. **RANSAC inlier threshold (ransacThresh)**: $std_{sl}(y)$

10. **Inlier threshold when evaluating best line search model (lineSearchThresh)**: $0.7 \cdot std_{sl}(y)$
11. **Permutation of the RANSAC initial variables (ransacInitPerm)**: 0% for phase, 5% for offset, offset inclination and frequency, 30% for amplitude.
12. **Certainty of picking an outlier free sample in the RANSAC loop (η_0)**: 0.9999
13. **Frequency line search values (freqVal)**: 0.8, 1.2, 1.6, 2 and 2.4
14. **Margin added to the lower jaw keypoint when comparing its horizontal location to that of the eye (eyeLjawMargin)**: 5 pixels

6.7 Respiration frequency extraction pipeline applied to experimental data

The salmon breathing frequency pipeline up until this point has extracted a single frequency for each individual fish from a temporally short recording. In the following, this will be leveraged to extract frequency evolutions for all fish through an entire downbreathing (see chapter 5). This frequency trajectory will be called a *full frequency evolution*.

Data from one downbreathing was provided as a set of 11:47 minute video clips, with a clip count between 23 and 45. These clips were sorted according to which battery shift they succeeded, such that the clips before the first battery change was grouped as hour one (the battery lasted for around one hour). For each of the clips, frame number 5000 to 7000 were extracted, passed through the pipeline developed in the previous sections of this chapter, before the output of the pipeline was stored (as a single frequency per fish individual, calculated as the median of F over each fish class). These measurements were then concatenated to provide a full salmon frequency evolution for all fish. Additionally, the average frequency at each time point (clip) in the downbreathing was calculated, resulting in an average trajectory called $fullTrajAvg$.

If a fish class, after the frequency analysis (see section 6.6), had less than $minFreqCount$ frequency estimates (stored in the F variable of algorithm 8), the fish was regarded as not present in the camera field of view at that time instance (clip).

Each downbreathing was filmed by two cameras, and each tank had two downbreathings, summing to 36 full downbreathing frequency evolutions. These trajectories are solving the initial problem statement P of chapter 3.

From the downbreathing data, three metrics were extracted from each full frequency evolution. These are presented below. When presenting the results of these metrics (see section 7), some statistical analysis were performed by the use of the Python library scipy.stats[139].

1. A shortened trajectory, that provided a single respiration frequency per fish per hour (instead of per clip, as described above). This trajectory is be called *hourly frequency evolution*, and was calculated the following way.

   (a) Discard the results from the first clip each hour.

   (b) Create a single array for each fish each hour, containing all the salmon respiration frequencies (output of algorithm 8) estimated that hour for that fish.

   (c) Take the median of all the arrays, each containing data from one fish over one hour.

   (d) In addition to these seven datapoints (one per fish) per hour, an hourly average trajectory was calculated ($hourlyTrajAvg$). This was accomplished by taking the average over all fish each hour.

2. The frequency increase when the cameras were inserted into the tanks. This insertion happened at the beginning of each hour, so the frequency increase can be calculated by reviewing how the frequency change between this clip, and the two clips on either side (assuming temporally ordered clips). Let clip number t be the first clip in an hour, then the ventilation increase after disturbances are calculated as shown in equation 6.1, where *vent* refers to the average trajectory of the full frequency evolution ($fullTrajAvg$). For all full frequency trajectories, an average ventilation frequency increase, called *respiration frequency increase after disturbance*, was calculated each hour. When calculating the respiration frequency increase the first hour of a downbreathing, only the frequency difference to the subsequent clip was considered, as there is no previous clip.

3. The personality of the fish, i.e. if the fish is trending towards breathing faster or slower than the average shoal respiration frequency. This is calculated by finding the difference between the hourly frequency evolution average ($hourlyTrajAvg$) and the hourly frequency evolution of an individual fish at each hour, and then averaging this over the full downbreathing.

$$vent_{avg} = \frac{(vent_t - vent_{t-1}) + (vent_t - vent_{t+1})}{2} \tag{6.1}$$

In this section, only one parameter were tuned. This was $minFreqCount$, which was set to two in the final salmon breathing frequency extraction pipeline.

Chapter 7

Results

7.1 Chapter introduction

The results of the different segments of the salmon breathing frequency pipeline are presented below. For some methods and algorithms, numerical assessment was deemed less informative than other evaluation techniques, such as visual inspection. To fully appreciate the text, the reader is advised to review an example video[2] displaying all pipeline components up until the time series operations (section 6.6).

7.2 Keypoint RCNN

The vast majority of the salmon detections and keypoint placements were done correctly by the network. When errors did happen, they mostly consisted of the following cases. Frames demonstrating the errors are displayed in appendix A.

1. False negative (the system failed to identify a fish that was actually present).
2. False positive (a fish was detected, but no fish was actually present).
3. Jaws detected on wrong fish.
4. Occluded jaws, which caused either jaw gape errors, or completely wrong keypoint detections.
5. Salmon turning away from the camera.
6. Dorsal fin keypoint sliding along the back of the fish.
7. Two fish detected very close to each other, resuting in the IoU bounding box pruning removing one of the detections.
8. Two fish contained inside the same bounding box, resulting in some keypoints placed on one fish, and some on the other.

When using the Keypoint RCNN model on recordings from other tanks than the one it was trained on (tank nine), the performance was slightly reduced. The types of errors were similar.

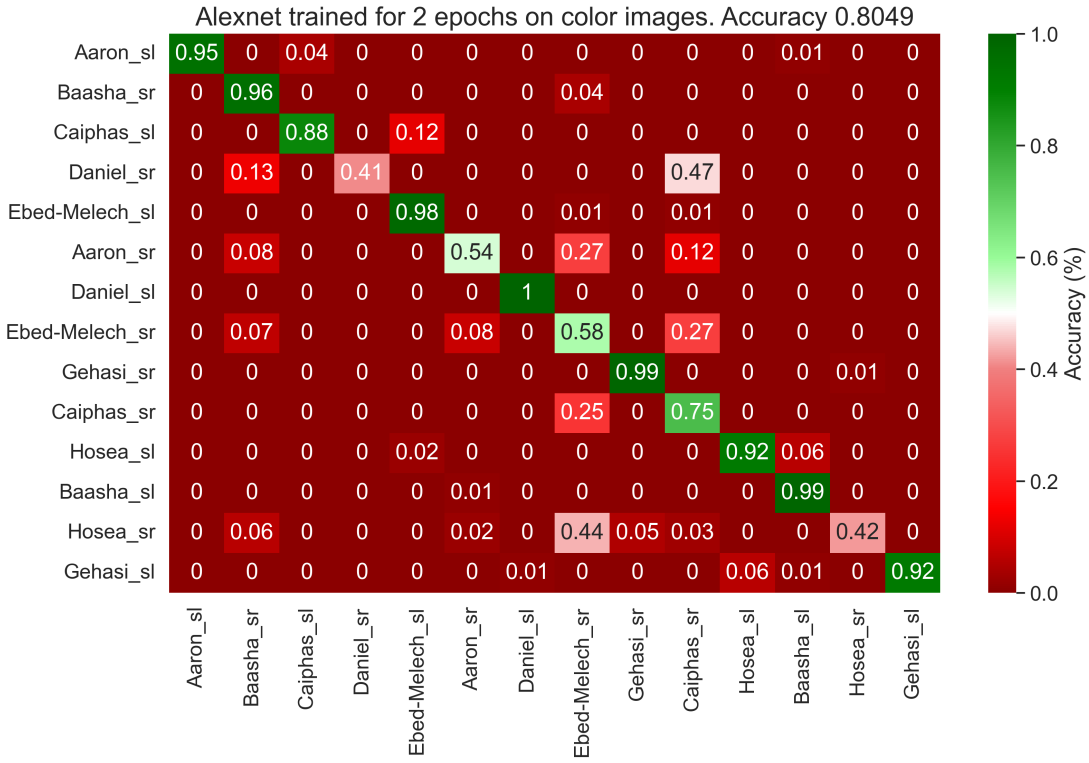
7.3 Tracking

The trackers work well, however some occasionally shift fish or are unable to retrieve a tracked fish after long frame sequences without detections. Examples of these errors are presented in appendix B.

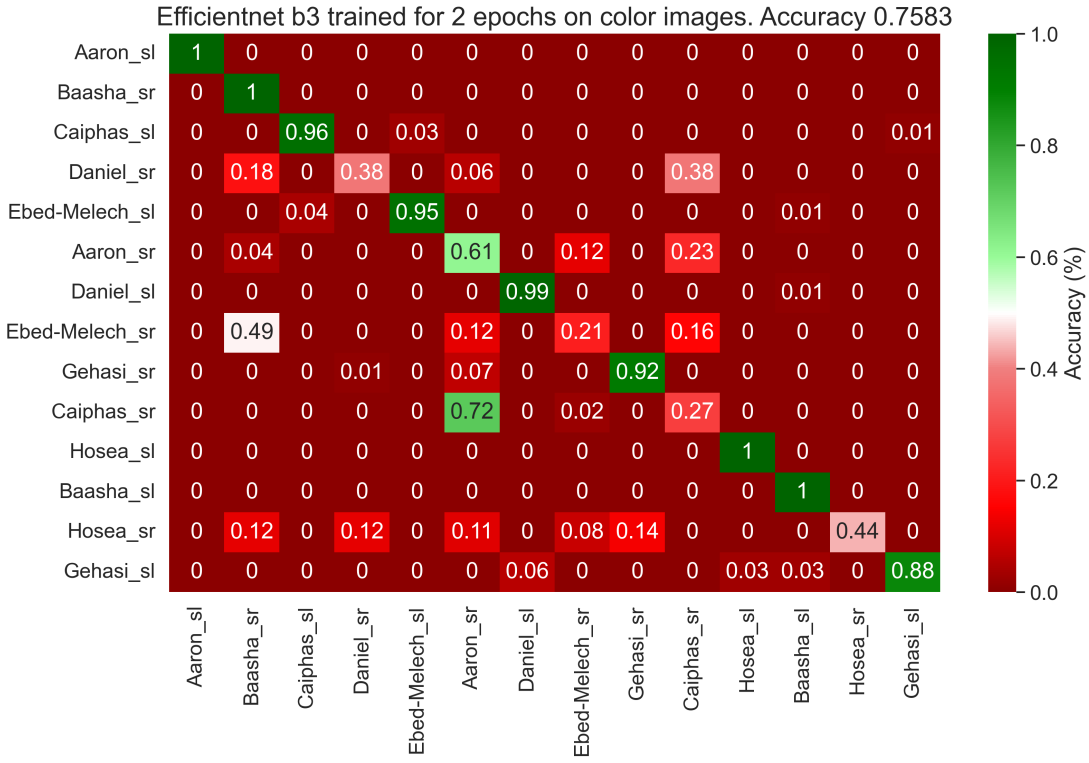
7.4 Salmon identification

In figure 7.1 and 7.2, confusion matrices from different deep neural network models trained for two epochs on the color image dataset are displayed. The Resnet models perform considerably better, beating Alexnet and Efficientnet with over 15 percent points on the test dataset. By training the two best models further, the highest performing network reaches an accuracy of 99.51 percent (figure 7.5). Some examples of the classification errors of this network are presented in figure 7.6.

In figure 7.3 and 7.4, the results after training different neural network models with the dot dataset are displayed. After two epochs of training, the Efficientnet model trained on the dot dataset beats the same network trained on color images by over 10 percent points. For the Resnet50 model, the color dataset beats the dot dataset by around three percent points after two epochs of training. By continuing training the Efficientnet model on dot images, the test performance reduces slightly.

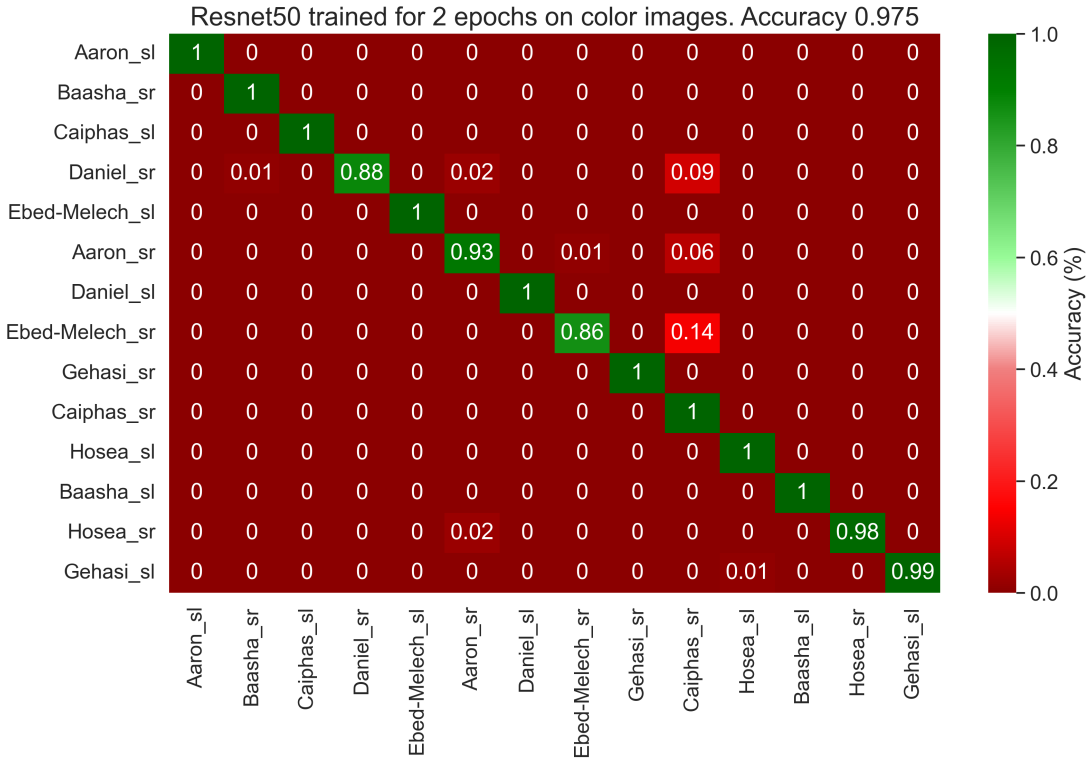


(a) Alexnet

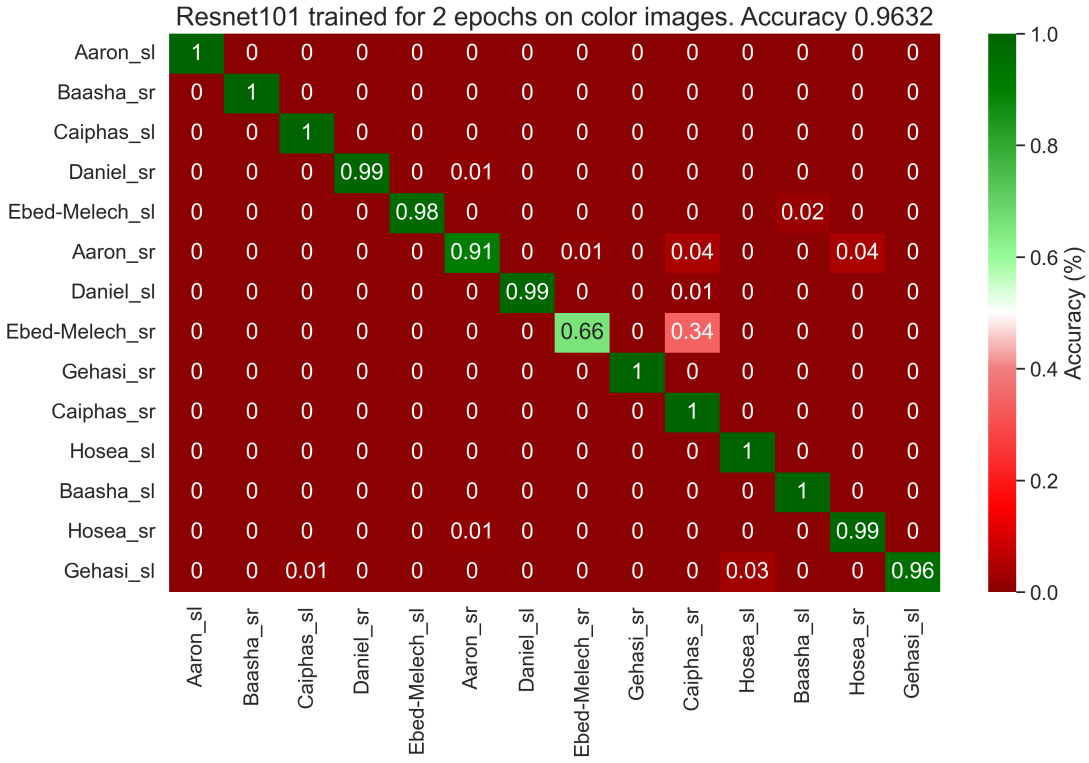


(b) Efficientnet

Figure 7.1: Salmon identification results for Alexnet and Efficientnet, trained on the color dataset for two epochs.

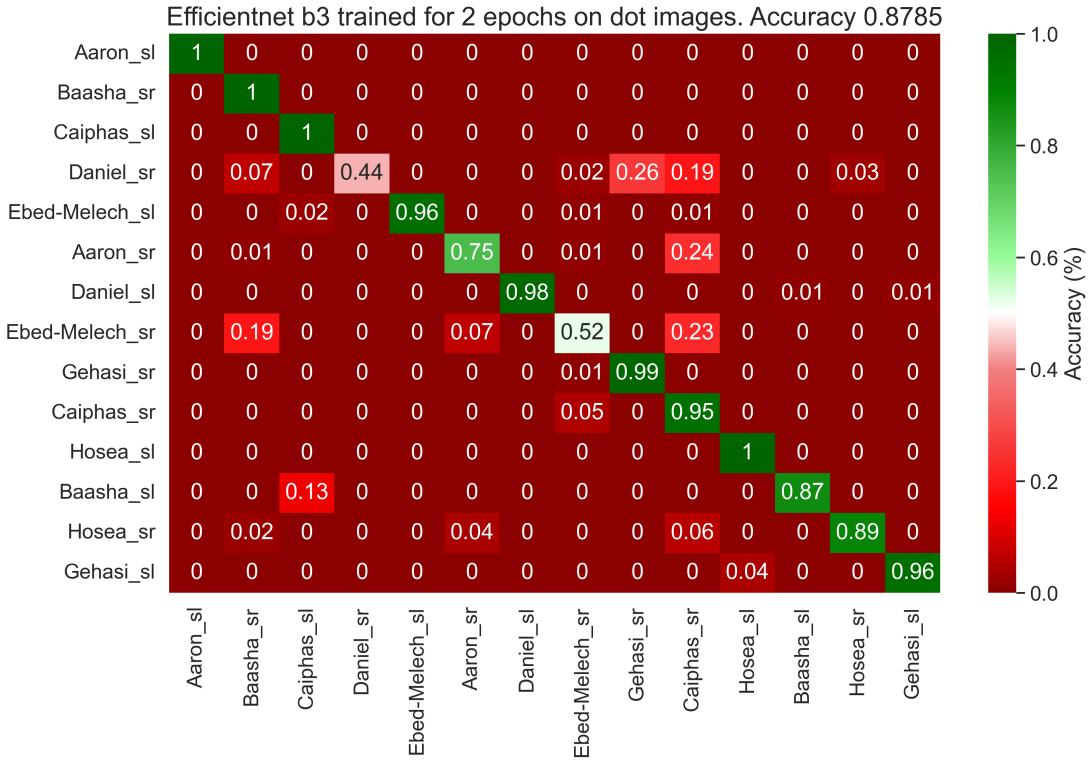


(a) Resnet50

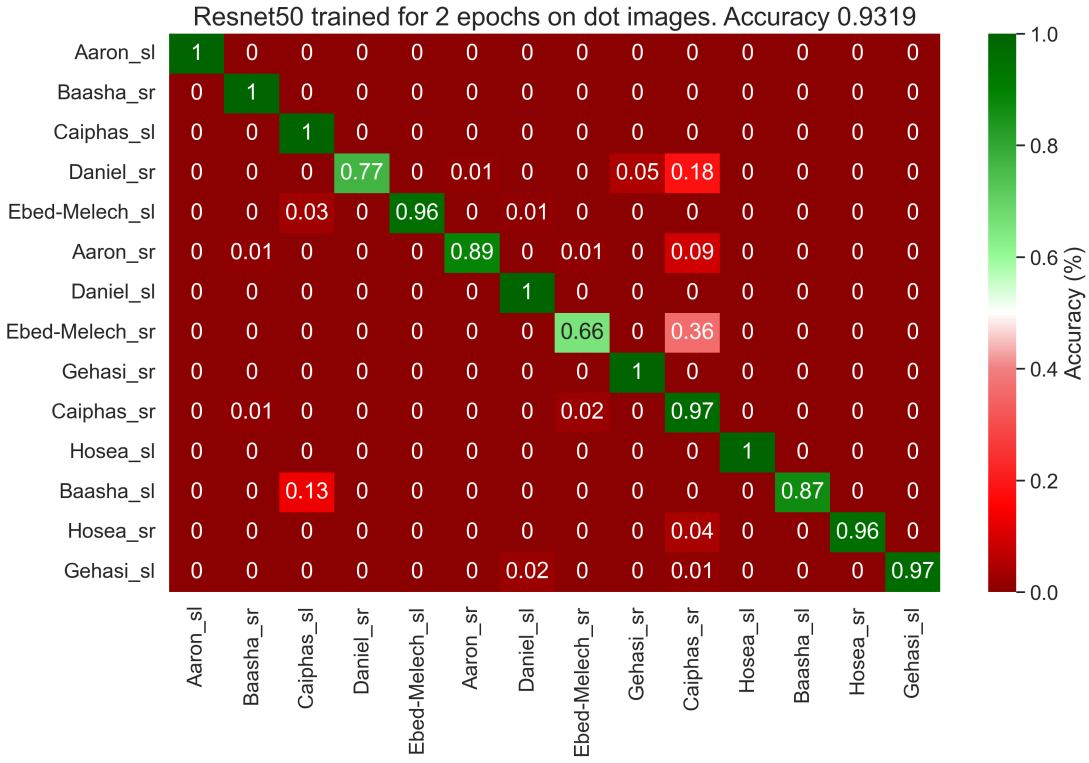


(b) Resnet101

Figure 7.2: Salmon identification results for two different Resnet models, trained on the color dataset for two epochs.



(a) Efficientnet, 2 epochs



(b) Resnet50, 2 epochs

Figure 7.3: Salmon identification results for two different networks trained on the dot dataset for two epochs.

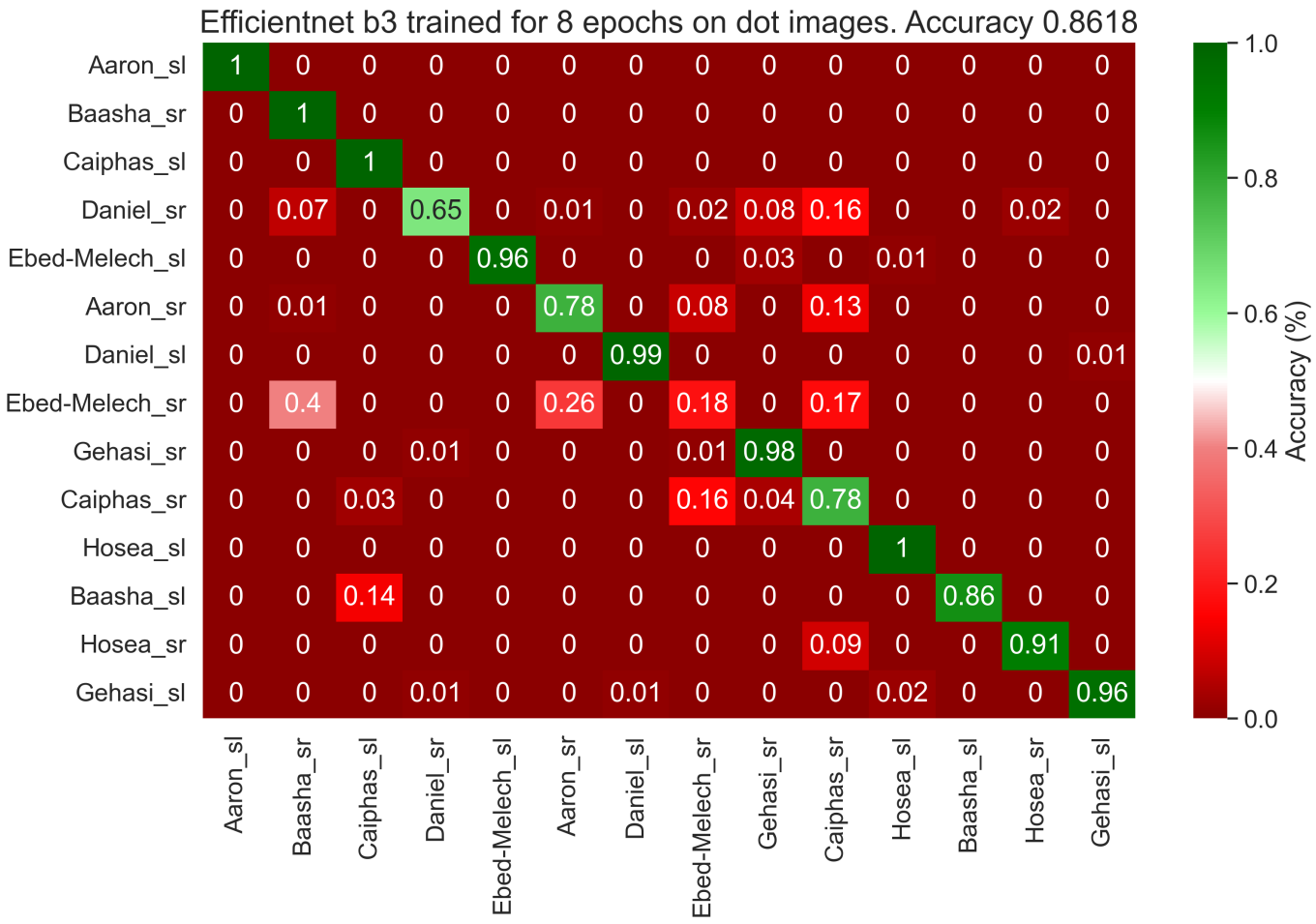
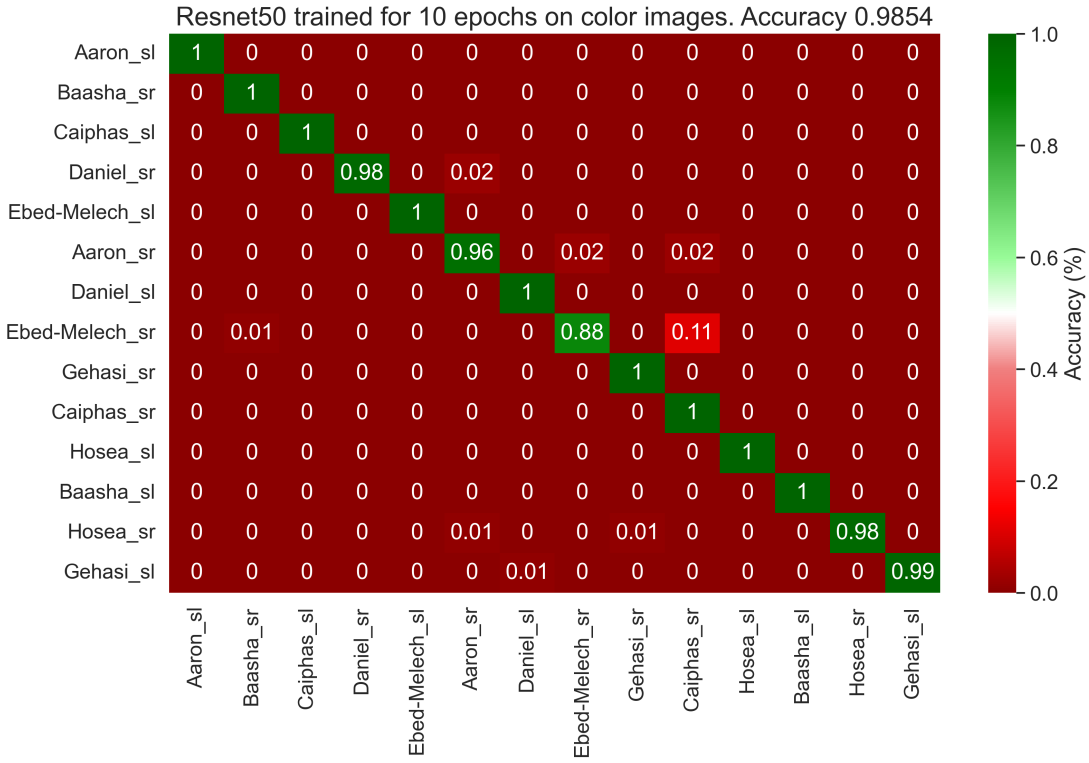
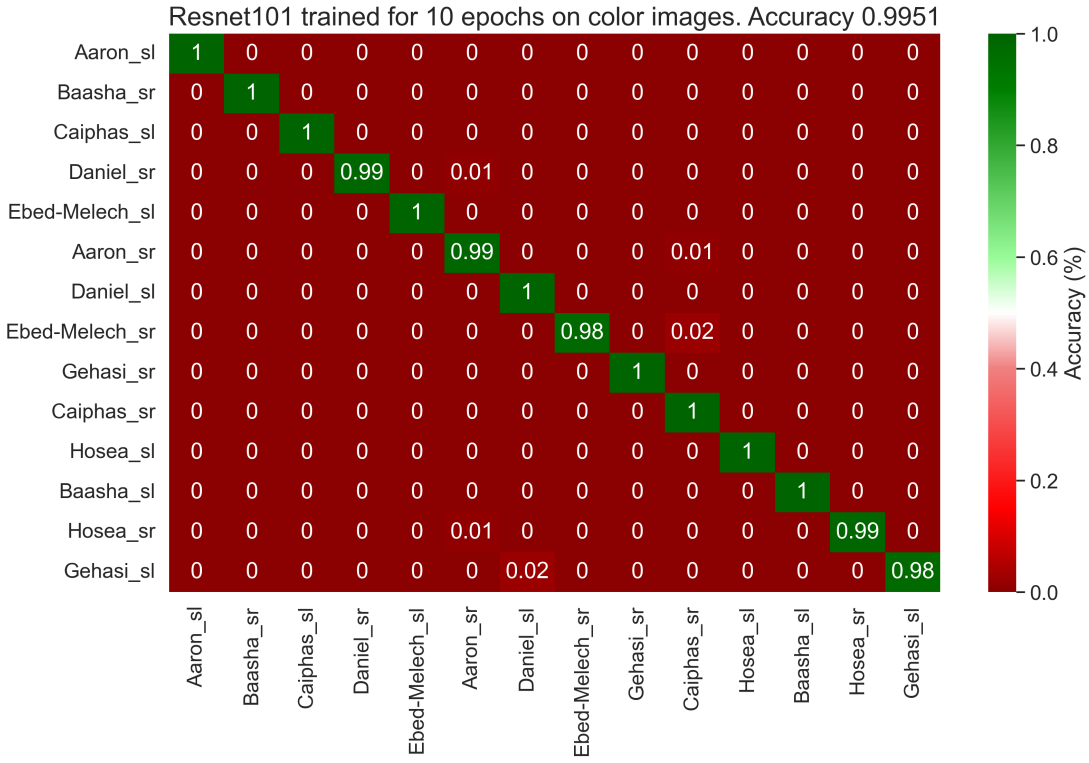


Figure 7.4: Salmon identification results for Efficientnet trained on the dot dataset for eight epochs.



(a) Resnet50, 10 epochs



(b) Resnet101, 10 epochs

Figure 7.5: Salmon identification results for two different ResNet models, trained on the color dataset for 10 epochs

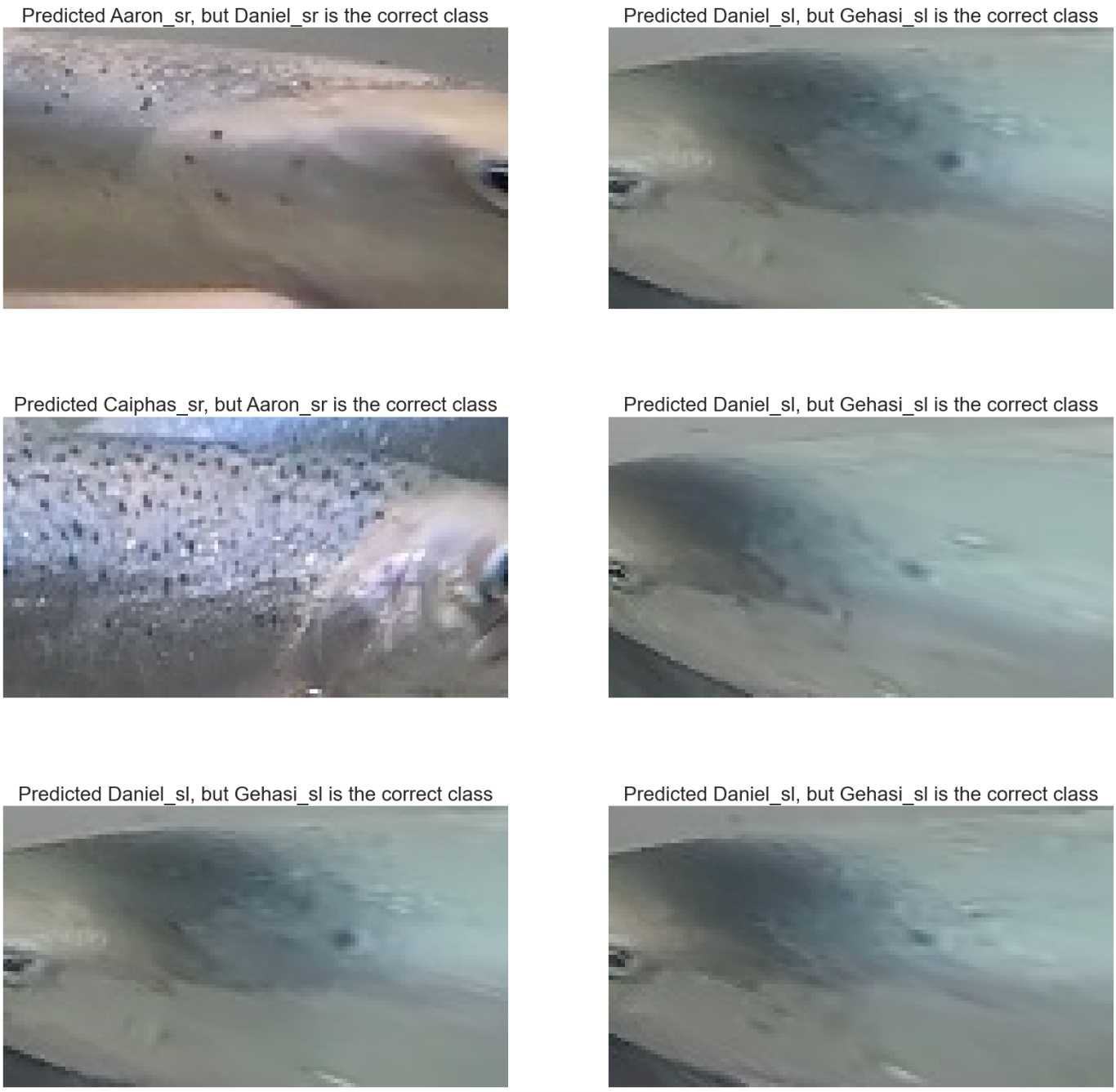


Figure 7.6: Examples of wrongly classified fish individuals for the Resnet101 model trained for 10 epochs on the color dataset.

7.5 Frequency analysis

In figure 7.7, a representative selection of jaw gape time series fitted to sinusoids are presented, while figure 7.8 displays the most common errors when fitting the sine wave models. The first case, figure 7.8a, displays a time series with a noisy signal, often received when salmon are turned towards the tank wall, making the salmon jaws difficult to distinguish for the Keypoint RCNN. Case two, shown in figure 7.8b, displays a signal that has captured the salmon body dynamics instead of the jaw dynamics. This happens when the mouth pose is difficult to detect, making the jaw keypoints more or less fixated to the salmon body, while the fish moves significantly. The final case, figure 7.8c, shows a salmon snapping (quickly opening and shutting its mouth). In this case, all of the time series gape estimates are correctly identified, however the true gape period is significantly different between two consecutive oscillations. The method employed in this report, fitting a sinusoidal model with fixed frequency, cannot deal with this kind of varying period.

In appendix C, a complete set of sine fitted time series from a 2000 frame video sequence is displayed.

A more concise representation of the extracted frequencies, in the form of a box plot, is presented in figure 7.9. In this plot, only the information used downstreams of the sinusoid fitting are presented, i.e. salmon class and respiration frequency.

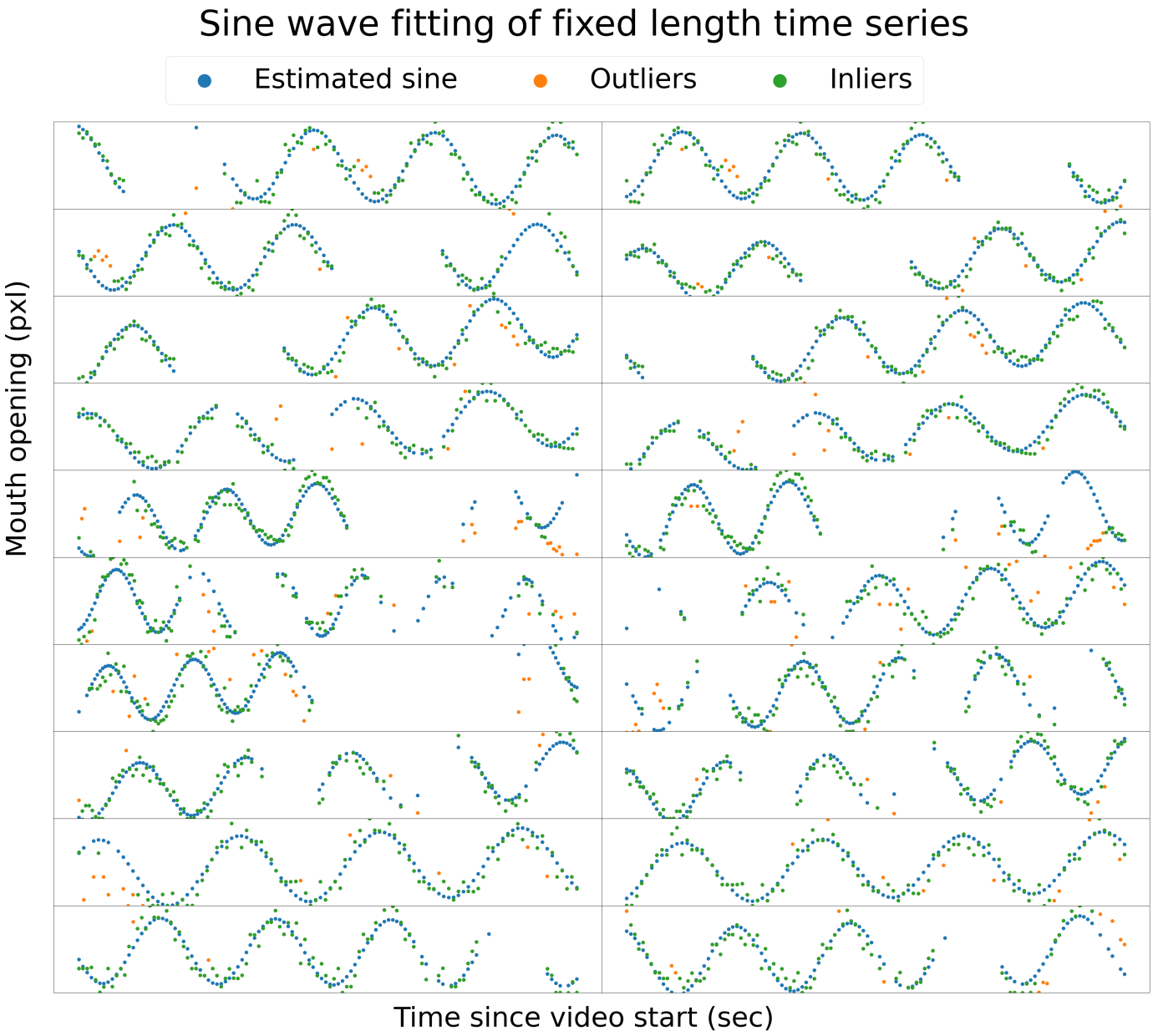
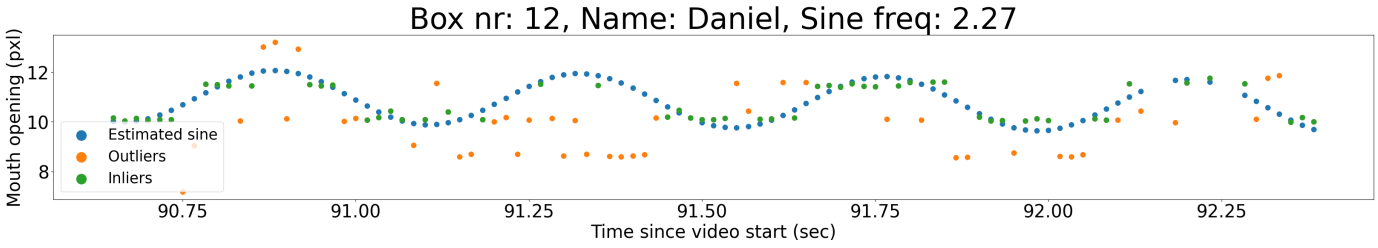
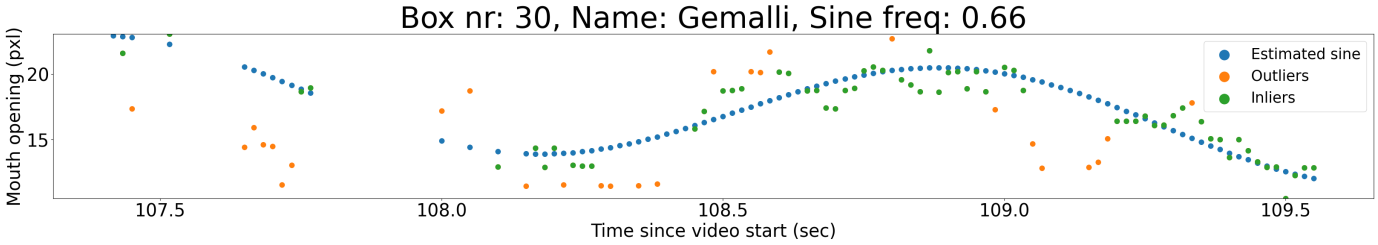


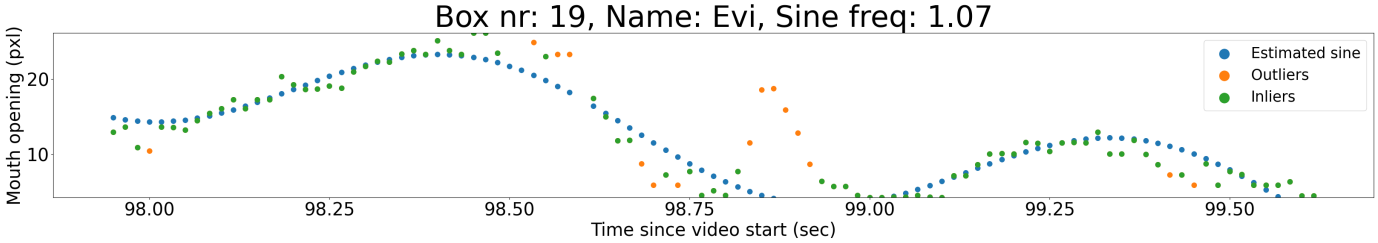
Figure 7.7: Time series function fitting. The x axis is describing the number of seconds since the start of the video clip, and the y axis is describing the mouth opening distance in pixels. All subplots axes are independent and broken, hence the scale of the oscillations, and the time they occur, are different between different time series. Each box is displaying one 100 frame time series. Ticks are omitted as their inclusion would take up space, without providing an increased understanding of the method.



(a) Noisy signal



(b) Signal displaying pronounced salmon dynamics



(c) Signal displaying snapping salmon

Figure 7.8: Some common error cases for time series model fitting

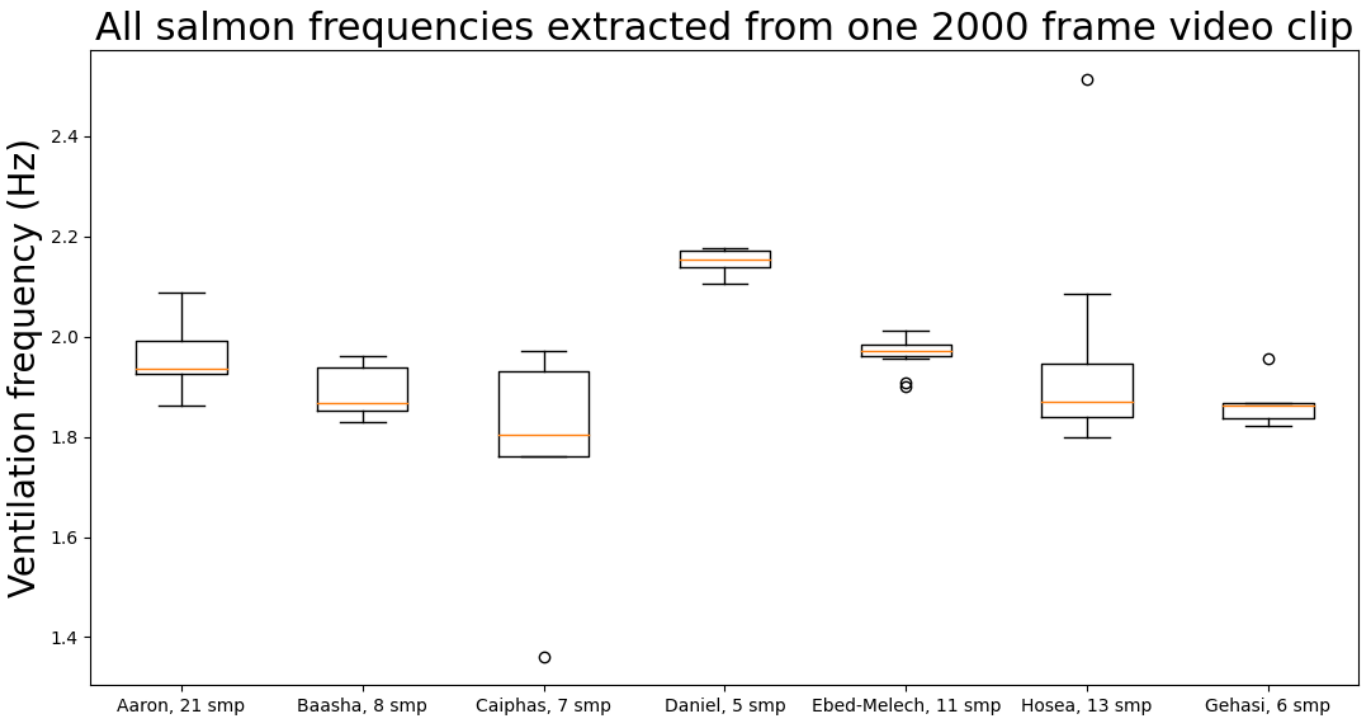


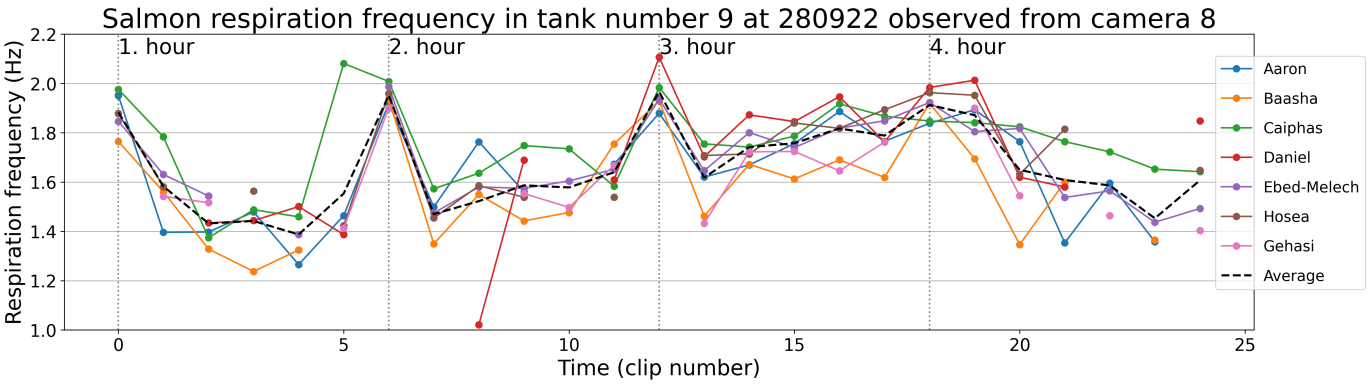
Figure 7.9: Box plot of salmon frequencies three hours into the first downbreathing of tank nine. The abbreviation smp. describes the number of frequency samples in each column, and the orange lines are the sample medians.

7.6 Respiration frequency extraction pipeline applied to experimental data

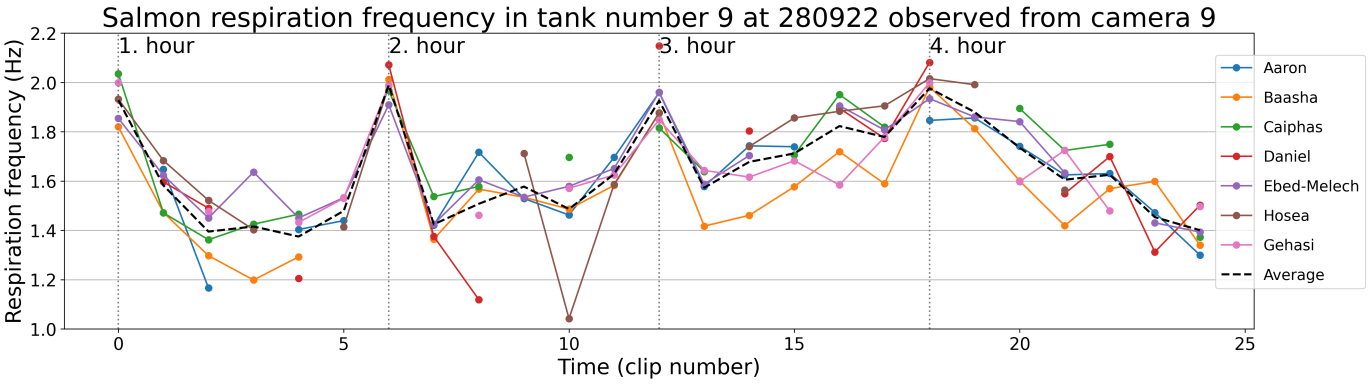
Frequency evolutions

In figure 7.10, the full frequency evolutions of all tank nine downbreathings are displayed, while the hourly respiration evolutions of tank nine are displayed in figure 7.11 and 7.12. For the hourly downbreathings, the data point at hour n contains all frequency data in the range [n-1, n], such that hour 1 is the first hour. The complete set of the full frequency evolutions of all tanks are presented in appendix E.

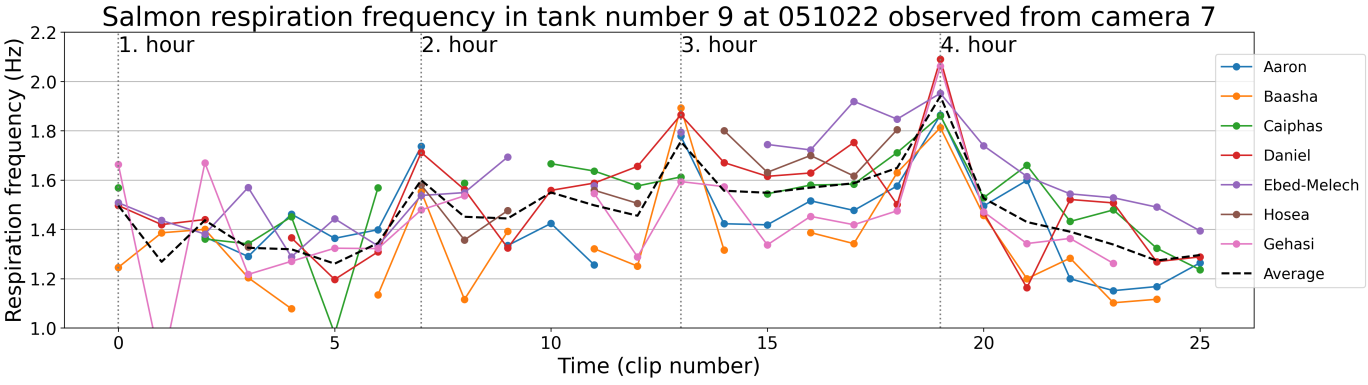
For both downbreathings, all hourly frequency trajectories were calculated, and plotted together in figures 7.13 and 7.14. The colors represent temperature, with cold tanks being drawn blue, tanks of intermediate temperature being drawn green, and warm tanks being drawn red. Furthermore, different cameras in the same tank at the same time are distinguished by being drawn with different line styles.



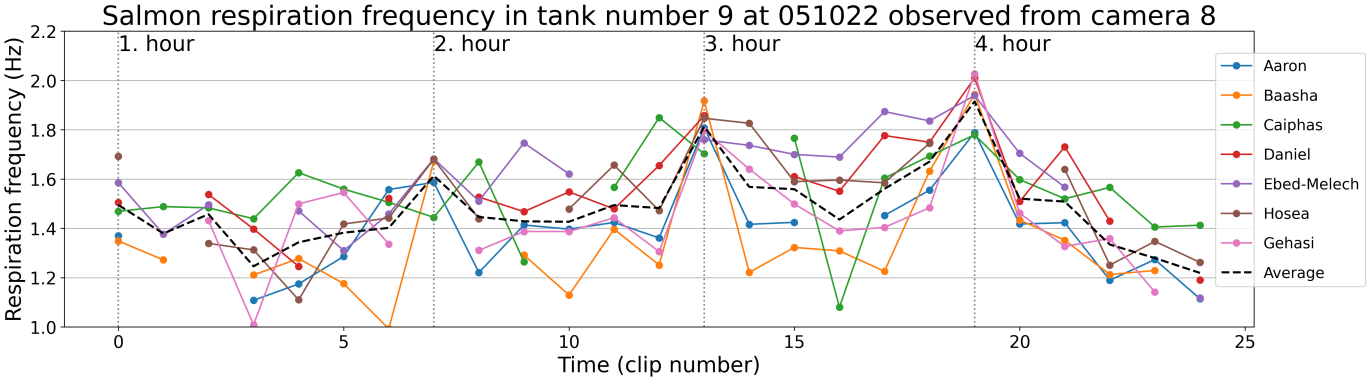
(a) Camera eight, first downbreathing



(b) Camera nine, first downbreathing

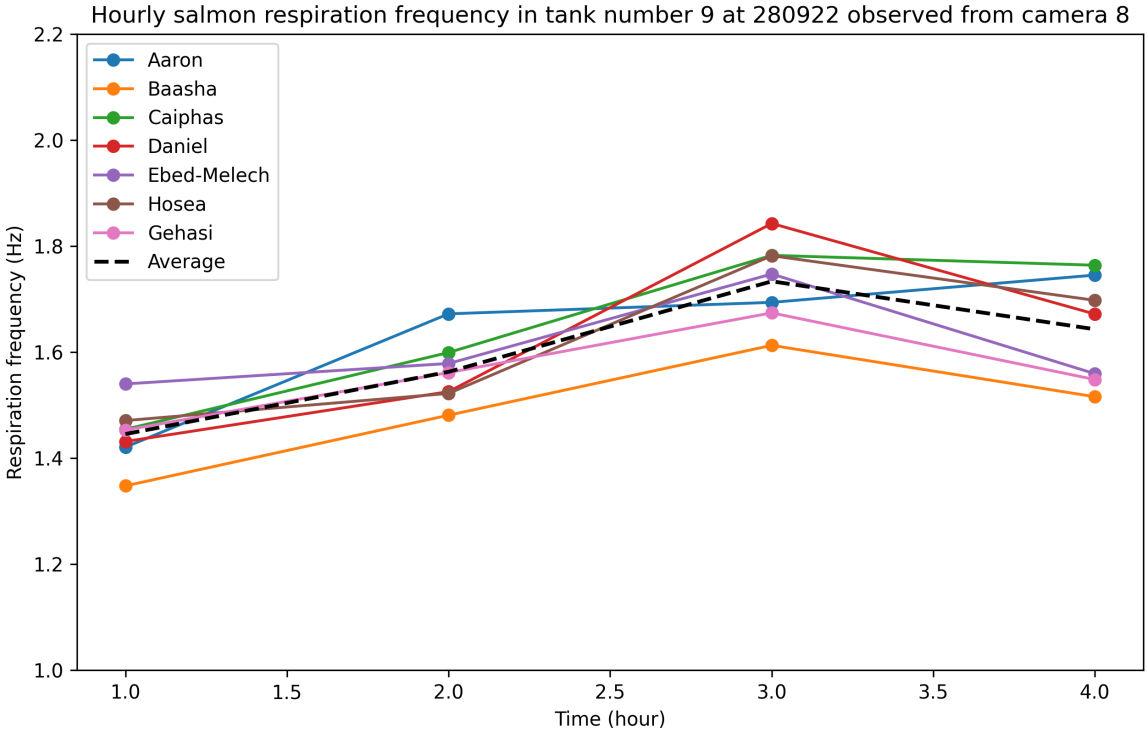


(c) Camera seven, second downbreathing

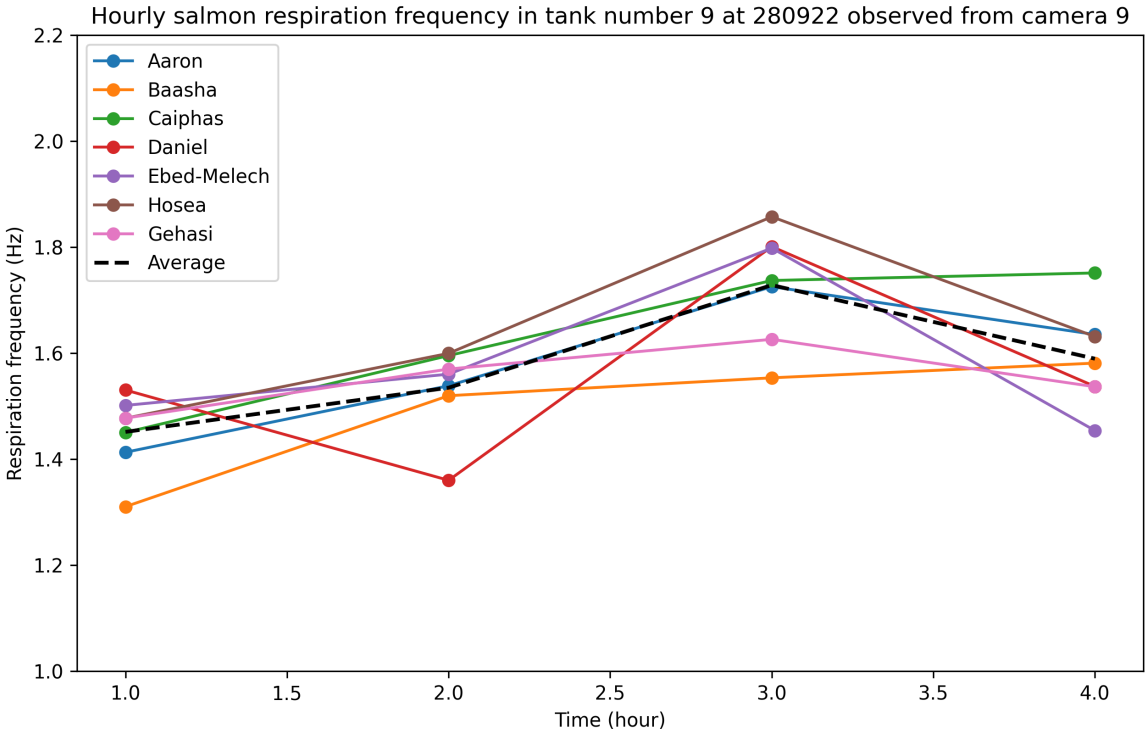


(d) Camera eight, second downbreathing

Figure 7.10: Respiration evolution for tank nine downbreathings. The vertical grey lines represent the time of camera insertion into the tank.

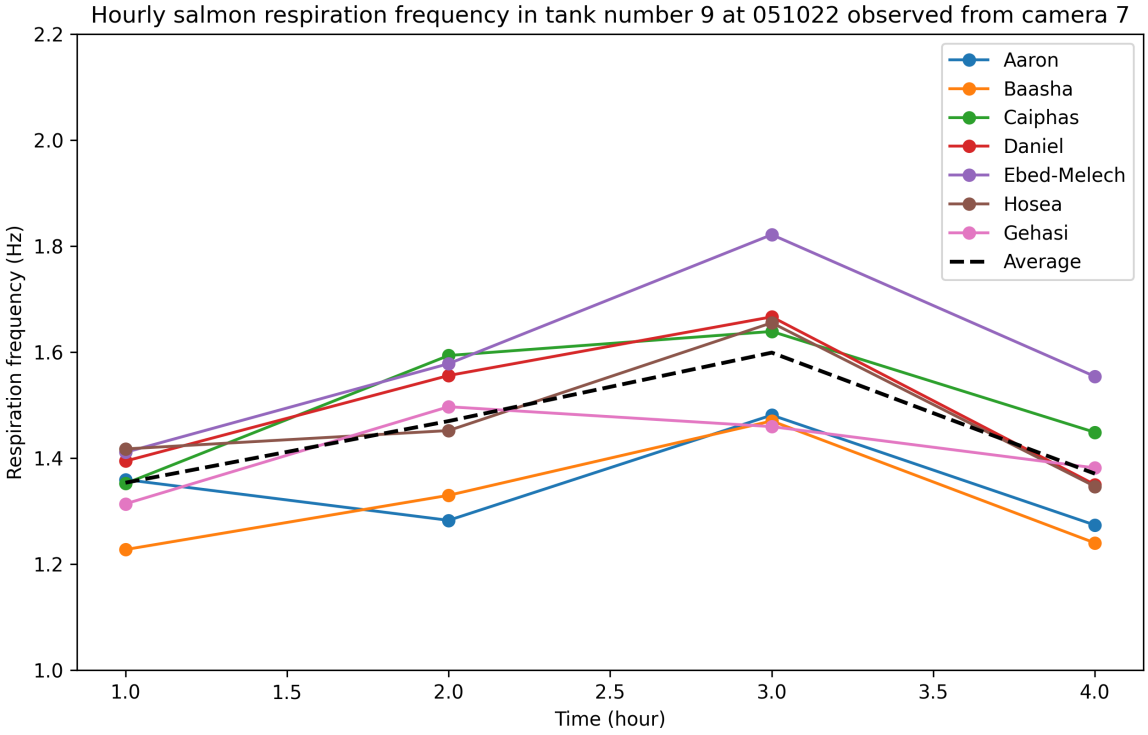


(a) Camera eight, first downbreathing

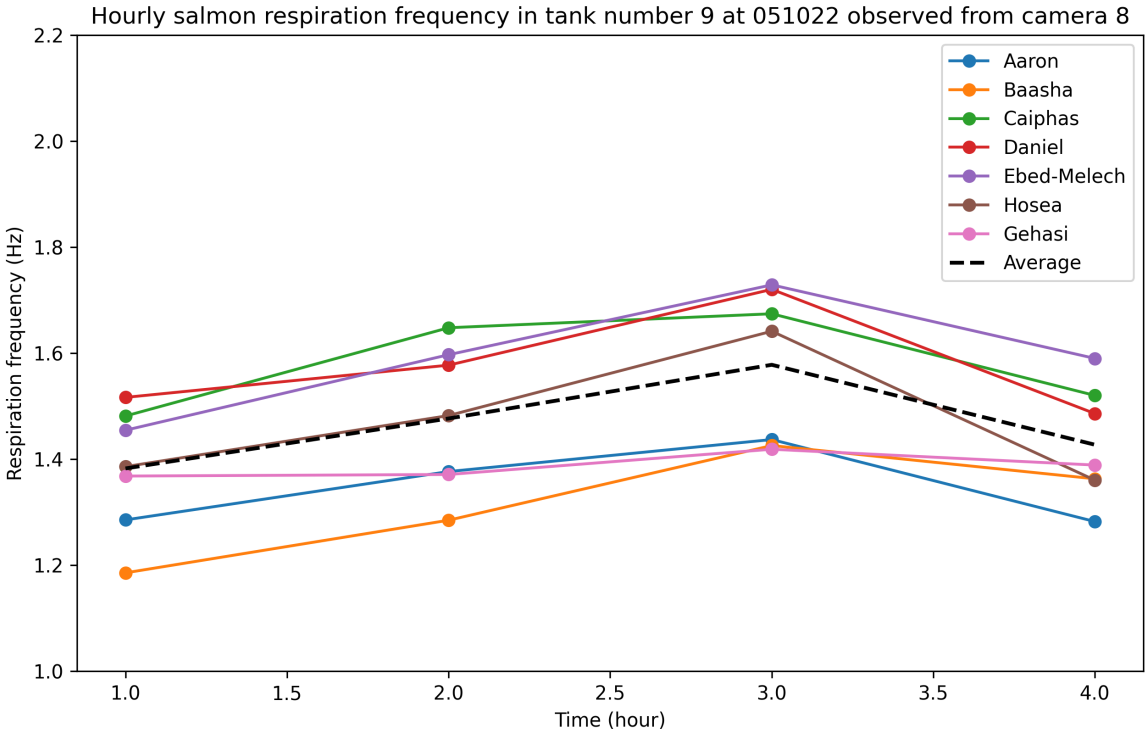


(b) Camera nine, first downbreathing

Figure 7.11: Hourly respiration evolutions for tank nine during downbreathing one



(a) Camera seven, second downbreathing



(b) Camera eight, second downbreathing

Figure 7.12: Hourly respiration evolutions for tank nine during downbreathing two

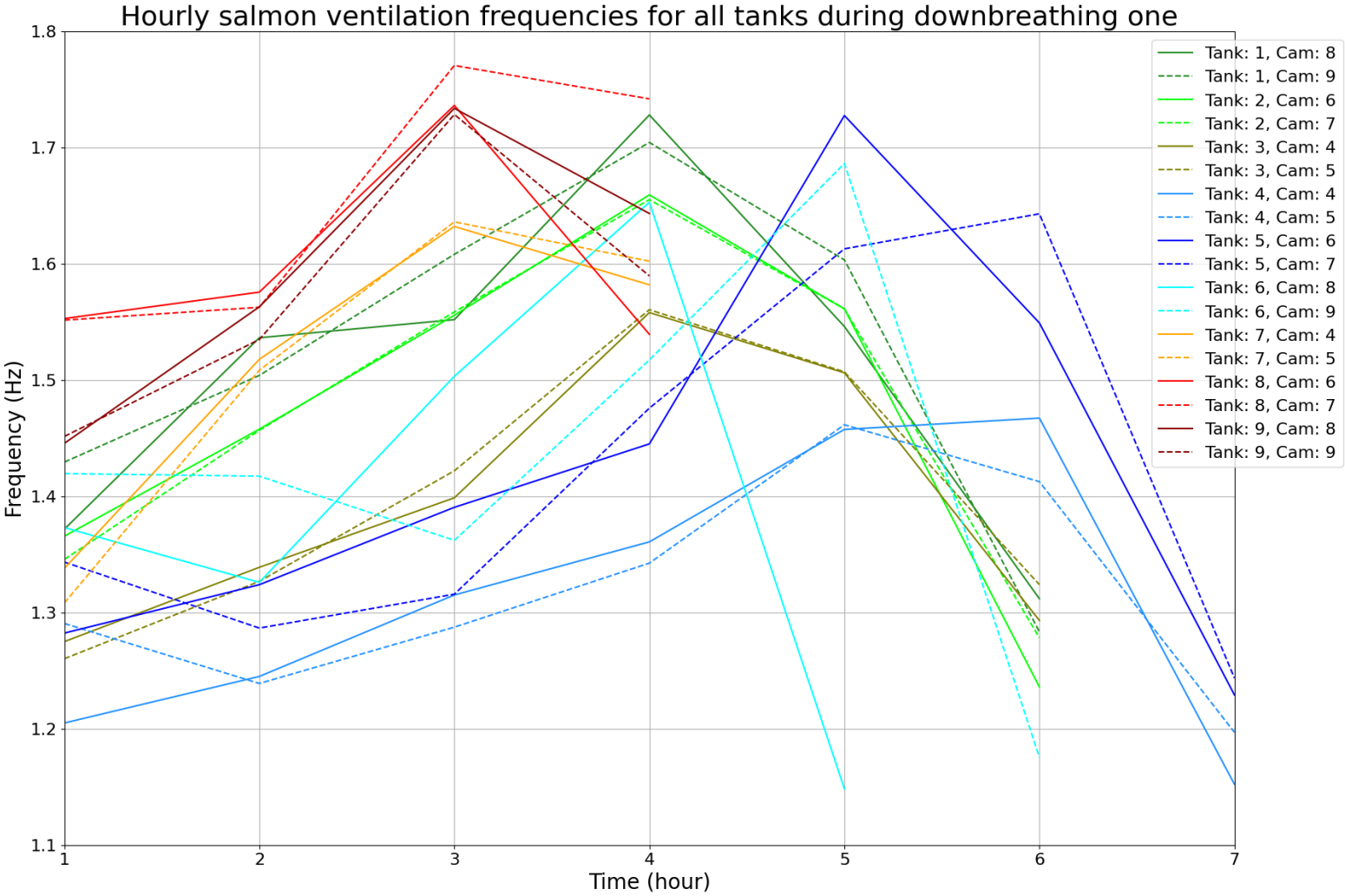


Figure 7.13: Hourly respiration frequencies for all first downbreathings (28.09.22 and 29.09.22)

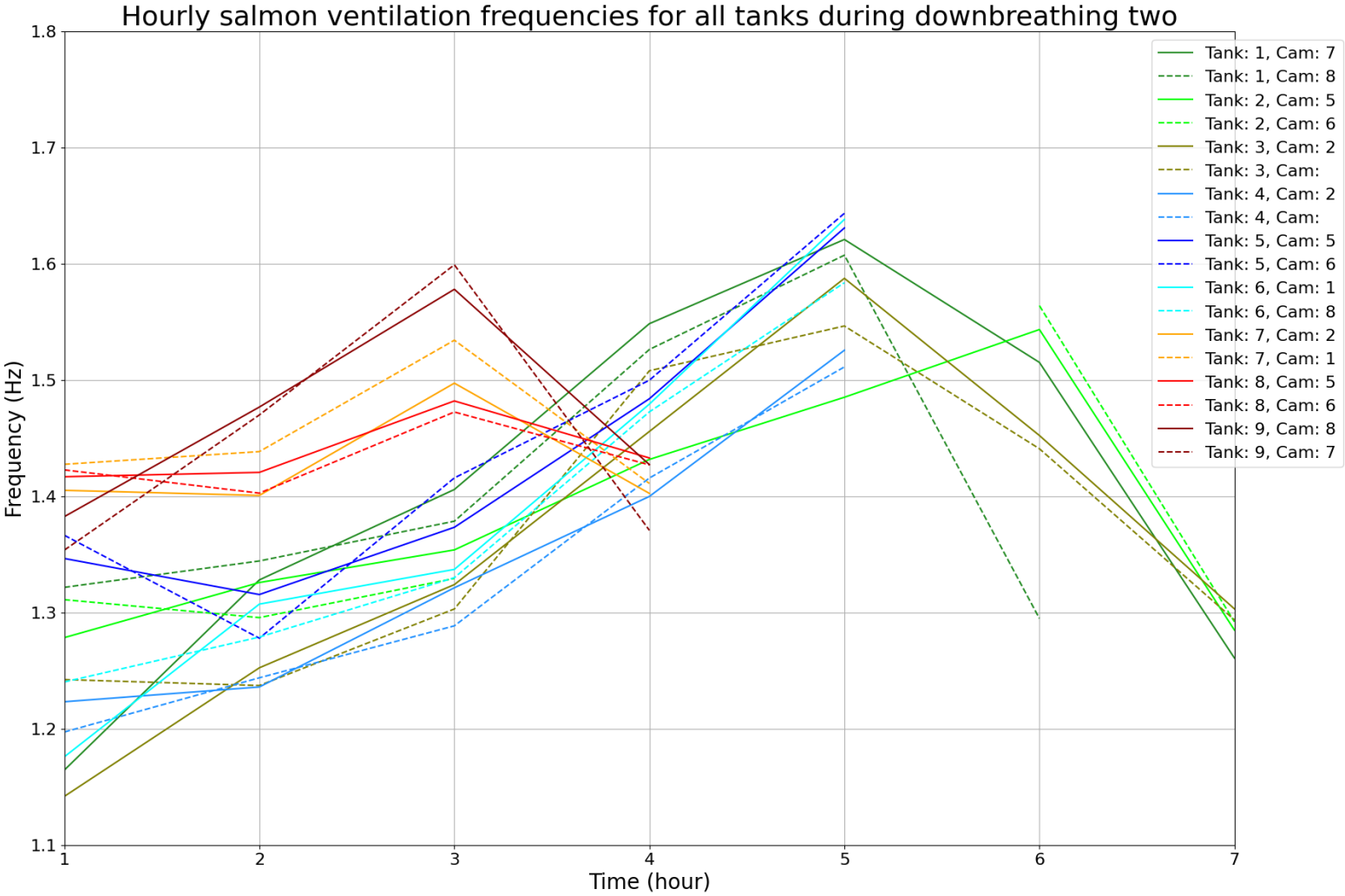


Figure 7.14: Hourly respiration frequencies for all second downbreathings (05.10.22 and 06.10.22)

Comparison of full ventilation frequencies and DO content

In figure 7.15, the average full ventilation frequencies ($fullTrajAvg$) of tank nine are plotted together with the negated DO content of tank nine. The following steps were performed to generate the plots.

1. All DO saturation values were negated, so that it could be meaningfully compared to the ventilation frequency.

2. Ventilation frequency clips with camera disturbance were omitted from the plot.

3. The lowest negated DO value was matched with the first hourly ventilation frequency, and the maximum negated DO value was matched with the highest value of the full frequency evolution (disregarding clips with camera disturbance). In this way, the two trajectories were scaled so that their values were in the same range.

4. Every salmon recording (clip) were assumed to be 12 minutes.

5. A Pearson correlation coefficient (r, see section 4.8.2) was calculated for each full frequency evolution using a dataset containing the $fullTrajAvg$ values (disregarding disturbance clips), and the DO saturation values at the times of the $fullTrajAvg$ measurements. In the title of the plots, both this r value, and its associated p value (probability that the correlation is different from zero), were included. Importantly, the r value was calculated from the DO content, not the negated DO content that is used in the plots.

In appendix F, the average full ventilation frequencies ($fullTrajAvg$) plotted together with the negated DO contents are displayed for all tanks.

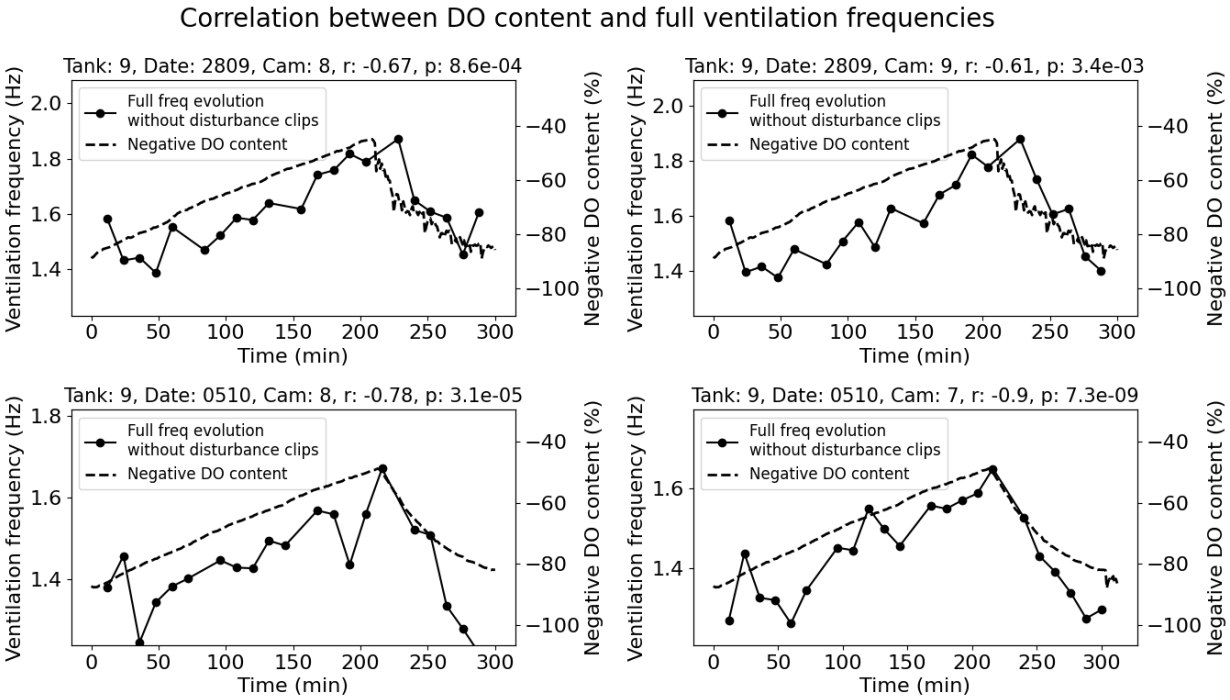


Figure 7.15: Negated DO content and full frequency evolutions of tank nine.

Statistical analysis of ventilation frequency extremes

The salmon stress experiment (chapter5) data contained six different temperatures; cold, intermediate and warm for each of the two downbreathings. At each of these temperatures, the following three extreme values were extracted from the ventilation frequency trajectories.

1. **Frequency first hour**: The average of all hour one respiration frequencies (item one in the hourly frequency evolution) at a given temperature.
2. **Max hourly frequency**: The average of all hourly maximum values (the item of max value in the hourly frequency evolution) at a given temperature.
3. **Max clip frequency**: The average of all maximum values in the full frequency evolutions (disregarding the clips containing camera disturbance) at a given temperature.

In table 7.1, all these extreme values are presented. The uncertainties, set to two standard deviations ($2 \cdot \sigma$), are included as well.

Temperature	Frequency first hour	Max hourly frequency	Max clip frequency
12.3 $°C$	1.32 $\pm$0.15 Hz	1.61 $\pm$0.23 Hz	1.69 $\pm$0.14 Hz
13.4 $°C$	1.26 $\pm$0.15 Hz	1.59 $\pm$0.11 Hz	1.73 $\pm$0.19 Hz
14.0 $°C$	1.34 $\pm$0.13 Hz	1.64 $\pm$0.14 Hz	1.77 $\pm$0.11 Hz
15.6 $°C$	1.24 $\pm$0.14 Hz	1.58 $\pm$0.06 Hz	1.69 $\pm$0.08 Hz
15.9 $°C$	1.44 $\pm$0.21 Hz	1.71 $\pm$0.12 Hz	1.85 $\pm$0.08 Hz
17.7 $°C$	1.40 $\pm$0.05 Hz	1.53 $\pm$0.10 Hz	1.62 $\pm$0.10 Hz

Table 7.1: Overview over maximum and initial ventilation frequencies for different temperatures. The uncertainty is set to twice the standard deviation ($2 \cdot \sigma$).

To evaluate whether any of the ventilation frequency extreme values of table 7.1 were statistically different, a t-test (see section 4.8) were performed between all the different temperatures for each of the three extreme values. The following temperatures were found to have statistically different (p value below 5%) ventilation frequency extremes. This is also displayed in graphical form in figure 7.16.

1. Frequency first hour:
   (a) 12.3 $°C$ measurements are different from those at 15.9 $°C$ (p = 0.043)
   (b) 12.3 $°C$ measurements are different from those at 17.7 $°C$ (p = 0.045)
   (c) 13.4 $°C$ measurements are different from those at 15.9 $°C$ (p = 0.007)
   (d) 13.4 $°C$ measurements are different from those at 17.7 $°C$ (p = 0.005)
   (e) 14.0 $°C$ measurements are different from those at 15.6 $°C$ (p = 0.036)
   (f) 15.6 $°C$ measurements are different from those at 15.9 $°C$ (p = 0.004)
   (g) 15.6 $°C$ measurements are different from those at 17.7 $°C$ (p = 0.003)
2. Max hourly frequency:
   (a) 13.4 $°C$ measurements are different from those at 15.9 $°C$ (p = 0.006)
   (b) 14.0 $°C$ measurements are different from those at 17.7 $°C$ (p = 0.010)
   (c) 15.6 $°C$ measurements are different from those at 15.9 $°C$ (p = 0.002)
   (d) 15.9 $°C$ measurements are different from those at 17.7 $°C$ (p = 0.000)
3. Max clip frequency:
   (a) 12.3 $°C$ measurements are different from those at 14.0 $°C$ (p = 0.049)
   (b) 12.3 $°C$ measurements are different from those at 15.9 $°C$ (p = 0.001)
   (c) 13.4 $°C$ measurements are different from those at 15.9 $°C$ (p = 0.031)
   (d) 14.0 $°C$ measurements are different from those at 15.6 $°C$ (p = 0.022)
   (e) 14.0 $°C$ measurements are different from those at 15.9 $°C$ (p = 0.021)
   (f) 14.0 $°C$ measurements are different from those at 17.7 $°C$ (p = 0.001)
   (g) 15.6 $°C$ measurements are different from those at 15.9 $°C$ (p = 0.000)
   (h) 15.6 $°C$ measurements are different from those at 17.7 $°C$ (p = 0.036)
   (i) 15.9 $°C$ measurements are different from those at 17.7 $°C$ (p = 0.000)

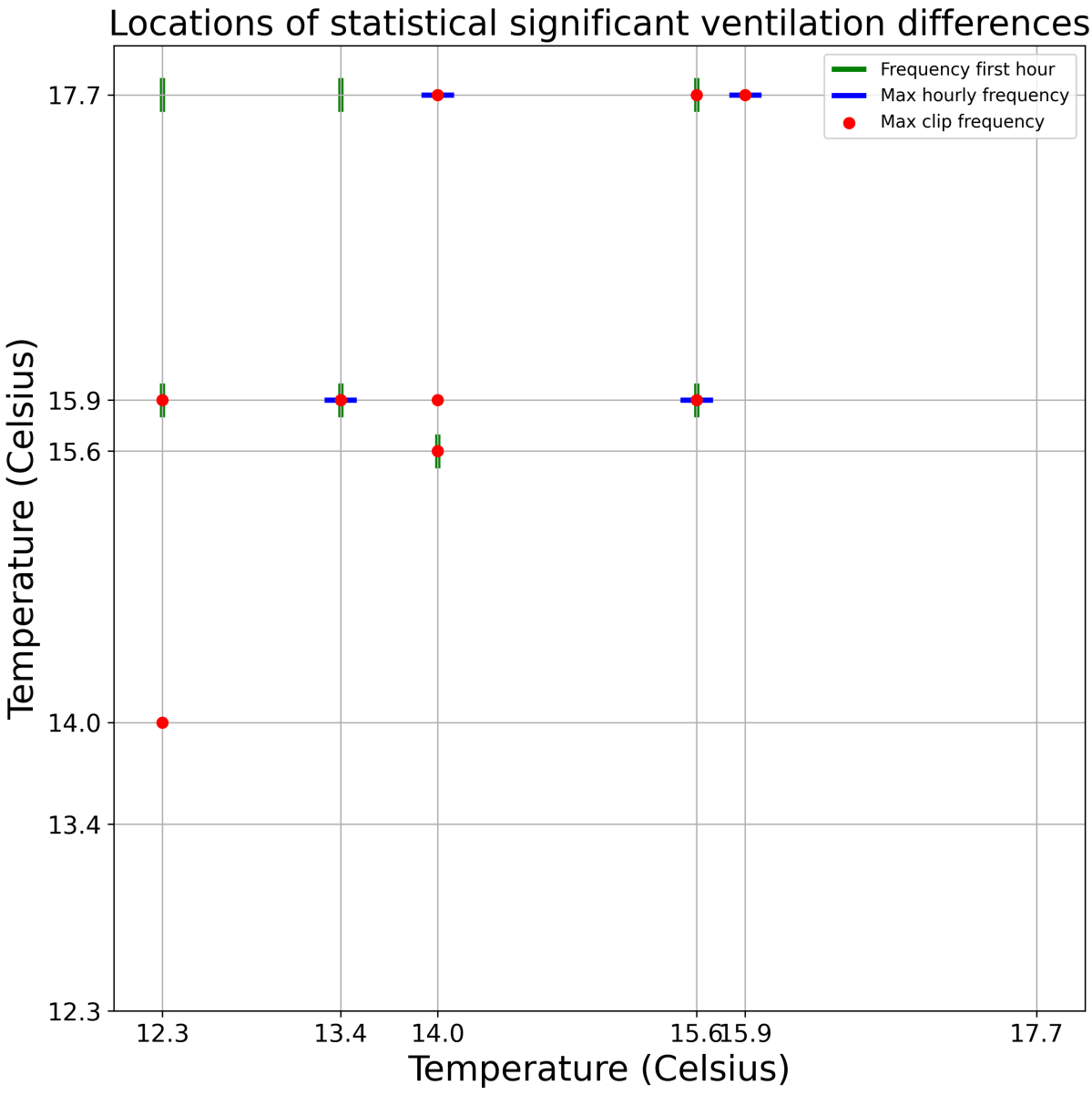


Figure 7.16: Visual presentation of the temperatures with significantly different ventilation frequency extremes (at a p value of 5 %). Only the upper left triangle of the plot is filled out, so each temperature comparison is only recorded once in the plot.

Fish personality

Figure 7.17 displays the personality of all fish, defined as the average deviation from the shoal ventilation frequency over a downbreathing (see section 6.7). All hourly frequency evolutions are used to evaluate fish personality, hence the total set of fish personalities will contain 4 samples for each fish individual (one for each downbreathing), totalling 252 samples. The histogram were fitted to a Gaussian curve using the scipy.stats.norm.fit() method[139].

To evaluate how fish personality changed between downbreathings, a bar plot was created (see figure 7.18) the following way.

1. Order each fish in a downbreathing according to their personality, such that the class number of the slowest breathing fish is placed at index zero in the personality order list.

2. Calculate how much each fish fall or rise on this ranking list between downberathing one and two, and put all these values into a list. Call a number in this list a *ranking change*.

3. Use the list of ranking changes to find the rate of ranking change for all possible ranking changes (zero to six).

4. Since some ranking changes are statistically more likely than others, calculate the rate of ranking change if all rankings were done randomly, by simulating 100000 random personality orderings.

5. For each possible ranking change (zero to six), subtract the rate of random ranking change from the data ranking change. This yields the rate of all ranking changes, corrected for random effects. In figure 7.18, these ranking change surpluses are displayed. Additionally, the uncorrected rate of ranking change (which sums to one) is written at the end of each bar in the plot.

Since each downbreathing had two cameras, one tank yielded four ranking comparisons, totalling 28 ranking changes (seven fish per comparison) for each tank.

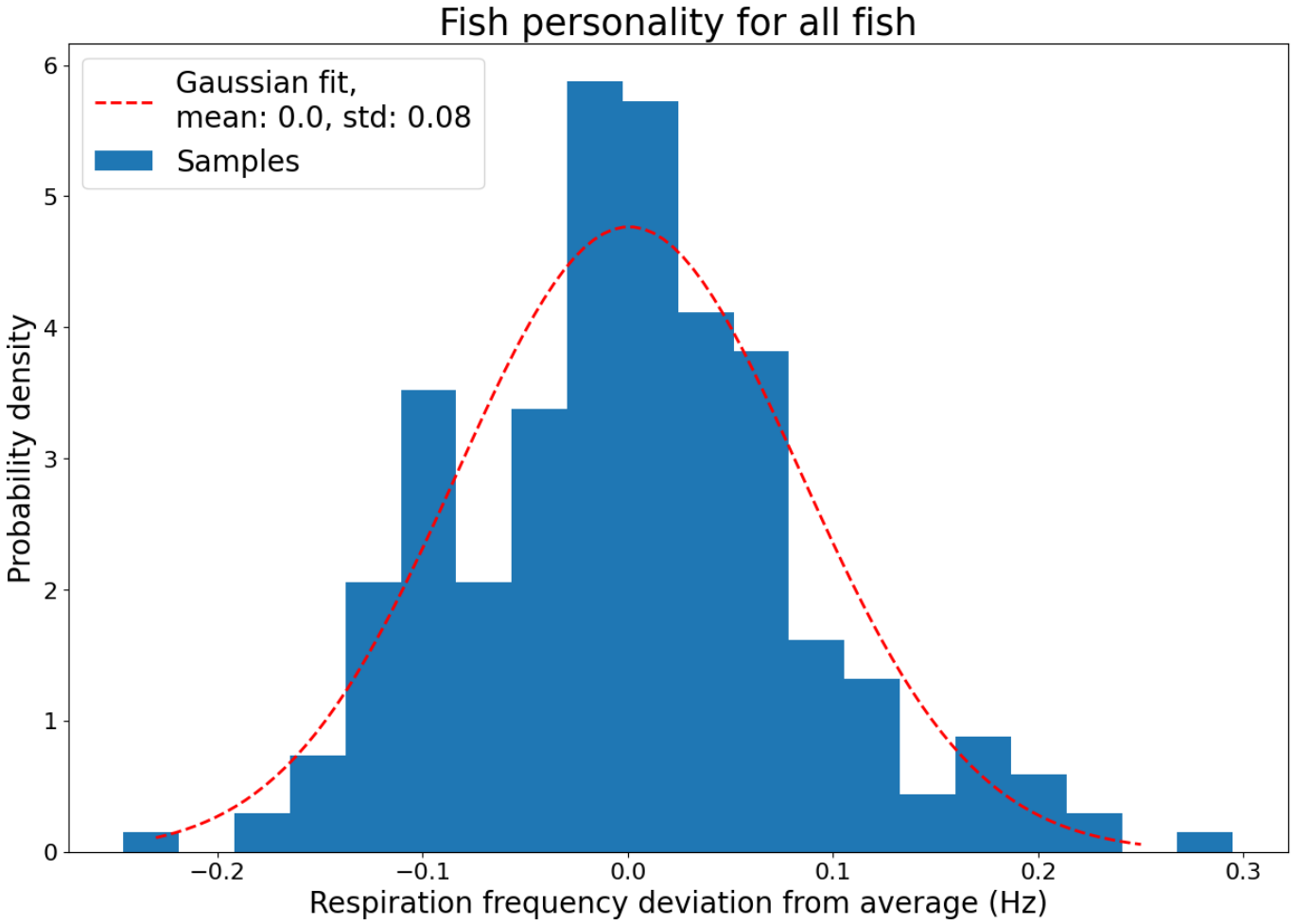


Figure 7.17: Histogram displaying all fish personalities

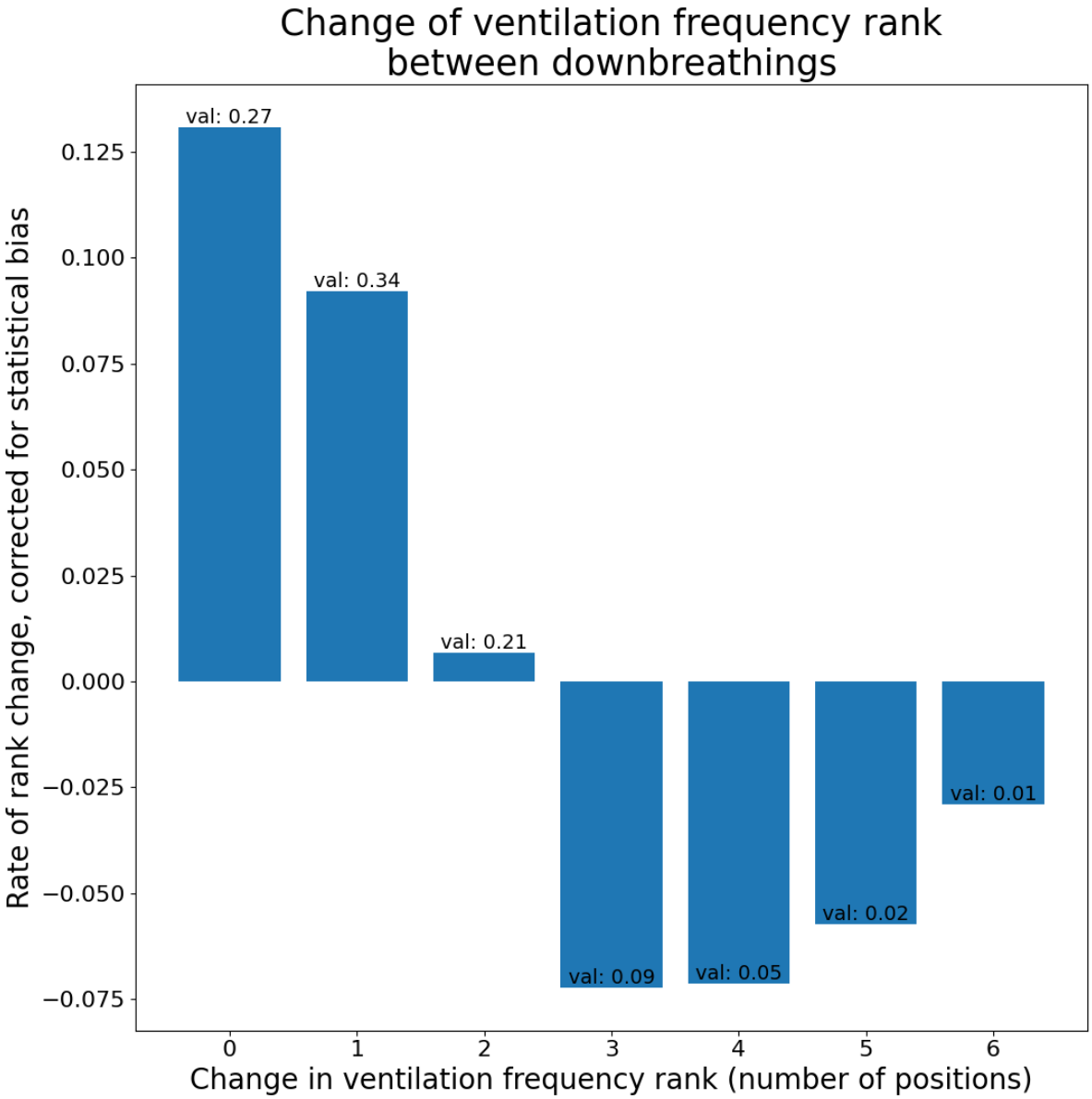


Figure 7.18: Figure displaying the change of fish personality ranking between downbreathings, corrected for how much ranking change a random fish ordering would induce. The text at each bar states the percentage of the given fish ranking change before correction of the statistical bias.

Respiration frequency increases after disturbance

Figure 7.19 displays all respiration frequency increases after camera disturbance of all downbreathings (see section 6.7). The mean of these samples was above zero with an infinitesimal p value ($9.05 \cdot 10^{-51}$).

To evaluate how these respiration increases are distributed over the downbreathing timeline, figure 7.20 was created, where all respiration frequency increases after camera insertions are displayed, separated into the hour following the respiration increase (so hour one displays the ventilation increase at the beginning of the downbreathing). Hour two, three and four are significantly different (at p value 5%) from hour one, five, six and seven.

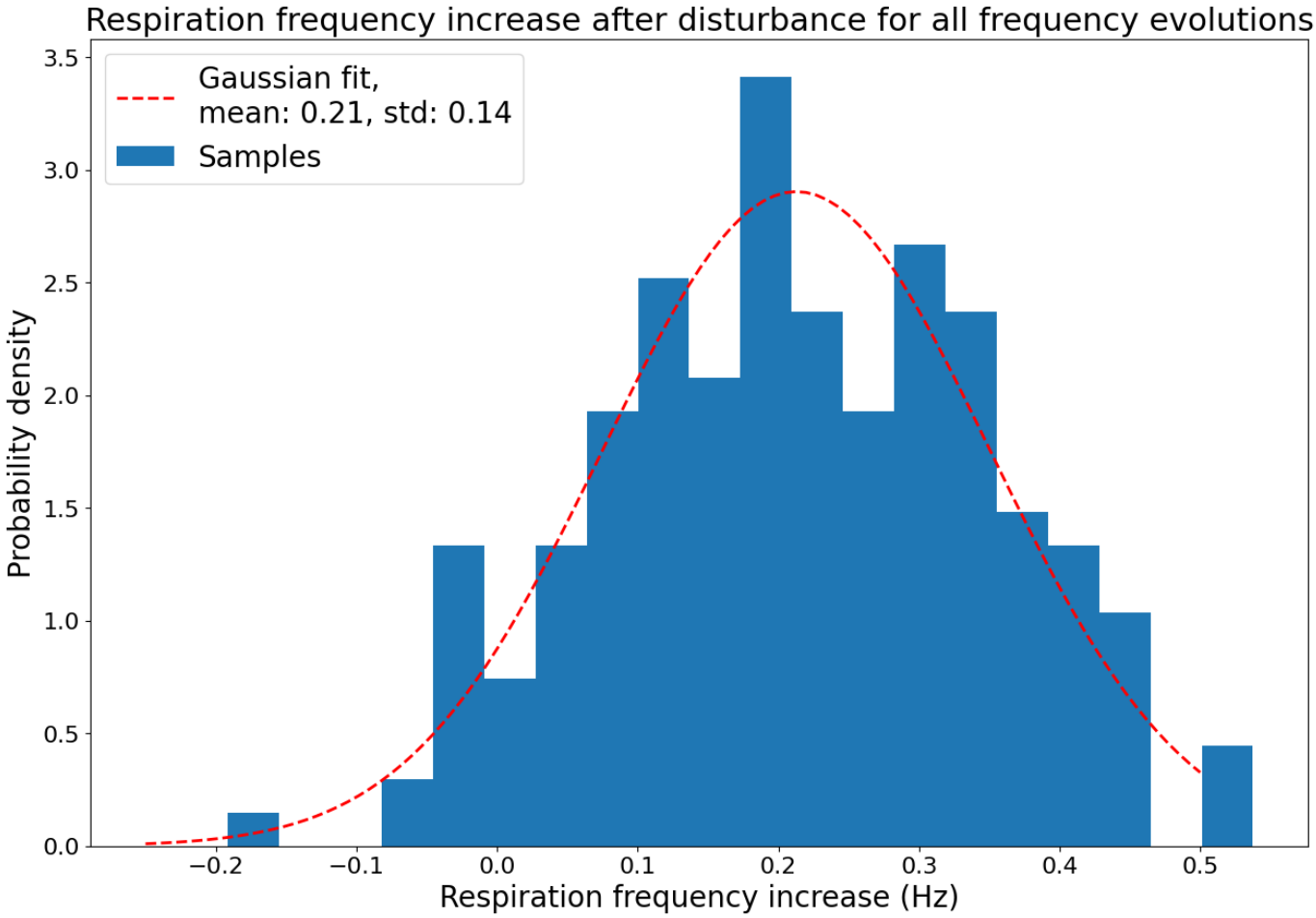


Figure 7.19: Histogram displaying all respiration frequency increases after camera disturbance.

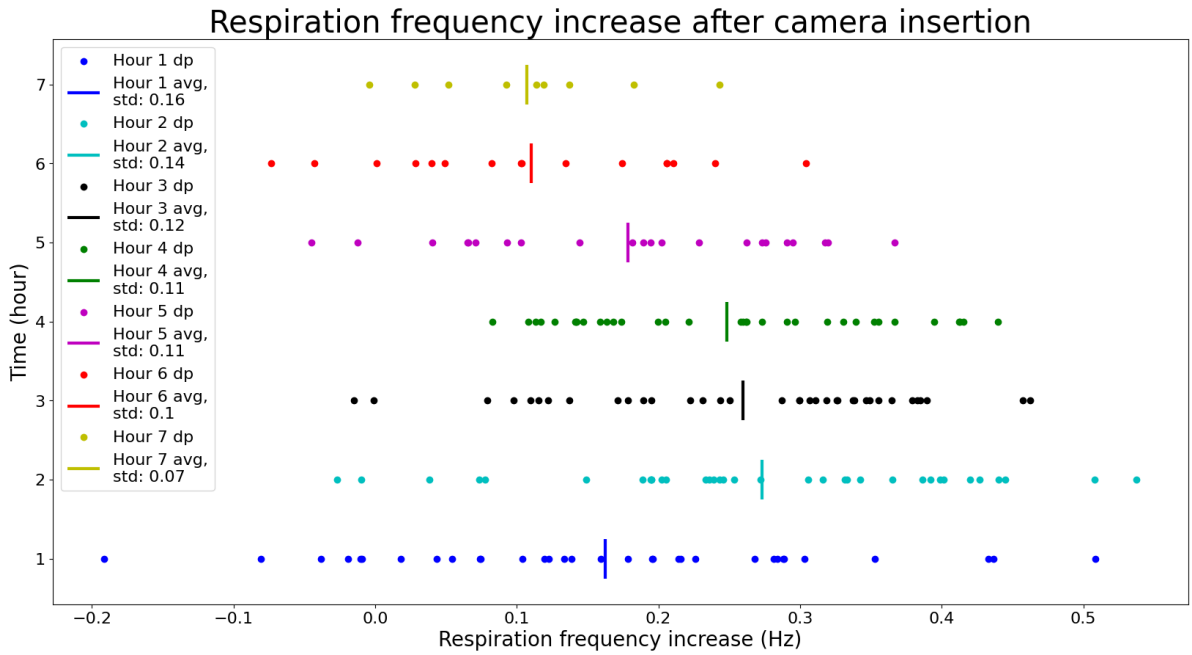


Figure 7.20: Figure displaying the increase in respiration frequency after camera insertion for all downbreathing hours. Each filled circle is a datapoint, while the bars are hourly averages.

Resource requirements

The processing time of generating jaw gape time series from one 2000 frame image sequence was around 01:30 hours, while the sinusoid fitting took around 2 minutes. As such, the total processing time for one downbreathing was between 30 and 70 hours on an Intel i7 processor.

Chapter 8

Discussion

8.1 Chapter introduction

This chapter presents a discussion of all pipeline segments, presented in the order they appear in the final salmon ventilation frequency method. Both methodology decisions and results are dealt with in detail.

8.2 Keypoint RCNN

8.2.1 Method

Keypoint placement

The locations of the keypoints were chosen so that they would be easy to distinguish, even on fish far away from the camera, and to ensure the fulfillment of the requirements stated in the numerated list in section 4.2.5 (extraction of salmon ventilation information and the calculation of a transformation matrix). The inclusion of a root jaw keypoint facilitated comparison of different jaw poses.

Selection of classes

In the salmon recordings, fish were overlapping or swimming towards the camera a significant amount of time. In this configuration, salmon identification is close to impossible, while the breathing frequency is detectable. By including the two classes *front* and *occluded* to the set of salmon individual classes, breathing frequency relevant keypoints (jaw keypoints) could be annotated even when classification was infeasible.

Augmentation

The augmentations employed can be separated into two categories, spatial-level and pixel-level. The spatial transformations are nested, such that rotation, scal-

ing and shearing are subsets of the perspective transformation. The reason for using transformations from several complexity levels was to have more control of the exact nature of the final image, in order to make the augmented frames as similar to the original ones as possible. The pixel-level transformations ensured robustness to different lighting conditions.

Label pruning

By removing any label with one occlusion, some fish that should be labeled *occluded* were instead labeled as *background*. This incoherent labelling increase the rate of false negatives at the frame borders, and might reduce the quality of the detections the network makes at these locations. One way to tackle this is to drop the *occluded* class altogether, and require a completely visible salmon torax for all annotations. This would reduce the amount, while increasing the reliability, of the extracted ventilation information, since partly hidden fish have a higher rate of erroneous keypoint placement than fish without occlusion.

Another approach to deal with annotation at the frame borders is to ensure that the labelling is coherent at these locations, by moving keypoints outside the frame to their closest projection on the image, to only do transformations that preserve all labels, or to use a network that accept missing keypoints. The two first suggestions would require writing own support functions, and the last one would require constructing a new Keypoint RCNN model.

8.2.2 Result

Classification

As expected, the classifications performed by the Keypoint RCNN worked poorly, most likely due to the insufficient size of the training dataset. The only classes with somewhat reasonable results were the two support classes *occluded* and *front*. As a result, the network class head was disregarded completely in the final salmon breathing frequency pipeline. Considering this, the classes can be thought of as an auxiliary target for the model performing salmon detection and keypoint prediction. Since class features are mainly dependent on salmon dots, while keypoint detection is dependent on anatomical landmarks, it would be interesting to compare whether training without the class information would give better or worse results.

Keypoint prediction and salmon detection

Since the Keypoint RCNN network was not able to learn keypoint occlusion, bad keypoint predictions for occluded fish are expected. Furthermore, some drop in accuracy for novel tanks is expected. The general picture is that the labelling should be slightly more conservative, as a false negative error is better than an erroneous true positive detection (that is, correct salmon detection, but wrong keypoint placement). A more restrictive labelling could mean not labelling fish turned away from the camera, and not labelling occluded fish.

The Keypoint RCNN network solves the task of keypoint prediction and salmon detection acceptably. By increasing the size of the training dataset, the performance could be improved even more.

8.3 Tracking

8.3.1 Method

Two important trade-offs, presented below, reveal themselves when it comes to tracker performance. The appropriate balance of these were found by tuning the relevant parameters through trial and error.

1. By accepting fewer consecutive frames without detections, and require closer proximity between eye detection and tracker, fewer trackers will jump between fish, however more trackers will loose its current fish assignment and be removed from the set of active trackers (T, see section 6.3). Components downstreams of the tracking algorithm in the breathing frequency pipeline, including splitting time series to a fixed size, and enforcing performance requirements of the time series classification, will ensure that jumping trackers is a fixable problem. To exemplify, if a tracker switch fish after 400 frames, the fixed-length time series before frame 400 will be accepted by the breathing frequnecy pipeline, while the time series including frame 400 will be rejected by the performance requirements of the time series classification (see the *acceptable_time_series*() function of section 6.6). Trackers with lost fish, however, have no later pipeline fixes, and is hence a larger issue.

2. The relative weight of the measurement noise (σ_m) compared to the model (acceleration) disturbance (σ_a) in the Kalman filter must be tuned appropriately. As the model of the fish dynamics is very simplified, the model uncertainty is larger than the measurement uncertainty, which is consistent with how the parameters were tuned (see section 6.3).

The reason for using the fish eye as the basis for the trackers, was because it was the keypoint with the most accurate predictions.

8.3.2 Results

From the results of the salmon tracker (see section 7.3) it can be observed that the Kalman filter based tracking worked well, despite the large simplifications done when constructing it (see section 4.4). The trackers are able to track salmon with sufficient accuracy for downstream tasks.

8.4 Semi-automated labelling

8.4.1 Method

Augmentation

Since the homogeneous mapping of the salmon torax image was used to remove spatial level variability, it should be ensured that this variability is not reintroduced to the images during the augmentation phase. Because of this, only one spatial transformation was applied the salmon image; optical distortion, in order to simulate small errors in the image mapping (homogeneous transformation). The other augmentations applied to the salmon classification dataset were used to train the network to not fixate on small details (dropout), ensure scale/resolution invariability (blur), and to make the network robust to changing lighting or color conditions (contrast, brightness and tone curve adjustment).

Dataset diversity

The inclusion of data from both downbreathings turned out to be necessary for the later network classification performance. Whether distinguishing features of the fish were dot intensity, scale loss or wounds, they tended to change significantly between downbreathings, making it hard, even for humans, to correctly re-identify salmon across the two periods of oxygen restriction.

Unlabelled data sampling

An improvement to the semi-automated labelling could be to sample only every tenth video frame when creating the frames annotated with the Keypoint RCNN predictions. This way, the constructed datasets would contain more diverse poses, without increasing the computing requirements of the unlabelled (not containing individual class labels) data generation. An issue with this is that the difficulty of, and time spent on, the manual labelling would increase, as the fish would move in and out of the camera field of view more frequently.

Salmon camera shyness

One problem encountered when generating the salmon classification datasets, was the case of what can be colloquially called *camera shy* fish[1]. Some fish tended to only swim in one direction along the side of the tank, resulting in difficulty of gathering data from the salmon side proximal to the tank rim. This required spending some time searching for salmon in the right pose, however all fish individuals eventually were represented by 500 images from each side in the final dataset.

[1] This name is meant to give the reader an intuition about the problem, and must not be understood as a suggestion that the fish is afraid of the camera.

A related issue is the case of consistent salmon trajectories. The salmon enjoyed following the same paths during its movements, causing lighting conditions and fish poses to correlate with salmon individuals. This could potentially confuse the neural network model trained on this dataset to learn the salmon behavior, instead of the salmon individual. This would lead to classification errors if salmon during evaluation of the network occupy novel configurations.

Resource requirements

The process of annotating the fine tuning dataset of one tank with seven fish took about eight to ten hours of manual labour. Hence, this process is feasible for experimental settings, but cannot be extended to industrial applications.

The dataset

The method of semi-automated labelling allowed the creation of a huge dataset, totalling 118492 images distributed over 126 classes. In this report the data was only used for finetuning deep neural networks, however it would be interesting to apply this training set to other few-shot frameworks. To the authors knowledge, no similar large scale salmon identification benchmark has been developed, and as such, this dataset can be considered a scientific contribution on its own.

8.5 Salmon identification

8.5.1 Method

Rejected salmon identification strategies

Two methods for salmon identification were briefly tried and rejected, before the deep learning approach described in section 6.5 was chosen.

The first strategy was the use of scale-invariant geometric relationships, such as euclidean distance relationships and angles, to distinguish salmon individuals, motivated by the way human re-identification is performed (see section 3.7.6). None of the tested relationships, calculated from the keypoint predictions of the Keypoint RCNN, showed any clear distinguishing patterns between intraindividual and interindividual measurements. This was likely due to the noise of fish pose and deformation being more pronounced than the individual variations.

The second strategy was to extract salmon dot locations of salmon torax images transformed to a predefined configuration, and then compare the similarity between two dot sets in order to determine whether they origin from the same fish. The dots were accurately detected by classical machine vision methods, however the homogeneous mapping was not accurate enough to allow for direct comparison between a novel and a prototype salmon dot mask.

Few-shot classification methods

Some few-shot classification methods were tested to see whether a large tank nine dataset together with small datasets for the other tanks would be enough to perform accurate individual salmon classifications in all tanks. It turned out that neither Siamese nor Prototypical networks were able to get accuracy close to the fine tuning approach, even with the best performing tank nine network as backbone. This is unfortunate, as Siamese and Prototypical networks can, unlike the fine tuning approach, easily be adapted to the zero-shot classification task (see section 3.7.7). The methods tested here were not adapted to the problem domain, instead they were general algorithms trained on the salmon classification datasets developed in this report. Methods tailored to the problem domain, such as ones showing increased attention to dot structures, will likely yield improved results.

Some work was spent on using only a few instances from each salmon class, inflated hundredfolds, as the finetuning dataset. The best validation accuracy with this method reached 94% with a single manually annotated frame for each salmon individual. The confusion matrix for this network is displayed in appendix D, and it shows that all classes had more than 80% accuracy. This accuracy, however, was not sufficient to provide reliable classifications for the salmon time series. Further increasing the manually annotated frames to 100 instances per class was still not sufficient, while 500 worked well. Note that increasing the number of manually annotated frames was compensated by a reduction of the dataset inflation ratio.

A possible explanation for why a high accuracy network was not able to provide reliable time series classifications, is that the augmentations applied to the few-shot dataset covers most, but not all, of the true variability inside a salmon individual class. When achieving 94% one-shot accuracy, a large amount of augmentations were tested, hence the final result is caused by overfitting on the augmentation pipeline. As such, the augmentations are tuned to cover the variability in the validation set (the set that yielded 94% accuracy), while the rest of the true variability confuse the network at evaluation time. The effect of these unseen variabilities have a large influence on some trackers, as the error dependence across temporally proximal detections of the same fish makes a single unseen variability influence large parts of a tracker.

Homogeneous transformation

An improvement that could reduce the necessary manual annotations of the finetuning network, as well as make the two rejected identification methods feasible, is to make the homogeneous transformation more accurate. This could be done by letting the Keypoint RCNN detect a larger number of keypoints, and use a loss function based method for calculating the homogeneous transformation matrix. By using more keypoints, it is also possible to use a transformation with more degrees of freedom than a homogeneous transform, for example one that allows progressive scaling.

All datasets created in this work used the pixels of a transformed salmon torax image as input features. It would be interesting to evaluate how a salmon identification network trained on raw (untransformed) salmon images would perform, and evaluate whether the homogeneous transformation yielded improved network performance.

Dot dataset

The motivation behind the dot dataset was to enhance informative salmon features, while removing information irrelevant for salmon re-identification. Gray scale images retain most of the information in the salmon images, and might have been sufficient for salmon classification on its own. The Pytorch network models, however, are built with three channel images, which made it reasonable to choose the same number of channels in the dot dataset, so that the dataset could be applied directly to the premade models. In a dot image, the two last channels (an adaptive thresholding image and a Laplacian grid) were used to simplify the salmon dot detection for the network, and to ensure dot structure was used for the individual salmon classification. The neural network could learn to calculate both adaptive thresholding images and Laplacian grids from a gray scale image without specifying these features in independent channels, however specifying them guides the network into focusing on dots, instead of geometry or color.

Network evaluation

Accuracy were used as optimality metric when evaluating the deep classification models. Due to the relatively large number of classes in each tank, and the fact that the dataset is balanced, this provides good insight into the model performance. Additionally, a confusion matrix was constructed to examine whether some classes were particularly difficult to separate.

8.5.2 Results

Comparing dot and color images

Comparing the dot dataset to the color dataset (see section 7.4), it seems like the models trained on the dot dataset find reasonable parameters quicker and easier than their color dataset counterparts, while they are not able to beat the best color image networks. One possible reason for this could be that information apart from the dots contribute to identifying salmon. This way, the manually engineered features facilitate quick and easy extraction of the identifying power of the dots, while other information that might be useful for salmon identification is lost. A counterargument to this position is that the gray scale image included in the dot dataset is almost indistinguishable to the color image by human inspection, so most of the information should be retained in the dot dataset.

Another possible explanation is that the imageNet dataset (used for pretraining the Resnet models) is made of color images, so the pretrained parameters extract features on the assumption of color images. The dot dataset might not be large enough to train entire new dot feature extractors in the network, resulting in suboptimal network parameters. Furthermore, the network is tailored for color images with depth (pixel resolution) of 255, and binary masks require different network structure for optimal learning.

Performance of the individual salmon classification networks

The best network achieves very high test accuracy, and when using the network to generate videos annotated with salmon names, close to all frames with well placed keypoints have correctly labeled salmon individuals (see video in [2]). This is also the case for the tanks finetuned with a dataset of 500 instances per fish class. By looking at the wrongly classified fish for the best performing network (figure 7.6), it can be seen that the erroneous classifications happen when the torax image is deformed (displaying an unusual pose). Furthermore, the most similar fish individuals to the human eye are the ones the network misclassifies (such as Gehasi and Daniel, see figure 6.3), which suggests that the network is using the same classification criterions as humans, such as dot placement. This indicates that the network is in fact classifying salmon individuals, and not salmon behaviour, which was one of the worries presented in the previous section (section 8.4).

8.6 frequency analysis

8.6.1 Method

Parameter tuning

A significant number of parameters were used in this section (see section 6.6). They were introduced and tuned to achieve a model fit that resembled the fit a human would deem accurate (L, s, $ransacThresh$, $lineSearchThresh$, $ransacInitPerm$, η_0, $freqVal$), or to remove time series that did not have the necessary quality to give rise to a proper model fit ($minValClsf$, p, $maxFreqStd$, $maxFrameDiff$, $minInlCnt$).

Jaw gape metric

Three jaw gape metrics were suggested in the literature review, raw euclidean distance, warped euclidean distance and angular opening (see section 3.7.5). After reviewing time series created with each of these, the raw euclidean distance ended up being the jaw gape metric incorporated into the salmon breathing frequency pipeline, due to its robustness (see tradeoff of different mouth pose metrics in table 3.1). The high outlier number of time series using this metric, stemming from the difficulty of discarding jaw gape measurements based on the

relative placement of only two keypoints, was not prohibitively large. Furthermore, the short time series length made the scale and rotational variance of the euclidean distance metric a problem of little concern, and the loss of amplitude information was deemed acceptable as the extraction of this was not the main focus of the work done in this report.

Frequency extraction method

The reason for choosing a RANSAC loop approach to extract the salmon breathing frequency was briefly motivated in the literature review section (section 3.7.5), however the choice of fitting a sinusoidal function to the time series deserves some further attention. The valuable information in the time series is the period between two true signal peaks (the time between two consecutive max mouth gapes of the physical salmon), and as such a local peak detector might seem more intuitive than fitting an oscillatory function. This would also allow handling snapping salmon appropriately (see figure 7.8c). Two main problems, however, interfere with such an approach.

The first problem is the small location variability that is present in all keypoint predictions, causing the time of the true signal peak to differ from that of the highest recorded measurement, and removing a strict correlation between increased measurement value and increased true mouth opening. Hence, the peak detection is inaccurate (if using a simple max() operation to determine the peak position), and it is not possible to check the strength of the peak detection by asserting monotonic decrease of measurement values on both sides of the peak.

Secondly, the large amount of outliers in the time series, caused by completely wrong keypoint placements, cause a large amount of false peaks (peaks in the measurement signal that does not correspond to true peaks), a problem that is amplified by the difficulty of peak strength estimation discussed above.

Neither of these problems are insurmountable, however they make a straight forward peak detector implementation insufficient.

Fitting a function with the LM method inside a RANSAC loop effectively solves the issues stated above, however, as mentioned previously, cannot handle varying peak period. An improvement of the salmon breathing frequency pipeline would be to find a frequency extraction method that harmonize the positive effects of both peak detection and function fitting with RANSAC.

Time series splitting

Using the moving window approach to split time series reduces complexity and length of the time series, and introduces more redundancy in the salmon frequency extraction pipeline, which ensures that a few imprecise RANSAC runs do not have a major impact on the final frequency estimates of a video clip. This effect is further boosted by using the median as a central tendency metric, since this metric is unaffected by the value of outliers. Another strength of splitting time series is that trackers that jumps from one fish to another during a frame

sequence do not need to be discarded entirely. Only the time series that overlap with the jump must be removed, which is controlled by the p parameter (see section 6.6).

The main weakness of the moving window approach is that it requires more resources than processing the entire time series at once. Since the analysis time of the time series is orders of magnitude lower than that of the video processing pipeline segments (section 6.2 to 6.5), this problem is not significant.

Rejection of data with erroneous keypoint configurations

The rejection of clearly erroneous keypoint predictions by only considering the keypoints (i.e. not considering the salmon image) turned out to be a difficult task, as the pose and deformation of the fish caused a large range of realizable keypoint configurations. An alternative approach to facilitate early rejection of faulty time series datapoints is to construct a dataset of erroneous keypoint configurations, and train a neural network to automatically detect which mouth gapes or homogeneous transforms should be rejected.

8.6.2 results

The fit of the RANSAC loop worked very well, and gave results similar to what humans would propose without *a priori* problem knowledge (see figure 7.7). The main difficulty was to detect and remove time series which did not contain sufficient information to decide the ventilation frequency, as shown in figure 7.8. The parameters used in the $acceptable\_time\_series()$ function introduced in section 6.6 were able to remove some of the poorest time series, e.g. the $minInlCnt$ detected a significant number of noisy signals.

The signals capturing the salmon dynamics instead of the jaw gape (figure 7.8b) often had low frequency standard deviation and high inlier count, which made them difficult to deal with. Luckily, this error case was not very common, so frequency redundancy provided reasonable error robustness.

As seen from the last time series error case, figure 7.8c, salmon gape frequency is not always fixed. To handle this error case, the sine fitting approach could be changed to a peak detector, as discussed in the previous subsection (section 8.6.1), which would allow elucidation of all jaw pose periods. Due to the similarity between the signals of snapping fish and those of salmon swimming past obstacles, the detection of snapping fish would still not be trivial. To satisfactory detect snapping and long gapes in salmon, a more accurate keypoint prediction model should be constructed, so that erroneous jaw detections approach zero.

No numerical evaluation of the exact rate of erroneous model fits were performed, due to the difficulty of determining whether the time series oscillations that the fitted sine wave followed were caused by respiration or not. What is clear is that the error count is very low; as an example the complete time series set in appendix C has no errors.

Looking at the box plot in figure 7.9, the assumption of low error rate is strengthened. Apart from two outliers, all of the estimates of each fish class fall inside a range of 0.3 Hz.

8.7 Respiration frequency extraction pipeline applied to experimental data

8.7.1 Method

The steps performed in this part of the salmon ventilation frequency extraction method leverage the previously developed algorithms to extract ventilation frequency information from the data gathered during the salmon stress experiment (see chapter 5). Only one parameter, $minFreqCount$, was tuned in this section, introduced to ensure that all ventilation estimates in the full frequency evolutions had some redundancy in case of occurrence of the errors discussed in section 8.6.

Full frequency evolution

The full frequency evolutions provide a good overview of how the fish is changing its respiration frequency in response to disturbances and changing levels of DO.

One problem with these trajectories is that the time between two consecutive video samples vary, making it difficult to compare different downbreathings. The following issues complicated the task of synchronizing the downbreathings to each other, and to synchronize downbreathings to DO content.

1. Some clips were shorter than 11:47 minutes, due to (accidentally) changed camera settings. The last clip before a battery change was always of reduced length, due to the camera being manually stopped.

2. Some periods of the downbreathings were not filmed. This was the case during battery change, if the battery ran out of power before battery change, or if the camera malfunctioned and turned off during filming.

3. If any clips were shorter than 7000 frames, they were removed from consideration, due to the sampling strategy (extracting frame 5000 to 7000).

Some of these effects are possible to mitigate by manual inspection of all timelines during analysis, while the effects of battery change, battery running out of power, and camera malfunctioning cannot be dealt with post experiment. As such, perfect synchronization of the full frequency trajectories was not achieved.

Another issue with the full frequency evolutions is that a single video clip sometimes lacks frequency estimates for some salmon individuals, causing the full average frequency trajectory ($fullTrajAvg$) to be biased towards fish that are frequently detected.

Hourly frequency evolution

The issues with the full frequency evolutions motivated the construction of hourly frequency evolutions. These are robust against some of the synchronization issues that affects the full frequency trajectories, such as camera malfunctioning or changed camera settings. Furthermore, the issue of bias in the full average trajectory ($fullTrajAvg$) is improved by calculating hourly frequency evolutions for the fish individuals, and then averaging over those, as all salmon normally have a lot of frequency estimates in a six clip period. Hence, the average hourly frequency evolution ($hourlyTrajAvg$) provides a precise, unbiased evolution of the average respiration frequency of the shoal.

The hourly frequency evolutions are still not synchronized, hence the length of *one hour* (which separate the video clips into periods between battery changes) of recording can vary by a significant amount of time, making direct comparison between downbreathings inaccurate, even when looking at the hourly frequency evolutions.

The main downside of the hourly frequency trajectories compared to the full frequency trajectories is the reduced time resolution of the former. Looking at the full frequency trajectories (see appendix E), the average ventilation frequency rise by up to 0.2 Hz in an hour, which is a significant increase. This is not a large problem in the middle of the downbreathings, as hourly sampling here is sufficient to catch the main dynamics. At the time of maximum ventilation, however, the crude time resolution creates an erroneous perception of the peak ventilation frequency and the curvature of the ventilation frequency evolution. As an example, if the peak of the full frequency evolution is early in the hour of maximum ventilation frequency, and the ventilation frequency starts declining in this hour, the estimated hourly frequency will be quite similar in the hour of maximum ventilation frequency, and the hour just before, causing a seemingly concave hourly frequency evolution, even if this does not match the true respiration frequency shape. If the downbreathing lasts only four hours, this datapoint significantly disturbs the hourly frequency evolution shape.

By removing the first clip each hour when calculating the hourly frequency evolutions, the influence of camera insertion disturbance on these trajectories is greatly diminished, ensuring the hourly frequency evolutions mainly elucidate the effects of varying temperature and DO content on salmon respiration frequency.

When calculating the average hourly frequency evolutions ($hourlyTrajAvg$), the calculation pipeline began with taking the intrafish median, followed by the interfish average, instead of the other way around (first take the intraclip average over all fish individuals, and then averaging over the averages to acquire the hourly average). This was to ensure a large pool of data points for each fish, to avoid bias due to missing fish individuals in some clips. The reason for using the median when finding a frequency central tendency for one fish was to increase outlier robustness, while the average was used when finding central frequency tendency over the fish individuals, as each fish should have the same influence on the average trajectory.

Correlation between DO content and ventilation frequencies

As discussed in the previous numerated list, under the full frequency evolution discussion, the exact timeline of the video recordings are not known, and can not be perfectly synchronized with the oxygen data. In order to still be able to compare the evolutions of salmon ventilation frequency and oxygen data, each video clip was assumed to take twelve minutes, implicitly saying that the cumulative effect of the synchronization issues resulted in 13 seconds (12:00 minus 11:47) of added time per video clip. This is a significant simplification, and must be kept in mind when comparing the full ventilation frequencies with the DO content.

The scaling approach when overlaying DO content and ventilation frequency was to map the first hourly frequency values to the highest DO contents (lowest negated DO content), because no significant positive ventilation frequency trends were observed during the first downbreathing hours. Hence, the hourly average frequency trajectories provide a robust, largely unbiased resting frequency estimate. For the maximum ventilation frequencies, however, there were a major difference between the hourly maximum and clip maximum ventilation frequencies (see table 7.1). Due to this, the maximum value of the full frequency evolutions were used when aligning the maximum ventilation frequency value to the lowest (highest negated) DO content. Scaling was only applied in the y (value) direction, and not in the x (time) direction.

Since the plots comparing DO saturation to ventilation frequency (figure 7.15) looked disorganized when the video clips with camera disturbance were included, and disturbance peaks are not interesting when comparing ventilation frequency and DO, these clips were removed from the plots.

Respiration frequency increase after disturbance

When calculating the respiration frequency increase after disturbance, single clip frequency estimates (instead of hourly estimates) had to be used, as the disturbance induced respiration increase is present in only one clip (each hour). This increases the uncertainty of the results, primarily due to some fish missing from some clips, which results in average frequency ($fullTrajAvg$) bias, as discussed above.

Fish personality

The fish personality could be calculated from both the full and the hourly frequency trajectories. Which of the two evolution types used did not have a large effect on the results. Since the hourly frequency trajectories had more robust individual estimates, this was the frequency evolution used in the final breathing frequency pipeline.

8.7.2 Results

Full frequency trajectories

In appendix E, all full frequency trajectories are displayed. The vast majority of frequency estimates are reasonably placed, however occasionally, large jumps in the ventilation frequency estimates for individual fish in two consecutive clips occur.

Looking at the tank nine ventilation frequency estimates, this is the case for Caiphas during clip five of the second downbreathing, recorded with camera seven. For this fish and clip, the other camera (camera eight during downbreathing two) has a frequency data point of Caiphas that is more consistent with the temporally proximal estimations, hence the camera seven estimate is likely erroneous. The reason for this error could be because the number of Caiphas ventilation frequency data points from camera seven during clip five are few, and these few frequency estimates exhibit one of the errors presented in figure 7.8.

Some seemingly erroneous frequency data points, like the frequency estimate of Daniel in the eight clip during downbreathing one in tank nine, is relatively similar for both cameras. This suggests that the frequency estimate is correct, while the fish, for some reason, had a large change in ventilation frequency at that clip.

Occasionally, the full salmon trajectories contain periods without salmon ventilation frequency estimates. This is the case for e.g. tank 5, camera 6, during the first hour of downbreathing one (see appendix E), where a salmon swam into the camera, knocking it out of place. Other periods of missing frequency data points have similar explanations.

Apart from the occasional large jumps in the frequency estimates (dealt with above) and the clips of camera insertion (which will be dealt with later), the difference between two consecutive frequency estimates is staying between zero and 0.3 Hz, which is consistent with the variability seen in the box plot of frequencies extracted from a video clip (figure 7.9). It is not unreasonable to assign this recorded frequency variation to short term ventilation frequency variability in the physical salmon.

The full frequency trajectories display ventilation frequencies ranging from 1 Hz for the slowest breathing resting fish, to 2 Hz for the fastest breathing salmon exposed to camera disturbance. This is correlating well with the table displaying common salmon ventilation frequencies (table 2.1) from the literature review section about fish ventilation (section 2.2).

Hourly ventilation frequency

Looking at the hourly ventilation frequencies of tank nine for individual salmon (figures 7.11 and 7.12), the trajectory shape for the different fish individuals is more or less the same, while the offset of the trajectory is changing. As such, averaging the hourly frequency trajectories over individual salmon is a good way

to make the data more concise, yielding an average hourly frequency trajectory ($hourlyTrajAvg$).

Most of the downbreathings have very similar trajectories ($hourlyTrajAvg$) for the right and left corner cameras (see figures 7.13 and 7.14). This is suggesting that the salmon breathing frequency pipeline is accurately identifying the ventilation frequency of the fish. The discrepancy between the two cameras have two potential sources, errors of the breathing frequency pipeline, or changing physical salmon breathing frequency as the fish swims between the field of view of the two cameras. To separate these are difficult, as both the physical salmon, and the frequency extraction pipeline, can display very different breathing frequencies in the same video clip. For the physical salmon, such a large frequency change can happen if it e.g. snaps.

The reason behind some of the large (above 0.1 Hz) camera discrepancies, however, are identifiable. Firstly, tank six during downbreathing one (see figures 7.13 and 7.14), has a large discrepancy between the two camera recordings, which seems to be caused by one being temporally skewed compared to the other. As the time of the two downbreathings are not synchronized, the camera discrepancy might be an error (a temporal mismatch) in the experimental data, and not a fault in the frequency extraction pipeline.

Secondly, some of the video clips had reduced quality, which could come from air stones in the camera field of view, or cameras pointing in an unfortunate direction. These poor recordings generally had detections of only a few fish, which led ta a bias in the hourly average frequency trajectories ($hourlyTrajAvg$). This was the case for tank eight during the first downbreathing, where one of the cameras (camera seven) had air bubbles in front of it during the final hour (hour 4).

In the hourly frequency trajectories, the reduced time resolution induced by a per-hour averaging decreases the precision of the peak value. This is more pronounced for higher temperatures, as these trajectories have a higher ventilation frequency rate of increase. As an example, the full frequency trajectories of the first tank nine downbreathing reaches approximately 1.8 Hz, while the hourly frequency trajectories only reach a bit above 1.7 Hz. Since this is more pronounced for high temperatures, this cause a slight bias in the results based on the hourly frequency trajectories.

Fish personality

As mentioned in section 6.7, the fish personality is describing whether a fish individual is breathing faster or slower than the shoal average. This is visualized in figure 7.11 and 7.12, where the different fish have ventilation frequency trajectories with different offsets from the average trajectory.

In figure 7.17, all these fish personalities are presented in a histogram, which shows that the personalities are more or less Gaussian distributed around the average frequency trajectory with a standard deviation (σ) of 0.08. This means that roughly two fish in each tank ($\frac{2}{7} = 29\%$) will have a breathing frequency that is, on average, more than 0.08 Hz away from the shoal average frequency

(68 % of the data falls in the range $\mu^{+}_{-}\sigma$[214]).

By looking at the figure comparing fish personality across downbreathings (figure 7.18), it can be seen that the fish order is consistent across downbreathings to some extent. There are, however, still a significant amount of ranking change, which can be seen from the text at the the end of the bars.

Respiration frequency increase after disturbance

By looking at the full frequency trajectories (appendix E), it can be seen that salmon elevate their respiration frequency when exposed to disturbances from cameras inserted into their tanks (the grey vertical lines represent time of camera insertion). This disturbance effect is primarily visible in the clip in which the camera was inserted, putting the disturbance between zero and 83 seconds (5000 frames) prior to the observed ventilation increase, depending on the exact time of camera insertion. The variability in the stop time of camera movement might explain some of the variance in the observed salmon ventilation frequency disturbance responses.

Figure 7.19 displays a histogram of the respiration frequency increases in the first video clip each hour, showing that almost all of the camera disturbances induce an increase in the salmon respiration frequency, with an average increase of over 0.2 Hz.

It was not found any clear trend between temperature and disturbance response, while the time since the beginning of the downbreathings had a significant correlation with the size of the disturbance response. Figure 7.20 displays a constant disturbance driven ventilation increase for hour two, three and four into the downbreathings, before the ventilation jump decreases to a smaller, constant value for the next three hours. Three possible biological mechanisms could explain this.

1. As the DO content decreases, the metabolic scope decreases as well, since less oxygen is available for metabolic activity. If the fish have a maximum ventilation frequency that they will not surpass, higher baseline ventilation frequency necessarily reduce the room for ventilation frequency increase in response to stressors. One full downbreathing trajectory that supports this view is the first tank nine downbreathing (figure 7.11), where the ventilation frequencies during disturbance recordings (the first clip each hour) are more or less constant, irrespective of the DO content.

2. During the last hours of a downbreathing, the DO content is often increasing, hence the fish are *happy* with its improved millieux. This might make them less susceptible to stressors, causing them to more or less ignore the camera disturbances.

3. In the fifth, sixth and seventh hour of the downbreathings, the fish have already been exposed to camera disturbances at least four times in the immediate past. This might cause the fish to learn that the camera is not dangerous, reducing the disturbance effect of inserting it.

DO

By looking into the hourly downbreathings, it is clear that the fish ventilate with low ventilation frequency during the initial high DO period, and that the ventilation frequency increases as the DO content decreases. This is quantitatively asserted by the negative linear correlations (r values) between the DO curves and full ventilation frequency trajectories during the downbreathings, as displayed in figure 7.15 and the figures of appendix F. When these two curves are more or less temporally aligned, the negative correlation is statistically significant, while the p value is above 5% for some of the curves with large temporal difference between the negated DO content peak and the ventilation frequency peak. The poor curve fits can most likely be attributed to the synchronization difficulties described above (subsection 8.7.1). As the peaks of the two curves often align well, it is reasonable to assume the peak of the downbreathing respiration frequencies to be caused by the release of oxygen back into the tanks.

The ventilation frequency decrease during the tank upbreathings might also be initialized because the critical DO content (p_{crit}) is reached (see section 2.5.1), and the fish become oxyconformers. In [85], p_{crit} (termed LOS in [85]) is estimated to around 39% for 12 $^{\circ}C$ and 55% for 18 $^{\circ}C$, which suggests that the experimental DO content (see chapter 5) has a maximum p_{crit} violation of up to 5 %. This would corresponds to around 30 minutes, or three video clips.

As p_{crit} and the bottom of the DO content curve is happening at approximately the same time, it is likely that the ventilation peaks of the full frequency evolutions are a consequence of the minimum DO content, as the violation of p_{crit} is not large enough to cause a significant ventilation frequency decrease on its own. The effects related to p_{crit} might instead have some influence on the rate of ventilation increase and decay on the two sides of the ventilation frequency peak caused by the minimum DO content.

In appendix F, all average full frequency evolutions are plotted together with the DO content. It is clear that the peaks of the full frequency evolutions are relatively temporally aligned with the minimum DO content, and that the ventilation frequency have an equal or more convex shape than the negated DO curve. This shows that the ventilation frequency rise faster than the DO content decays, which is consistent with a stress effect in the salmon (see section 2.5.1).

In the literature review (section 2.5.1), a sigmoid salmon ventilation response were suggested, with the concave shape at low DO saturations coming from salmon becoming oxyconformers. This respiration frequency concavity close to the ventilation frequnecy peak was not clearly observed in the figures of appendix F, which might be due to the DO content staying mostly above p_{crit}. The hourly frequency trajectories in figures 7.13 and 7.14 shows a more sigmoid shape, but this is mainly due to the averaging effects discussed previously. It does not seem like temperatures have any effect on the shape of the hourly salmon trajectories.

Temperature

In [85], the salmon ventilation frequency ranges from 1.1 Hz to 1.7 Hz for a temperature of 12 $^{\circ}C$, and 1.6 Hz to 2.05 Hz for a temperature of 18 $^{\circ}C$, as the DO changes from full saturation to p_{crit}. The breathing frequency data from this report (table 7.1), displays salmon respiration frequencies of approximately the same values (see table 7.1), however the distribution of the data of this report is not as coherent as the one in [85] (see e.g. the uncertainty of the elements of table 7.1).

Table 7.1 shows the relationship between temperatures and downbreathing extreme values, averaged over downbreathings holding the same temperatures. For the six coldest tanks (tanks one to six, which hold temperatures of 12.3 $^{\circ}C$ to 15.6 $^{\circ}C$), the hourly frequency trajectories averaged over downbreathing temperatures starts in the range [1.24-1.34] Hz, and peaks in the range [1.58-1.64] Hz (see table 7.1). The maximum clip frequencies goes a bit higher than the hourly maximum frequencies (as explained earlier in this subsection), up to the range [1.69-1.77] Hz. In these tanks (one to six), the only significant ventilation frequency difference was found between the 14 $^{\circ}C$ downbreathings and the 12.3 $^{\circ}C$ and 15.6 $^{\circ}C$ downbreathings (see figure 7.16). Since the 14.0 $^{\circ}C$ downbreathings show the highest ventilation frequencies, higher than both colder and warmer tanks, it is unlikely that temperature is the driver of the ventilation difference in the cold and intermediate temperature tanks.

For the three warmest tanks (with temperatures of 15.9 $^{\circ}C$ and 17.7 $^{\circ}C$), two differences to the rest of the tanks can be observed. Firstly, the resting ventilatory frequency is around 0.1 Hz higher at the warmer tanks than in the cold and medium temperature tanks. This difference is significant for the ventilation frequencies at the downbreathings with temperatures 12.3 $^{\circ}C$, 13.4 $^{\circ}C$ and 15.6 $^{\circ}C$, while the downbreathings with temperature of 14.0 $^{\circ}C$ have p values between seven and eight percent. A resting ventilation increase in response to elevated temperature is coherent with the findings in section 2.5.2.

Secondly, the average maximum ventilation frequencies at the 15.9 $^{\circ}C$ downbreathings are over 0.07 Hz higher than the rest of the tanks, while the average maximum ventilation frequencies at 17.7 $^{\circ}C$ are over 0.05 Hz lower than the rest of the tanks (see table 7.1). These effects are significant for the following cases.

1. The maximum hourly ventilation frequency at the 15.9 $^{\circ}C$ downbreathings is statistically higher than the ventilation frequencies for the downbreathings with temperatures of 13.4 $^{\circ}C$, 15.6 $^{\circ}C$.

2. The maximum clip frequency at the 15.9 $^{\circ}C$ downbreathings is statistically higher than all other tanks.

3. The maximum hourly frequency at the 17.7 $^{\circ}C$ downbreathings is statistically lower than the ventilation frequencies for the downbreathings with temperature of 14.0 $^{\circ}C$.

4. The maximum clip frequency at the 17.7 $°C$ downbreathings is statistically lower than the downbreathings with temperatures of 14.0 $°C$ and 15.6 $°C$.

5. All maximum ventilation frequencies for the downbreathings at temperatures 15.9 $°C$ and 17.7 $°C$ are statistically different from from each other.

The ventilation frequency increase at the 15.9 $°C$ downbreathings is expected, as the literature review (see section 2.5) uncovered that higher temperatures cause higher ventilation frequencies in fish. As such, the results at 17.7 $°C$ stands out as an unexpected datapoint.

One possible reason for the low maximum ventilation of the 17.7 $°C$ downbreathings is that the DO content at the colder tanks reached a lower concentration (see table 5.1), hence the difference in peak ventilation frequency might be caused by different DO levels, and not the temperature per se. A weaknesses with this explanation is that the full frequency evolutions of the warmest tanks, (see e.g. tank seven during the second downbreathing in appendix E) are almost completely flat, while the first downbreathing trajectories of tank seven show a steady increase, suggesting other mechanisms than the final (minimum) DO content at play.

One alternative explanation is that the high temperature effects discussed in section 2.5 starts becoming significant at 17.7 $°C$, i.e. that the metabolic scope is reduced, and the rate of oxyconformers increase, at this temperature.

The information in chapter 5 provide additional information about the influence of temperature on salmon. In table 5.1, it is observed that higher temperatures correlate with quicker downbreathings, with the difference between tanks one to six (DB time between 300 and 371 minutes), and tanks seven to nine (DB time between 204 and 231 minutes), being the clearest change of downbreathing length. The reason for the quicker downbreathing time of tank seven to nine is twofold, including both a quicker rate of DO content decrease, and a higher minimum DO content before reintroducing oxygen to the tanks. The quicker rate of DO content decrease suggests tanks seven to nine contain salmon with high metabolic rate, which is coherent with the ventilation frequency data from the downbreathings holding 15.9 $°C$, but not from those holding 17.7 $°C$.

This suggests a third explanation of the low ventilation frequency at the 17.7 $°C$ downbreathings. If both the measurement of the downbreathing time and the salmon breathing frequency pipeline results are correct, the ventilation frequency of the salmon is low, while its metabolic rate is high, which is only possible if other effects than ventilation frequency boosts the metabolic rate of the fish. Increased oxygen uptake to sustain this metabolic rate could come from increased ventilation amplitude, or a switch to ram ventilation (see section 2.2)

Stress and welfare

There are three clear welfare related effects that the salmon frequency extraction pipeline is able to capture.

Firstly, the ventilation frequency increase after camera disturbance was clearly evident in the experimental results (see figures 7.19 and 7.20). This relates directly to acute stress, and subsequently welfare (see section 2.6).

Secondly, the shape of the ventilation evolutions were slightly convex, advocating for a DO mediated stress response (see section 2.5.1). Even if the ventilation increase in response to DO decrease would be linear, which cannot be dismissed from the results of this report, decreasing DO content is still associated with reduced salmon welfare (see e.g. [9]).

Thirdly, the respiration frequency of individual salmon compared to the average shoal ventilation frequency was discernible in the hourly salmon evolutions, and these differences were consistent between downbreathings. This can be used to classify the personality of individual salmon (proactive or reactive, see section 2.3.4), which relates to the welfare of this particular individual.

When it comes to the final variable in the salmon stress experiment, temperature, no unequivocal links between water temperature and ventilation frequency were found, however some weak trends were observed. The two most interesting are a resting ventilation frequency increase between the six coldest and three warmest tanks, and a possible change in salmon oxygen uptake strategy in response to reduced DO when the temperature changes. Under the current salmon breathing frequency pipeline, the only detectable temperature effect was the elevated resting ventilation increase.

Resource requirements

As mentioned in section 7.6, analysing one downbreathing (ten to 25 minutes of video in total) took several days, making the salmon ventilation frequency pipeline developed in this report inapplicable to real time applications in its current form.

During development of the salmon respiration frequency method, no effort was spent on improving the efficiency of the method. As such, there are options for making it more effective, the most obvious being GPU accelerated neural network evaluation during data analysis. Due to limited GPU availability, only the neural network training used GPU acceleration in this report.

Chapter 9

Conclusion

In this thesis, a salmon respiration frequency deep learning workflow has been developed, successfully estimating ventilation frequency on identified salmon individuals. The three main technical contributions of the thesis are stated below, and to the authors knowledge, none of them have been presented in previous publications.

1. A method for estimating salmon ventilation frequency by tracking keypoints attached to salmon jaws.
2. A large-scale salmon identification dataset counting 118492 semi-manually annotated images over 126 classes (originating from both sides of 63 fish that were equally distributed over 9 different tanks).
3. A method for salmon re-identification using underwater fish recordings.

Applying this breathing frequency pipeline to data from a salmon stress experiment yielded the following results.

1. Salmon ventilation increases with decreasing dissolved oxygen (DO) content, at least for temperatures between 12 and 18 degrees Celsius, and DO content above p_{crit}. The curve describing ventilation frequency as a function of negated DO content is close to linear, with a slightly convex shape.
2. Salmon increase their ventilation frequency in response to acute disturbances in the form of camera inserted into their tanks.
3. If fish individuals are ranked by their ventilation frequency, the ranking order tend to be constant over both hours and weeks, i.e. the fish breathing fastest at the start of the downbreathing is also the fish breathing fastest at the end of the downbreathing, and this fish is often the same in downbreathing one and two.

4. There might be a positive correlation between temperature and resting ventilation frequency.

From this it can be concluded that an automated pipeline for extracting respiration frequency has been successfully constructed, and that this pipeline can be used to detect acute disturbance stress which lower the salmon welfare. The method is also capable of detecting decreased DO content in the water by an increase of salmon ventilation frequency, which can be linked to decreased salmon welfare. On individual level, the algorithm can detect the ventilation frequency of a single fish in relationship to the larger shoal, which can inform on the personality type of the fish (proactive or reactive), which is related to its individual welfare.

Chapter 10

Future directions

Every part of the complete breathing frequency extraction pipeline could be improved on, in order to create even more reliable frequency estimates. Possible improvements includes a better Keypoint RCNN model that handles occlusion and frame border detections more robustly, adding more keypoints to the model in order to improve the salmon transformations, finding a way to reject more of the faulty jaw gape time series and to create a method to detect variable breathing frequency periods.

Considering the good performance of the algorithm, the most interesting future work is to extend the method to industrial applications. This would mainly require a speed up of the processing pipeline, and a zero-shot model that can re-identify salmon not seen before model deployment.

Another avenue to pursue is the continued analysis of the experimental data, with the most pressing improvements being to synchronize the data to allow clip-by-clip comparisons across different downbreathings, and analysis of a larger portion of the gathered video recordings. Also linking the result of the ventilation frequency analysis to other data from the same experiment would be interesting. This could be to estimate whether fish personality (consistent high or low ventilation frequency) can be linked to hierarchical positions, scale loss or other welfare indicators, as the literature review would suggest.

Furthermore, the temperature effects on salmon ventilation would be interesting to examine further. Some of the trends that were observed, but could not be ascertained, might show themselves more definite in a new experiment with more temperature datapoints, and larger temperature differences. This could inform about whether higher temperatures change the salmon oxygen uptake strategy.

Bibliography

[1] Espen Høgstedt. *Using deep learning to evaluate salmon welfare by respiration frequency [Unpublished]*. June 2022.

[2] Espen Høgstedt. *Code to develop a deep learning pipeline for automated salmon welfare analysis by respiration frequency*. 2023. URL: https://github.com/espenbh/salmon_welfare_analysis_by_automated_respiration_frequency_extraction.

[3] Tony Pitcher. "Shoaling and Schooling in Fishes". In: June 1998, pp. 748–760.

[4] Jillian P Fry et al. "Feed conversion efficiency in aquaculture: do we measure it correctly?" In: *Environmental Research Letters* 13.2 (Feb. 2018), p. 024017. DOI: 10.1088/1748-9326/aaa273. URL: https://dx.doi.org/10.1088/1748-9326/aaa273.

[5] Ole Torrissen et al. "Atlantic Salmon (Salmo salar): The "Super-Chicken" of the Sea?" In: *Reviews in Fisheries Science* 19 (July 2011), pp. 257–278. DOI: 10.1080/10641262.2011.597890.

[6] O Torrissen et al. "Salmon lice – impact on wild salmonids and salmon aquaculture". In: *Journal of Fish Diseases* 36.3 (2013), pp. 171–194. DOI: https://doi.org/10.1111/jfd.12061. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jfd.12061. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jfd.12061.

[7] Heidi Moe Føre and Trine Thorvaldsen. "Causal analysis of escape of Atlantic salmon and rainbow trout from Norwegian fish farms during 2010–2018". In: *Aquaculture* 532 (2021), p. 736002. ISSN: 0044-8486. DOI: https://doi.org/10.1016/j.aquaculture.2020.736002. URL: https://www.sciencedirect.com/science/article/pii/S0044848620315684.

[8] FRANK ASCHE et al. "The Salmon Disease Crisis in Chile". In: *Marine Resource Economics* 24.4 (2009), pp. 405–411. DOI: 10.1086/mre.24.4.42629664. eprint: https://doi.org/10.1086/mre.24.4.42629664. URL: https://doi.org/10.1086/mre.24.4.42629664.

[9] Lars Stien 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”. In: *Reviews in Aquaculture* 5 (Mar. 2013), pp. 33–57. DOI: 10.1111/j.1753-5131.2012.01083.x.

[10] Jostein Pettersen et al. “Salmon welfare index model 2.0: An extended model for overall welfare assessment of caged Atlantic salmon, based on a review of selected welfare indicators and intended for fish health professionals”. In: *Reviews in Aquaculture* 6 (June 2013). DOI: 10.1111/raq.12039.

[11] Renato A. Quiñones et al. “Environmental issues in Chilean salmon farming: a review”. In: *Reviews in Aquaculture* 11.2 (2019), pp. 375–402. DOI: https://doi.org/10.1111/raq.12337. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/raq.12337.

[12] Wendelaar Bonga SE. “The stress response in fish”. In: *Physiol Rev.* 77.3 (1997), pp. 591–625. DOI: 10.1152/physrev.1997.77.3.591.

[13] T. Ellis et al. “A non-invasive stress assay based upon measurement of free cortisol released into the water by rainbow trout”. In: *Journal of Fish Biology* 65.5 (2004), pp. 1233–1252. DOI: https://doi.org/10.1111/j.0022-1112.2004.00499.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0022-1112.2004.00499.x.

[14] Carl Schreck and Lluis Tort. “The Concept of Stress in Fish”. In: vol. 35. Dec. 2016, pp. 1–34. ISBN: 9780128027288. DOI: 10.1016/B978-0-12-802728-8.00001-1.

[15] Ziyi Liu et al. “Measuring feeding activity of fish in RAS using computer vision”. In: *Aquacultural Engineering* 60 (2014), pp. 20–27. ISSN: 0144-8609. DOI: https://doi.org/10.1016/j.aquaeng.2014.03.005. URL: https://www.sciencedirect.com/science/article/pii/S0144860914000211.

[16] Dorith Israeli and Eitan Kimmel. “Monitoring the behavior of hypoxia-stressed Carassius auratus using computer vision”. In: *Aquacultural Engineering* 15.6 (1996), pp. 423–440. ISSN: 0144-8609. DOI: https://doi.org/10.1016/S0144-8609(96)01009-6. URL: https://www.sciencedirect.com/science/article/pii/S0144860996010096.

[17] T.H. Pinkiewicz, G.J. Purser, and R.N. Williams. “A computer vision system to analyse the swimming behaviour of farmed fish in commercial aquaculture facilities: A case study using cage-held Atlantic salmon”. In: *Aquacultural Engineering* 45.1 (2011), pp. 20–27. ISSN: 0144-8609. DOI: https://doi.org/10.1016/j.aquaeng.2011.05.002. URL: https://www.sciencedirect.com/science/article/pii/S014486091100029X.

[18] Boaz Zion. "The use of computer vision technologies in aquaculture – A review". In: *Computers and Electronics in Agriculture* 88 (2012), pp. 125–132. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2012.07.010. URL: https://www.sciencedirect.com/science/article/pii/S0168169912001950.

[19] Per Pethon and Bente Olesen Nyström. *Aschehougs store fiskebok: Norges fisker i farger*. Aschehoug, 1998.

[20] David Eidsvoll. *NIVA Research Facility Solbergstrand*. [Online; accessed May 10, 2023]. 2023. URL: https://www.niva.no/en/contact/solbergstrand-research-facility.

[21] Elise Thovsland Sannes and Tosca B. L. Koers. *Layout of experimental setup*. 2023.

[22] *Google Scholar*. 2022. URL: https://scholar.google.com/.

[23] *Science Direct*. 2022. URL: https://www.sciencedirect.com.

[24] David Randall. "THE CONTROL OF RESPIRATION AND CIRCULATION IN FISH DURING EXERCISE AND HYPOXIA". In: *J. exp. Biol* 100.1 (1982), pp. 275–288. DOI: 10.1242/jeb.100.1.275.

[25] John Fleng Steffensen. "The Transition between Branchial Pumping and Ram Ventilation in Fishes: Energetic Consequences and Dependence on Water Oxygen Tension". In: *Journal of Experimental Biology* 114.1 (Jan. 1985), pp. 141–150. ISSN: 0022-0949. DOI: 10.1242/jeb.114.1.141.

[26] Michael L. Kelly et al. "Behavioural sleep in two species of buccal pumping sharks (Heterodontus portusjacksoni and Cephaloscyllium isabellum)". In: *Journal of Sleep Research* 30.3 (2021), e13139. DOI: https://doi.org/10.1111/jsr.13139. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jsr.13139. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jsr.13139.

[27] E.W. Taylor. "CONTROL OF RESPIRATION — Generation of the Respiratory Rhythm in Fish". In: (2011). Ed. by Anthony P. Farrell, pp. 854–864. DOI: https://doi.org/10.1016/B978-0-12-374553-8.00250-1.

[28] Kathleen M Gilmour. "The CO2/pH ventilatory drive in fish". In: *Comparative Biochemistry and Physiology Part A: Molecular Integrative Physiology* 130.2 (2001), pp. 219–240. ISSN: 1095-6433. DOI: https://doi.org/10.1016/S1095-6433(01)00391-9.

[29] William K. Milsom. "New insights into gill chemoreception: Receptor distribution and roles in water and air breathing fish". In: *Respiratory Physiology Neurobiology* 184.3 (2012). New Insights into Structure/Function Relationships in Fish Gills, pp. 326–339. ISSN: 1569-9048. DOI: https://doi.org/10.1016/j.resp.2012.07.013.

[30] Chris M. Wood and R. S. Munger. "CARBONIC ANHYDRASE INJECTION PROVIDES EVIDENCE FOR THE ROLE OF BLOOD ACID–BASE STATUS IN STIMULATING VENTILATION AFTER EXHAUSTIVE EXERCISE IN RAINBOW TROUT". In: *J. exp. Biol.* 194 (1994), pp. 225–253. DOI: 10.1242/jeb.194.1.225.

[31] S. F. Perry and J. E. McKendry. "The relative roles of external and internal CO2versus H+ in eliciting the cardiorespiratory responses of Salmo salar and Squalus acanthias to hypercarbia". In: *Journal of Experimental Biology* 204.22 (Nov. 2001), pp. 3963–3971. ISSN: 0022-0949. DOI: 10.1242/jeb.204.22.3963. eprint: https://journals.biologists.com/jeb/article-pdf/204/22/3963/1239471/3963.pdf.

[32] Camilla Diesen Hosfeld et al. "Physiological effects of normbaric environmental hyperoxia on Atlantic salmon (Salmo salar L.) presmolts". In: *Aquaculture* 308.1 (2010), pp. 28–33. ISSN: 0044-8486. DOI: https://doi.org/10.1016/j.aquaculture.2010.08.003. URL: https://www.sciencedirect.com/science/article/pii/S0044848610005053.

[33] M.B. Knoph. "Gill ventilation frequency and mortality of Atlantic salmon (Salmo salar L.) exposed to high ammonia levels in seawater". In: *Water Research* 30.4 (1996), pp. 837–842. ISSN: 0043-1354. DOI: https://doi.org/10.1016/0043-1354(95)00233-2. URL: https://www.sciencedirect.com/science/article/pii/0043135495002332.

[34] K. J. Millidine, N. B. Metcalfe, and J. D. Armstrong. "The use of ventilation frequency as an accurate indicator of metabolic rate in juvenile Atlantic salmon (Salmo salar)". In: *Canadian Journal of Fisheries and Aquatic Sciences* 65.10 (2008), pp. 2081–2087. DOI: 10.1139/F08-118. eprint: https://doi.org/10.1139/F08-118. URL: https://doi.org/10.1139/F08-118.

[35] Ulf Erikson et al. "Live transport of Atlantic salmon in open and closed systems: Water quality, stress and recovery". In: *Aquaculture Research* 53.11 (2022), pp. 3913–3926. DOI: https://doi.org/10.1111/are.15895. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/are.15895. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/are.15895.

[36] Ulf Erikson et al. "Crowding of Atlantic salmon in net-pen before slaughter". In: *Aquaculture* 465 (2016), pp. 395–400. ISSN: 0044-8486. DOI: https://doi.org/10.1016/j.aquaculture.2016.09.018.

[37] Ken M. Jeffries et al. "Consequences of high temperatures and premature mortality on the transcriptome and blood physiology of wild adult sockeye salmon (Oncorhynchus nerka)". In: *Ecology and Evolution* 2.7 (2012), pp. 1747–1764. DOI: https://doi.org/10.1002/ece3.274. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.274. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.274.

[38] Nicholas J. Bernier and Paul M. Craig. "CRF-related peptides contribute to stress response and regulation of appetite in hypoxic rainbow trout". In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 289.4 (2005). PMID: 15932968, R982–R990. DOI: 10.1152/ajpregu.00668.2004. eprint: https://doi.org/10.1152/ajpregu.00668.2004. URL: https://doi.org/10.1152/ajpregu.00668.2004.

[39] Bruce A. Barton. "Stress in fishes: a diversity of responses with particular reference to changes in circulating corticosteroids". In: *Integr Comp Biol.* 42.3 (2002), pp. 517–25. DOI: 10.1093/icb/42.3.517.

[40] Stephen G. Reid and Steve F. Perry. "Peripheral O2 chemoreceptors mediate humoral catecholamine secretion from fish chromaffin cells". In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 284.4 (2003). PMID: 12511426, R990–R999. DOI: 10.1152/ajpregu.00412.2002.

[41] Davison W, Axelsson M, and Nilsson S. Forster M. "Cardiovascular responses to acute handling stress in the Antarctic fish Trematomus bernacchii are not mediated by circulatory catecholamines". In: *Fish Physiol Biochem.* 14.3 (1995), pp. 253–7. DOI: 10.1007/BF00004316.

[42] A. P. Farrell. "Cardiovascular changes in the unanaesthetized lingcod (Ophiodon elongatus) during short-term, progressive hypoxia and spontaneous activity". In: *Canadian Journal of Zoology* 60.5 (1982), pp. 933–941. DOI: 10.1139/z82-127. eprint: https://doi.org/10.1139/z82-127.

[43] Andreas Ekström, Fredrik Jutfelt, and Erik Sandblom. "Effects of autonomic blockade on acute thermal tolerance and cardioventilatory performance in rainbow trout, Oncorhynchus mykiss". In: *Journal of Thermal Biology* 44 (2014), pp. 47–54. ISSN: 0306-4565. DOI: https://doi.org/10.1016/j.jtherbio.2014.06.002.

[44] A. P. Farrell, K. R. MacLeod, and B. Chancey. "Intrinsic mechanical properties of the perfused rainbow trout heart and the effects of catecholamines and extracellular calcium under control and acidotic conditions". In: *Journal of Experimental Biology* 125.1 (Sept. 1986), pp. 319–345. ISSN: 0022-0949. DOI: 10.1242/jeb.125.1.319. eprint: https://journals.biologists.com/jeb/article-pdf/125/1/319/1226130/319.pdf.

[45] Richard S. Farrar and Kenneth J. Rodnick. "Sex-dependent effects of gonadal steroids and cortisol on cardiac contractility in rainbow trout". In: *Journal of Experimental Biology* 207.12 (May 2004), pp. 2083–2093. ISSN: 0022-0949. DOI: 10.1242/jeb.00996. eprint: https://journals.biologists.com/jeb/article-pdf/207/12/2083/1248388/2083.pdf.

[46] J. Isaia, J. P. Girard, and P. Payan. "Kinetic study of gill epithelium permeability to water diffusion in the fresh water trout,Salmo gairdneri: Effect of adrenaline". In: *The Journal of Membrane Biology* 41.4 (1978), pp. 337–347. DOI: 10.1007/BF01871998.

[47] Mikko Nikinmaa. "Membrane Transport and Control of Hemoglobin-Oxygen Affinity in Nucleated Erythrocytes". In: *The American Physiological Society* 72.2 (1992), pp. 301–321. DOI: 10.1152/physrev.1992.72.2.301.

[48] José A. Zadunaisky. "Chloride cells and osmoregulation". In: *Kidney International* 49.6 (1996), pp. 1563–1567. ISSN: 0085-2538. DOI: https://doi.org/10.1038/ki.1996.225.

[49] Pierre Laurent. "Effects of cortisol on gill chloride cell morphology and ionic uptake in the freshwater trout,Salmo gairdneri". In: *Cell and Tissue Research* 259.3 (1990), pp. 429–442. DOI: 10.1007/BF01740769.

[50] Jean-Louis Gallis, Pierre Lasserre, and Francis Belloc. "Freshwater adaptation in the euryhaline teleost, Chelon labrosus: I. Effects of adaptation, prolactin, cortisol and actinomycin D on plasma osmotic balance and (Na+K+)ATPase in gill and kidney". In: *General and Comparative Endocrinology* 38.1 (1979), pp. 1–10. ISSN: 0016-6480. DOI: https://doi.org/10.1016/0016-6480(79)90081-9.

[51] Peter Langhorne and Thomas H. Simpson. "The interrelationship of cortisol, Gill (Na + K) ATPase, and homeostasis during the Parr-Smolt transformation of atlantic salmon (Salmo salar L.)" In: *General and Comparative Endocrinology* 61.2 (1986), pp. 203–213. ISSN: 0016-6480. DOI: 10.1016/0016-6480(86)90198-X.

[52] S. Aota et al. "A Possible Role for Catecholamines in the Ventilatory Responses Associated with Internal Aciosis or External Hypoxia in Rainbow Trout Oncorhynchus Mykiss". In: *Journal of Experimental Biology* 151.1 (July 1990), pp. 57–70. ISSN: 0022-0949. DOI: 10.1242/jeb.151.1.57.

[53] David J. Randall and E.W. Taylor. "Evidence of a role for catecholamines in the control of breathing in fish". In: *Rev Fish Biol Fisheries* 1 (1991), pp. 139–157. DOI: 10.1007/BF00157582.

[54] Richard Kinkead and Steve F. Perry. "The effects of catecholamines on ventilation in rainbow trout during hypoxia or hypercapnia". In: *Respiration Physiology* 84.1 (1991), pp. 77–92. ISSN: 0034-5687. DOI: 10.1016/0034-5687(91)90020-J.

[55] S.D McCormick et al. "Repeated acute stress reduces growth rate of Atlantic salmon parr and alters plasma levels of growth hormone, insulin-like growth factor I and cortisol". In: *Aquaculture* 168.1 (1998), pp. 221–235. ISSN: 0044-8486. DOI: https://doi.org/10.1016/S0044-8486(98)00351-2.

[56] Rolf Olsen et al. “Acute stress alters the intestinal lining of Atlantic salmon, Salmo salar L.: An electron microscopical study”. In: *Fish Physiology and Biochemistry* 26 (May 2002), pp. 211–221. DOI: 10.1023/A:1026217719534.

[57] Thomas P. Mommsen, Mathilakath M. Vijayan, and Thomas W. Moon. “Cortisol in teleosts: dynamics, mechanisms of action, and metabolic regulation”. In: *Reviews in Fish Biology and Fisheries* 9.3 (1999), pp. 211–268. DOI: 10.1023/A:1008924418720.

[58] Muthuraman Pandurangan et al. “Effects of stress hormone cortisol on the mRNA expression of myogenenin, MyoD, Myf5, PAX3 and PAX7”. In: *Cytotechnology* 66.5 (2014), pp. 839–844. DOI: 10.1007/s10616-013-9635-6.

[59] A. D. Pickering and A. Stewart. “Acclimation of the interrenal tissue of the brown trout, Salmo trutta L., to chronic crowding stress”. In: *Journal of Fish Biology* 24.6 (1984), pp. 731–740. DOI: https://doi.org/10.1111/j.1095-8649.1984.tb04844.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1095-8649.1984.tb04844.x.

[60] Bruce A Barton, Car1 B Schreck, and Lesley D Barton. “Effects of chronic cortisol administration and daily acute stress on growth, physiological conditions, and stress responses in juvenile rainbow trout”. In: *Diseases of aquatic organisms* 2.3 (1987), pp. 173–185.

[61] Mark D. Fast et al. “Cortisol response and immune-related effects of Atlantic salmon (Salmo salar Linnaeus) subjected to short- and long-term stress”. In: *Fish Shellfish Immunology* 24.2 (2008), pp. 194–204. ISSN: 1050-4648. DOI: https://doi.org/10.1016/j.fsi.2007.10.009.

[62] Aleksei Krasnov et al. “Gene expression in Atlantic salmon skin in response to infection with the parasitic copepod Lepeophtheirus salmonis, cortisol implant, and their combination”. In: *BMC Genomics* 13.1 (). DOI: 10.1186/1471-2164-13-130.

[63] Iger Y et al. “Cortisol induces stress-related changes in the skin of rainbow trout (Oncorhynchus mykiss)”. In: *Gen Comp Endocrinol.* 97.2 (1995), pp. 188–98. DOI: 10.1006/gcen.1995.1018.

[64] Nicolas R. Bury et al. “Cortisol protects against copper induced necrosis and promotes apoptosis in fish gill chloride cells in vitro”. In: *Aquatic Toxicol.* 40.1 (1998), pp. 193–202.

[65] M. S. Eriksen et al. “Prespawning stress in farmed Atlantic salmon Salmo salar: maternal cortisol exposure and hyperthermia during embryonic development affect offspring survival, growth and incidence of malformations”. In: *Journal of Fish Biology* 69.1 (2006), pp. 114–129. DOI: https://doi.org/10.1111/j.1095-8649.2006.01071.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1095-8649.2006.01071.x.

[66] T.G. Pottinger and T.R. Carrick. "Modification of the Plasma Cortisol Response to Stress in Rainbow Trout by Selective Breeding". In: *General and Comparative Endocrinology* 116.1 (1999), pp. 122–132. ISSN: 0016-6480. DOI: https://doi.org/10.1006/gcen.1999.7355. URL: https://www.sciencedirect.com/science/article/pii/S0016648099973550.

[67] Marcel T.M. van Raaij et al. "Behavioral Strategy and the Physiological Stress Response in Rainbow Trout Exposed to Severe Hypoxia". In: *Hormones and Behavior* 30.1 (1996), pp. 85–92. ISSN: 0018-506X. DOI: https://doi.org/10.1006/hbeh.1996.0012. URL: https://www.sciencedirect.com/science/article/pii/S0018506X96900124.

[68] Øyvind Øverli et al. "Stress coping style predicts aggression and social dominance in rainbow trout". In: *Hormones and Behavior* 45.4 (2004), pp. 235–241. ISSN: 0018-506X. DOI: https://doi.org/10.1016/j.yhbeh.2003.12.002. URL: https://www.sciencedirect.com/science/article/pii/S0018506X0400011X.

[69] Øyvind Øverli et al. "Differences in behaviour between rainbow trout selected for high- and low-stress responsiveness". In: *Journal of Experimental Biology* 205.3 (Feb. 2002), pp. 391–395. ISSN: 0022-0949. DOI: 10.1242/jeb.205.3.391. eprint: https://journals.biologists.com/jeb/article-pdf/205/3/391/1243208/391.pdf. URL: https://doi.org/10.1242/jeb.205.3.391.

[70] Rodrigo E Barreto and Gilson L Volpato. "Ventilation rates indicate stress-coping styles in Nile tilapia". In: *Journal of biosciences* 36.5 (2011), pp. 851–855.

[71] J. J. Woodward. "Plasma catecholamines in resting rainbow trout, Salmo gairdneri Richardson, by high pressure liquid chromatography*". In: *Journal of Fish Biology* 21.4 (1982), pp. 429–432. DOI: https://doi.org/10.1111/j.1095-8649.1982.tb02848.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1095-8649.1982.tb02848.x.

[72] T. V. Basrur, R. Longland, and R. J. Wilkinson. "Effects of repeated crowding on the stress response and growth performance in Atlantic salmon (Salmo salar)". In: *Fish Physiology and Biochemistry* 36.3 (2010), pp. 445–450. DOI: 10.1007/s10695-009-9314-x.

[73] Agnes Beate Holden. "Effects of crowding intensity on stress responses in Atlantic salmon (Salmo salar)".

[74] Eva Veiseth et al. "Accelerated recovery of Atlantic salmon (Salmo salar) from effects of crowding by swimming". In: *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 144.3 (2006), pp. 351–358. ISSN: 1096-4959. DOI: https://doi.org/10.1016/j.cbpb.2006.03.009.

[75] Frances Bonier et al. "Do baseline glucocorticoids predict fitness?" In: *Trends in Ecology Evolution* 24.11 (2009), pp. 634–642. ISSN: 0169-5347. DOI: https://doi.org/10.1016/j.tree.2009.04.013.

[76] Rodrigo Egydio Barreto and Gilson Luiz Volpato. "Caution for using ventilatory frequency as an indicator of stress in fish". In: *Behavioural Processes* 66.1 (2004), pp. 43–51. ISSN: 0376-6357. DOI: https://doi.org/10.1016/j.beproc.2004.01.001.

[77] Brittany D. Kammerer, Joseph J. Cech, and Dietmar Kültz. "Rapid changes in plasma cortisol, osmolality, and respiration in response to salinity stress in tilapia (Oreochromis mossambicus)". In: *Comparative Biochemistry and Physiology Part A: Molecular Integrative Physiology* 157.3 (2010), pp. 260–265. ISSN: 1095-6433. DOI: https://doi.org/10.1016/j.cbpa.2010.07.009.

[78] Rodrigo Egydio Barreto and Gilson Luiz Volpato. "Ventilatory frequency of Nile tilapia subjected to different stressors". In: *Journal of Experimental Animal Science* 43.3 (2006), pp. 189–196. ISSN: 0939-8600. DOI: https://doi.org/10.1016/j.jeas.2006.05.001.

[79] Fazli Shabani et al. "Live transport of rainbow trout (Onchorhynchus mykiss) and subsequent live storage in market: Water quality, stress and welfare considerations". In: *Aquaculture* 453 (2016), pp. 110–115. ISSN: 0044-8486. DOI: https://doi.org/10.1016/j.aquaculture.2015.11.040.

[80] J. Altimiras and E. Larsen. "Non-invasive recording of heart rate and ventilation rate in rainbow trout during rest and swimming. Fish go wireless!" In: *Journal of Fish Biology* 57.1 (2000), pp. 197–209. DOI: https://doi.org/10.1111/j.1095-8649.2000.tb00786.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1095-8649.2000.tb00786.x.

[81] R. L. Oswald. "The use of telemetry to study light synchronization with feeding and gill ventilation rates in Salmo trutta". In: *Journal of Fish Biology* 13.6 (1978), pp. 729–739. DOI: https://doi.org/10.1111/j.1095-8649.1978.tb03487.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1095-8649.1978.tb03487.x.

[82] Yoshimi Kawai and Akiyoshi Wada. "Diurnal sea surface temperature variation and its impact on the atmosphere and ocean: A Review". In: *J. Oceanogr* 63 (Oct. 2007), pp. 721–744. DOI: 10.1007/s10872-007-0063-0.

[83] Gesa Milzer et al. "Spatial distribution of benthic foraminiferal stable isotopes and dinocyst assemblages in surface sediments of the Trondheimsfjord, central Norway". In: *Biogeosciences* 10 (July 2013). DOI: 10.5194/bg-10-4433-2013.

[84] Jannicke Vigen. "Oxygen variation within a seacage". June 2008. URL: http://hdl.handle.net/1956/6128.

[85] Mette Remen. "The oxygen requirement of Atlantic salmon (Salmo salar L.) in the on-growing phase in sea cages". Oct. 2012. URL: https://hdl.handle.net/1956/9535.

[86] Odd Inge Forsberg. “Modelling oxygen consumption rates of post-smolt Atlantic salmon in commercial-scale, land-based farms”. In: *Aquaculture International* 2 (1994), pp. 180–196.

[87] Angelico Madaro et al. “Effects of acclimation temperature on cortisol and oxygen consumption in Atlantic salmon (Salmo salar) post-smolt exposed to acute stress”. In: *Aquaculture* 497 (2018), pp. 331–335. ISSN: 0044-8486. DOI: https://doi.org/10.1016/j.aquaculture.2018.07.056. URL: https://www.sciencedirect.com/science/article/pii/S0044848617322597.

[88] Imants G. Priede. “Metabolic Scope in Fishes”. In: *Fish Energetics: New Perspectives*. Ed. by Peter Tytler and Peter Calow. Dordrecht: Springer Netherlands, 1985, pp. 33–64. ISBN: 978-94-011-7918-8. DOI: 10.1007/978-94-011-7918-8_2. URL: https://doi.org/10.1007/978-94-011-7918-8_2.

[89] Hans-Otto Pörtner and M.K. Grieshaber. “Critical Po(s) in oxyconforming and oxyregulating animals: Gas exchange, metabolic rate and the mode of energy production2”. In: *The Vertebrate Gas Transport Cascade: Adaptations to Environment and Mode of Life* (Jan. 1993), pp. 330–357.

[90] E Don Stevens, Arnold Sutterlin, and Todd Cook. “Respiratory metabolism and swimming performance in growth hormone transgenic Atlantic salmon”. In: *Canadian Journal of Fisheries and Aquatic Sciences* 55.9 (1998), pp. 2028–2035. DOI: 10.1139/f98-078. eprint: https://doi.org/10.1139/f98-078. URL: https://doi.org/10.1139/f98-078.

[91] Robin (Katersky) Barnes, Harry King, and Chris G. Carter. “Hypoxia tolerance and oxygen regulation in Atlantic salmon, Salmo salar from a Tasmanian population”. In: *Aquaculture* 318.3 (2011), pp. 397–401. ISSN: 0044-8486. DOI: https://doi.org/10.1016/j.aquaculture.2011.06.003. URL: https://www.sciencedirect.com/science/article/pii/S0044848611004790.

[92] Bjørn Olav Kvamme et al. “Modulation of innate immune responses in Atlantic salmon by chronic hypoxia-induced stress”. In: *Fish Shellfish Immunology* 34.1 (2013), pp. 55–65. ISSN: 1050-4648. DOI: https://doi.org/10.1016/j.fsi.2012.10.006. URL: https://www.sciencedirect.com/science/article/pii/S1050464812003622.

[93] Brankica Djordjevic et al. “Comparison of Circulating Markers and Mucosal Immune Parameters from Skin and Distal Intestine of Atlantic Salmon in Two Models of Acute Stress”. In: *International Journal of Molecular Sciences* 22.3 (2021). ISSN: 1422-0067. DOI: 10.3390/ijms22031028. URL: https://www.mdpi.com/1422-0067/22/3/1028.

[94] Edwin J. Niklitschek and David H. Secor. “Dissolved oxygen, temperature and salinity effects on the ecophysiology and survival of juvenile Atlantic sturgeon in estuarine waters: II. Model development and testing”. In: *Journal of Experimental Marine Biology and Ecology* 381 (2009).

Ecological Impacts of Hypoxia on Living Resources, S161–S172. ISSN: 0022-0981. DOI: https://doi.org/10.1016/j.jembe.2009.07.019. URL: https://www.sciencedirect.com/science/article/pii/S0022098109002883.

[95] R. Mallekh and J. P. Lagardère. "Effect of temperature and dissolved oxygen concentration on the metabolic rate of the turbot and the relationship between metabolic scope and feeding demand". In: *Journal of Fish Biology* 60.5 (2002), pp. 1105–1115. DOI: https://doi.org/10.1111/j.1095-8649.2002.tb01707.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1095-8649.2002.tb01707.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1095-8649.2002.tb01707.x.

[96] Nicholas Wegner et al. "Whole-body endothermy in a mesopelagic fish, the opah, Lampris guttatus". In: *Science (New York, N.Y.)* 348 (May 2015), pp. 786–9. DOI: 10.1126/science.aaa8902.

[97] Francis G. Carey. "FISHES WITH WARM BODIES". In: *Scientific American* 228.2 (1973), pp. 36–45. ISSN: 00368733, 19467087. URL: http://www.jstor.org/stable/24922978 (visited on 03/06/2023).

[98] Lefrancois Christelle and Guy Claireaux. "Influence of ambient oxygenation and temperature on metabolic scope and scope for heart rate of the sole (Solea solea)". In: *Marine Ecology Progress Series (0171-8630) (Inter-research), 2003 , Vol. 259 , P. 273-284* 259 (Sept. 2003). DOI: 10.3354/meps259273.

[99] Frederick Beamish. "Influence of Temperature and Salinity Acclimation on Temperature Preferenda of the Euryhaline Fish Tilapia nilotica". In: *Journal of the Fisheries Research Board of Canada* 27 (Apr. 2011), pp. 1209–1214. DOI: 10.1139/f70-143.

[100] J. M. Elliott and J. A. Elliott. "Temperature requirements of Atlantic salmon Salmo salar, brown trout Salmo trutta and Arctic charr Salvelinus alpinus: predicting the effects of climate change". In: *Journal of Fish Biology* 77.8 (2010), pp. 1793–1817. DOI: https://doi.org/10.1111/j.1095-8649.2010.02762.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1095-8649.2010.02762.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1095-8649.2010.02762.x.

[101] Selina M Stead and Lindsay Laird. *The handbook of salmon farming*. Springer Science & Business Media, 2002.

[102] Kilian M Stehfest et al. "Response of Atlantic salmon Salmo salar to temperature and dissolved oxygen extremes established using animal-borne environmental sensors". In: *Scientific reports* 7.1 (2017), pp. 1–10.

[103] J. M. ELLIOTT. "Tolerance and resistance to thermal stress in juvenile Atlantic salmon, Salmo salar". In: *Freshwater Biology* 25.1 (1991), pp. 61–70. DOI: https://doi.org/10.1111/j.1365-2427.1991.tb00473.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2427.1991.tb00473.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2427.1991.tb00473.x.

[104] Emily Corey et al. "Physiological effects of environmentally relevant, multi-day thermal stress on wild juvenile Atlantic salmon (Salmo salar)". In: *Conservation Physiology* 5.1 (Feb. 2017). cox014. ISSN: 2051-1434. DOI: 10.1093/conphys/cox014. eprint: https://academic.oup.com/conphys/article-pdf/5/1/cox014/11266056/cox014.pdf. URL: https://doi.org/10.1093/conphys/cox014.

[105] Sebastian Boltana et al. "Behavioural fever is a synergic signal amplifying the innate immune response". In: *Proceedings of the Royal Society B: Biological Sciences* 280 (July 2013). DOI: 10.1098/rspb.2013.1381.

[106] Sonia Rey et al. "Fish can show emotional fever: stress-induced hyperthermia in zebrafish". In: *Proceedings of the Royal Society B: Biological Sciences* 282.1819 (2015), p. 20152266. DOI: 10.1098/rspb.2015.2266. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspb.2015.2266. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2015.2266.

[107] H.-O. Pörtner. "Oxygen- and capacity-limitation of thermal tolerance: a matrix for integrating climate-related stressor effects in marine ecosystems". In: *Journal of Experimental Biology* 213.6 (Mar. 2010), pp. 881–893. ISSN: 0022-0949. DOI: 10.1242/jeb.037523. eprint: https://journals.biologists.com/jeb/article-pdf/213/6/881/1272853/881.pdf. URL: https://doi.org/10.1242/jeb.037523.

[108] Camille J. Macnaughton et al. "Using meta-analysis to derive a respiration model for Atlantic salmon (Salmo salar) to assess bioenergetics requirements of juveniles in two Canadian rivers". In: *Canadian Journal of Fisheries and Aquatic Sciences* 76.12 (2019), pp. 2225–2234. DOI: 10.1139/cjfas-2018-0436. eprint: https://doi.org/10.1139/cjfas-2018-0436. URL: https://doi.org/10.1139/cjfas-2018-0436.

[109] Claes Vernerback. *Effect of incubation temperature on Atlantic salmon metabolism as indicated by ventilation rate*. 2016.

[110] Yuki Hayashi et al. "High temperature causes masculinization of genetically female medaka by elevation of cortisol". In: *Molecular Reproduction and Development* 77.8 (2010), pp. 679–686. DOI: https://doi.org/10.1002/mrd.21203. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrd.21203. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrd.21203.

[111] John F. Cockrem, Mohammad A. Bahry, and Vishwajit S. Chowdhury. "Cortisol responses of goldfish (Carassius auratus) to air exposure, chasing, and increased water temperature". In: *General and Comparative Endocrinology* 270 (2019), pp. 18–25. ISSN: 0016-6480. DOI: https://doi.org/10.1016/j.ygcen.2018.09.017. URL: https://www.sciencedirect.com/science/article/pii/S0016648018303721.

[112] M. A. Delaney, P. H. Klesius, and R. A. Shelby. "Cortisol Response of Nile Tilapia, Oreochromis niloticus (L.), to Temperature Changes". In: *Journal of Applied Aquaculture* 16.3-4 (2005), pp. 95–104. DOI: 10.1300/J028v16n03_06. eprint: https://doi.org/10.1300/J028v16n03_06. URL: https://doi.org/10.1300/J028v16n03_06.

[113] J. Alejandro Buentello, Delbert M. Gatlin, and William H. Neill. "Effects of water temperature and dissolved oxygen on daily feed consumption , feed utilization and / growth of channel catfish Ictalurus punctatus". In: 2000.

[114] D. CHABOT and J.-D. DUTIL. "Reduced growth of Atlantic cod in non-lethal hypoxic conditions". In: *Journal of Fish Biology* 55.3 (1999), pp. 472–491. DOI: https://doi.org/10.1111/j.1095-8649.1999.tb00693.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1095-8649.1999.tb00693.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1095-8649.1999.tb00693.x.

[115] Catarina I. M. Martins et al. "Behavioural indicators of welfare in farmed fish". In: *Fish Physiology and Biochemistry* 38.1 (2012), pp. 17–41. DOI: 10.1007/s10695-011-9518-8.

[116] Tricia S. Clement et al. "Behavioral coping strategies in a cichlid fish: the role of social status and acute stress response in direct and displaced aggression". In: *Hormones and Behavior* 47.3 (2005), pp. 336–342. ISSN: 0018-506X. DOI: https://doi.org/10.1016/j.yhbeh.2004.11.014.

[117] Jonatan Nilsson et al. "Laksvel-Standardisert operasjonell velferdsovervåking for laks i matfiskanlegg". In: *Rapport fra havforskningen* (2022).

[118] Szeliski Richard. "Computer Vision: Algorithms and Applications". In: vol. 35. Springer London Ltd, 2010. ISBN: 9781848829343.

[119] Etienne Perot et al. "Learning to Detect Objects with a 1 Megapixel Event Camera". In: (Sept. 2020).

[120] *Samsung Galaxy S22 Ultra 5G*. Radius connect solutions. 2022. URL: https://www.radiusconnectsolutions.com/app/uploads/2022/04/samsung-galaxy-s22-ultra-5g_datasheet.pdf.

[121] John G Proakis. *Digital signal processing: principles, algorithms, and applications, 4/E*. Pearson Education India, 2007.

[122] Pin Wang, En Fan, and Peng Wang. "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning". In: *Pattern Recognition Letters* 141 (2021), pp. 61–67. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2020.07.042. URL: https://www.sciencedirect.com/science/article/pii/S0167865520302981.

[123] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019. ISBN: 9781492032618. URL: https://books.google.no/books?id=HHetDwAAQBAJ.

[124] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9 (1997), pp. 1735–1780.

[125] Xingjian Shi et al. "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting". In: *CoRR* abs/1506.04214 (2015). arXiv: 1506.04214. URL: http://arxiv.org/abs/1506.04214.

[126] João Carreira and Andrew Zisserman. "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset". In: *CoRR* abs/1705.07750 (2017). arXiv: 1705.07750. URL: http://arxiv.org/abs/1705.07750.

[127] Sudhakar Kumawat et al. "Depthwise Spatio-Temporal STFT Convolutional Neural Networks for Human Action Recognition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.9 (2022), pp. 4839–4851. DOI: 10.1109/TPAMI.2021.3076522.

[128] Kaiming He et al. *Mask R-CNN*. 2017. DOI: 10.48550/ARXIV.1703.06870. URL: https://arxiv.org/abs/1703.06870.

[129] Bruce Lucas and Takeo Kanade. "An Iterative Image Registration Technique with an Application to Stereo Vision (IJCAI)". In: vol. 81. Apr. 1981.

[130] J.R. Bergen et al. "A three-frame algorithm for estimating two-component image motion". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.9 (1992), pp. 886–896. DOI: 10.1109/34.161348.

[131] Adria Recasens et al. "Learning to Zoom: a Saliency-Based Sampling Layer for Neural Networks". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Sept. 2018.

[132] Angel Lopez-Molinero et al. "Feasibility of digital image colorimetry—application for water calcium hardness determination". In: *Talanta* 103 (2013), pp. 236–244.

[133] David F. Crouse. "On implementing 2D rectangular assignment algorithms". In: *IEEE Transactions on Aerospace and Electronic Systems* 52.4 (2016), pp. 1679–1696. DOI: 10.1109/TAES.2016.140952.

[134] George B Dantzig. "Application of the simplex method to a transportation problem". In: *Activity analysis and production and allocation* (1951).

[135] Victor Klee and George J. Minty. “HOW GOOD IS THE SIMPLEX ALGORITHM”. In: 1970.

[136] Harold W. Kuhn. “The Hungarian method for the assignment problem”. In: *Naval Research Logistics (NRL)* 52 (1955).

[137] James Munkres. “Algorithms for the Assignment and Transportation Problems”. In: *Journal of the Society for Industrial and Applied Mathematics* 5.1 (1957), pp. 32–38. DOI: 10.1137/0105003. eprint: https://doi.org/10.1137/0105003. URL: https://doi.org/10.1137/0105003.

[138] Mark Levedahl. “Performance comparison of 2D assignment algorithms for assigning truth objects to measured tracks”. In: *Signal and Data Processing of Small Targets 2000*. Ed. by Oliver E. Drummond. Vol. 4048. International Society for Optics and Photonics. SPIE, 2000, pp. 380–389. DOI: 10.1117/12.392018. URL: https://doi.org/10.1117/12.392018.

[139] Pauli Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[140] N. Tomizawa. “On some techniques useful for solution of transportation network problems”. In: *Networks* 1.2 (1971), pp. 173–194. DOI: https://doi.org/10.1002/net.3230010206. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/net.3230010206. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/net.3230010206.

[141] Edsger W. Dijkstra. “A note on two problems in connexion with graphs”. In: *Numerische Mathematik* 1 (1959), pp. 269–271.

[142] Roy Jonker and A. Volgenant. “A shortest augmenting path algorithm for dense and sparse linear assignment problems”. In: *Computing* 38 (1987), pp. 325–340.

[143] PC Mahalanobis. “Mahalanobis distance”. In: *Proceedings National Institute of Science of India*. Vol. 49. 2. 1936, pp. 234–256.

[144] R. E. Kalman. “A New Approach to Linear Filtering and Prediction Problems”. In: *Journal of Basic Engineering* 82.1 (Mar. 1960), pp. 35–45. ISSN: 0021-9223. DOI: 10.1115/1.3662552. eprint: https://asmedigitalcollection.asme.org/fluidsengineering/article-pdf/82/1/35/5518977/35_1.pdf. URL: https://doi.org/10.1115/1.3662552.

[145] S. Y. Chen. “Kalman Filter for Robot Vision: A Survey”. In: *IEEE Transactions on Industrial Electronics* 59.11 (2012), pp. 4409–4420. DOI: 10.1109/TIE.2011.2162714.

[146] Martin Føre et al. “Modelling of Atlantic salmon (Salmo salar L.) behaviour in sea-cages: Using artificial light to control swimming depth”. In: *Aquaculture* 388-391 (2013), pp. 137–146. ISSN: 0044-8486. DOI: https://doi.org/10.1016/j.aquaculture.2013.01.027. URL: https://www.sciencedirect.com/science/article/pii/S0044848613000422.

[147] Martin Føre et al. “Modelling growth performance and feeding behaviour of Atlantic salmon (Salmo salar L.) in commercial-size aquaculture net pens: Model details and validation through full-scale experiments”. In: *Aquaculture* 464 (2016), pp. 268–278. ISSN: 0044-8486. DOI: https://doi.org/10.1016/j.aquaculture.2016.06.045. URL: https://www.sciencedirect.com/science/article/pii/S0044848616303490.

[148] Martin Føre et al. “Modelling how the physical scale of experimental tanks affects salmon growth performance”. In: *Aquaculture* 495 (2018), pp. 731–737. ISSN: 0044-8486. DOI: https://doi.org/10.1016/j.aquaculture.2018.06.057. URL: https://www.sciencedirect.com/science/article/pii/S0044848617303277.

[149] S.J. Julier and J.K. Uhlmann. “Unscented filtering and nonlinear estimation”. In: *Proceedings of the IEEE* 92.3 (2004), pp. 401–422. DOI: 10.1109/JPROC.2003.823141.

[150] D. Reid. “An algorithm for tracking multiple targets”. In: *IEEE Transactions on Automatic Control* 24.6 (1979), pp. 843–854. DOI: 10.1109/TAC.1979.1102177.

[151] I.J. Cox and S.L. Hingorani. “An efficient implementation of Reid’s multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18.2 (1996), pp. 138–150. DOI: 10.1109/34.481539.

[152] Rahul Raguram et al. “USAC: A Universal Framework for Random Sample Consensus”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (2013), pp. 2022–2038.

[153] Martin A. Fischler and Robert C. Bolles. “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography”. In: *Commun. ACM* 24.6 (June 1981), pp. 381–395. ISSN: 0001-0782. DOI: 10.1145/358669.358692. URL: https://doi.org/10.1145/358669.358692.

[154] Darren R. Myatt et al. “NAPSAC: High Noise, High Dimensional Robust Estimation - it’s in the Bag”. In: *BMVC*. 2002.

[155] O. Chum and Jiri Matas. “Matching with PROSAC - progressive sample consensus”. In: vol. 1. July 2005, 220–226 vol. 1. ISBN: 0-7695-2372-2. DOI: 10.1109/CVPR.2005.221.

[156] J. Matas and O. Chum. “Randomized RANSAC with sequential probability ratio test”. In: *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*. Vol. 2. 2005, 1727–1732 Vol. 2. DOI: 10.1109/ICCV.2005.198.

[157] Ondřej Chum, Jiřı Matas, and Josef Kittler. “Locally optimized RANSAC”. In: *Joint Pattern Recognition Symposium*. Springer. 2003, pp. 236–243.

[158] P.H.S. Torr and A. Zisserman. "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry". In: *Computer Vision and Image Understanding* 78.1 (2000), pp. 138–156. ISSN: 1077-3142. DOI: https://doi.org/10.1006/cviu.1999.0832. URL: https://www.sciencedirect.com/science/article/pii/S1077314299908329.

[159] Amit Mitra and Debasis Kundu. "Genetic algorithms based robust frequency estimation of sinusoidal signals with stationary errors". In: *Engineering Applications of Artificial Intelligence* 23.3 (2010), pp. 321–330. ISSN: 0952-1976. DOI: https://doi.org/10.1016/j.engappai.2009.07.001. URL: https://www.sciencedirect.com/science/article/pii/S0952197609001201.

[160] Petr Cisar et al. "Computer vision based individual fish identification using skin dot pattern". In: *Scientific Reports* 11 (Aug. 2021). DOI: 10.1038/s41598-021-96476-4.

[161] Renato Dala-Corte, Júlia Moschetta, and Fernando Becker. "Photo-identification as a technique for recognition of individual fish: A test with the freshwater armored catfish Rineloricaria aequalicuspis Reis Cardoso, 2001 (Siluriformes: Loricariidae)". In: *Neotropical Ichthyology* 14 (Apr. 2016). DOI: 10.1590/1982-0224-20150074.

[162] Enrico Bondi et al. "Long Term Person Re-identification from Depth Cameras Using Facial and Skeleton Data". In: *Understanding Human Activities Through 3D Sensors*. Ed. by Hazem Wannous et al. Cham: Springer International Publishing, 2018, pp. 29–41. ISBN: 978-3-319-91863-1.

[163] Michal Balazia and Petr Sojka. "You are how you walk: Uncooperative MoCap gait identification for video surveillance with incomplete and noisy data". In: *2017 IEEE International Joint Conference on Biometrics (IJCB)*. 2017, pp. 208–215. DOI: 10.1109/BTAS.2017.8272700.

[164] Javier Lorenzo-Navarro, Modesto Castrillón Santana, and Daniel Hernández-Sosa. "On the Use of Simple Geometric Descriptors Provided by RGB-D Sensors for Re-Identification". In: *Sensors (Basel, Switzerland)* 13 (July 2013), pp. 8222–8238. DOI: 10.3390/s130708222.

[165] Patrice Y Simard, David Steinkraus, John C Platt, et al. "Best practices for convolutional neural networks applied to visual document analysis." In: *Icdar*. Vol. 3. 2003. Edinburgh. 2003.

[166] Diego Marcos, Michele Volpi, and Devis Tuia. "Learning rotation invariant convolutional filters for texture classification". In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. 2016, pp. 2012–2017. DOI: 10.1109/ICPR.2016.7899932.

[167] Yifang Chen et al. "A Rotation-Invariant Convolutional Neural Network for Image Enhancement Forensics". In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2018, pp. 2111–2115. DOI: 10.1109/ICASSP.2018.8462057.

[168] Jinlong Fan et al. *Wide-angle Image Rectification: A Survey*. 2021. arXiv: 2011.12108 [cs.CV].

[169] Guosheng Hu et al. "When Face Recognition Meets With Deep Learning: An Evaluation of Convolutional Neural Networks for Face Recognition". In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*. Dec. 2015.

[170] Peyman Bateni et al. "Improved Few-Shot Visual Classification". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2020.

[171] Jake Snell, Kevin Swersky, and Richard Zemel. "Prototypical networks for few-shot learning". In: *Advances in neural information processing systems* 30 (2017).

[172] Flood Sung et al. *Learning to Compare: Relation Network for Few-Shot Learning*. 2017. DOI: 10.48550/ARXIV.1711.06025. URL: https://arxiv.org/abs/1711.06025.

[173] Zhongheng Zhang. "Introduction to machine learning: k-nearest neighbors". In: *Annals of Translational Medicine* 4.11 (2016). ISSN: 2305-5847. URL: https://atm.amegroups.com/article/view/10170.

[174] Gregory R. Koch. "Siamese Neural Networks for One-Shot Image Recognition". In: 2015.

[175] Jean-Bastien Grill et al. *Bootstrap your own latent: A new approach to self-supervised Learning*. 2020. DOI: 10.48550/ARXIV.2006.07733. URL: https://arxiv.org/abs/2006.07733.

[176] Jason Yosinski et al. "How transferable are features in deep neural networks?" In: (2014). DOI: 10.48550/ARXIV.1411.1792. URL: https://arxiv.org/abs/1411.1792.

[177] Chelsea Finn, Pieter Abbeel, and Sergey Levine. "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks". In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, June 2017, pp. 1126–1135. URL: https://proceedings.mlr.press/v70/finn17a.html.

[178] James Requeima et al. *Fast and Flexible Multi-Task Classification Using Conditional Neural Adaptive Processes*. 2020. arXiv: 1906.07697 [stat.ML].

[179] Boris N. Oreshkin, Pau Rodriguez, and Alexandre Lacoste. *TADAM: Task dependent adaptive metric for improved few-shot learning*. 2019. arXiv: 1805.10123 [cs.LG].

[180] Ayoub Benali Amjoud and Mustapha Amrouch. "Convolutional neural networks backbones for object detection". In: *Image and Signal Processing: 9th International Conference, ICISP 2020, Marrakesh, Morocco, June 4–6, 2020, Proceedings 9*. Springer. 2020, pp. 282–289.

[181] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Neural Information Processing Systems* 25 (Jan. 2012). DOI: 10.1145/3065386.

[182] Andrew J. Dunnings and Toby P. Breckon. "Experimentally Defined Convolutional Neural Network Architecture Variants for Non-Temporal Real-Time Fire Detection". In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. 2018, pp. 1558–1562. DOI: 10.1109/ICIP.2018.8451657.

[183] Md Mafiul Hasan Matin et al. "An efficient disease detection technique of rice leaf using AlexNet". In: *Journal of Computer and Communications* 8.12 (2020), p. 49.

[184] Mingxing Tan and Quoc V. Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". In: *CoRR* abs/1905.11946 (2019). arXiv: 1905.11946. URL: http://arxiv.org/abs/1905.11946.

[185] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *CoRR* abs/1512.03385 (2015). arXiv: 1512.03385. URL: http://arxiv.org/abs/1512.03385.

[186] Yongqin Xian et al. *Zero-Shot Learning – A Comprehensive Evaluation of the Good, the Bad and the Ugly*. 2020. arXiv: 1707.00600 [cs.CV].

[187] John A Hartigan and Manchek A Wong. "Algorithm AS 136: A k-means clustering algorithm". In: *Journal of the royal statistical society. series c (applied statistics)* 28.1 (1979), pp. 100–108.

[188] Kamran Khan et al. "DBSCAN: Past, present and future". In: *The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014)*. 2014, pp. 232–238. DOI: 10.1109/ICADIWT.2014.6814687.

[189] Xiaofei He et al. "Laplacian Regularized Gaussian Mixture Model for Data Clustering". In: *IEEE Transactions on Knowledge and Data Engineering* 23.9 (2011), pp. 1406–1418. DOI: 10.1109/TKDE.2010.259.

[190] Yutian Lin et al. "A bottom-up clustering approach to unsupervised person re-identification". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 8738–8745.

[191] Farhana Sultana, Abu Sufian, and Paramartha Dutta. "Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey". In: *Knowledge-Based Systems* 201-202 (2020), p. 106062. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2020.106062. URL: https://www.sciencedirect.com/science/article/pii/S0950705120303464.

[192] Elaine Nicpon Marieb and Katja Hoehn. *Human anatomy & physiology, ninth edition*. Pearson education, 2020. ISBN: 9781292314471.

[193] Sheth BR and Young R. "Two Visual Pathways in Primates Based on Sampling of Space: Exploitation and Exploration of Visual Information". In: *Front Integr Neurosci.* (2016). DOI: 10.3389/fnint.2016.00037.

[194] Alzayat Saleh, Marcus Sheaves, and Mostafa Rahimi Azghadi. "Computer vision and deep learning for fish classification in underwater habitats: A survey". In: *Fish and Fisheries* 23.4 (2022), pp. 977–999. DOI: https://doi.org/10.1111/faf.12666. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/faf.12666. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/faf.12666.

[195] Junyan Chen et al. *Make Every feature Binary: A 135B parameter sparse neural network for massively improved search relevance*. https://www.microsoft.com/en-us/research/blog/make-every-feature-binary-a-135b-parameter-sparse-neural-network-for-massively-improved-search-relevance/. Accessed: 2022-11-08. 2021.

[196] Osman Semih Kayhan and Jan C. van Gemert. "On Translation Invariance in CNNs: Convolutional Layers can Exploit Absolute Spatial Location". In: *CoRR* abs/2003.07064 (2020). arXiv: 2003.07064. URL: https://arxiv.org/abs/2003.07064.

[197] David E. Rumelhart and James L. McClelland. "Learning Internal Representations by Error Propagation". In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. 1987, pp. 318–362.

[198] Adam Paszke et al. "PyTorch: An Imperative Style, High Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[199] Espen Høgstedt, Hielke Walinga, and Tommaso Tofacchi. *An exploration of fish nets [Unpublished]*. June 2022.

[200] Shaoqing Ren et al. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. 2015. DOI: 10.48550/ARXIV.1506.01497. URL: https://arxiv.org/abs/1506.01497.

[201] Ross Girshick. *Fast R-CNN*. 2015. DOI: 10.48550/ARXIV.1504.08083. URL: https://arxiv.org/abs/1504.08083.

[202] Jonathan Long, Evan Shelhamer, and Trevor Darrell. *Fully Convolutional Networks for Semantic Segmentation*. 2015. arXiv: 1411.4038 [cs.CV].

[203] Misha Gromov. *Metric structures for Riemannian and non-Riemannian spaces, english ed., Modern Birkhäuser Classics*. 2007.

[204] Stephen Wright, Jorge Nocedal, et al. "Numerical optimization". In: *Springer Science* 35.67-68 (1999), p. 7.

[205] H. W. Kuhn and A. W. Tucker. "Nonlinear programming". In: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*. Berkeley and Los Angeles: University of California Press, 1951, pp. 481–492.

[206] László Lovász and Michael D Plummer. *Matching theory*. Vol. 367. American Mathematical Soc., 2009.

[207] Brent Perreault. "Introduction to the Kalman Filter and its Derivation". In: *Concordia College, Moorhead Minnesota* (2012).

[208] Roger Labbe. "Kalman and bayesian filters in python". In: *Chap* 7.246 (2014), p. 4. URL: https://drive.google.com/file/d/0By_SW19c1BfhSVFzNHc0SjduNzg/view?resourcekey=0-41olC9ht9xE3wQe2zHZ45A.

[209] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Second. Cambridge University Press, ISBN: 0521540518, 2004.

[210] Elan Dubrofsky. "Homography estimation". In: *Diplomová práce. Vancouver: Univerzita Britské Kolumbie* 5 (2009).

[211] G. Bradski. "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools* (2000).

[212] Rafael C Gonzales and Paul Wintz. *Digital image processing*. Addison-Wesley Longman Publishing Co., Inc., 1987.

[213] Irwin Sobel. "An Isotropic 3x3 Image Gradient Operator". In: *Presentation at Stanford A.I. Project 1968* (Feb. 2014).

[214] R.E. Walpole et al. *Probability and Statistics for Engineers and Scientists*. Pearson, 2017. ISBN: 9780134115856. URL: https://books.google.no/books?id=aOKHrgEACAAJ.

[215] B. L. Welch. "The Generalization of 'Student's' Problem when Several Different Population Variances are Involved". In: *Biometrika* 34.1/2 (1947), pp. 28–35. ISSN: 00063444. URL: http://www.jstor.org/stable/2332510 (visited on 05/29/2023).

[216] Michael Allwood. "The Satterthwaite formula for degrees of freedom in the two-sample t-test". In: *The College Board* (2008).

[217] Ananth Ranganathan. "The levenberg-marquardt algorithm". In: *Tutoral on LM algorithm* 11.1 (2004), pp. 101–110.

[218] Haiyan Wang and Jinyan Fan. "Convergence rate of the Levenberg-Marquardt method under Hölderian local error bound". In: *Optimization Methods and Software* 35.4 (2020), pp. 767–786. DOI: 10.1080/10556788.2019.1694927. eprint: https://doi.org/10.1080/10556788.2019.1694927. URL: https://doi.org/10.1080/10556788.2019.1694927.

[219] Miao Li, Yukun Huang, and S p Gong. "Assessing the risk of potentially hazardous asteroids through mean motion resonances analyses". In: *Astrophysics and Space Science* 364 (Apr. 2019). DOI: 10.1007/s10509-019-3557-5.

[220] Precision Measurement Engineering. *miniDOT Logger user's manual*. 2021. URL: https://www.pme.com/products/minidot.

[221] Kentaro Wada. *Labelme: Image Polygonal Annotation with Python*. DOI: 10.5281/zenodo.5711226. URL: https://github.com/wkentaro/labelme.

[222] Alice Gibb et al. "The Teleost Intramandibular Joint: A mechanism That Allows Fish to Obtain Prey Unavailable to Suction Feeders". In: *Integrative and Comparative Biology* (June 2015). DOI: 10.1093/icb/icv042.

[223] Walter J Rainboth. *Fishes of the cambodian mekong*. Food & Agriculture Org., 1996.

[224] A. Buslaev et al. "Albumentations: fast and flexible image augmentations". In: *ArXiv e-prints* (2018). eprint: 1809.06839.

[225] J. Deng et al. "ImageNet: A Large-Scale Hierarchical Image Database". In: *CVPR09*. 2009.

[226] Rahmad Sadli. *Object Tracking: 2-D Object Tracking using Kalman Filter in Python*. https://machinelearningspace.com/2d-object-tracking-using-kalman-filter/. Accessed: 2023-05-23. 2020.

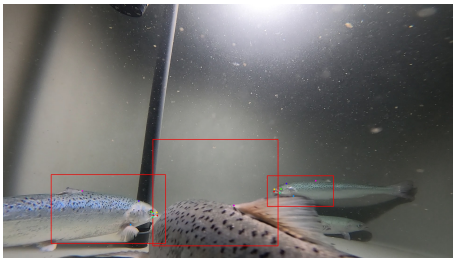
Chapter 11

Appendix

Appendix A

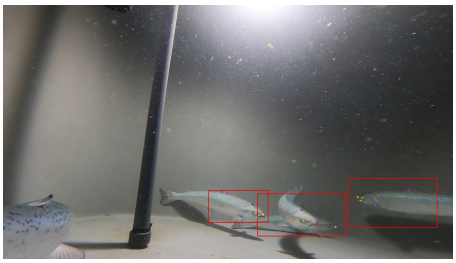
Examples of Keypoint RCNN errors

(a) False negative

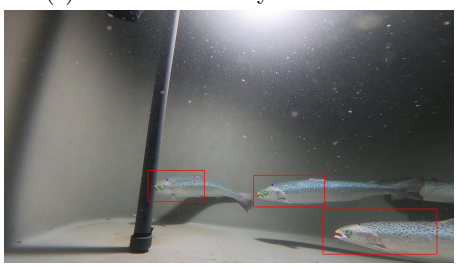


(b) False positive

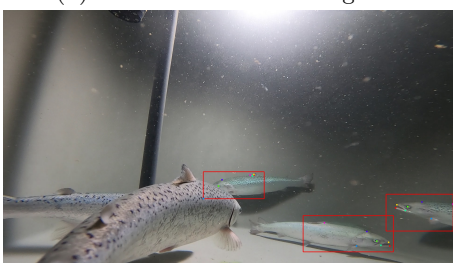
(c) Fish turned away from camera



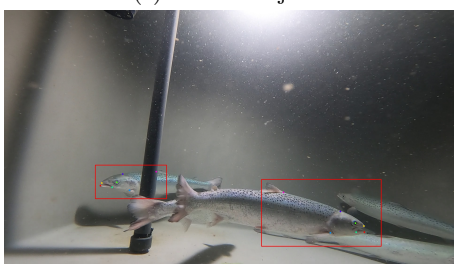
(d) Jaws detected on wrong fish



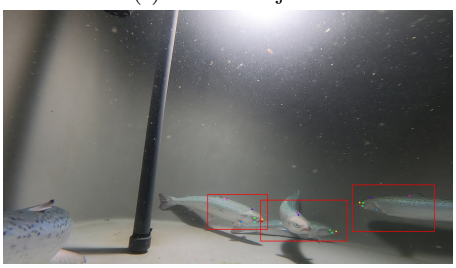
(e) Occluded jaws



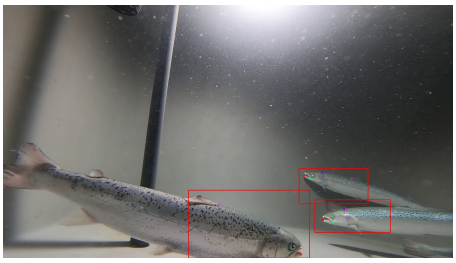
(f) Occluded jaws



(g) Excessive IoU pruning



(h) Two fish sharing one bounding box

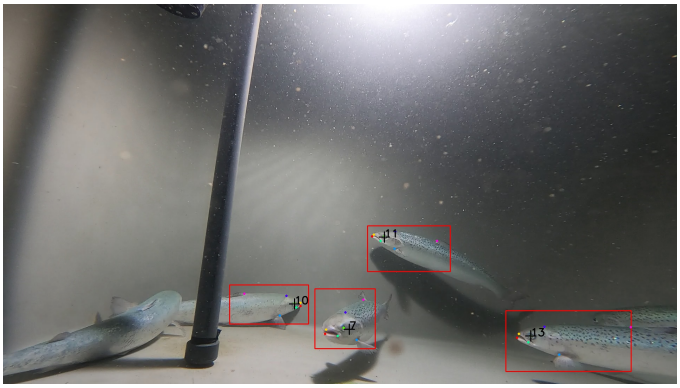


(i) Dorsal fin keypoint sliding along the fish back

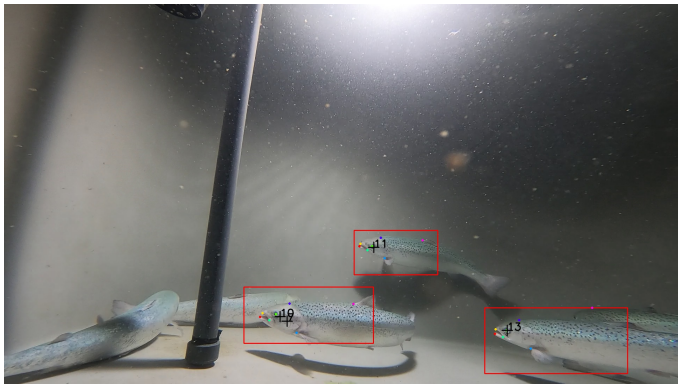
Figure A.1: The most common Keypoint RCNN errors

Appendix B

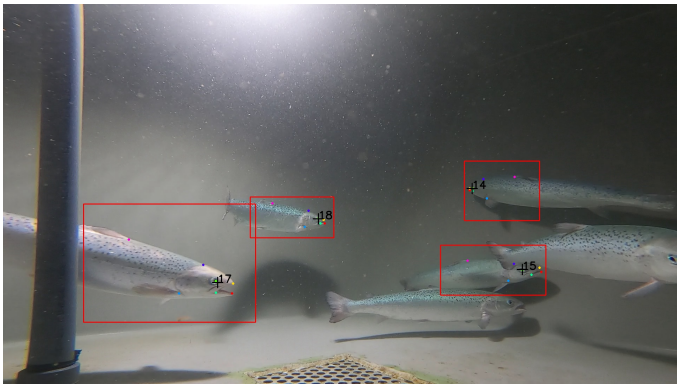
Examples of tracker errors



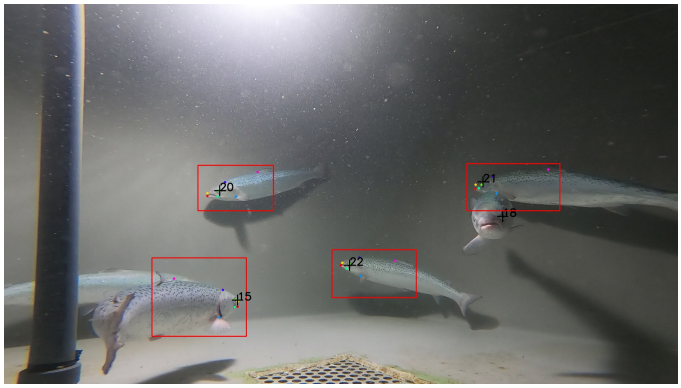
(a) Tracker 10 shifts fish



(b) Tracker 10 shifts fish



(c) Tracker 14 is replaced by tracker 20



(d) Tracker 14 is replaced by tracker 20

Figure B.1: Two examples of tracker errors

Appendix C

Model fit for complete time series set

Below, a complete set of timeseries with corresponding model fit is displayed. The video sequence used to generate the time series consisted of frame 5000 to 7000 from the first snippet (clip) after three hours of the first downbreathing in tank nine, recorded with camera nine.

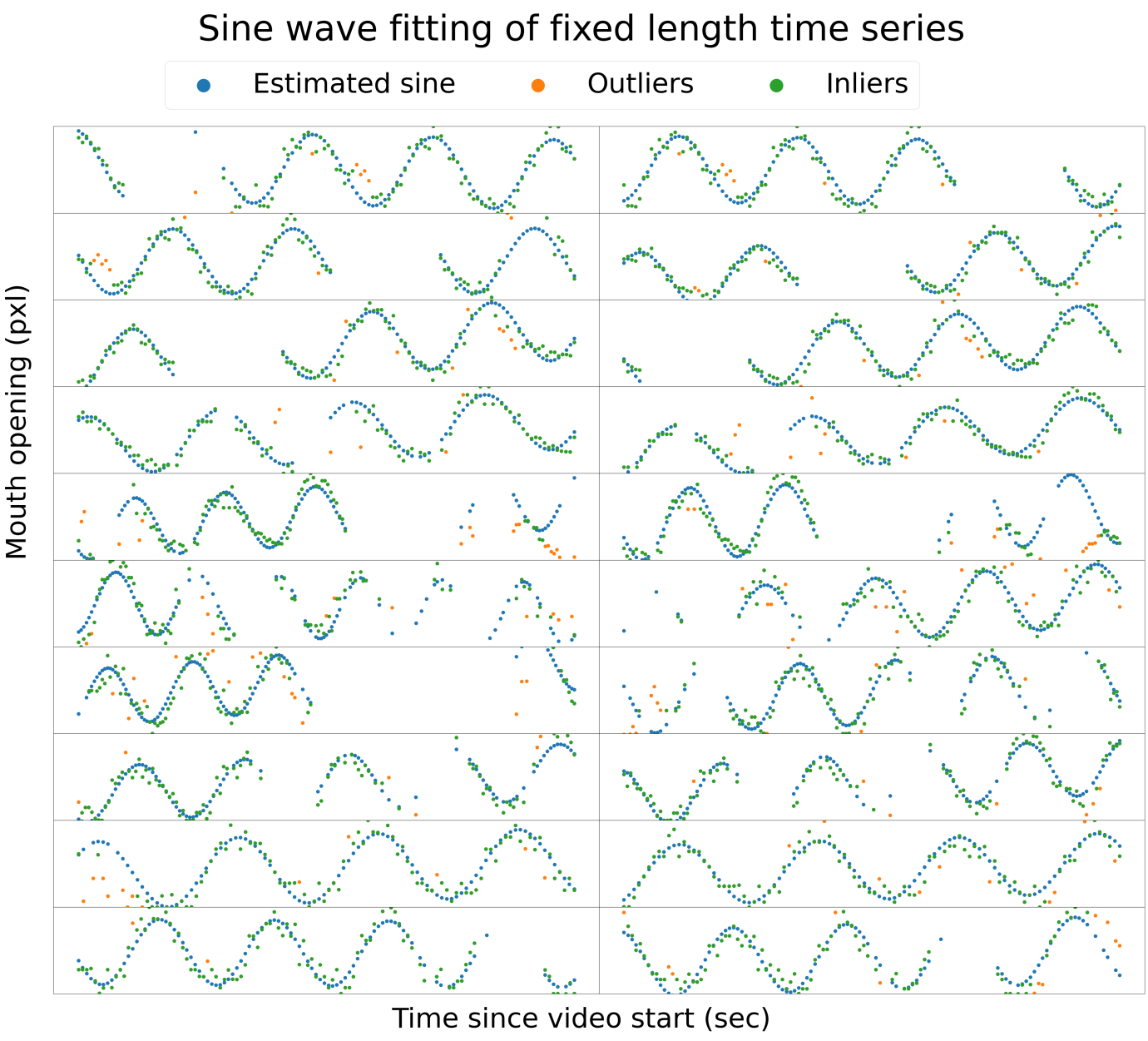


Figure C.1: Time series function fitting part 1

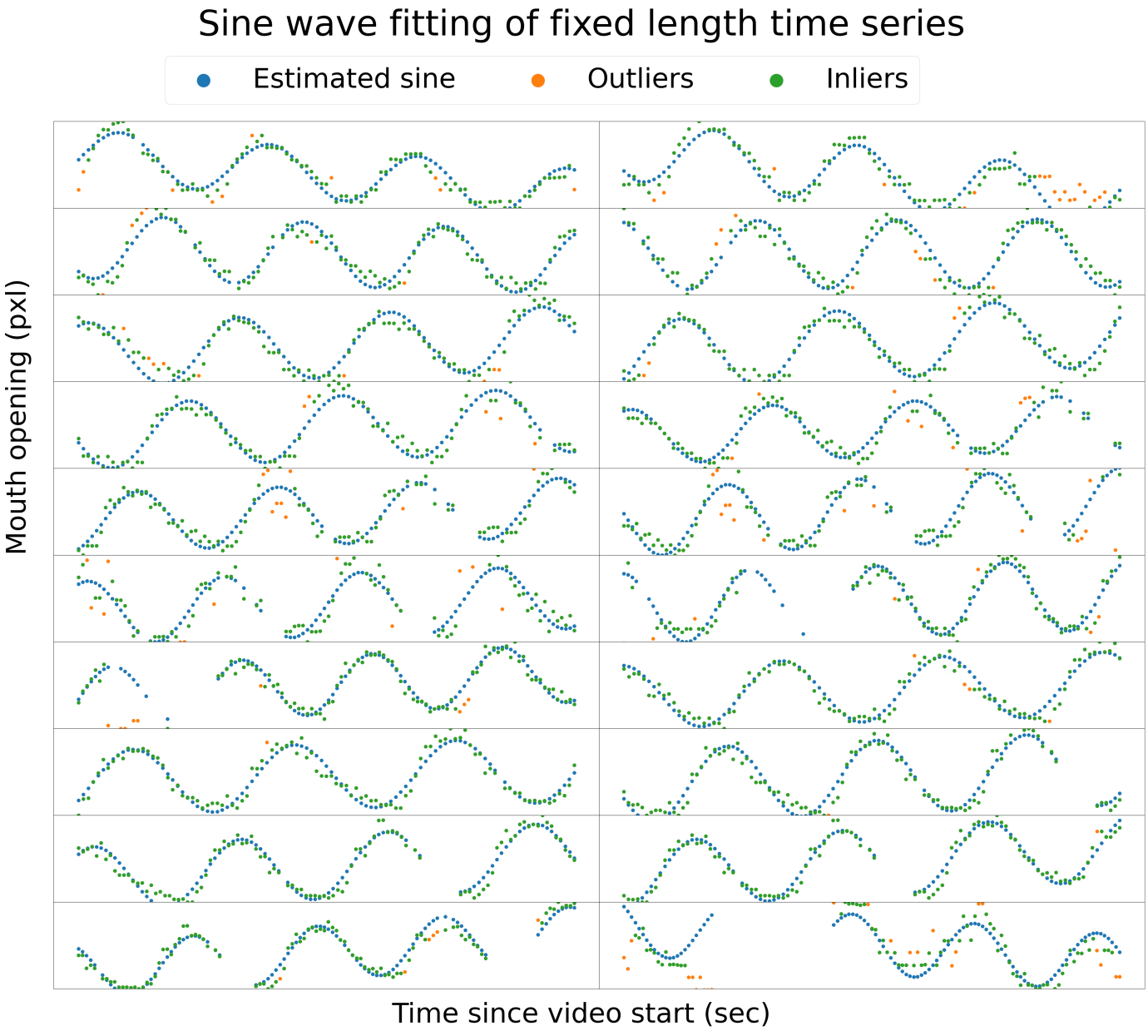


Figure C.2: Time series function fitting part 2

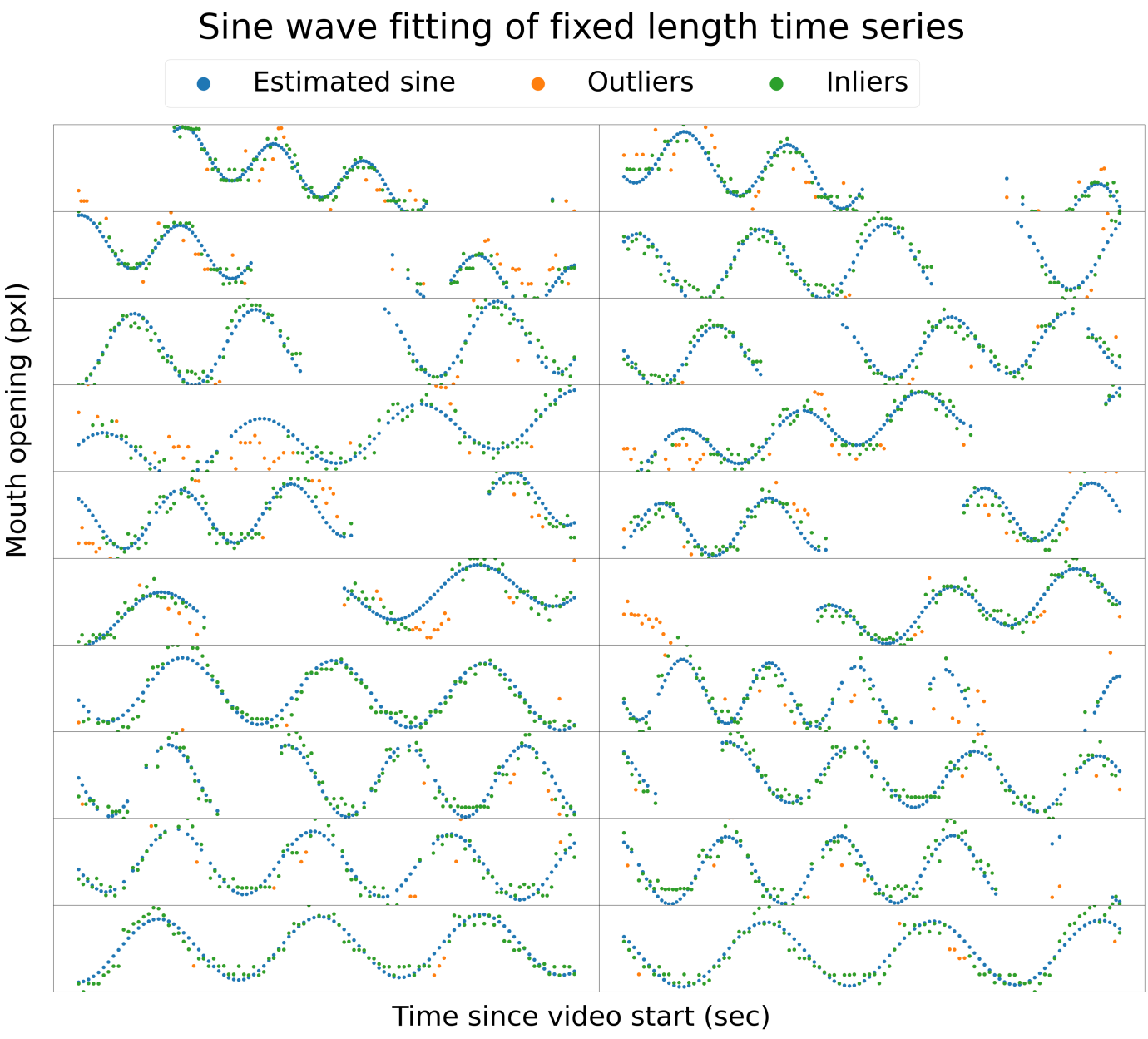


Figure C.3: Time series function fitting part 3

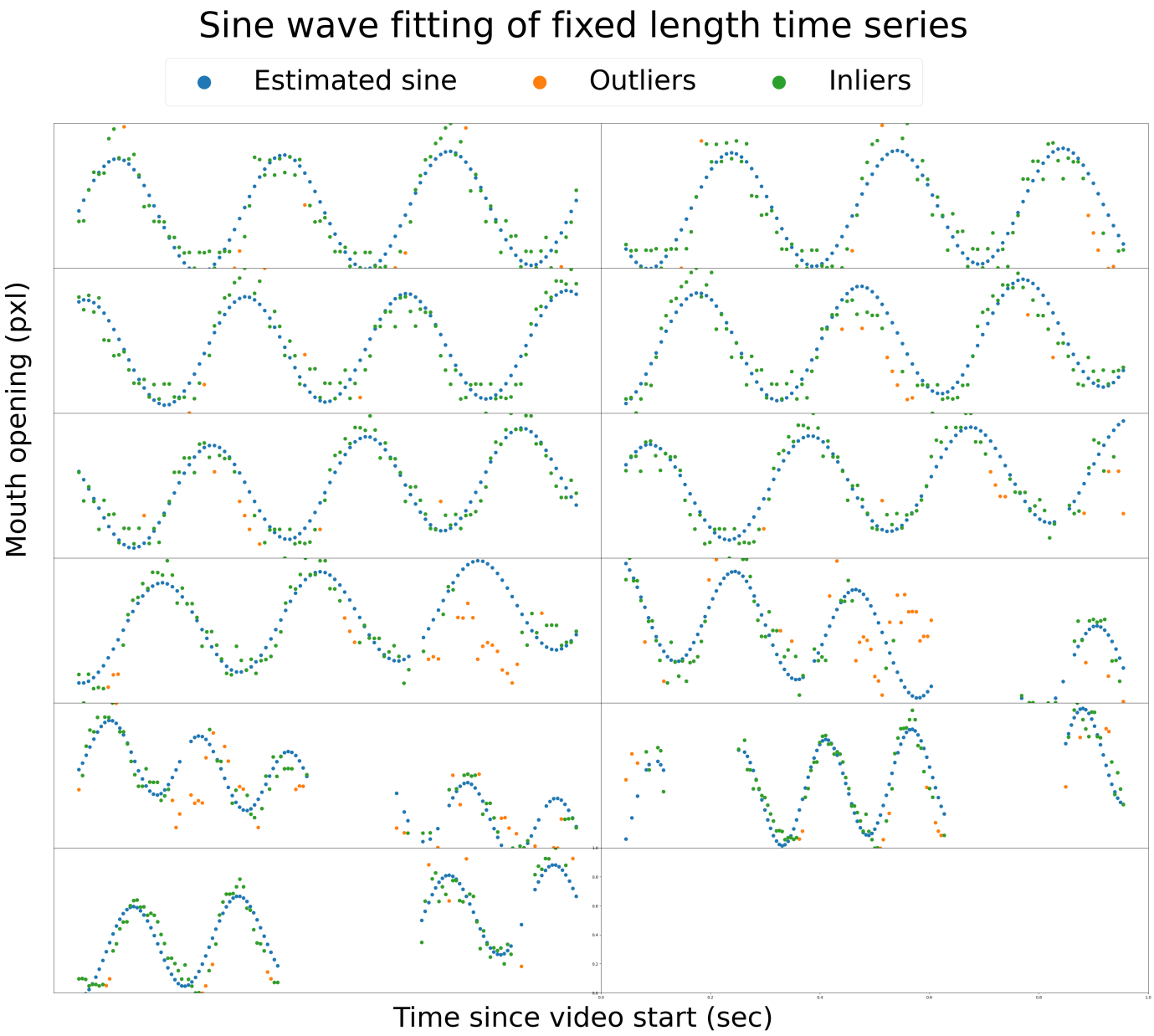
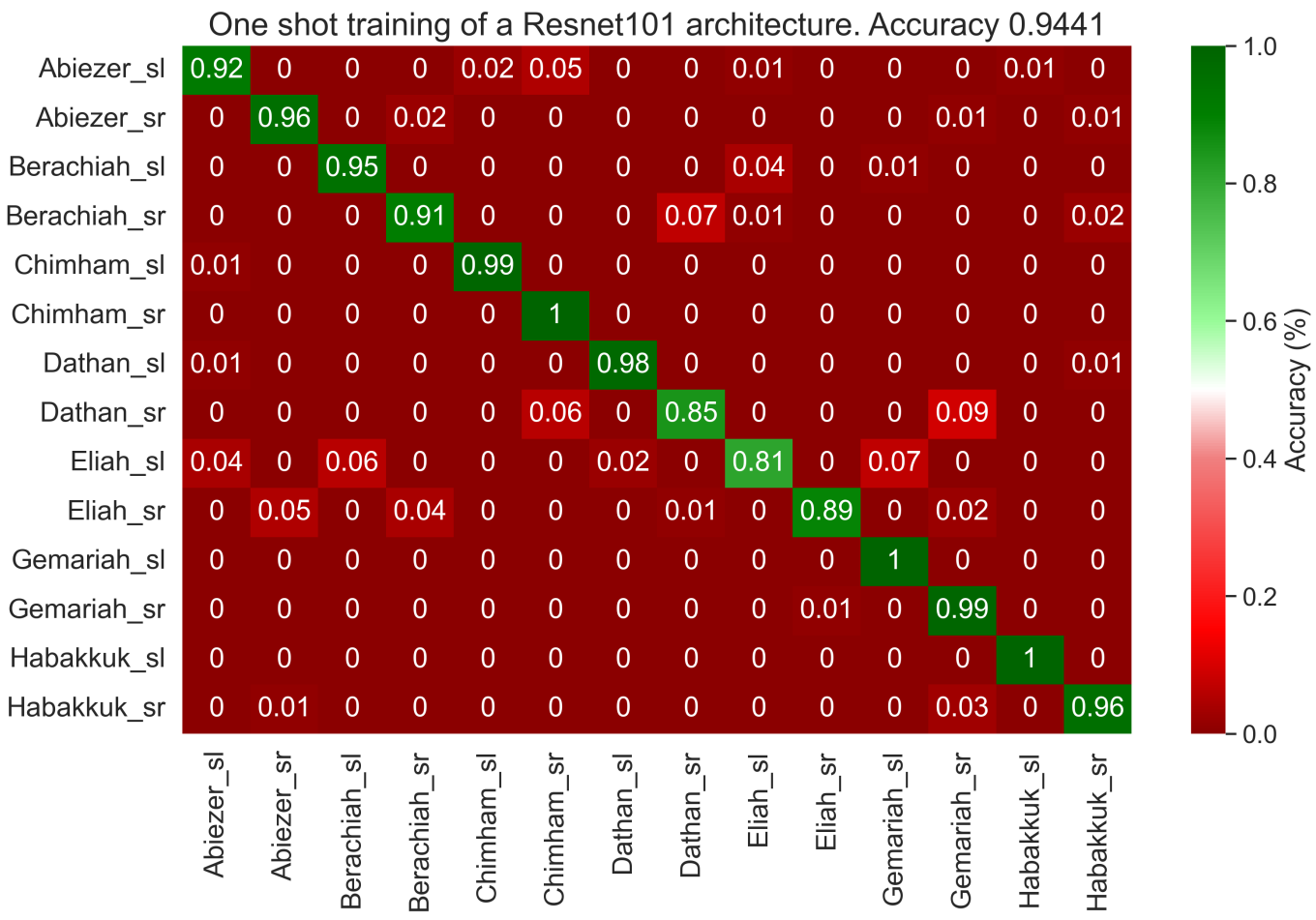


Figure C.4: Time series function fitting part 4

Appendix D

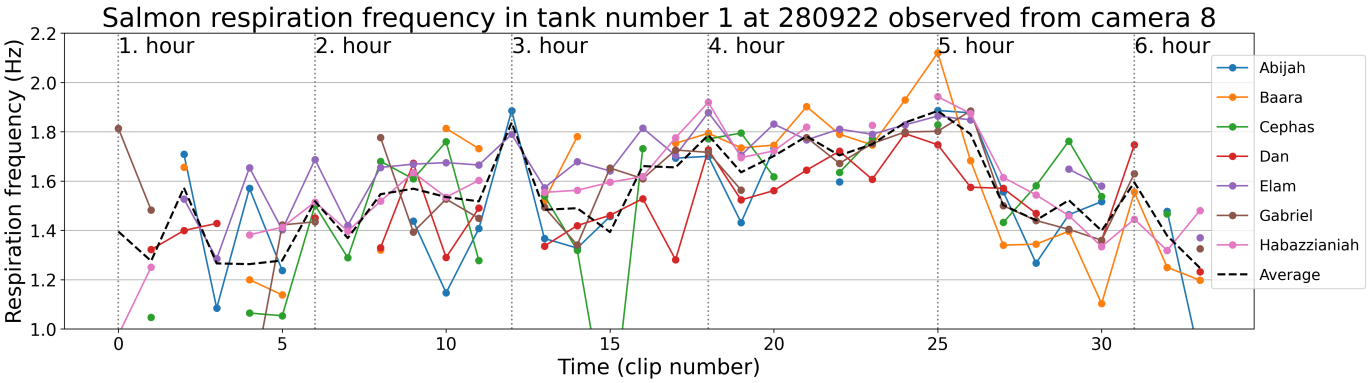
One-shot classification confusion matrix



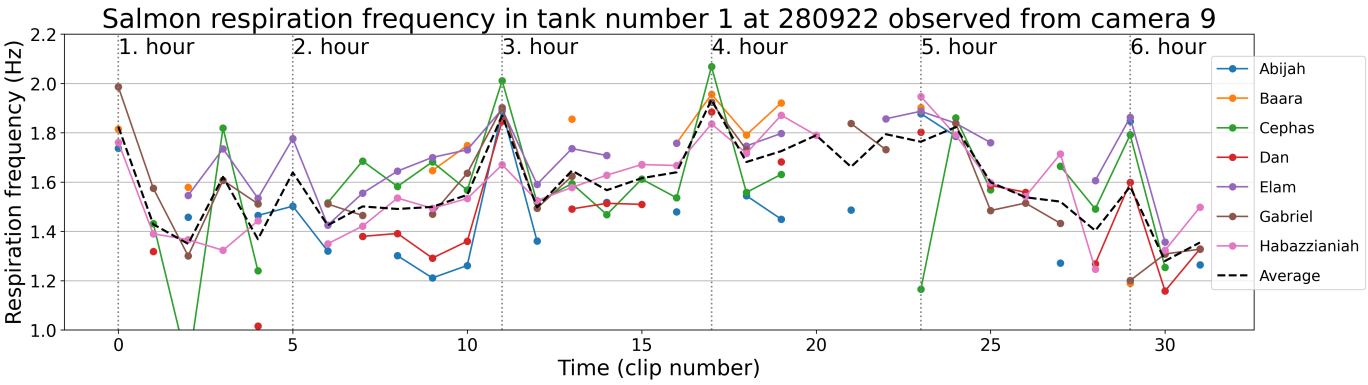
(a) Confusion matrix for one shot training of tank three classes

Appendix E

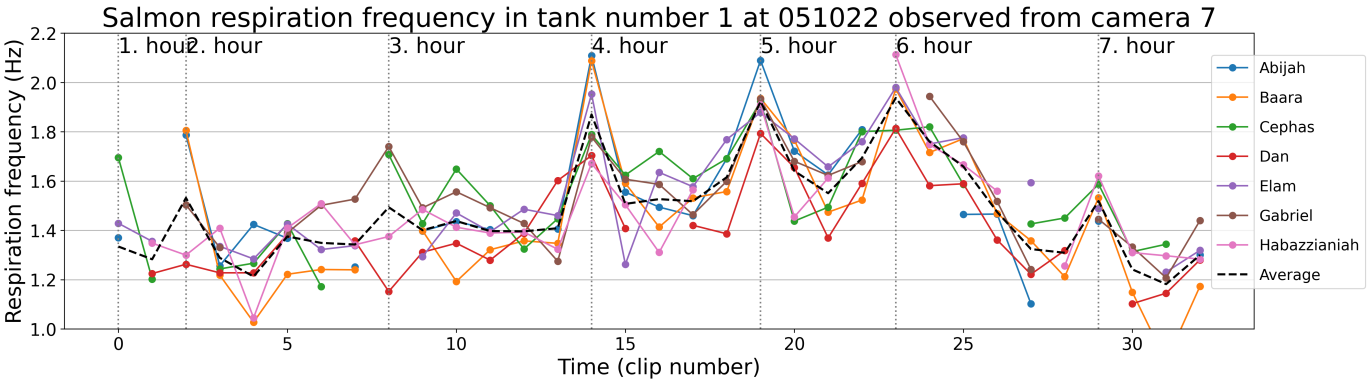
All full downbreathing evolutions



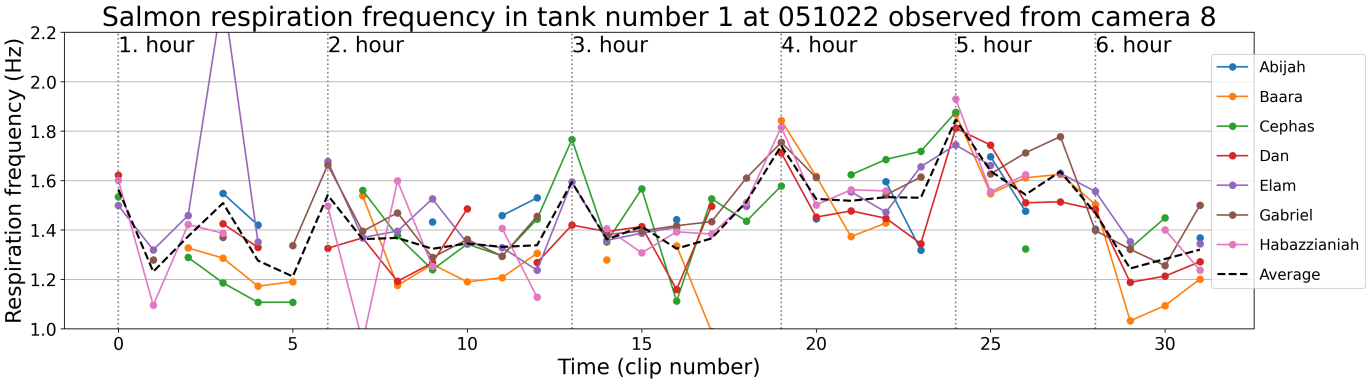
(a) Camera eight, first downbreathing



(b) Camera nine, first downbreathing

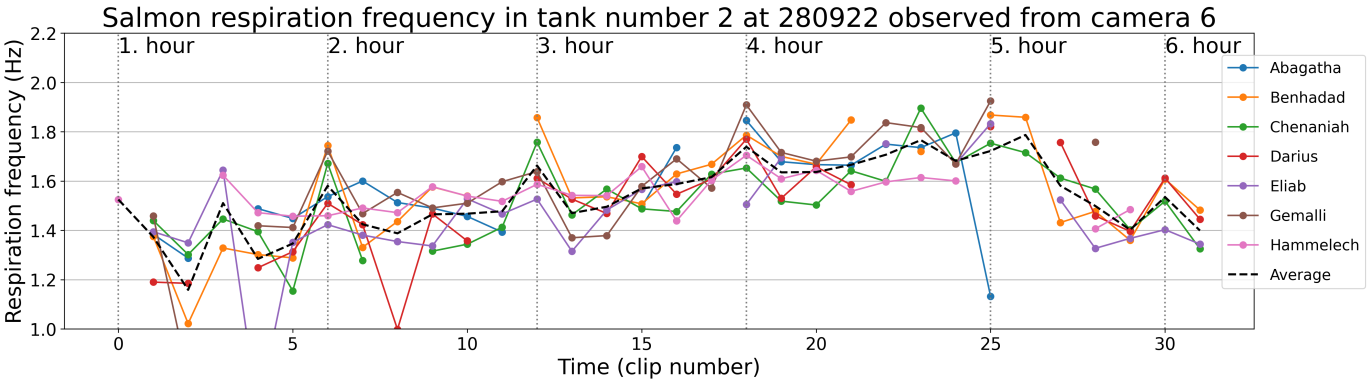


(c) Camera seven, second downbreathing

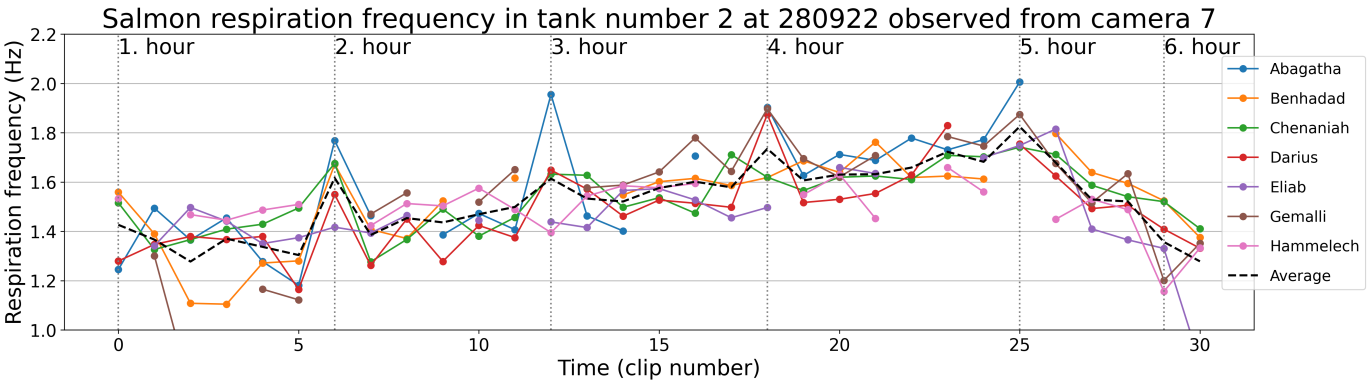


(d) Camera eight, second downbreathing

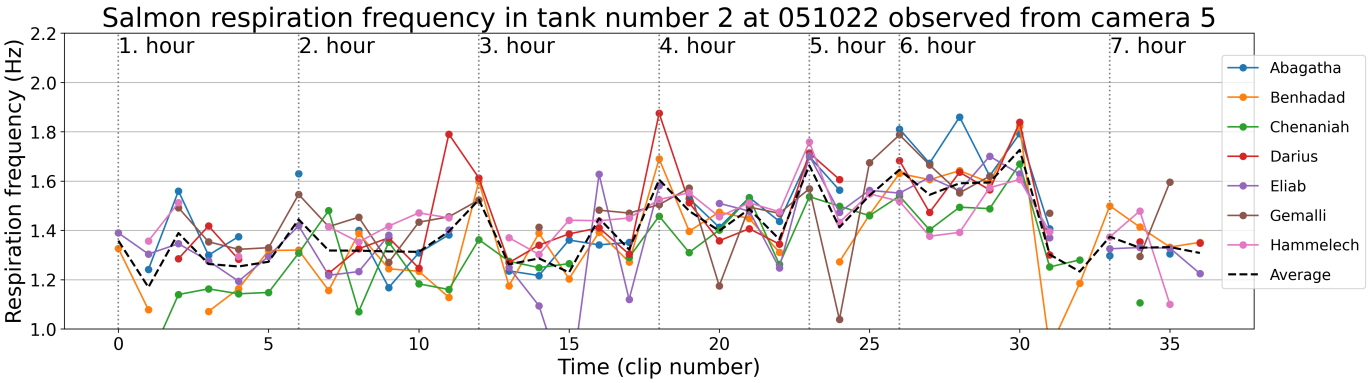
Figure E.1: Respiration evolution for tank one downbreathings



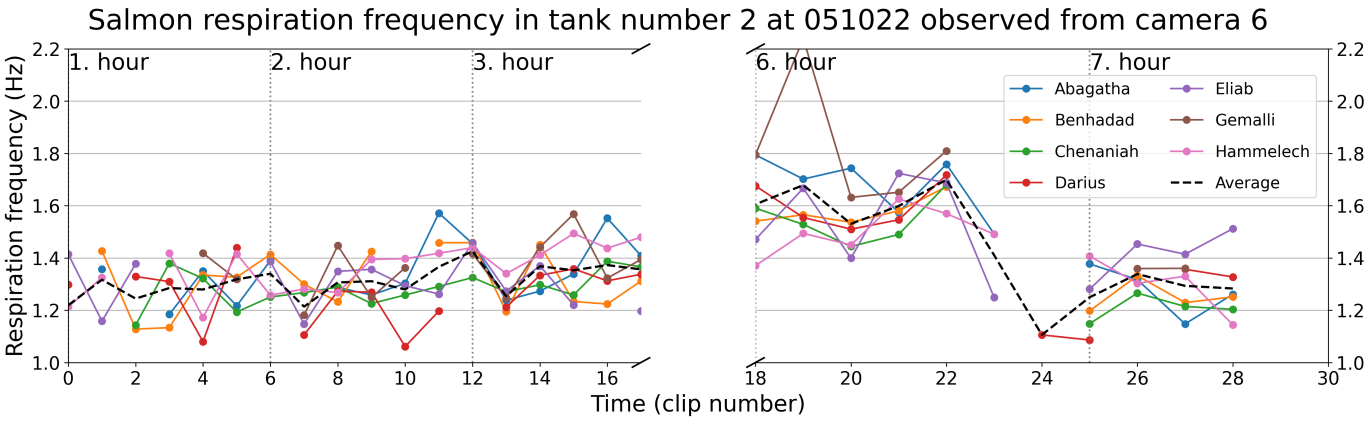
(a) Camera six, first downbreathing



(b) Camera seven, first downbreathing

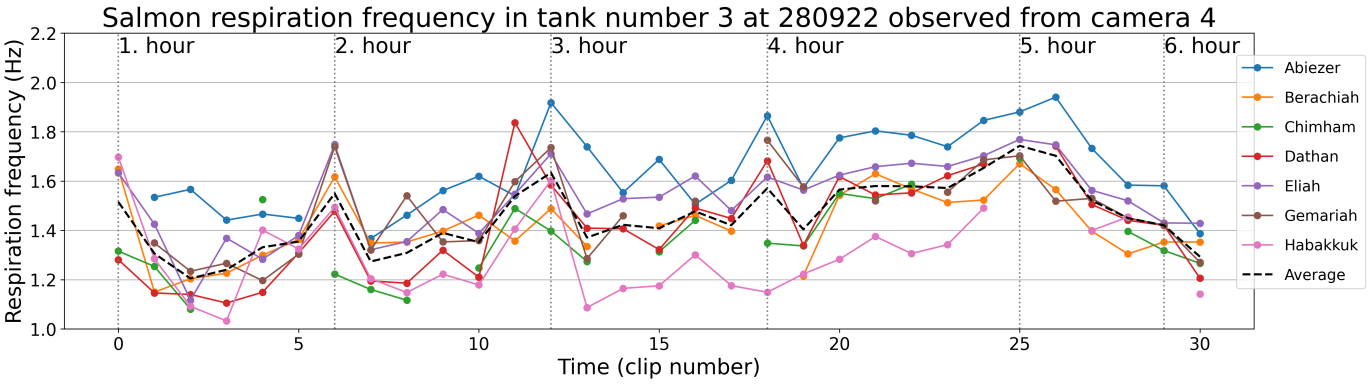


(c) Camera five, second downbreathing

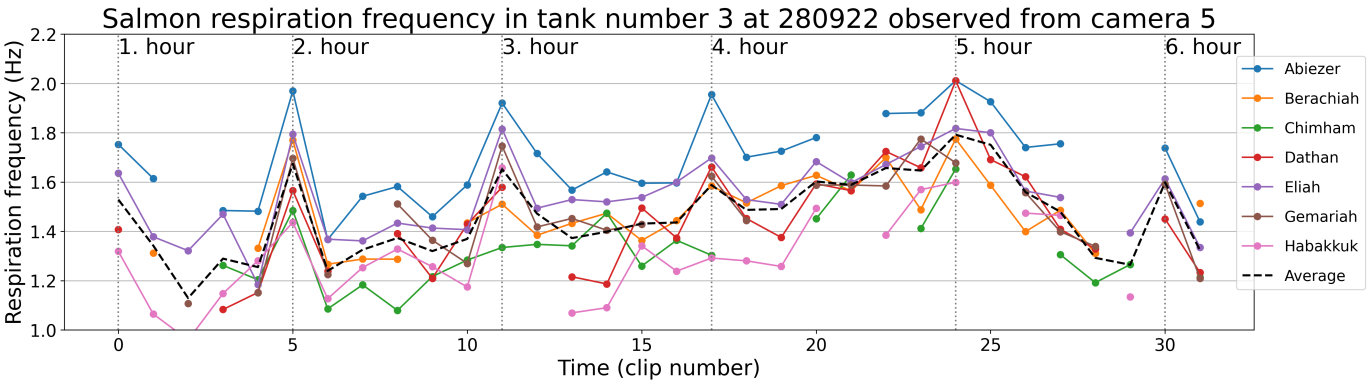


(d) Camera six, second downbreathing

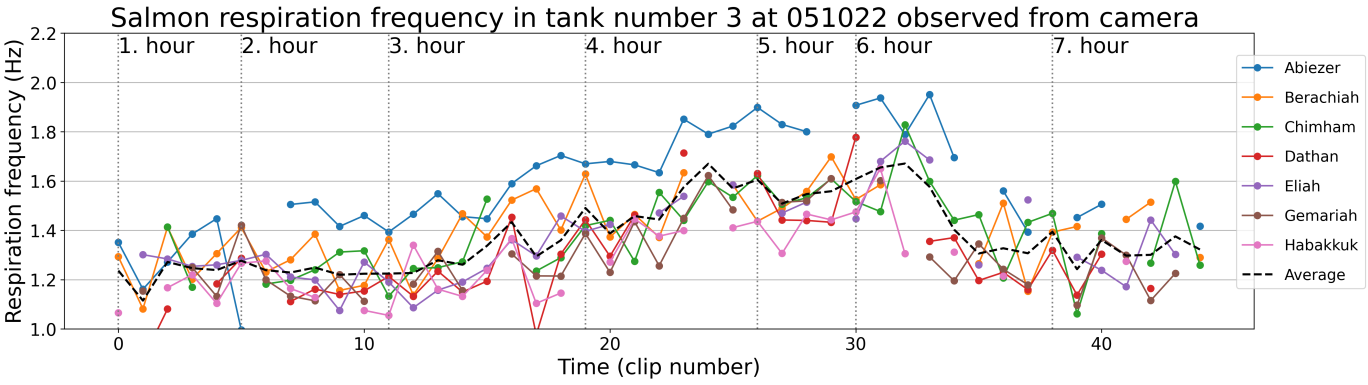
Figure E.2: Respiration evolution for tank two downbreathings



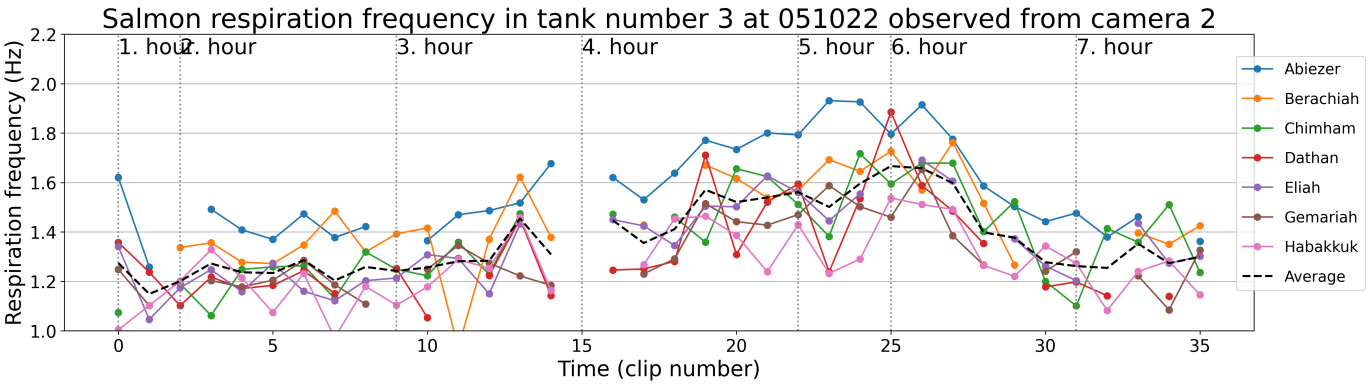
(a) Camera four, first downbreathing



(b) Camera five, first downbreathing

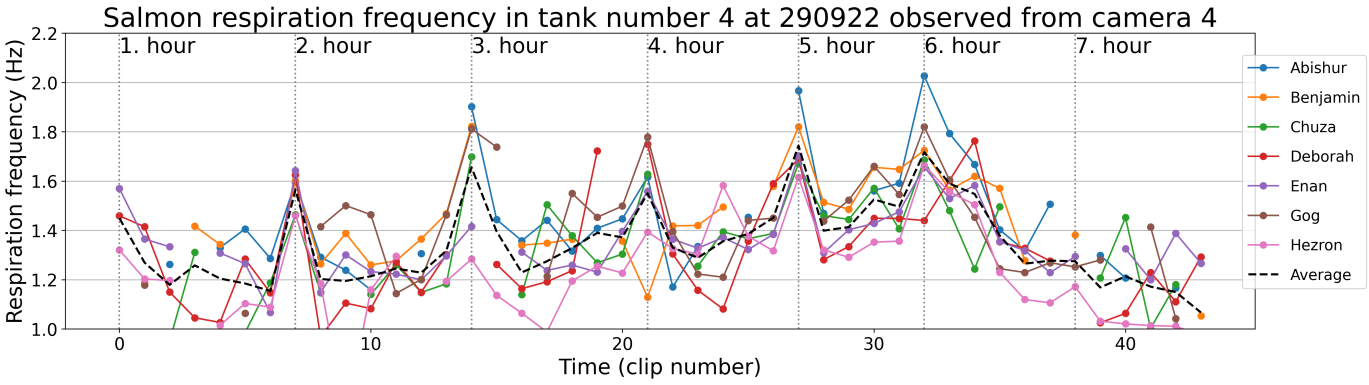


(c) Camera uncertain, second downbreathing

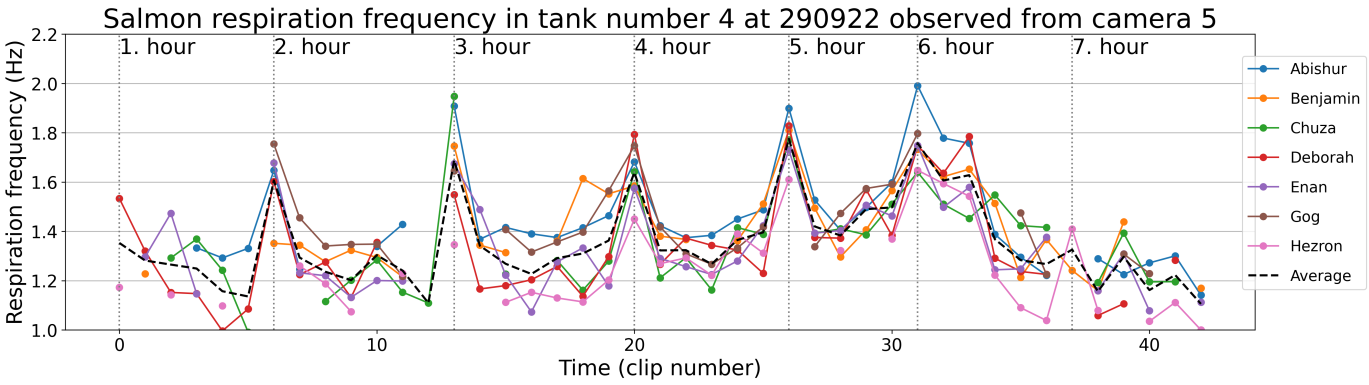


(d) Camera two, second downbreathing

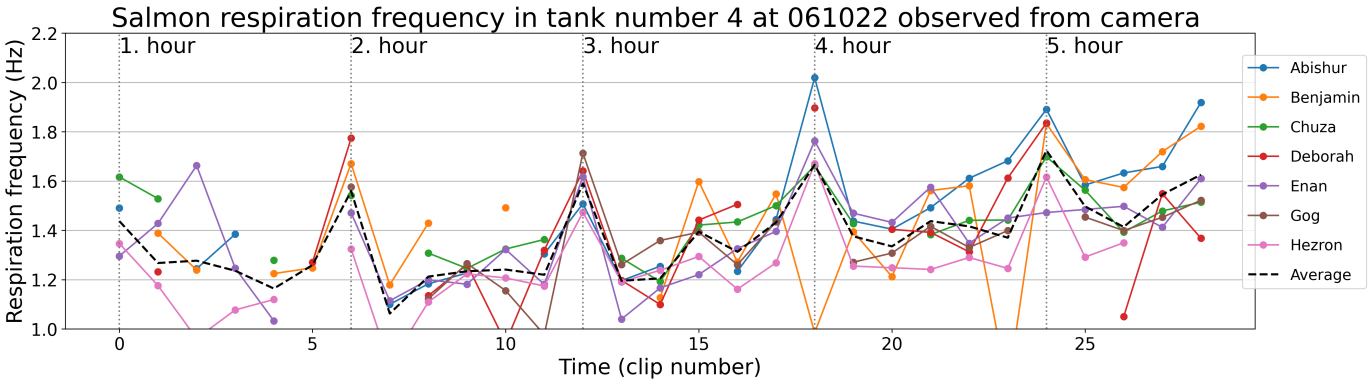
Figure E.3: Respiration evolution for tank three downbreathings



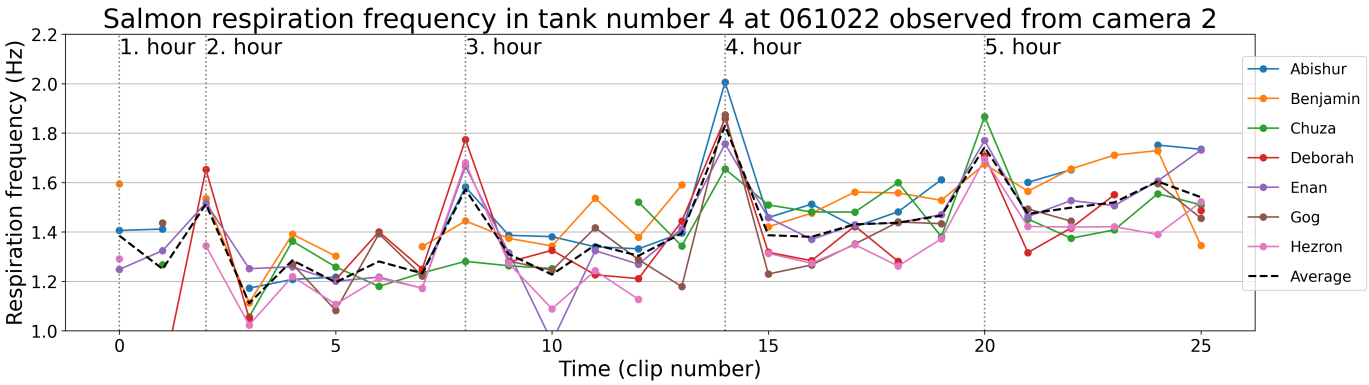
(a) Camera four, first downbreathing



(b) Camera five, first downbreathing

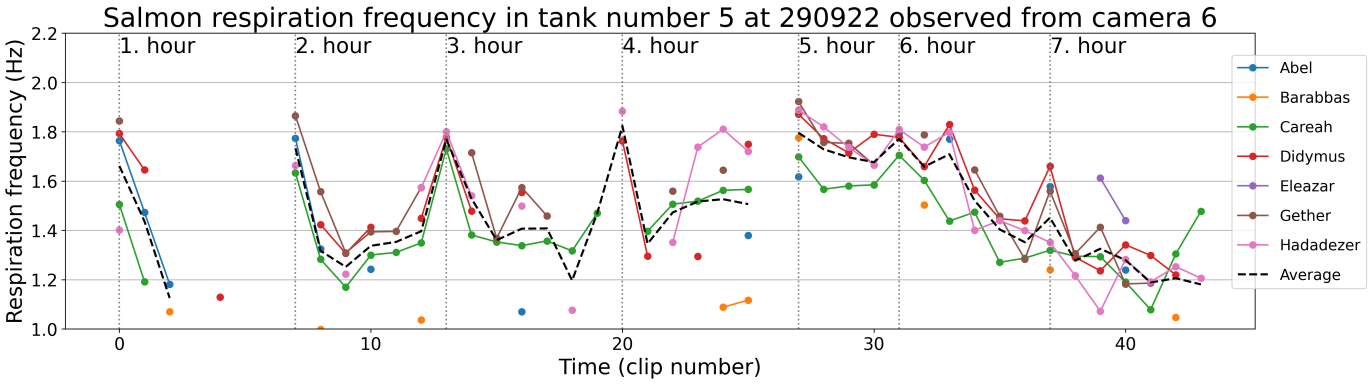


(c) Camera uncertain, second downbreathing

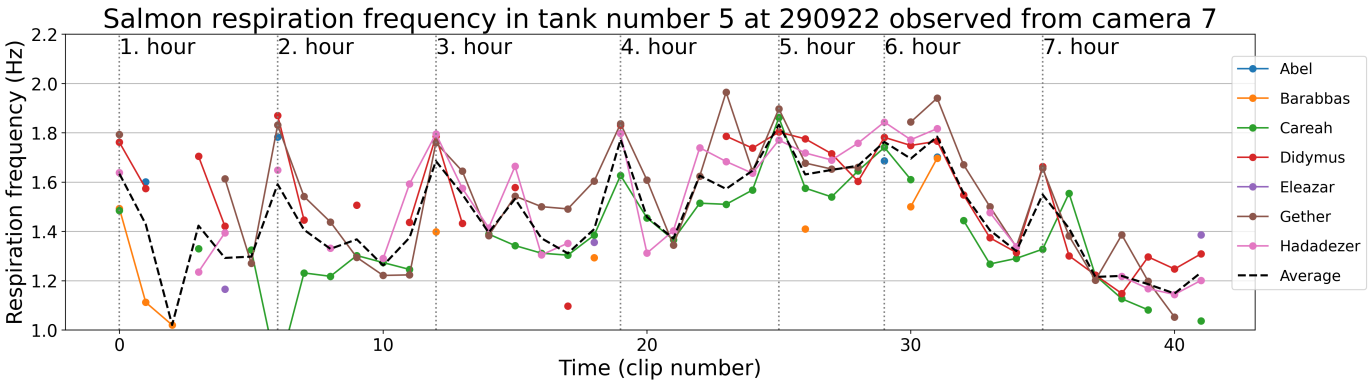


(d) Camera two, second downbreathing

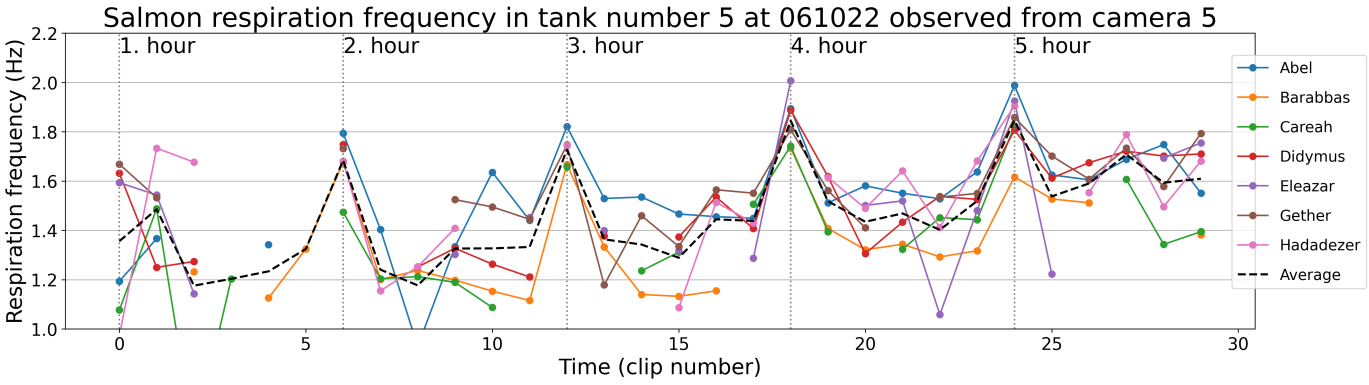
Figure E.4: Respiration evolution for tank four downbreathings



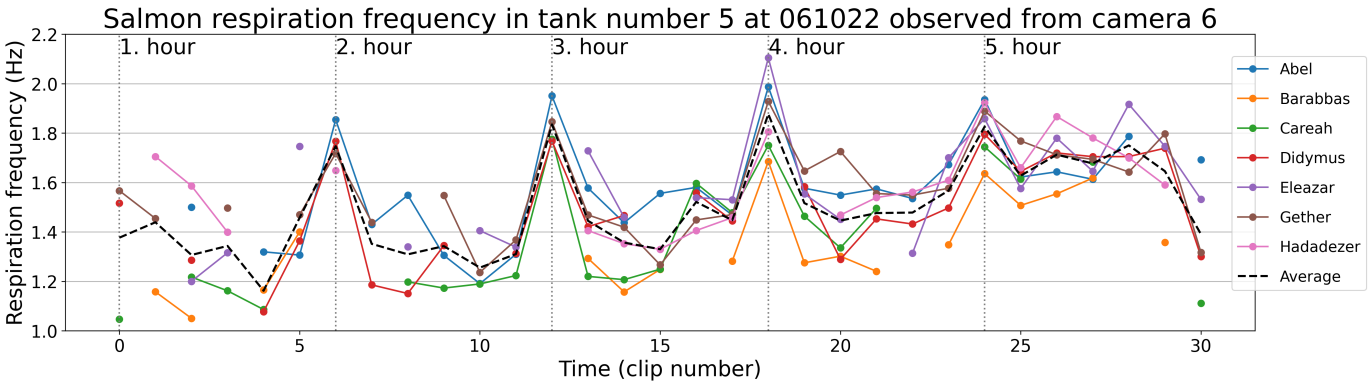
(a) Camera six, first downbreathing



(b) Camera seven, first downbreathing

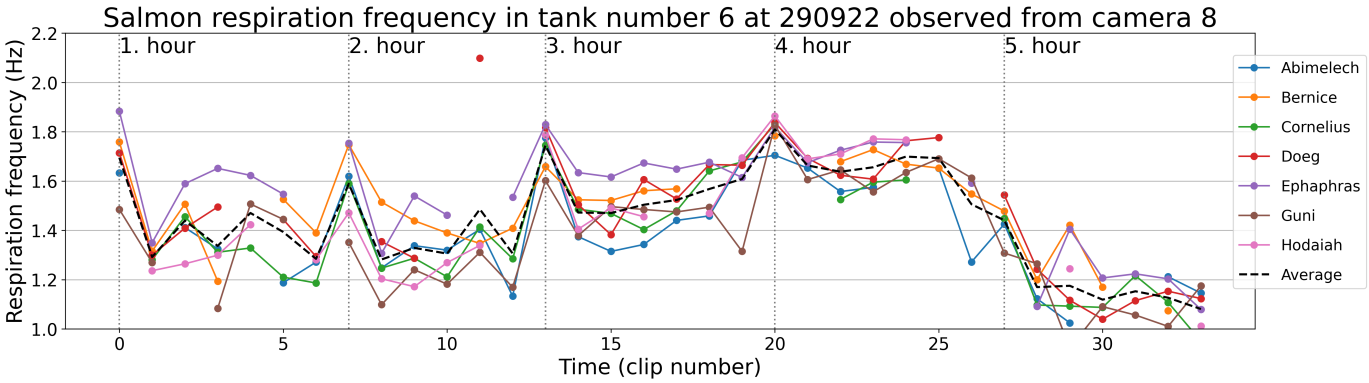


(c) Camera five, second downbreathing

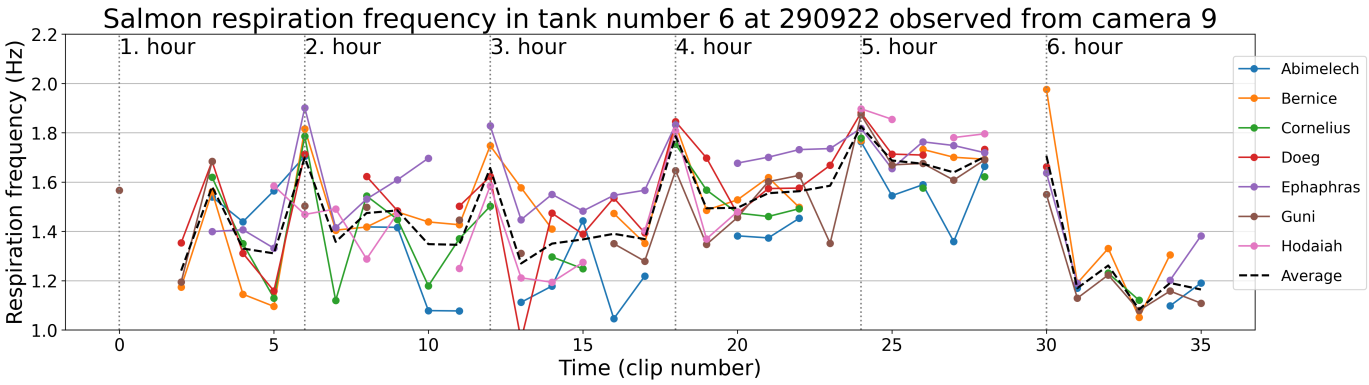


(d) Camera six, second downbreathing

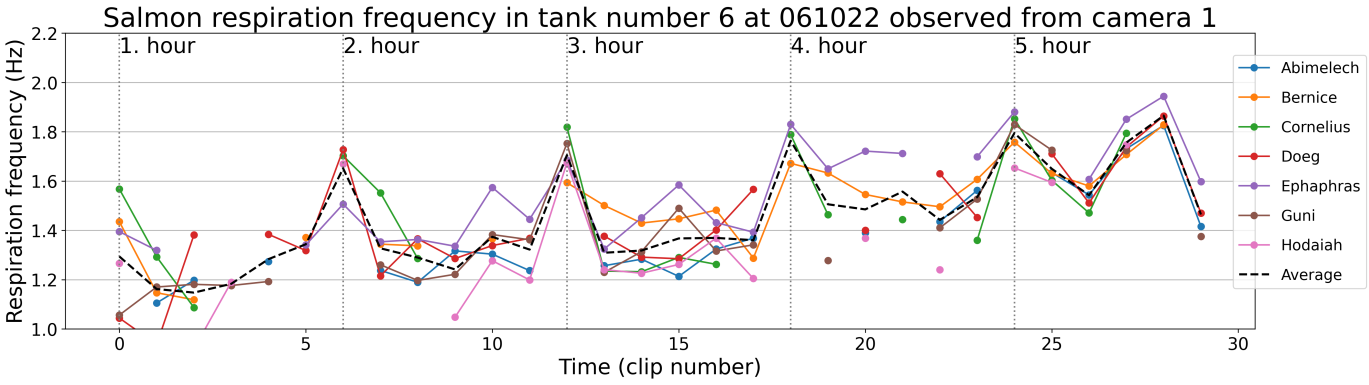
Figure E.5: Respiration evolution for tank five downbreathings



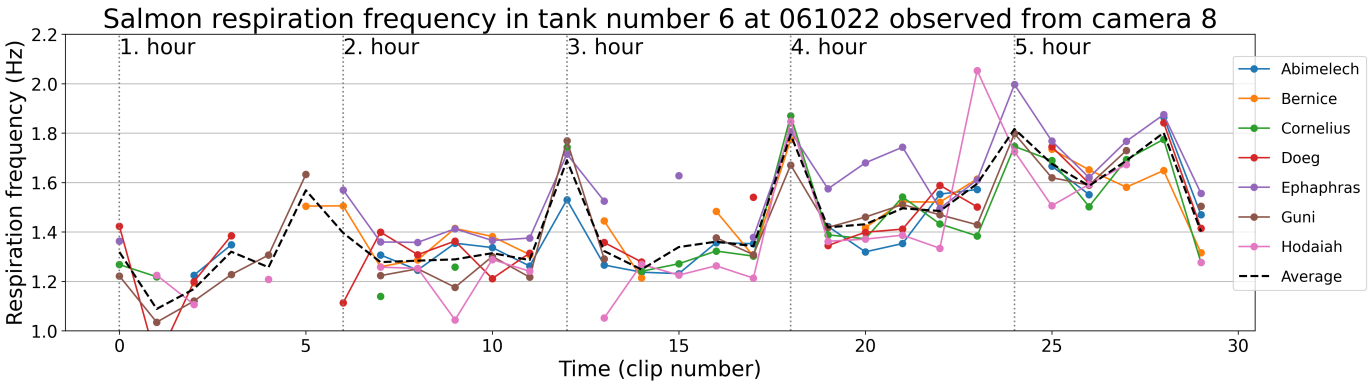
(a) Camera eight, first downbreathing



(b) Camera nine, first downbreathing

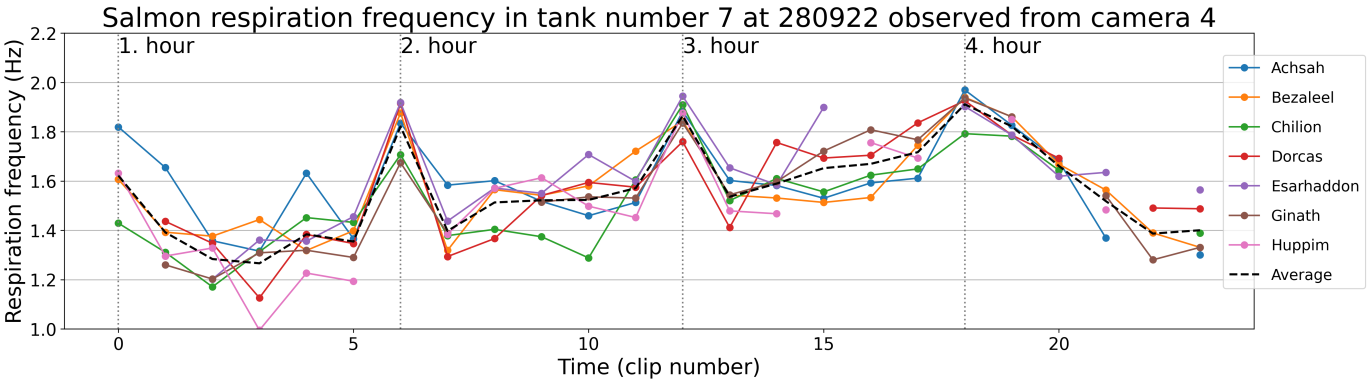


(c) Camera one, second downbreathing

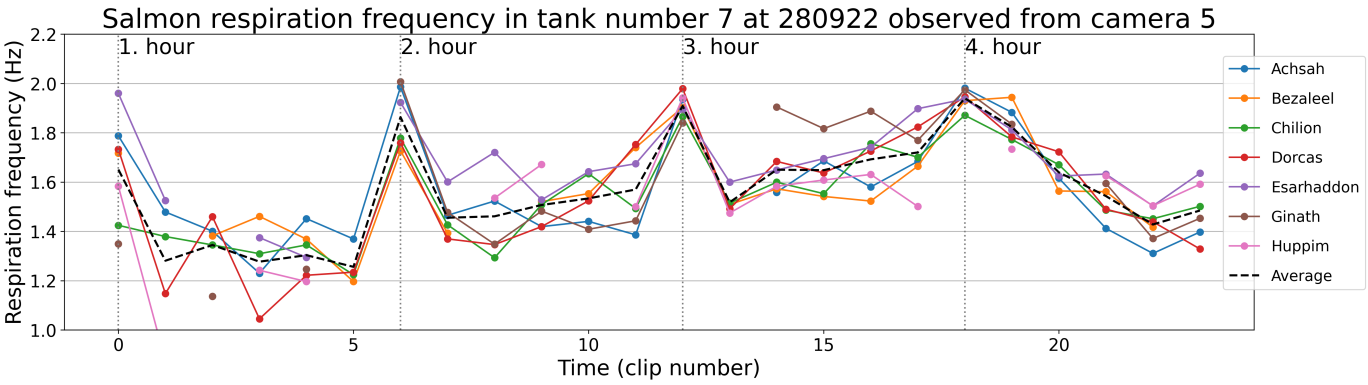


(d) Camera eight, second downbreathing

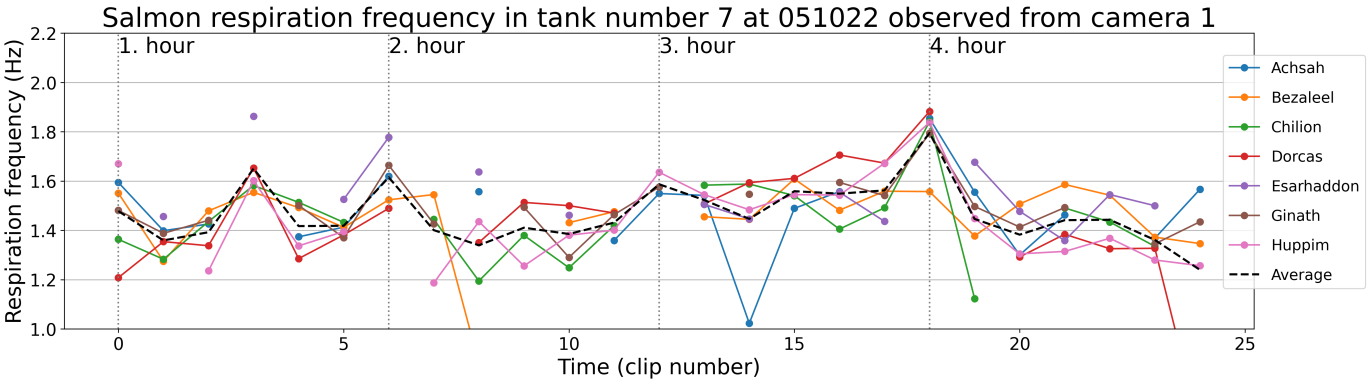
Figure E.6: Respiration evolution for tank six downbreathings



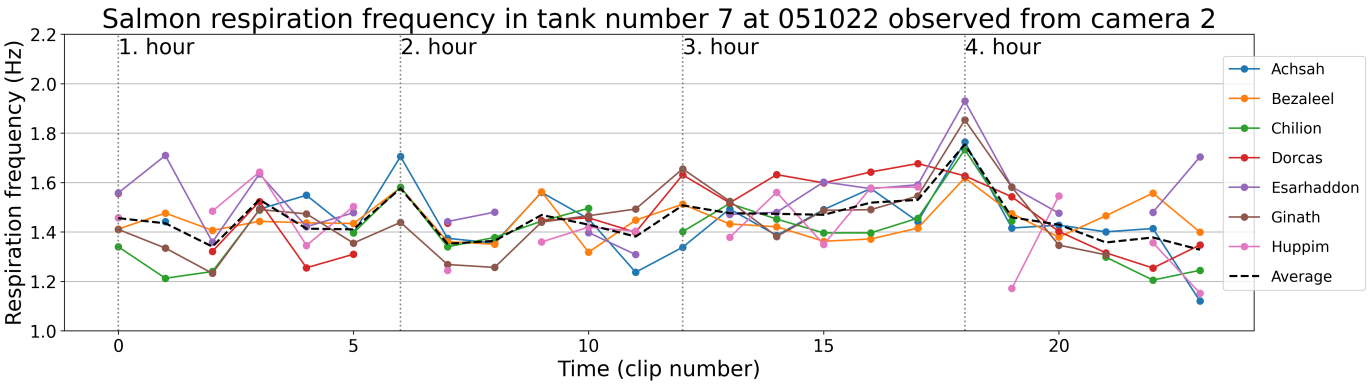
(a) Camera four, first downbreathing



(b) Camera five, first downbreathing

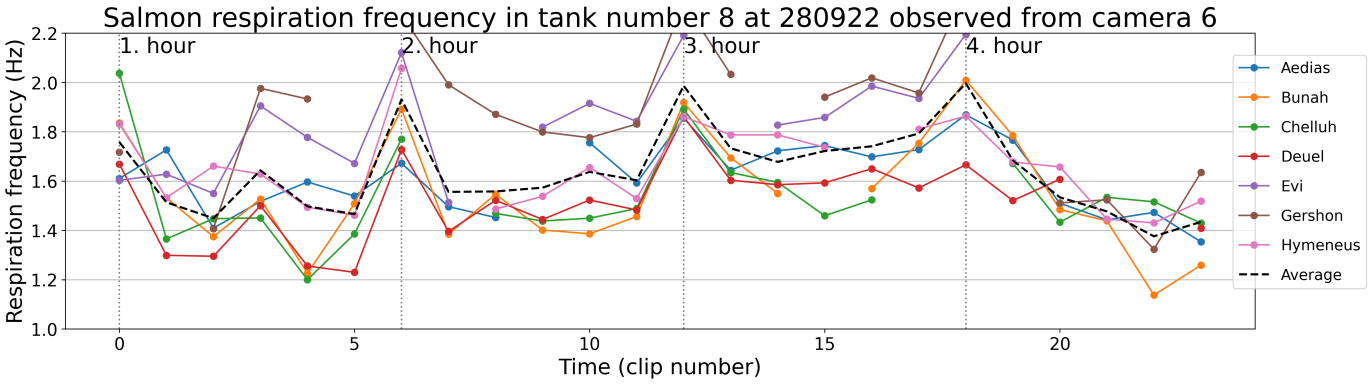


(c) Camera one, second downbreathing

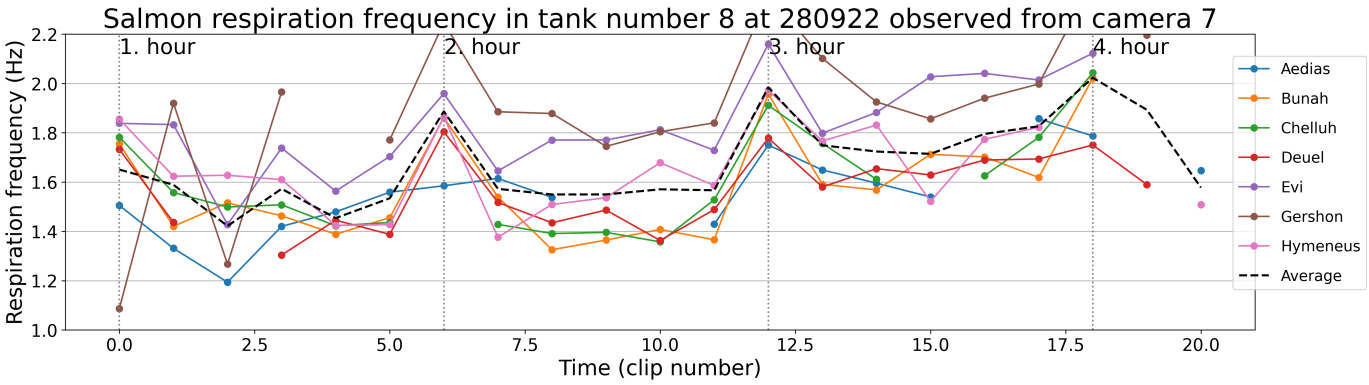


(d) Camera two, second downbreathing

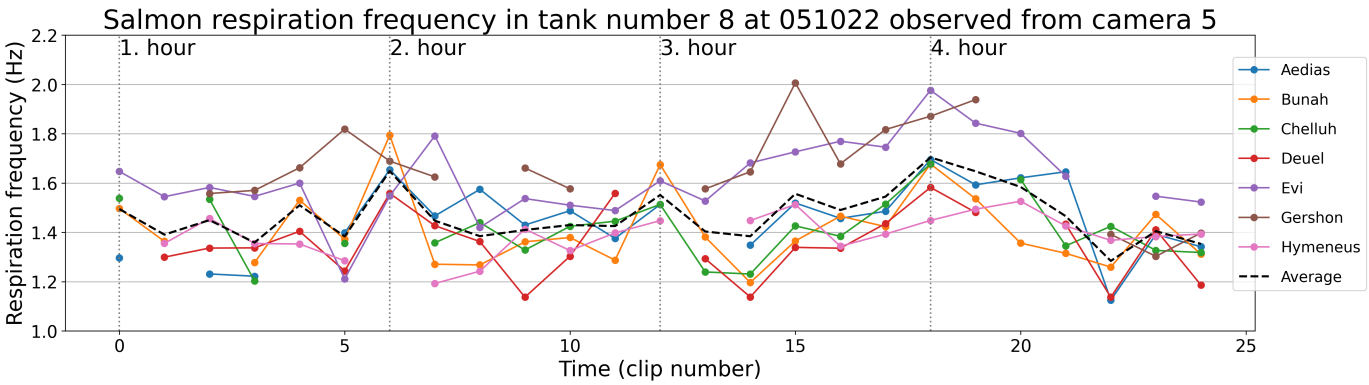
Figure E.7: Respiration evolution for tank seven downbreathings



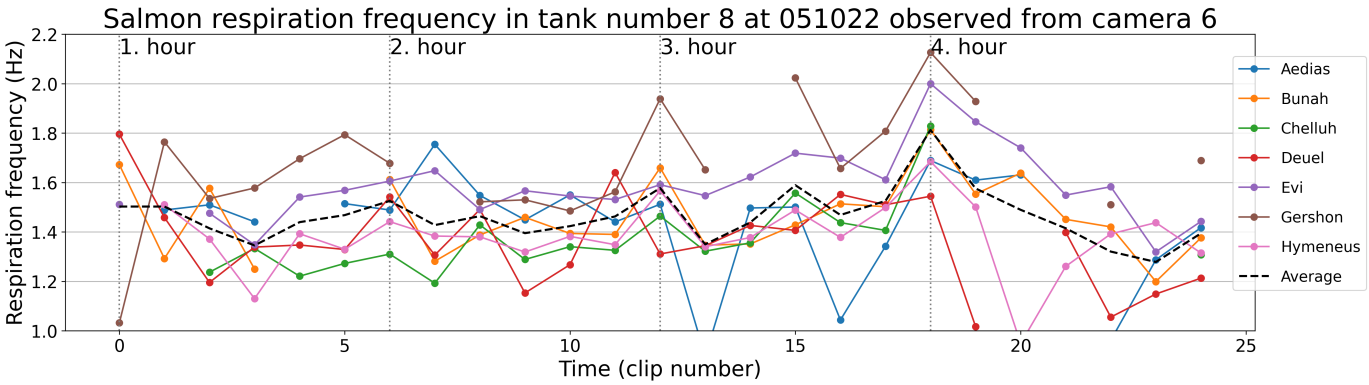
(a) Camera six, first downbreathing



(b) Camera seven, first downbreathing

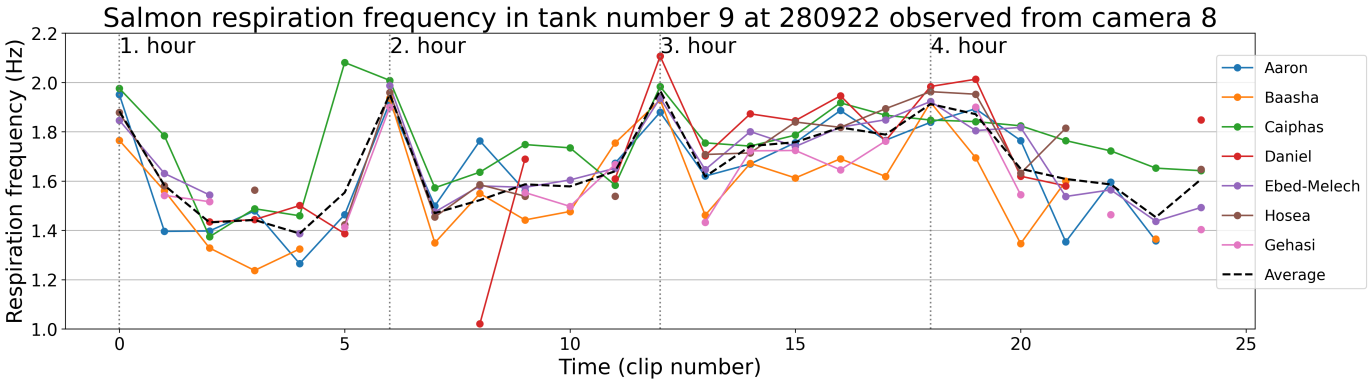


(c) Camera five, second downbreathing

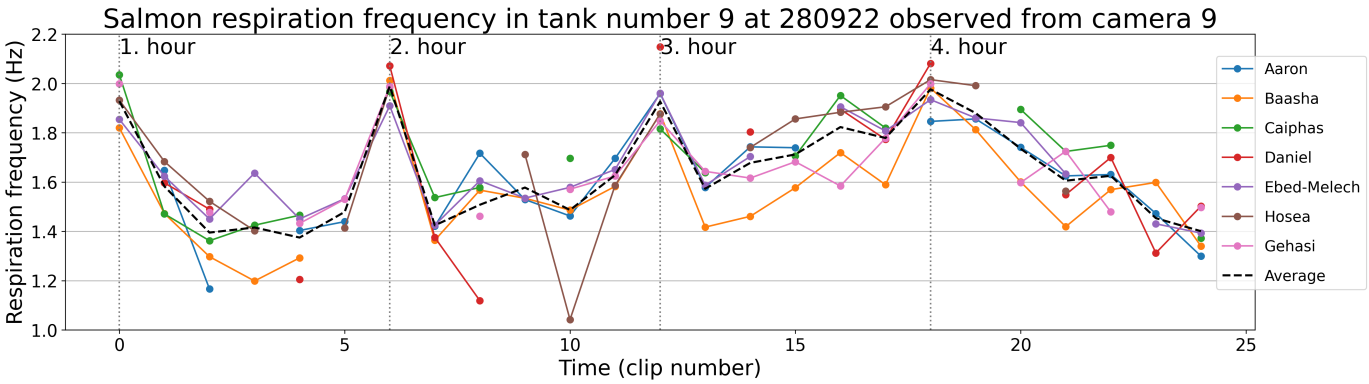


(d) Camera six, second downbreathing

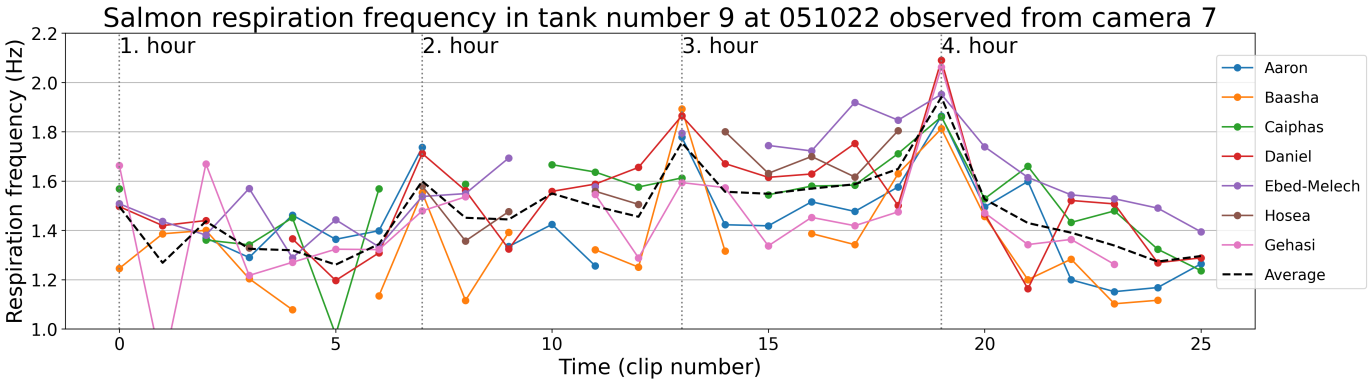
Figure E.8: Respiration evolution for tank eight downbreathings



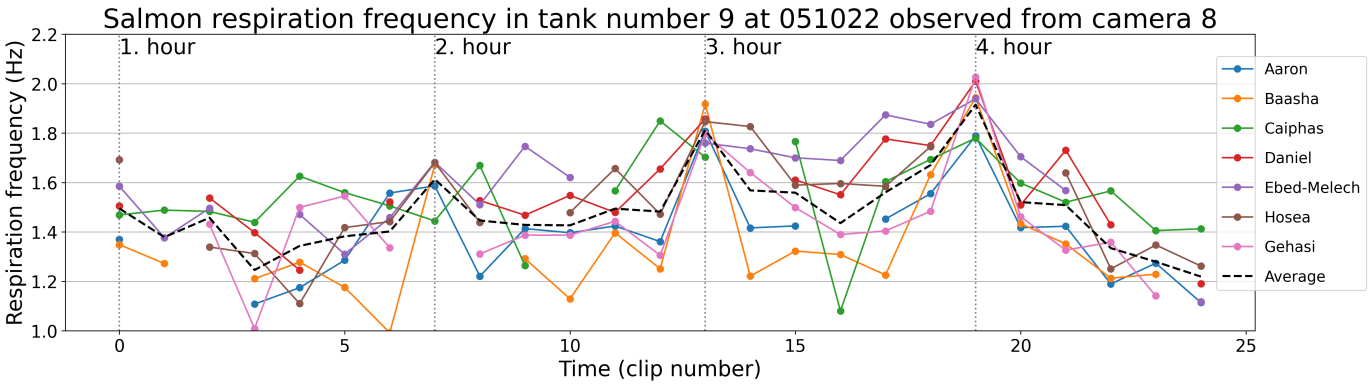
(a) Camera eight, first downbreathing



(b) Camera nine, first downbreathing



(c) Camera seven, second downbreathing

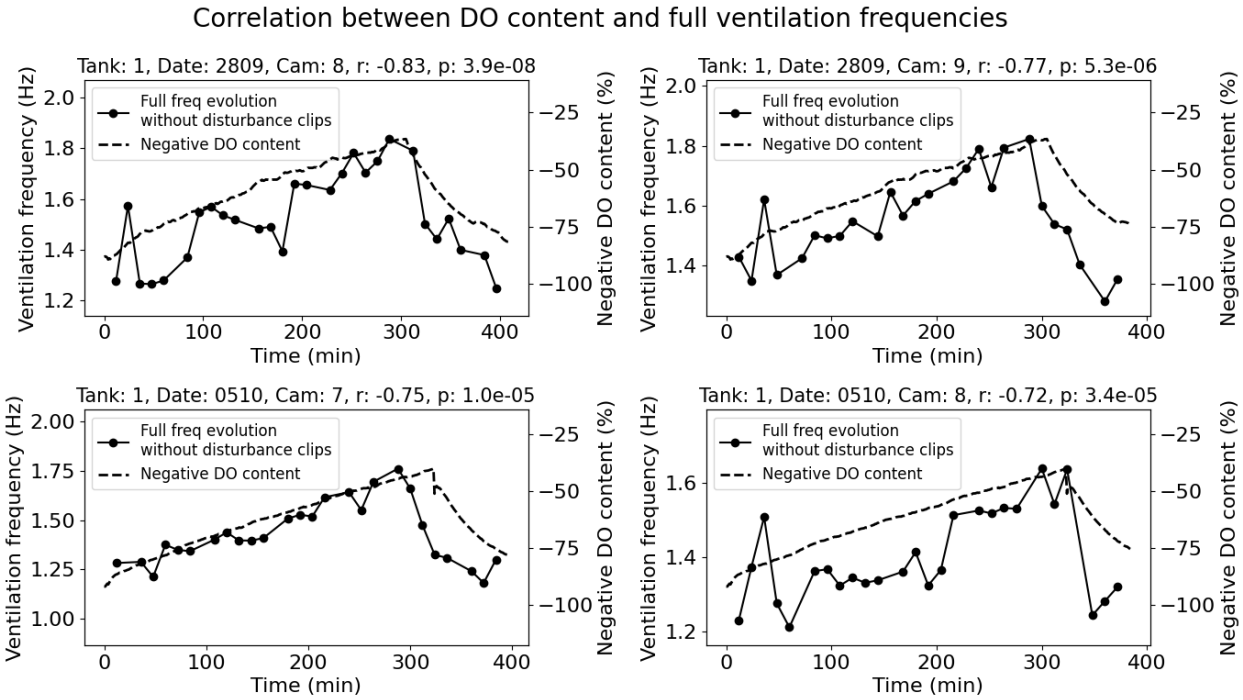


(d) Camera eight, second downbreathing

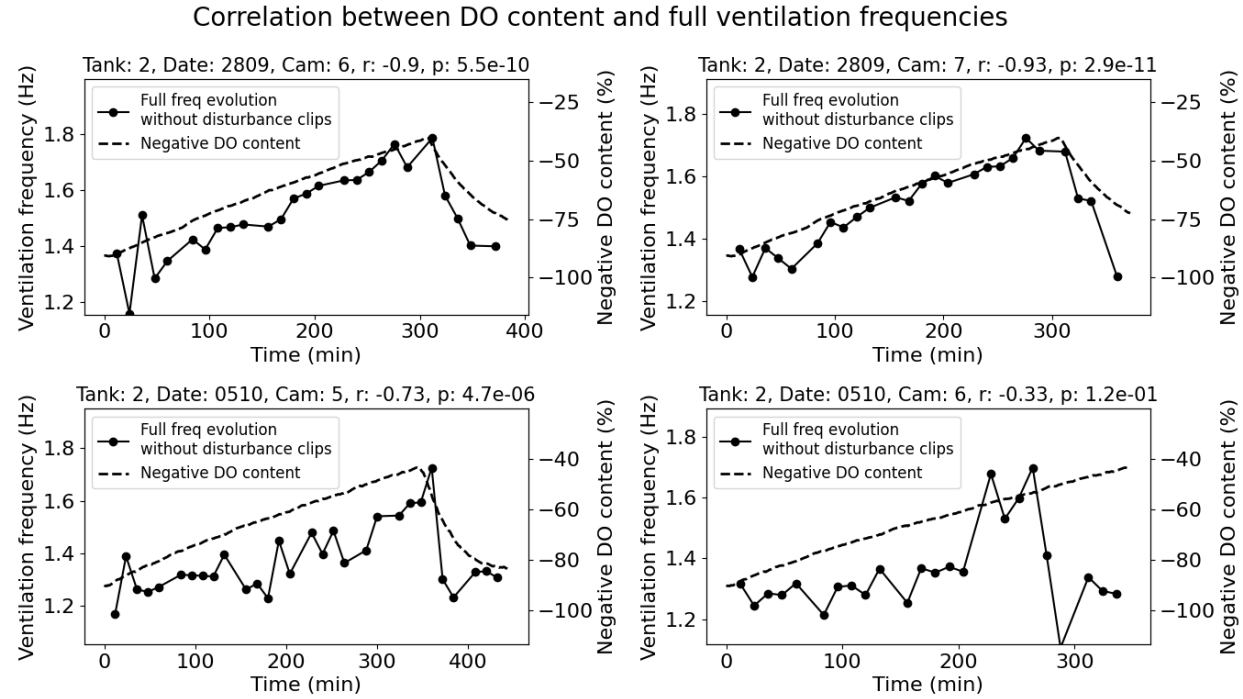
Figure E.9: Respiration evolution for tank nine downbreathings

Appendix F

All average full frequency evolutions plotted together with DO content

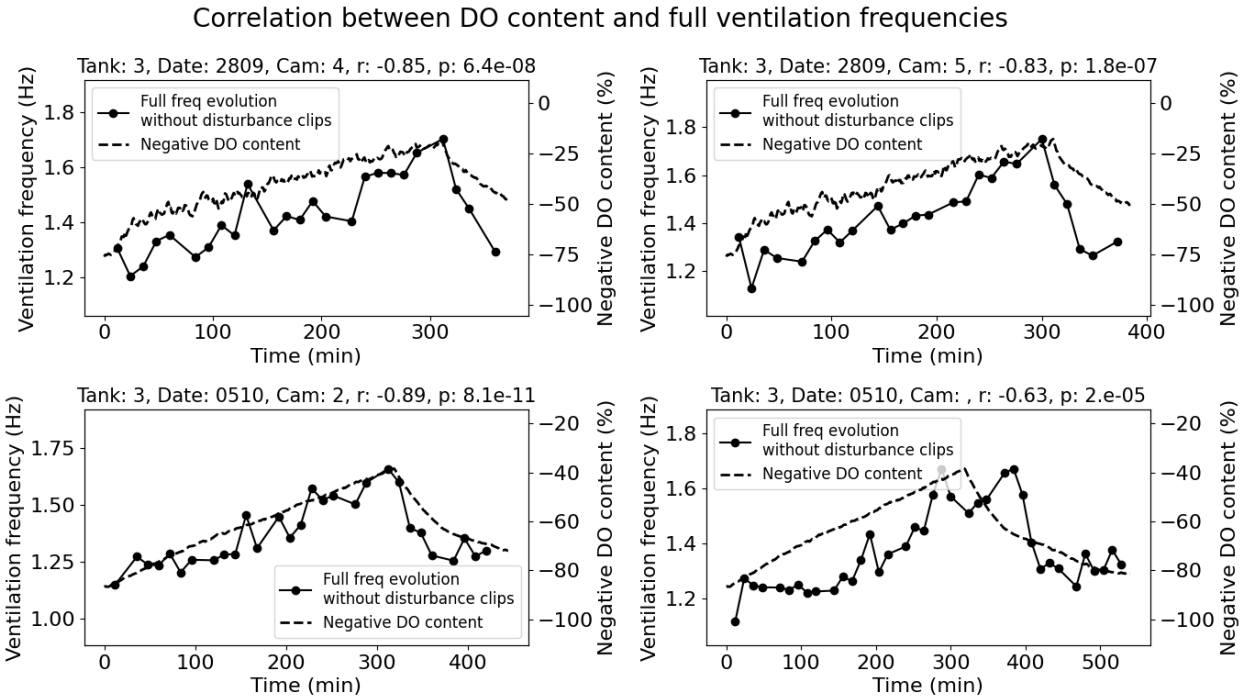


(a) DO content and full ventilation frequency of tank one.

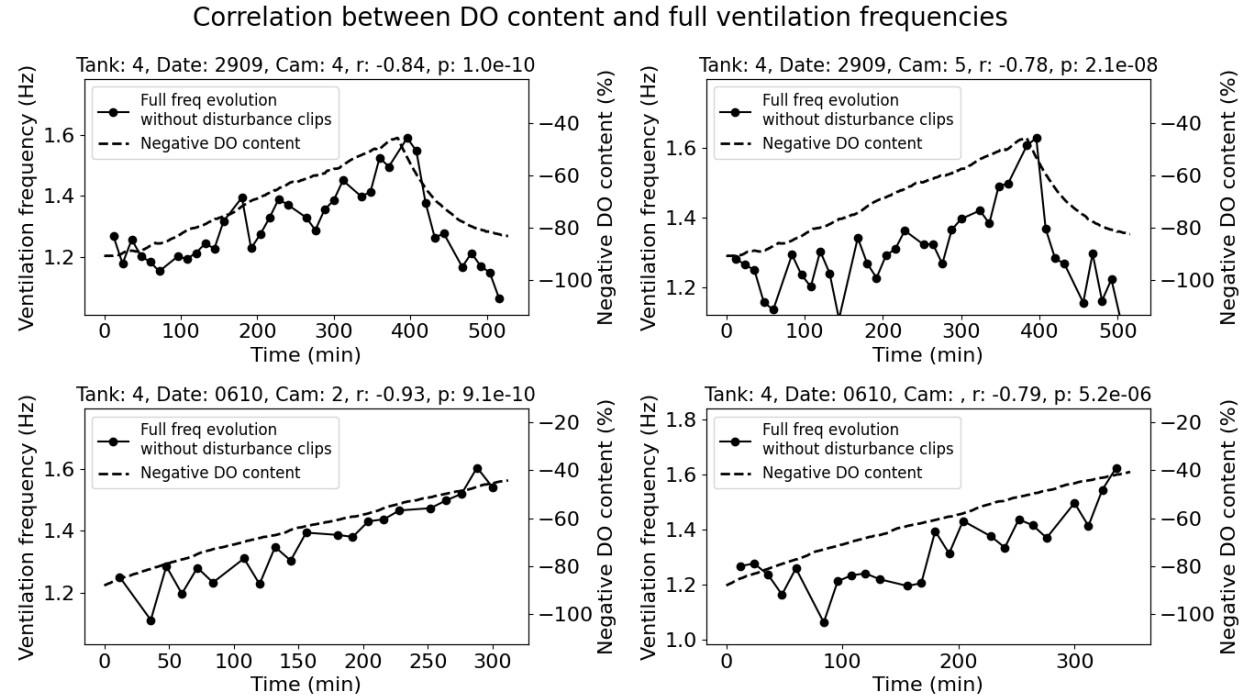


(b) DO content and full ventilation frequency of tank two.

Figure F.1: DO content and full ventilation frequency of tank one and two.

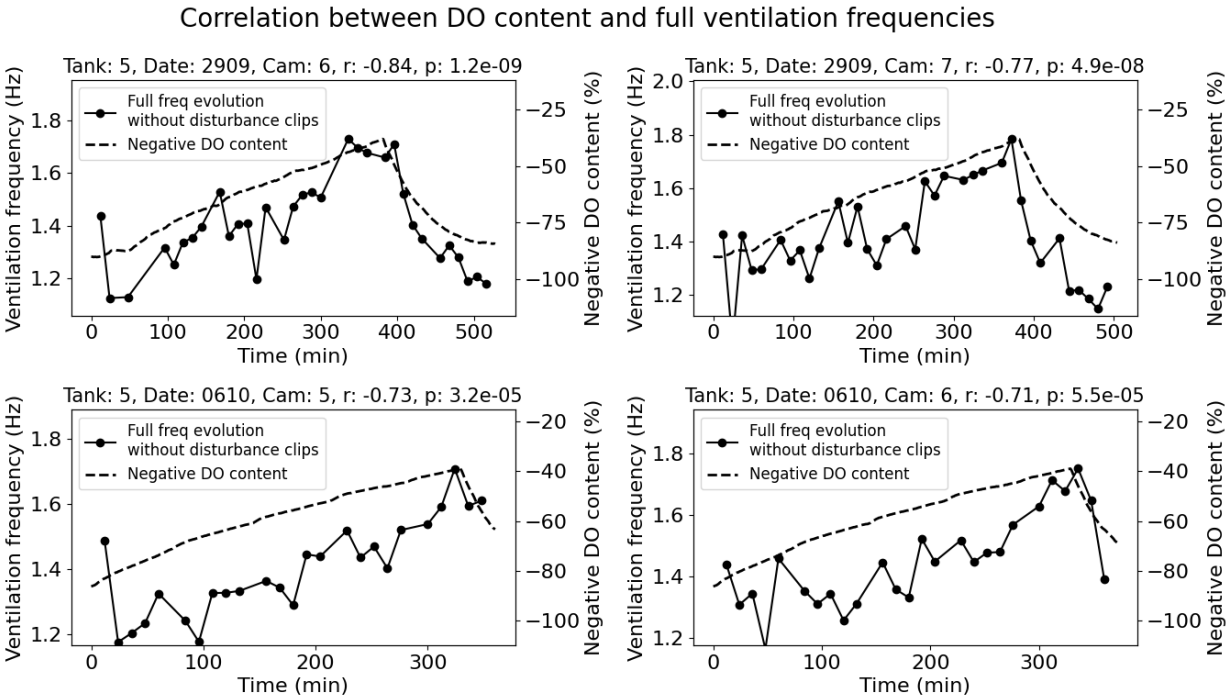


(a) DO content and full ventilation frequency of tank three.

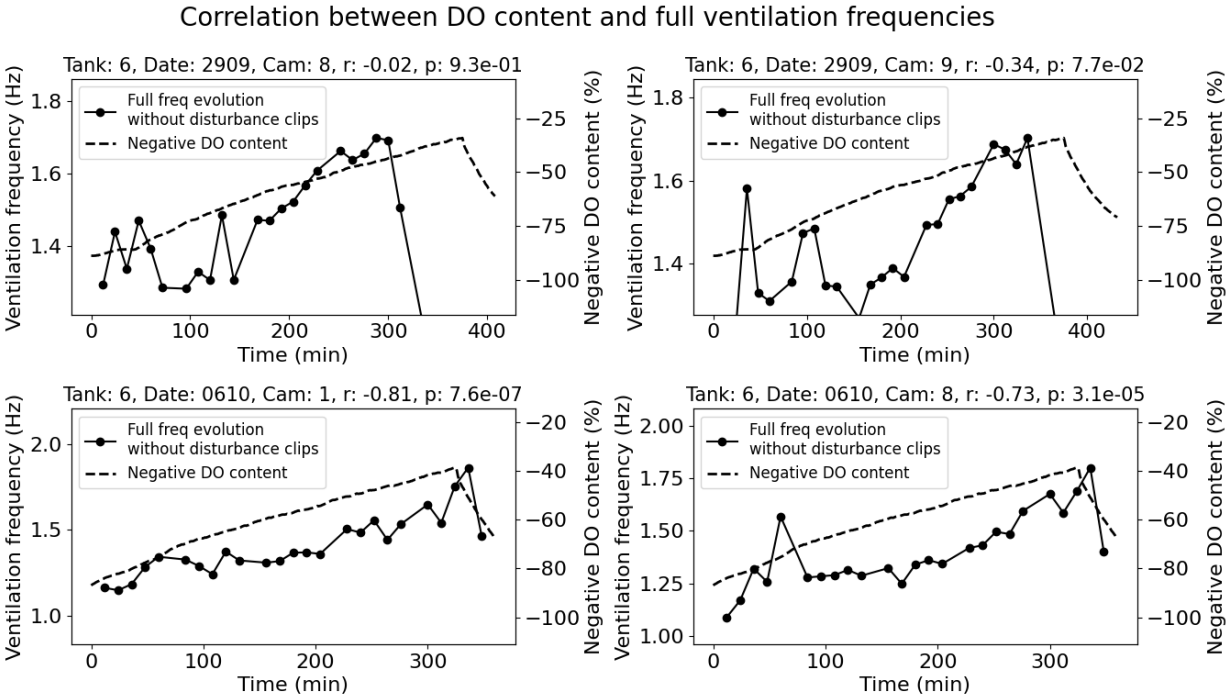


(b) DO content and full ventilation frequency of tank four.

Figure F.2: DO content and full ventilation frequency of tank three and four.

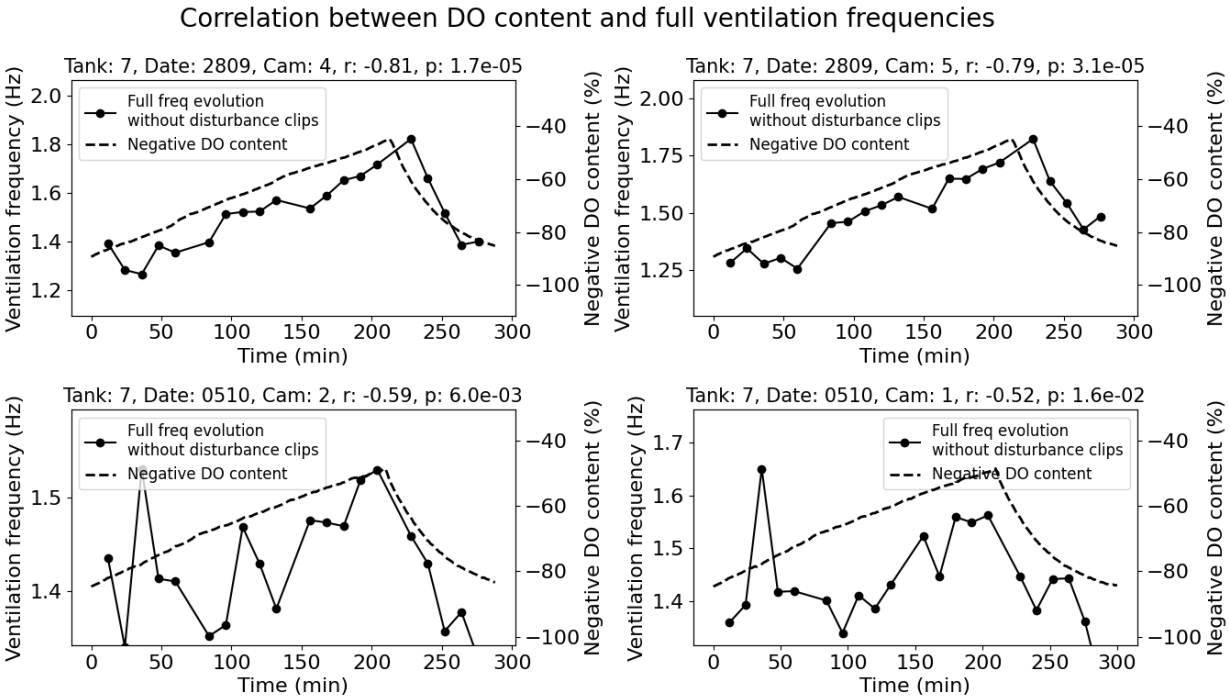


(a) DO content and full ventilation frequency of tank five.

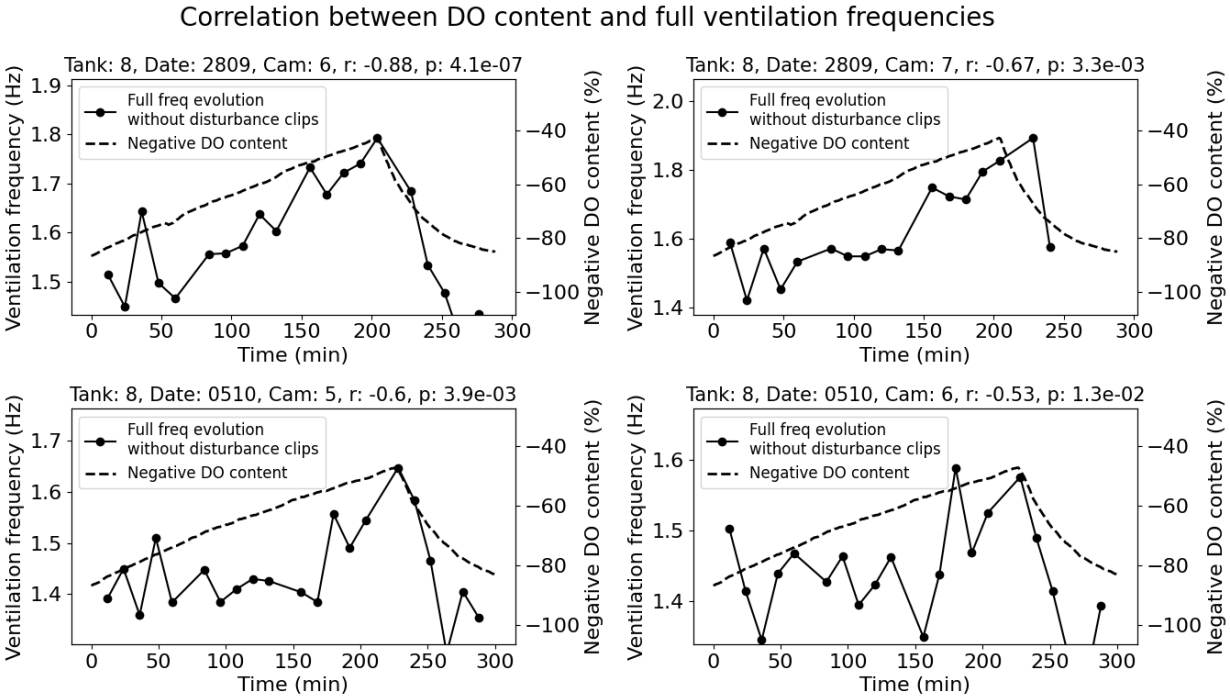


(b) DO content and full ventilation frequency of tank six.

Figure F.3: DO content and full ventilation frequency of tank five and six.



(a) DO content and full ventilation frequency of tank seven.



(b) DO content and full ventilation frequency of tank eight.

Figure F.4: DO content and full ventilation frequency of tank seven and eight.

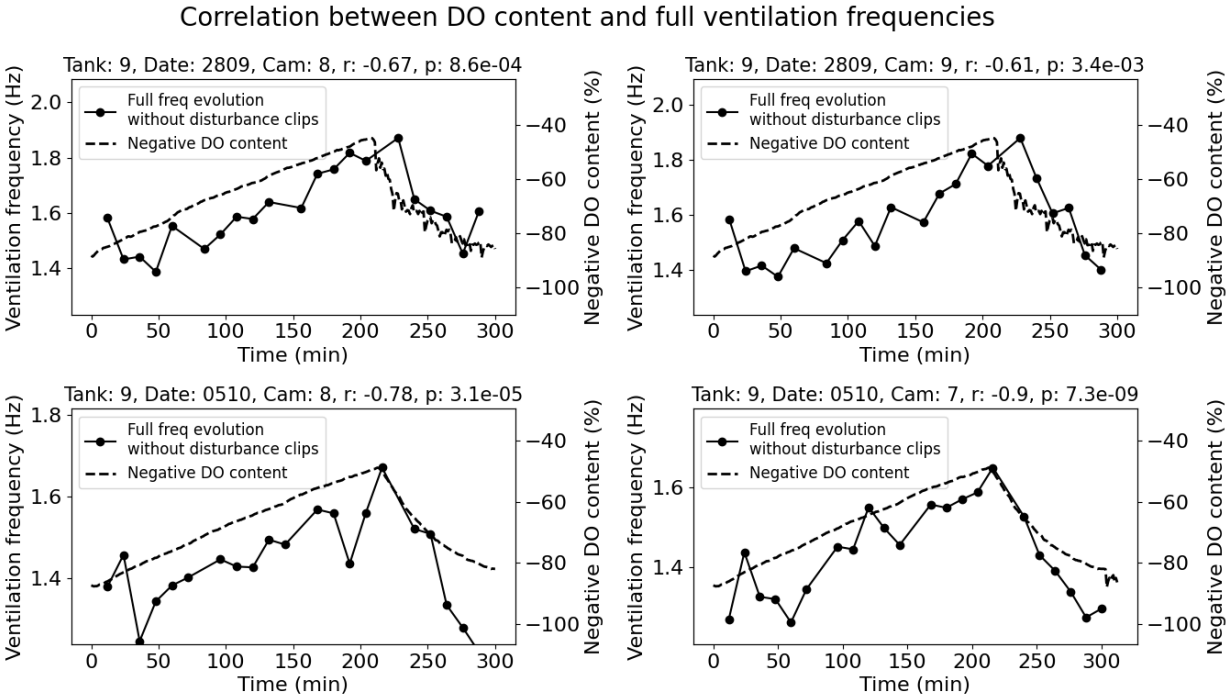
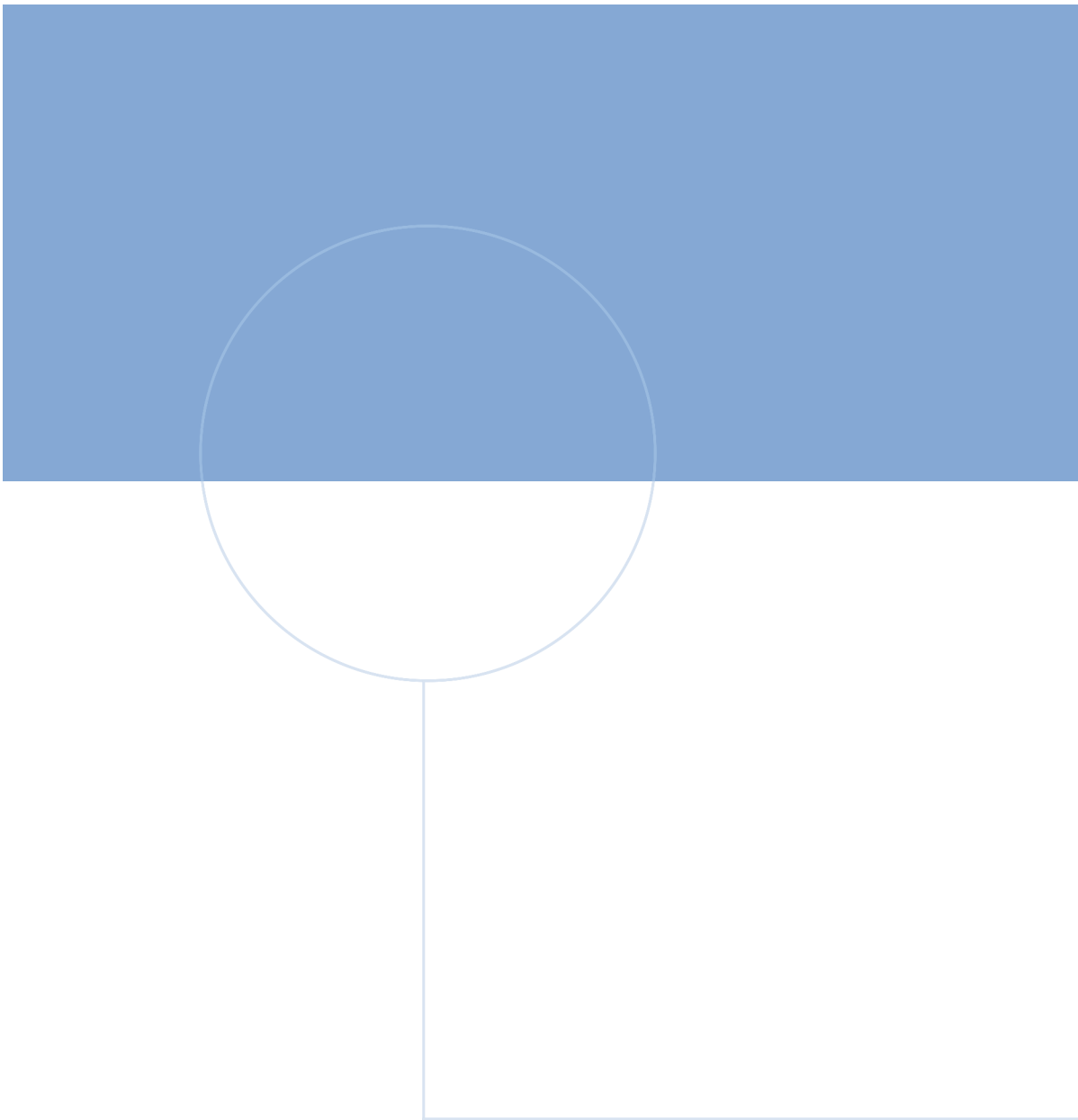


Figure F.5: DO content and full ventilation frequency of tank nine.



NTNU

Norwegian University of
Science and Technology