NTNU
Kunnskap for en bedre verden

Department of Engineering Cybernetics

TTK4550 - Engineering Cybernetics,
Specialization Project

Using deep learning to evaluate salmon welfare by respiration frequency

Author: Espen Berntzen Høgstedt

Supervisor external organization: Christian Schellewald, SINTEF
Supervisor NTNU: Annette Stahl, Department of Engineering Cybernetics

Date: December 18, 2022

Abstract

In this report we have thoroughly examined stress and welfare in salmon, how it relates to breathing, and how ventilation can be inferred by computer vision methods. The validity of ventilation frequency as a stress metric has been demonstrated by the construction of a pipeline capable of coarsely discerning shoals exposed to different levels of Dissolved Oxygen (DO) from video streams. To construct and evaluate the algorithm, an experiment was performed in which salmon was recorded at different temperatures and oxygen levels.

Preface

This document is a preparatory work for my master thesis at the Engineering Cybernetics faculty at the Norwegian University of Science and Technology (NTNU). The work was conducted between 20^{th} of September and 19^{th} of December 2022. Supervision was performed by Christan Schellewald from Sintef and Annette Stahl from NTNU. Fish experiments were done 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 were constructed in Python. All code is available on my GitHub page[1].

This preproject, and my later 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)*.

Acknowledgement

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

Contents

List of Figures

List of Tables

Acronyms

2D 2 Dimensional. 23

3D 3 Dimensional. 24

A/D Analog to Digital. 20

ACTH AdrenoCorticoTropic Hormone. 13

CNN Convolutional Neural Network. 29

CRH Corticotropin Releasing Hormone. 13

DETR DEtection TRansformer. 49

DO Dissolved Oxygen. 1, 10, 13, 39, 51, 53–55

FHF Norwegian Seafood Research Fund. 1

GPUs Graphics Processing Units. 24, 37, 52

HOG Histogram of Oriented Gradients. 22

HPI Hypothalamic-Pituitary-Interrenal. 12, 13

HSC Hypothalamic-Sympathetic-Chromaffin cell. 12, 13

ID IDentifier. 41

IoU Intersection over Union. 31, 32, 49

LM Levenberg-Marquardt. 34, 37, 41

LSTM Long Short-Term Memory. 23, 24, 54

NIVA Norwegian Institute for Water Research. 1

NMBU Norwegian University of Life Sciences. 1

NMS Non Maximum Suppression. 32, 37, 41, 49

NTNU Norwegian University of Science and Technology. 1

P2P Peak to Peak. 50

RANSAC RAndom SAmple Consensus. 2, 32, 35

RCNN Region-based CNN. 2, 25, 29–31, 36, 41, 48, 54

RGB Red Green Blue. 22, 36

ROI Region Of Interest. 29–31

SOTA State Of The Art. 21, 29

SWIM Salmon Welfare Index Model. 18, 19

Notation

1. Salmon refers to atlantic salmon (*Salmo salar*).
2. We will follow the convention of calling neural networks with single image inputs 2D, even if the images have multiple color channels. The term 3D will be reserved for networks with multiple different images in the same tensor.

Chapter 1

Introduction

Aquaculture is the fastest growing food animal sector in the world[2], 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[3]. Like all success stories, however, the industry is not without its problems. Lice infestation[4], escaping fish[5], sick fish[6], fish mortality[7], retarded growth[8] and pollution[9] are some of the issues researchers and farmers are faced with, causing economic losses and reduced salmon welfare.

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[10], causing both acute and chronic changes in the physiology of the fish. These 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[11]. All of these methods have their downsides ([12], 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 and effective way of performing this automation. It has been successfully applied to a number of related tasks, such as measuring feeding activity in salmon[13], analysing hypoxia response in goldfish[14], measuring salmon welfare from behaviour indicators[15], counting fish[16], estimation of fish size, quality assessment and species identification. All of the projects related to welfare and stress used behaviour indicators like speed, direction and location as features. By doing so, they are neglecting an easily discernible and possibly important parameter; respiration. The aim of this project is therefore to investigate how respiration relates to stress and welfare in fish, how this can be elucidated by computer vision methods, and to perform a proof of concept that displays a pipeline capable of automatic welfare assessment by computer vision based respiration analysis.

The report will start with a survey of related works, in which the first part covers fish physiology. This will handle teleost ventilation, stress, as it relates to endocrines, homeostasis and chronic responses, and finally evaluation and

quantification of fish stress and welfare. The emphasis will be on salmon and respiration due to our problem statement, although a more general scope is required occasionally to understand the broader context.

The second part of the literature review will cover computer vision methods for respiration analysis. The main focus will be on deep learning, together with methods that accommodate and support such a pipeline.

After looking into previous work, an end-to-end breathing frequency pipeline is developed. The workings of this will be presented, starting with theoretical background, before covering the practical implementation. Thereafter, a salmon respiration experiment is described, in which fish are subjected to different levels of DO and temperature. The data gathered from this experiment is used to develop and assess the ventilation frequency algorithm.

The report will end with a discussion of the main findings of the project, and an outline for subsequent work that will be performed in my master thesis.

Chapter 2

Literature survey

2.1 Survey methodology

This literature survey is split into two, where the first part covers ventilation, stress and welfare mechanisms in fish, and the second part covers computer vision methods that can be used for respiration estimation. Both inquiries start with material directly related to the problem statement, and auxiliary literature is gathered to expand on interesting findings discovered during the survey. For the biological examination, the main keywords used during searches were salmon, aquaculture, stress, catecholamines, cortisol, welfare and ventilation/respiration/breathing. The technological part was organized a bit differently, as I already had knowledge of interesting neural networks and traditional methods. Due to this, literature was mainly used to expand and correct loose initial ideas. The primary search engines used during the survey were google, google scholar[17] and Science Direct[18].

2.2 Fish ventilation

Teleosts absorb oxygen for metabolism by gas exchange between erythrocytes and water in lamellae attached to the gill arch[19]. Two modes of breathing have been observed; buccal pumping and ram-ventilation[20]. 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[21]. In the latter, high 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 nerves, respectively[22]. 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 forma-

tion and the cranial nerves, and is through this capable of altering respiration. In addition to efferent pathways, the cranial nerves posses 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 chemoreceptors[23]. As the O_2 balance of the fish gets disturbed, these organs transmit afferent signals to higher brain centers, causing altered ventilation frequency and amplitude[24]. It seems different population of receptors are responsible for different ventilation responses[25]. Modulation of breathing frequency is mainly caused by branchial receptors, with orientation both towards the water, and towards the blood. The ventilation amplitude, however, is also altered by chemoreceptors that are located at extrabranchial sites, such as the orobranchial cavity. These receptors are usually oriented externally.

In later years, it has also been found that CO_2 sensitive chemoreceptors are active in modulation of the respiration of fish[24]. Milsom found in his review that these were exclusively located at the gill arches, facing externally[25]. Some studies have found resiratory response to internal stimuli with stable arterial O_2 levels, such as Wood, which discovered carbonic anhydrase injected in fish attenuates post-excercise hyperventilation with stable O_2 levels[26]. This is likely caused by PH sensitive structures close to the fish arteries.

Looking specifically into salmon, the chemoreceptive sensors linked to hypercarbia is primarily responding to CO_2 concentration, as opposed to H^+[27]. CO_2 causes a significant increase in both ventilation frequency and amplitude, while H^+ only results in a minor amplitude increase. Furthermore, O_2 chemoreceptors in coho salmon are only located in the first gill arch, and innervated by the IX cranial nerve[28].

Putting this together, salmon regulate their ventilation by a complex interplay of neural processes, receiving information about about oxygen and CO_2 levels external and internal to the salmon from the ascending tract. The breathing frequency this network gives rise to depends on stressors, environment and size, as shown in the summary presented in table 2.1. Most salmon seem to fall in the ventilation range [1, 2.5] Hz, with 1 Hz being the resting respiratory frequency.

2.3 Stress in fish

Fish, like humans, get stressed when subjected to real or perceived stressors. A common definition of "stress" has been the nonspecific response of the body to any demand made upon it[34], however this might be a too narrow description. As we shall see, stressors like high temperature and hypoxia elicit different physiological responses in fish. Even with some ambiguity of what stress is, researchers are largely agreeing on the main endocrine stress pathways, the Hypothalamic-Sympathetic-Chromaffin cell (HSC) axis and the Hypothalamic-Pituitary-Interrenal (HPI) axis. The cascade of mechanisms following the acti-

Author	Weight (g)	resp. freq. (num/min)	Stressor
Hosfeld[29]	29.2	73	control
Hosfeld[29]	29.2	65	125% DO
Hosfeld[29]	29.2	57	145% DO
Hosfeld[29]	29.2	53	178% DO
Knoph[30]	430	56	control
Knoph[30]	430	61	30 mg/l ammonia
Knoph[30]	430	60	56.2 mg/l ammonia
Millidine[31]	1.8-12.6	61-140	Diverse
Erikson[32]	1641g	161	Commercial transport
Erikson[32]	1641g	137	Open transport
Erikson[32]	1641g	153	Closed transport
Erikson[33]	4410g	55 and 64	Control
Erikson[33]	4410g	80 and 81	Crowding

Table 2.1: Overview over salmon ventilation frequency

vation of these is highly diverse, and conflicting findings are not uncommon. In the following we will review the most important stress mechanisms in fish by looking into the three main layers of physiological responses[35].

2.3.1 Primary responses

The primary responses describe the endocrine changes that occur in the fish immediately after being exposed to a stressor. The main agents are catecholamines and corticosteroids, which are effectuated by the HSC axis and the HPI axis, respectively[10].

Cathecolamines, of which ephinephrine and norephinephrine are the main actors, are controlled by direct action of the hypothalamus on the chromaffin cells of the head kidneys by the sympathetic branch of the autonomous nervous system[10]. The chromaffin cells and head kidneys are homologues to the adrenal medulla and adrenal gland, respectively[36]. When stimulated, the chromaffin cells release cathecolamines into circulation. Catecholamines can also be excreted from chromaffin cells in direct response to reduced oxygen levels, without being stimulated by the hypothalamus[37].

The HPI axis works a bit differently, as it involves hormonal control[10]. It 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, which is homologous to the adrenal cortex, to synthesize and release cortisol. Beyond the main pathway described here, several other endocrines have been shown to support, inhibit or replace CRH and ATCH in the cortisol pathway of fish.

Some fish circumvent these main endocrine axes, and rely on sympathetic nerve activity as a response to stress instead[38]. Salmon is not in this group,

so we will not pursue this peculiarity 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[35].

Cardiovascular response

Starting with the cardiovascular system, stress has a pronounced modulatory effect on the heart, however the response depends on the stressor. Aerobic exercise in lingcod[39] and temperature increase in rainbow trout[40] cause tachycardia, while bradycardia emerge from hypoxia[19] and burst swimming in lingcod[39]. The heart rate of teleosts are mainly controlled by cholinergic and adrenergic neural pathways ([12], chapter 7), so the observed frequency modulations are effectuated through the nervous system. Other cardiovascular effects are instead caused by direct endocrine actions. Both catecholamines[41] and cortisol[42] increase cardiac contractility in rainbow trout, and catacholamines induce vasoconstriction in the systemic circuit, and vasodilation in the branchial capillaries[10] of fish. The dilation and constriction synergy cause higher gill perfusion rate, increasing oxygenation of the blood. Furthermore, adrenaline has been shown to increase the permeability of gill epithelium[43], boosting gas exchange additionally. In salmon, CO_2 has been shown to elevate heart rate and systemic resistance[27].

Catecholamines also increase the blood transport capacity in fish. This is caused by β-adrenergic receptors that boost the Na^+/H^+ exchange of erythrocytes, moving charge from the extracellular to the intracellular fluid. This blood acidosis increases the hemoglobin saturation, boosting plasma oxygen[44]. The blood is further oxygenized by a raise of hematocrit by catecholamines acting on the spleen[10]. Salmon shows a particularly strong increase in hemoglobin saturation due to its large amount of β receptors on the blood cells.

Metabolism

With regards to metabolism, catecholamines are the primary actors. They cause elevated glycogenolysis, slightly elevated gluconeogenesis, and might cause an increase of free fatty acids[10]. Cortisol has some glycemic effects, although less prominent than the ones of catecholamines. It seems to promote gluconeogenesis in particular, as well as contributing to some extent in lipolysis. The result of these metabolic actions is hyperglycaemia of the fish.

Ventilatory response

The relationship between stress, hormones and ventilation in fish is controversial[10]. Aota[45] found evidence for a positive correlation between cat-

echolamines and ventilation response under acidotic and hypoxic conditions, while Kinkhead[46] measured a depression of ventilation frequency during catecholamine administration. Randall[47] also strongly advocated for a catecholamine mediated ventilatory increase in response to stress. Both Aota and Randall note that the ventilation increase is suppressed by propanolol injection, which is a known β-adrenergic antagonist. This strengthens the hypothesis of a positive correlation between catecholamines and breathing. Furthermore, the higher metabolic rate during stress necessitates oxygen, hence an increase in respiration seems obvious[47]. Catecholamines seem to have an effect also in the case of severe hypoxia, where feeding muscles are recruited in an attempt to stave off further oxygen deficit[22].

Hydromineral balance

As previously seen, catecholamines cause increased diffusion through the gills, disturbing the delicate hydromineral equilibrium of teleosts. Cortisol seems to have some effect on this balance, but the picture is unclear. Langhorne found no stimulatory effect of cortisol on sodium-potassium ATPase activity[48], while Gallis found both cortisol dependency and cortisol independence of ATPase in mullets acclimating to water with different osmolarity[49]. Laurent looked into the chloride cells, and found that they proliferate when stimulated with cortisol[50]. The chloride cells are responsible for osmolarity regulation in salt water fish, and can adapt quickly in the event of change in the ion concentration[51].

2.3.3 Tertiary responses

This kind of responses are 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.

Growth

Fish subjected to stress shows several growth-inhibiting responses. Nutrient assimilation is reduced by actions on the appetite and gut. This has been demonstrated on salmon, which showed reduced appetite and growth rate after handling stress[52], and gut damage and reduction of gut microflora after being stressed for 15 minutes[53].

As both catecholamines and cortisol promote metabolism and hyperglycaemia, they necessarily remove energy available for investment activities ([12], chapter 5). Cortisol seems to be the most important growth inhibiting factor, both promoting proteolysis[54] and reducing myogenesis[55].

There are some conflicting findings with regards to the relationship between cortisol, stress and growth. Mckormick found a reduction in plasma cortisol after chronic stress in salmon[52], which suggests lowered basal plasma cortisol when exposed to stressful environments. Pickering[56] similarly found that after

40 days, plasma cortisol levels was not elevated in brown trouts exposed to handling stress. He still measured reduced growth, leading him to hypothesize that growth rate is inhibited by stress, but that cortisol only plays a minor role. Barton, however, got an opposite result of Pickering in an experiment with rainbow trout, where he found growth inhibition in fish fed cortisol, but not in stressed fish[57]. Fast[58] similarly found no growth modulation in stressed salmon. In addition he measured, similary to Mckormick, that acute stress raised plasma cortisol, while chronic stress had no effect on cortisol levels. Mckormick suggests the differences observed to be due to species differences, as well as level of domestication[52]. This raises the question of whether different fish strains, even in the same subfamily (Salmoninae) or species (salmon), might have fundamentally different stress responses.

Immune response

There are a plethora of research aiming at investigating the immune response of fish as it relates to stress ([12], chapter 10), here it will suffice to give a brief overview. 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. Looking at the complete picture, the depressive effects far outweigh the enhancing ones.

This is coherent with the findings in salmon. After acute stress, salmon shows enhanced expression of inflammatory genes, while chronic stress leads to decreased stimulation and survival of leukocytes[58]. Furthermore, cortisol seems to prevent skin growth[59].

Accelerated ageing

Also accelerated aging has been observed in teleosts as a stress response. Cortisol is the main agent, having been shown to promote apoptosis in pavement and mucous cells of rainbow trout[60], as well as pavement and chloride cells in Mozambique Tilapia[61].

Reproduction system

Lastly, we mention the reproductive system. Stress affects all its levels, with a clear detrimental effect ([12], chapter 8). In salmon, elevated maternal cortisol levels lead to increased mortality, reduced size and increased morphological malfunction in offsprings[62].

2.4 Stress evaluation

As demonstrated above, stress can have a large impact on fish, both chronic and acute. To properly research and draw inference on this property, some method of stress observation and quantification is necessary. In general, any primary, secondary or tertiary response could be used as a stress indicator, however

some are more or less suited. Main considerations are ease of analysis, *id est* how easy it is to acquire a sample that can be used for stress quantification, and specificity and robustness of indicator, that is how much the indicator is responding to non-stress phenomena, and how well it is correlating with stress phenomena ([12], chapter 11). Referring to our previous discussion, cortisol and catecholamines are the most precise stress metrics. They are considered primary responses, and the initiator of later modulation of physiological systems. As we shall see, however, in some situations the evaluation of secondary responses is preferred, due to easier data access. Tertiary responses are valuable for evaluating chronic stress, however they are too slow to provide information during short-term homeostatic imbalances. In this section, the focus will be on describing the gold standards for stress evaluation, and comparing this to the use of ventilatory response as an indicator instead.

2.4.1 Catecholamines

Catecholamine concentration in blood plasma is intimately linked to stress, having a quick response after stressor exposure ([12], chapter 11). This quick response, however, makes 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 specialized equipment, in fact researchers were not able to quantify resting catecholamine levels in fish until 40 years ago[63].

2.4.2 Cortisol

Cortisol is better fit as a stress metric than catecholamines, due to its slower response and easier means of analysis. Cortisol can even be measured in water round the fish[11]. Furthermore, a correlation between crowding stress and blood cortisol has been firmly established[64] [65] [11] [66], although, as previously seen, some accommodation effect might occur after long term exposure. There are some evidence that cortisol can predict chronic stress as well as acute stress, however the relationship is not clear[67].

2.4.3 Respiration

Stress alters the ventilation of fish, although the exact mechanisms are debated. The use of ventilation as a stress indicator is equally controversial. The quick modulatory effects on both respiration frequency and amplitude to several stressors seem certain, although significant species differences exist[25]. Barreto argues against ventilation frequency as a stress metric, as he found the respiration response to be insensitive to stress intensity[68]. The same goes for Holden, which noted that both temperature and feeding has a pronounced effect on ventilation[65]. On the other side, Erikson found that ventilation rate and cortisol was the only two metrics clearly able to differentiate between stressed and calm salmon, finding indicators such as behaviour, lactate, PH, and onset of rigor

mortis to be less indicative[33]. Also Kammerer[69] claimed increased ventilation frequency as a response to salinity stress when studying tilapia.

One of the great advantages of ventilation rate for stress evaluation is that it is very simple to measure. With a camera and a stop watch, opercular oscillations, which accurately match ventilation frequency, can be counted and analysed[70]. It is also possible to measure the frequency of premaxilla-kype closure, however this behaviour can also happen during feeding. Other ways of measuring ventilation have been developed as well. One interesting method is to take advantage of the great conductivity of water to measure bioelectrical signals from the opercular and buccal muscles on electrodes external to the fish[71]. This requires appropriate placement of electrodes, and the signal is influenced by noise from locomotive muscles, so the method is most relevant in experimental settings. Another way to measure ventilation rate is to insert an embedded system transmitting ultrasound signals dependent on the activity of adductor mandibula into the fish. A receiver can then pick up these signals, and later processing can refine the information. This method provide signals accurate enough to distinguish feeding and ventilation behaviour, but is expensive and time consuming to set up[72].

An additional strength of respiration frequency as a stress metric is that it might say something about the type of stressor the fish is exposed to[73].

Some downsides to this approach also needs to be addressed. Firstly, only if the fish is purely breathing by buccal pumping can ventilation frequency be applied as a stress indicator, as ram ventilation leads to constant opercular and jaw gape. Secondly, both amplitude and frequency of respiration is altered during stress, and it seems different stressors trigger one or the other to different degrees[25]. Hence, breathing frequency alone is not sufficient to unambiguously determine stress.

2.5 Fish welfare

Welfare in fish is perhaps more of a philosophical than scientific question. In humans, we generally consider welfare as *the state of doing well especially in respect to good fortune, happiness, well-being, or prosperity*[74], but this is not directly applicable to fish. Traditionally, three different definitions of fish welfare have been proposed; feelings-based, function-based and comparison with natural lives ([12], chapter 12). 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 last method is evaluating in what manner the fish is exhibiting behaviour similar to that of a wild fish. 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[7][8].

If we compare 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 indexes cause stress. This is the case for temperature, stocking density and disturbances.

One important aspect of welfare not addressed in the SWIM index deserves some attention. When fish are subjected to reduced well-being, it 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[75]. As we have seen, cortisol is linked to both feeding behaviour and oxygen requirements, hence these responses might be a tertiary stress response. In african cichlid fish, dominant males tend to defend their territories, while subordinates tend to respond to intruders by directing their aggression at subordinate males[76]. A correlation is observed between cortisol levels and hierarchical position, suggesting a stress mediated pathway in displacement aggression. These behaviour indicators have been used for welfare analysis in earlier computer vision based works[15].

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.

2.6 Computer vision methods

Having explored stress, welfare and respiration in fish, we will now explore how computer vision methods can aid in elucidating these traits.

2.6.1 Problem statement

Human visual system

The human vision system is extraordinary. It is able to classify animal pictures within a few hundred of milliseconds[77] and identify tens of thousands of different objects and scenes[78]. This is done by photoreceptive cells in our retina firing electrical pulse trains towards the occipital lobe in the telencephalon based on reflected light in the surrounding scene[79]. From there, a complex set of scene understanding processes are initiated, recruiting up to 30 percent of the cortex[80] to segment and classify the millieux, estimate motion of self and others, and interpret the environment.

Cameras

The way computers see are not very different from that of a human. In the most common type of cameras, 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 ([81], chapter 2). This is analogous to what the eye does in a human. One issue with this kind of camera, is that information between frames is lost. This is fixed in the event camera, where each pixel acts as an integrator that fires when a certain amount of intensity is reached. This allows for continuous time registration, however the spatial resolution is reduced, and usually they only provide gray-scale pictures[82].

Viable end features

As motivated in the previous section, we wish to find some sort of indicator of stress in salmon. Since informative features manifest themselves externally, and can be evaluated by human perception, computer vision algorithms should be able to use tensors received from cameras to infer on salmon welfare. It is mainly two avenues to explore in this regard; behaviour, consisting of swimming speed, location, aggression and directional change, and respiration, consisting of frequency and magnitude of opercular and jaw movement. To limit the scope of this review, we will focus on the ventilation frequency. This has, to the authors knowledge, not yet been explored through computer vision, even tough it is quantifiable, provides important information on the physiological state of the fish, and has technical advantages over other approaches.

Motivation for respiration frequency as end feature

The main advantage of respiration frequency as a stress metric, is that no 3 dimensional reconstruction is necessary. A video of salmon has clearly distinguishable maximum and minimum jaw gape, which a proper algorithm can exploit to find breathing frequency. Most behaviour metrics, like swimming speed and location, on the other hand, requires 3 dimensional scene understanding. This can be constructed, but require resource intensive video prepossessing.

Another advantage of using respiration frequency is that for low salmon velocities, ventilation rate is always present, easily detectable and has a rather short time constant. Behaviour indicators, like aggression, only happens sporadically, require biological expertise to distinguish, and is more dependent on historical information. Some behaviour indicators have even longer time constants, like scale loss and fin damage.

Lastly, behaviour indicators are often a result of chronic stress, while ventilation frequency can effectively capture acute stress. This is useful for farmers, which can respond quickly to stressful events.

Even with these advantages, the salmon environment pose several challenges for any computer vision algorithm. Salmon lives under water, individuals look

similar to the human eye, they swim fast, they deform and they live in large shoals.

Problem type

We have already specified the input to the computer vision pipeline as a stream of tensors, and the end feature as respiration frequency. The problem of estimating ventilation frequencies, altough, can be posed in a couple of ways. If we track individual fish, we can frame the problem as a classification problem, putting each fish in a frequency range, or as a regression problem, estimating a numeric respiration rate. It is also possible to look at the whole shoal together, and estimate a common respiration. With this approach, we loose some information about hierarchies and individual differences, however the algorithm is more robust to fish swimming out of the camera frame, and inaccuracies. In the next section we will look into algorithms and State Of The Art (SOTA) methods for generating these input-output maps.

2.6.2 Algorithm proposals

Computer vision can be broadly separated into two categories, traditional approaches and deep learning[83]. In the former, we are transforming the input in a known and specific way to acquire a small set of features, before using a shallow function approximator to generate the output. Deep learning, on the other hand, exploits the descriptive power of deep neural networks to let a single large function approximator estimate outputs from raw inputs.

A trend in later years is the increased use of deep learning in the computer vision pipeline. The reason is that deep learning, under assumptions of a large dataset and high computing power, performs better than traditional approaches[84], do not require feature engineering, and is able to learn some form of optimal map between input and output. Even with all these strengths, traditional methods are still used for some tasks. Deep networks tend to overfit the training data and produce black box functions that are impossible for a human to understand [83], while traditional methods are more interpretable, easier to construct in a way that avoids overfitting. and require less data. Further, pitching these two methods against each other is partly a false dichotomy, as modern pipelines often use traditional methods to highlight certain features or perform data augmentation, before a deep learning system generates an input-output map. Further, traditional approaches are often employed on the output of the networks, to further improve and specialize the estimates.

Feature extraction and preprocessing

As traditional methods tend to aid deep learning ones, we will start our discussion with some feature extraction methods. Digital images today, like the ones taken from a cellular phone, are often composed of several million pixels[85].

When we are trying to separate ten respiration classes, there are a lot of irrelevant information contained in these pixels. By manually selecting features from the images, we can remove irrelevant information, while enhancing the effects of relevant information. This enables a smaller and more precise subsequent network.

One way to reduce the complexity of the image tensor, is to create masks, so each pixel is either on or off. By masking the operculum or the jaws of salmon, the subsequent classification or regression is easier, as irrelevant color nuances are removed. A popular masking operator is the canny edge detector[86], efficiently registering the edges of objects. If a simple binary mask looses too much information, the Histogram of Oriented Gradients (HOG) descriptor can be used, which is a soft edge detector[87]. It computes the weighted sum of gradient directions in windows in the picture, in order to describe intensity changes.

The issue with both of these descriptors, is that they are providing information frame by frame, while our problem is aiming at inference about frame changes. Optical flow is addressing this, where it tracks the changes of each pixel between frames. A popular method was developed in 1981, assuming local constant flow [88]. This breaks down for the case of salmon recording, were both swimming and breathing are happening simultaneously. Bergen[89] suggested a solution to this, by assuming two coherent motion patterns in a local area.

A method tightly linked with optical flow, tracking of region of interest, has been applied in human ventilation estimation[90]. A window on the chest were registered at each frame, and the vertical movement of this corresponded to ventilation. This method is transferable to salmon, *id est* by tracking the upper and lower jaw as a region of interest. The complexity of this task, however, is greatly increased when moving from humans to salmon, because of the movements and deformations of the latter.

The regions with large optical flow are the most interesting areas for respiration analysis, so enhancing these areas in the image before using deep learning would boost our task performance. One way to do this is to warp the image grid to magnify locations with useful information, like Recasens did in his work on saliency based resampling to solve a gaze following problem[91]. One potentially downside with this approach is inconsistent warping, which would cause difficulty when comparing frames.

Precision livestock farming research has also suggested some ways to extract vital parameter features. In cows, infrared cameras were used to assess temperature changes around the nostrils to estimate ventilation rate, and Red Green Blue (RGB) cameras were used to record color changes to predict cardiovascular action[92]. As fish are cold-blooded with less transparent skin than humans, these methods are not directly applicable to salmon, however related opportunities exists. Lopez[93] discovered that colorimetric properties of water change based on calcium content, hence stress mediated hydromineral disturbances might be observable by color cameras.

2 Dimensional (2D) Neural networks

After an acceptable feature set has been created, a map between input features and output values must be constructed. One possibility is to use shallow machine learning methods. Popular frameworks are support vector machines, able to handle classification, regression and nonlinear boundaries in the feature space ([94], chapter 5), decision trees for regression or classification, and shallow neural networks. Although these algorithms have proven effective for certain computer vision tasks[95], the general trend is that deep neural networks have better accuracy and robustness[96], especially when using pretrained weights. With only two simple building blocks, fully connected layers and convolutional layers, they are able to learn state of the art automated medical image segmentation[97], visual recognition/object classification[98] and object detection[99]. Convolutional neural networks are also able to learn object detection together with keypoint classification[100]. This can be used to estimate mouth gape of individual salmon, frame by frame, which could discern breathing frequency with appropriate post processing.

To further increase the descriptive power of neural networks, a temporal dimension can be incorporated. A naive approach is to make a connection backwards in the network, so the next data batch travelling through the network is summed together with the previous state ([94], chapter 15). This type of network, called recurrent neural network, forgets fast, which is problematic for a network that should remember the last respiration cycle of the salmon, up to 100 frames ago, depending on frame rate and respiration.

To improve on this, a more sophisticated Long Short-Term Memory (LSTM) cell has been proposed[101]. This introduces a forget mechanism, so that both short term and long term memory can be achieved. Later, this has been refined for use in image applications[102], and employed in an event-camera object detection algorithm by Perot[82]. He estimates bounding boxes by considering historical data, and one could imagine this network being expanded to include frequency estimation heads linked to these bounding boxes.

When estimating mouth gape of fish far away from the camera, the lengths become small. One proposed solution to enhance detection of small objects in neural networks is to include global context by frequency domain convolution, as well as using skip connections to avoid loosing detailed features in long signal paths[103]. Global information will likely be of little use in detecting jaw keypoints, however skip connections are worth considering. Other work have focused on magnifying subpixel motions by tracking color gradients[104], also in the presence of background motion[105]. This was accomplished by a traditional computer vision pipeline, but later work improved on the results by deep learning[106]. It is unlikely that jaw motion is on sub-pixel scale, and due to deformation and translation of salmon, comparing pixels in different frames are difficult. Hence, this method does not seem useful for the problem discussed here.

3 Dimensional (3D) neural networks

The task of estimating respiration frequency can be framed as an action classification task, in which one action consists of breathing with one frequency. Carreira[107] presented five different architectures for action recognition, Long Short-Term Memory (LSTM), like reviewed above, 3D-ConvNet, Two-stream, 3D-Fused Two-Stream and Two-Stream 3D-ConvNet. The last four incorporates temporal information by stacking frames into a 3D tensor, and they differ among each other by how they incorporate optical flow into their framework.

The extension of neural networks into 3 dimensions cause a large increase in tunable parameters, and consequently resource requirements. Carreira uses 32 to 64 parallel Graphics Processing Units (GPUs) for training, and a dataset of over 160000 videos. This is with reasonable sized data dimensions, up to images of size 224x224 pixels and a temporal history of 64 frames. Comparing this to our salmon dataset, which would need around 100 frames to guarantee capturing a respiration cycle, together with images of dimension 1920x1080 pixels, the resource requirements become prohibitively large. By decreasing image resolution, performing random cropping and sub sampling frames, this approach could still be feasible.

There are two ways a 3 dimensional neural network could be labelled for estimating frequency. One possibility is to crop a 2 second snippet from the salmon video stream, count the respiration of the fish, and use the average of this as a frequency label. The other possibility is to label each frame with bounding boxes, and then learn both fish object detection and fish-specific respiration frequency. The former would be very easy to annotate, but only extract part of the possible information in the video. The other option would be more informative, but very time consuming to label. A large amount of actions, at least 50, would need to be labelled, with 100 frames in each action, and around 5 bounding boxes in each frame. This accounts to 25'000 bounding boxes, and due to the time dependency of the network, artificial inflation of data is not trivial.

Options exists to limit the amount of parameters and resource requirements of 3 dimensional neural networks. One approach is to change the token mixer module[108], which Kumawat[109] did with great success. He applied a set of fixed Fourier transforms on proximal time and space pixel in order to extract local frequency information. This is of particular interest for analysing respiration, as the computed features will be tightly related to frequency.

Fish tracking

If a non-temporal network is used, it is necessary with some sort of tracking method to match individual fish in different frames. The easiest way to do this is to find the Intersection over Union (IoU) of bounding boxes in two consecutive frames, and accept two detections as the same fish if this metric is above a certain threshold. Other more advanced assigning algorithms could be employed, such as the Hungarian method[110].

The main issue with this simplistic method, is that a fish not registered for one frame is assumed lost. A solution to this would be to use a Kalman filter[111], and reiterate prediction steps until a new detection is registered inside the probability distribution of the filter representation of the fish location. This would 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[112][113][114].

Another way to track the fish is to use instance recognition. This has been popularized as a human identification tool[115], and would allow us to recognize individual fish independently of their time outside the camera frame. The use of deep learning object recognition for salmon identification has been done with impressive accuracy by Cisar, where the salmon dot structure were leveraged to recognize individual fish both short term and long term[116].

Instance recognition is usually performed in four steps; instance detection, instance alignment, feature extraction and classification[117]. Mask RCNN[100], which has already been established as useful in breathing frequency estimation, can be modified to accommodate for this. If we introduce a new bounding box and class head, and let each fish be a separate class, the model can learn to differentiate fish based on the features in the identified bounding box. To do this, a database of individual fish data must be available during training, which pose a difficulty. It requires building a database of all fish that is to be identified, which reduces generalizability and increase annotation workload. If the network is to be used during farming operations, building such a database becomes infeasible.

Adapting the network to new fish populations can be done with few training instances. By using methods such as Model Agnostic Meta Learning[118], which specifically trains on quick adaption, or fine tuning with most parameters frozen, only a couple of data points are necessary for each new class. This does not remove the issue with very large populations, however, which motivates the use of unsupervised re-identification.

It is possible to determine individuals without building a database first. Normally this is done by learning a discriminative feature map, which can be exploited for instance clustering at evaluation time. The discrimintative feature maps can be trained by ordering the data into similar pairs, by manual annotation or data augmentation, and dissimilar pairs, before a contrastive[119] network maximizes the separability between classes. A late approach is able to generate a descriptive feature map without negative pairs, and without convergence to the trivial solution[120]. The feature map inference can be done by unsupervised clustering[121], or by calculating a distance metric to other instances[122].

One intriguing possibility is to teach a network to generate a mask over the dot structure on the salmon. This structure can then be parameterized to a low dimensional space, and used to cluster salmon of the same instance. One objection to this is that salmon has two sides with different dot structure, so the fish will be classified differently depending on the direction it swims.

Frequency estimation

By allowing a neural network to estimate mouth gape of individual salmon, and use some method to track the fish, a one dimensional signal oscillating with the respiration frequency of salmon should result. These signals will be noisy, contain offset drift, and have varying amplitude and frequency, making the task of estimating period non-trivial.

Two obvious candidates to extract oscillation periods are the Fourier transform and the autocorrelation function. These transformations are specifically designed for inference on frequency, and efficient algorithms have been designed for calculating them[123]. The problem with these approaches, is that they are vulnerable to noise. It is possible to set up a pipeline that remove outliers and filter signals before passing them downstream, however this might disturb the frequency content of the signals in undesirable and surprising ways.

Another solution is to fit a function, like a sinusoid, to the time series. This is a flexible method capable of handling offset and amplitude dynamics. It is sensitive to outliers, however outlier pruning beforehand will work well, as it is possible to move, or remove, the outliers in a way that have very little effect on the final estimate.

The final possibility is to use a random sample consensus (RANSAC) type of algorithm. This algorithm handles outlier detection and function fitting simultaneous by sampling a subset of points, generating a model from the sampled points, and finally evaluating the model fit by counting inliers[124]. The basic algorithm does the initial sampling of points randomly[125], while later works suggest picking points close together in the feature space[126], or points with high quality[127]. Also the evaluation step of the algorithm has been proposed improved, from an iteration over all points in the dataset for all proposed models, to more effective validations. One such method is the sequential probability ratio test[128], which iteratively checks whether points are consistent with a good model, enabling quick discarding of bad models.

A main assumption in RANSAC is that a model computed from outlier-free samples is consistent with all inliers. This is not true, as inliers are also contaminated with noise. A proposed solution is to refine the current best RANSAC estimate by some sort of local optimization, like estimating a new model by least square fitting, or performing nonminimal RANSAC iterations on the estimated inliers[129].

Another critique of RANSAC is that it treats outliers and inliers with different, but constant, penalty. A proposed solution is to let inliers receive a penalty equal to the sum of squared errors, or to minimize the negative log likelihood of the samples[130]. When detecting mouth gape outliers, a reasonable assumption is that they are uniformly distributed over the bounding box. If this is the case, errors will normally be large, and an inlier weighting is not necessary.

Chapter 3

Theoretical background

Having established possible methods for salmon respiration estimation, this section will provide a detailed theoretical overview of algorithms and mathematics relevant for the pipeline developed. It will start with an overview of layers and training of neural networks, move on to a specific deep learning implementation, and finally discuss methods for frequency estimation from time series.

3.1 Neural networks

3.1.1 Introduction

Deep neural networks draw inspiration from the way our cortex process visual information from the retina. They use huge amounts of nodes (neurons) and connections (synapses) to create a generalized function with impressive descriptiveness. Up to 135 billion such parameters have been used in a single function approximator[131]. These neurons are organized in layers, with the simplest one being a fully connected layer. Such a module calculates a weighted sum of all nodes in the previous layer, passed through a nonlinear function. If we let σ be an elementwise nonlinear function, X be a row vector containing the inputs to the layer, W being a square matrix containing the weights between all nodes, b being a row bias vector and Y being the output of the layer, we can represent a fully connected layer as shown in equation 3.1 ([94], chapter 10).

$$Y = \sigma(XW + b) \tag{3.1}$$

3.1.2 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 solve this issue by letting filters scan previous layers for learned items, facilitating both detection and localization

with a single weight set ([94], chapter 14). If the location of an object is irrelevant for the task, pooling layers can reduce complexity and leave the network almost[132] 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 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 X that is captured in one node in the output layer Y, and the stride of a layer is the size of pixel shift between each filter computation. Then, equation 3.2 ([94], chapter 14) describes how a node in the output layer is related to the previous feature tensor. Note that the dimensions are extended compared to the fully connected layer representation, and no stride is applied in the depth direction.

$$Y_{i,j,k} = b_k + \sum_{u=0}^{f_h-1} \sum_{v=0}^{f_w-1} \sum_{z=0}^{f_d-1} X_{i\times s_h+u, j\times s_w+v, z} \times W_{u,v,z,k} \tag{3.2}$$

3.1.3 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. If a function p defines some pooling operation, one element in the output of a pooling layer is defined in equation 3.3.

$$Y_{i,j,k} = p(X_{[s_h\cdot i, s_h\cdot i+f_h], [s_w\cdot i, s_w\cdot i+f_w], [s_d\cdot i, s_d\cdot i+f_d]}) \tag{3.3}$$

3.1.4 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. The training is done by applying the generalized delta rule (backpropagation)[133], 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 some 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 we wish to adjust a weight $w^l_{1,1}$ that goes from node 1 in layer $l-1$ to node 1 in layer l. Let superscript describe layers, subscript describe node connections, w describe weights, v describe node outputs and n describe the size of hidden layer l+1. Then, the gradient of the loss with regards to node $w^l_{1,1}$ is given by equation 3.4. All gradients can be unwrapped based on deeper nodes this way, until a full loss gradient is acquired.

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial w^l_{1,1}} = (&\frac{\partial \mathcal{L}}{\partial v^{l+1}_1}\frac{\partial v^{l+1}_1}{\partial v^l_1} + \frac{\partial \mathcal{L}}{\partial v^{l+1}_2}\frac{\partial v^{l+1}_2}{\partial v^l_1} + ... \\ &+ \frac{\partial \mathcal{L}}{\partial v^{l+1}_n}\frac{\partial v^{l+1}_n}{\partial v^l_1})\frac{\partial v^l_1}{\partial w^l_{1,1}}\end{aligned} \tag{3.4}$$

Any gradient descent method can iteratively improve the function approximation after backpropagation has provided $\frac{\partial \mathcal{L}}{\partial W}$. A good option is a stochastic version with momentum and weight decay, allowing batch training and ensuring some robustness for local minima and noisy gradients ([94], chapter 11). Let **b** with parameter μ be the momentum, 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, and parameters updated, as displayed in algorithm 1[134].

Algorithm 1 Stochastic gradient descent

$g_t \leftarrow \nabla_W \mathcal{L}(W_{t-1}) + \lambda W_{t-1}$
$b_t \leftarrow \mu b_{t-1} + g_t$
$g_t \leftarrow b_t$
$W_t \leftarrow W_{t-1} - \gamma g_t$

3.1.5 Keypoint Region-based CNN (RCNN)

Construction

The layers and training methods described above have been incorporated into a SOTA network called keypoint RCNN, which is described in the original mask RCNN paper[100][1]. This network is based upon faster RCNN[136], which again draws inspiration from fast RCNN[137]. To understand the former, therefore, we must begin with the latter. 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 neural network uses a conventional Convolutional Neural Network (CNN) to extract a feature map of the input

[1]The explanations related to the mask RCNN construction is partly based on unpublished work of the author[135].

image, crops the feature map according to the ROIs, and then pool the feature map to a fixed size. Then, the fixed-size feature maps are passed through fully connected layers to estimate a softmax class vector and a per-class bounding box. The ROIs are processed individually, each leading to a bounding box and class probability for all classes.

The main problem 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, and then the feature map and ROIs are fed into the ROI pooling layer of the fast RCNN.

Now, the mask RCNN[100] comes along to add segmentation to the network. To do this, it adds a third branch after the ROI pooling layer. This branch uses several fully convolutional layers to generate a pixel mask for each object. Further, when fast RCNN maps ROIs from the input image to the feature map, it introduces quantization noise, greatly impairing segmentation performance. Mask RCNN fixes this by using floating numbers and bipolar interpolation to do the mapping, resulting in a proper alignment of the feature map.

Finally, keypoint prediction is achieved by introducing k number of on-hot masks, each representing one keypoint. The final network is now a complete end-to-end instance detection and keypoint net, capable of finding premaxilla, dentary, eye and head area of the salmon.

Training

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

Keypoint RCNN is trained from images annotated with bounding boxes and bounding box specific keypoints, which it learns to predict on images during training. By labeling the images in such a way that respiration relevant states for each salmon can be constructed, a ventilation signal can be generated by concatenating these states.

The bounding boxes could capture the whole salmon. This is easy for the network to learn, since a salmon is large and has a distinguishable shape. The problem with this, however, is that salmon in tanks and nets often overlap with each other, and features from the stomach and tail of salmon is irrelevant when evaluating breathing frequency. A better solution is to only capture the salmon head, which provides the same amount of respiration relevant information as the whole salmon, while reducing keypoint head size, and thereby resource requirements. When using this approach, it should be ensured that both the operculum and the tip if the jaws are contained in the bounding box.

Two types of respiration states can be extracted from keypoint locations. The first is the euclidean distance between keypoints at the top and bottom jaw. The second is the fish mouth opening angle, which can be calculated by additionally labelling the root of the jaw.

In addition to the task-specific keypoints, it is a good idea to label points not used in subsequent ventilation state calculations, but which instead forces the

network to learn the right relationships (auxiliary targets). One such keypoint is the eye, which is easy to predict, and can be used by the network to learn relative jaw positioning.

By using the coordinates of the upper jaw, lower jaw and eye, all bounding boxes will contain $K = 3$ keypoints, each represented by two numbers (x, y). Furthermore, the number of bounding boxes will be equal to the number of fish in the frame, and each box will have dimension four, two numbers (x, y) for two of the corners. Having described the network labels, we now move on to the loss function.

For simplicity, we assume batches of one frame. The ground truth targets consists of R bounding boxes contained in a $v \in \mathbb{R}^{Rx4}$ tensor and a set of K different keypoints contained in an $s \in \mathbb{R}^{RxKx2}$ tensor. As 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 and h_k. 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 ($\hat{s} \in \mathbb{R}^{R_{ROI},K,C,h_k,w_k}$).

The total loss function (equation 3.5) for a frame is the sum of the losses for 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[136]. Localization loss and keypoint loss are only calculated for ROIs that are associated with a positive (salmon) class, hence we introduce an encoding function λ that is 1 when a ROI is associated with a positive ground truth class, and 0 otherwise.

The classification loss is estimated as a log loss over the two classes object and not object (equation 3.5b), the localization loss is calculated as a smooth L1 loss (equation 3.5c)[134] and the keypoint loss is calculated as the cross-entropy over an $K \cdot h_k \cdot w_k$ way softmax output (equation 3.5d). In these equations, 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.

$$\mathcal{L}(p, v, s, \hat{p}, \hat{v}, \hat{s}) = \sum_{r=0}^{R_{ROI}} \frac{1}{N_{cls}} \mathcal{L}_{cls}(p_r, \hat{p}_r) + \frac{1}{N_{loc}} \lambda_r \mathcal{L}_{loc}(v_r, \hat{v_r}) + \frac{1}{N_{key}} \lambda_r \mathcal{L}_{key}(s_r, \hat{s}_r) \tag{3.5a}$$

$$\mathcal{L}_{cls} = -(p \cdot log(\hat{p}) + (1-p) \cdot log(1-\hat{p})) \tag{3.5b}$$

$$\mathcal{L}_{loc} = \begin{cases} 0.5 \cdot (v - \hat{v})^2, & if \|v - \hat{v}\| < 1 \\ \|v - \hat{v}\| - 0.5, & otherwise \end{cases} \tag{3.5c}$$

$$\mathcal{L}_{key}(s, \hat{s}) = -\frac{1}{K \cdot h_k \cdot w_k} \sum_{k=0}^{K} \sum_{i=0}^{h_k} \sum_{j=0}^{w_k} (s_{k,1,i,j} * log(\hat{s}_{k,1,i,j})) + (1 - s_{k,1,i,j}) log(1 - \hat{s}_{k,1,i,j})) \tag{3.5d}$$

3.2 Frequency estimation

Frequency estimation of a one dimensional signal will be an important part of the final pipeline. Due to this, we will here explain methods of outlier rejection, non-linear function fitting, autocorrelation and RAndom SAmple Consensus (RANSAC). First, however, we will describe Non Maximum Suppression (NMS) and IoU.

3.2.1 NMS and IoU

NMS and IoU are fundamental terms in any computer vision pipeline, and will be briefly explained below.

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 3.6.

$$IoU = \frac{A \cap B}{A \cup B} = \frac{true\,positives}{true\,positives + false\,negatives + false\,positives} \tag{3.6}$$

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, T a threshold and U the set of non-overlapping bounding boxes. NMS is then displayed in algorithm 2.

Algorithm 2 NMS

Require: B, U: $\emptyset$, $C : B \rightarrow [0, 1]$, T
  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

3.2.2 Outlier detection

Several outlier detection algorithms have been developed in the past couple of decades, some specializing in soft detection by finding a local outlier factor[138], others in effective and quick separation by generating an isolation forest[139]. Here, I will describe a novel and intuitive method instead, the frame jump limiter.

Let T_d be a distance threshold for separating inliers and outliers, ws (window size) determine the neighborhood over which the algorithm will search and N the length of the signal x. Then, algorithm 3 calculates a local average which it draws points to if they are too far away from other local points. One of the main strengths of this algorithm is that even outlier clusters larger than the neighborhood will be handled, since the signal is dynamically altered. A weakness is that prolonged periods of outlier rejection results in horizontal lines overwriting the signal, and these lines will continue after the outlier cluster has passed if the difference between the recorded signal and the horizontal line exceeds T_d.

Algorithm 3 Frame jump limiter

Require: ws, T_d
  for i=ws to N **do**
    $avg \leftarrow \frac{x[i-ws,i]}{ws}$
    if $\|avg - x[i]\| > T_d$ **then**
      $x[i] \leftarrow avg$
    end if
  end for

3.2.3 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 us define a parameter vector p (3.7a)

determining the function $f(x,p)$ (3.7b), and let $\hat{p}$ be the optimal p, defined as the parameter vector that minimizes the sum of squared errors (3.7d). An effective and robust algorithm to find $\hat{p}$ is Levenberg-Marquardt (LM), which effectively combines the advantages of gradient descent and Gauss-Newton[140] by finding the $\mathbf{s}_k$ that solves equation 3.7e, and then updates p in direction of s_k. This algorithm will converge to a local minima, and under some conditions achieve locally quadratic convergence[141].

$$p = \begin{bmatrix} A & f & \phi & O & a \end{bmatrix} \tag{3.7a}$$

$$f(x,p) = A \cdot sin(2\pi \cdot f \cdot x + \phi) + o + a \cdot x \tag{3.7b}$$

$$\mathcal{L}(x,y,p) = \frac{1}{2}\sum_{i=1}^{N}(f(x_i,p) - y_i)^2 = \frac{1}{2}\sum_{i=1}^{N} r_i^2 \tag{3.7c}$$

$$\hat{p} = min_p \mathcal{L}(x,y,p) \tag{3.7d}$$

$$(\mathbf{J}_k^T\mathbf{J}_k + \mu_k\mathbf{I})\mathbf{s}_k = -\mathbf{J}_k^T r_i(p_k) \tag{3.7e}$$

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

1. **Offset**: This describe the vertical bias of the data, hence should be initialized to the average of the datapoints. $O_0 = \mu = \sum_{i=1}^{N} \frac{y_i}{N}$.
2. **Amplitude**: The standard deviation of a perfectly sampled sinusoid is $\frac{A}{\sqrt{2}}$[142], 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**: This should be initialized to an expected respiration frequency.
4. ϕ: This could be initialized by looking at the numerical gradient, or the absolute gape size, at the beginning of the signal.
5. **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.

3.2.4 Autocorrelation

The autocorrelation function is a description of how similar a signal is with how it looked k timesteps ago (eq: 3.8). The local peaks of this signal, therefore, corresponds to a possible period of the mouth gape.

$$\mathcal{R}[k] = \sum_{m=-\infty}^{\infty} x[m]x[m-k] \tag{3.8}$$

3.2.5 RANSAC

The RANSAC algorithm is an approach 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 $P \in \mathbb{R}^N$, and then calculating a set of inliers I containing points close to a model $M := f(S)$ fitted on the minimal sample. Close is defined as any point with loss function $\mathcal{L}(M|_x$, y) smaller than a threshold T. The random sampling is done k_{max} times, where k_{max} can be found by the relationship shown in equation 3.9, giving a 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[124]. Since we do not know the inlier ratio *a priori*, it can be dynamically updated every time a new maximum inlier set is found. The full method is displayed in algorithm 4[124].

Algorithm 4 RANSAC

Require: $N > n, k_{max}, \eta_0$
  $I_{max}, k \leftarrow 0$
  while $k < k_{max}$ **do**
    $S_k \leftarrow \{A : dim(A) = n, A \subset P\}$
    $M_k \leftarrow f(S_k)$
    $I_k \leftarrow \{A : \mathcal{L}(M|_x, y) < T, A \subset P\}$
    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

$$g(\epsilon, \eta_0, n) = \frac{log(1 - \eta_0)}{log(1 - \epsilon^n)} \tag{3.9}$$

Chapter 4

Method

After looking into possibilities for computer vision based frequency estimation of salmon, including a detailed theoretical description of some of them, we will here establish a baseline algorithm for end-to-end salmon ventilation regression. The proposed algorithm is a two-step pipeline, in which the first step is a neural network to estimate the mouth gape of the fish, and the second step is to extract a frequency from the resulting mouth gape trajectory. We will describe how the neural network were constructed and trained, and how the output of the deep learning network were used to find respiration rate. For development of the method, a 10 minute video recording from the 4^{th} hour of a salmon downbreathing was available. For the evaluation of the algorithm, additional recordings from the 1^{st} and 2^{nd} downbreathing hour were included.

4.1 Deep learning

4.1.1 Training data

The 10 minute salmon recording were split into frames, and 35 images spaced out over the video recording were annotated with bounding boxes going from the attachment of the pectoral fin to a bit in front of the snout. Each salmon head were further annotated with three keypoints; one eye, the kype (dentary prominence) and the tip of the premaxilla. Partly occluded fish were not annotated.

To inflate the dataset, an albumentations[143] pipeline was constructed, performing random crops, horizontal and vertical flips, contrast and gamma adjustment and RGB shifts 60 times per image. This resulted in a 2100 frame dataset.

4.1.2 Network and training

The keypoint RCNN implementation from PyTorch[134], constructed as explained in the previous chapter, was used for salmon detection and keypoint prediction. The backbone was pretrained on imagenet[144], while the keypoint

and bounding box heads were randomly initialized. All parameters in the model were adjustable, and the model was trained for 10 epochs on 2 T4 GPUs with the 2100 frame augmented dataset.

At the output of the model, all bounding boxes were pruned by the PyTorch NMS method when making predictions.

4.2 Frequency analysis

The mouth gape of the salmon was extracted simplistically, by calculating the euclidean distance between the kype and premaxilla keypoints. Thereafter, fish in different frames were matched by an IoU (eq: 3.6) threshold algorithm, and wrongly detected keypoints removed by the frame jump limiter algorithm (alg: 3). After this, all mouth gape time series under 50 frames were removed, and long signals were split into two second long snippets. Finally, the autocorrelation function and the Levenberg-Marquardt algorithm were used to estimate the frequency of the resulting one dimensional signals by the statsmodels[145] and scipy[146] libraries, respectively. For the autocorrelation function, only correlations above a threshold T_1, detected with at least T_2 frames delay, were considered valid frequency estimates.

4.3 Complete model

The sine fitting approach was found to be the best frequency extraction method, and therefore used in the full model. Levenberg-Marquardt outputs both frequency and fitting error for each signal, so the output of the full pipeline were two vectors, one with the estimated frequency of all trajectories, and one with the uncertainty of the estimate of all trajectories. The most erroneous sinusoids were removed from the vector, and the rest were used to plot a weighted histogram over respiration frequencies. Also the weighted average of the frequencies were computed.

4.4 Parameters

The parameters used in the pipeline were tuned from a set of mouth gape signals generated from a 1000 frame snippet of the 10 minute salmon recording. Below is an overview of the parameters and how they were determined.

1. Outlier detection
   (a) **Averaging window size (ws)**: Visual inspection was performed to maximize the human visibility of the respiration oscillations.
   (b) **Distance threshold for outlier classification (T_d)**: Visual inspection was performed to maximize the human visibility of the respiration oscillations.

2. Autocorrelation

   (a) **Lowest valid correlation (T_1)**: Visual inspection was performed to put the threshold such that only clear oscillation peaks, not noise or flat correlations, would be estimated.

   (b) **Lowest valid delay (T_2)**: Cutoff was placed such that frequencies above natural salmon range would not be detected.

3. Sine fitting

   (a) **Initial parameter estimate (p_0)**: Frequency was initialized to a reasonable value, amplitude and offset were initialized from dataset statistics. The rest of the initial parameters were initialized to zero.

4. Complete model

   (a) **Max fitting error**: Visual inspection was performed to set a threshold that removes sinusoids that are unlikely to match physical respiration.

4.5 Testing

4.5.1 Quantitative evaluation of frequency correctness

The same 1000 frame video snippet as used for tuning the parameters were used for constructing a validation set. It was developed the following way.

1. Create a movie with annotations from the deep neural network drawn on top.

2. Note which bounding boxes have valid autocorrelation estimates.

3. Note down the frames in the generated video of max mouth gape for all valid bounding boxes.

4. Use the period between max mouth gape to calculate frequencies for the valid bounding boxes.

After this, frequencies estimated by the model were compared with the ground truth frequencies.

4.5.2 Qualitative evaluation of shoal respiration frequency as it relates to stressors

In the experiment section we will describe how recordings of salmon exposed to different kind of stressors are obtained. These recordings are passed through the model to generate shoal frequency estimates, which is qualitatively analysed.

Chapter 5

Experiment

The following description will be restricted to video recordings, as the full experimental data and setup material is not obtained at the time of report delivery.

Nine tanks, each with seven fish, were subjected to changing oxygen and temperature millieux. 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 downbreathing, the oxygen supply to the tanks were restricted, causing a gradual decline in Dissolved Oxygen (DO) content. This was enabled by removal of air stones and water flow to the tanks. When the DO levels reached 50%, oxygen were 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 was a new downbreathing similar to the first one.

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 downbreathing, the fish

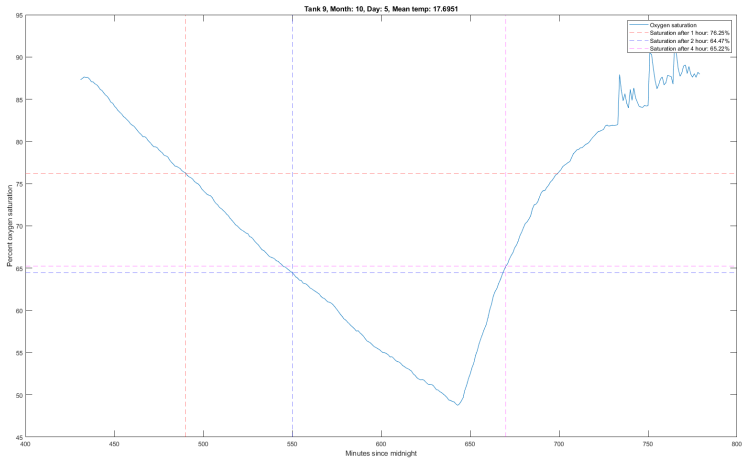


Figure 5.1: Dissolved oxygen content during a downbreathing

were filmed constantly from a bottom corner, with minor breaks during battery and memory card switching.

Dissolved oxygen and temperature were measured continuous in tank 2, 5 and 7 with the $miniDO_2T\,Logger$[147]. During the downbreathing, oxygen and temperature were measured each hour with a handheld device additionally.

For this project, three small clips from a downbreathing phase in tank 9 were used, recorded one, two and four hours after the start of the oxygen cutoff. All clips had a length of 42420 frames (707 seconds). For the model evaluation, 1000 frame (16,7 seconds) snippets at different times during these clips were used. In figure 5.1, a representative oxygen trajectory during a downbreathing is displayed. Due to the nature of the preliminary data, some metadata is missing, and it is uncertain whether the displayed oxygen recording is from the same downbreathing as the video recordings. The oxygen trajectories look relatively similar during both downbreathings, the only relevant variation is how close the oxygen levels after two and four hours are. In figure 5.1, the oxygen content is almost equal.

Chapter 6

Results

The detections of the keypoint RCNN are shown in figure 6.1, where frames are displayed to highlight the difficulties of a fish detection network. In frame 832, both false negatives and false positives are present due to similarities between fish and other objects, as well as the deformation of the fish, in frame 835 the fish to the bottom left of box 67 has lost its bounding box due to overlap between two fish and NMS pruning, in frame 890 occlusion of the jaw leads to a major keypoint localization error, and in frame 952 a short period of occlusion caused a new bounding box IDentifier (ID) (72 to 97).

Figure 6.2 shows the processed mouth gape trajectories together with the Levenberg-Marquardt fitted sinusoides. Most signals show clear oscillations, which the fitted model captures well. On some signals, like box 67, the oscillations are barely visible, and the model fit is poor. This is captured by the standard deviation of the frequency parameter, which is second largest on box 67.

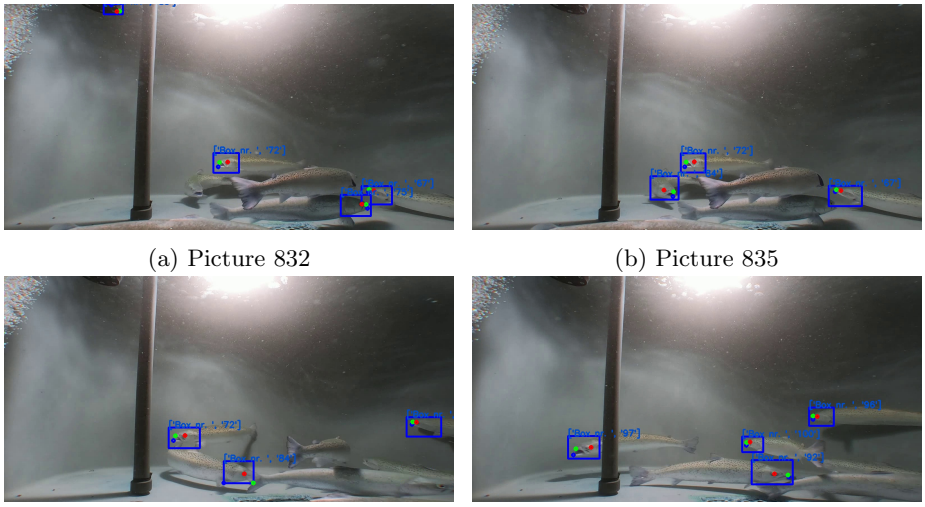
Looking at the autocorrelation function (figure 6.3), the time series with the clearest oscillations are captured well. Some of the more noisy signals, and signals with offset drift and amplitude dynamics, however, show no peak in the autocorrelation function. Also the ground truth periods are visible in the autocorrelation plot, showing that the generated estimates are of the same magnitude as the true respiration. Furthermore, when both have detections, the autocorrelation estimates are similar to the sine fitting estimates.

In figure 6.4 three frequency histograms for each of the three clips at different times during the downbreathing are displayed. All of the histograms have a clear peak around a reasonable frequency, and the four hour histogram peaks are clearly distinguishable from the peaks of the rest of the histograms. The one hour and two hour histograms, however, estimate approximately the same frequency.

In the next histogram plot, figure 6.5, high error estimates are removed. As with the other histogram plot, the average frequency after four hours is clearly distingusihable from those after one and two hours.

At evaluation the algorithm spent 2209 seconds on an i7-11700 Intel Core

CPU to process a 16,7 second video snippet. 2201 of these seconds were used in applying the deep neural network, track bounding boxes and generate an annotated movie. Of these three tasks, the deep neural network was by far the most resource intensive.



(a) Picture 832

(b) Picture 835

(c) Picture 890

(d) Picture 952

Figure 6.1: Illustration of different issues with the neural network

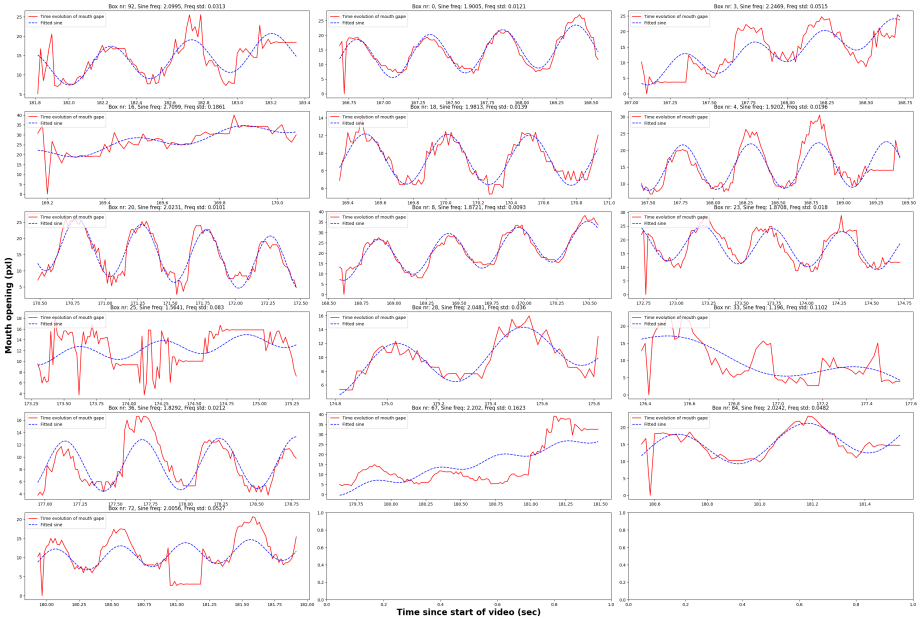


Figure 6.2: Time series of mouth gape

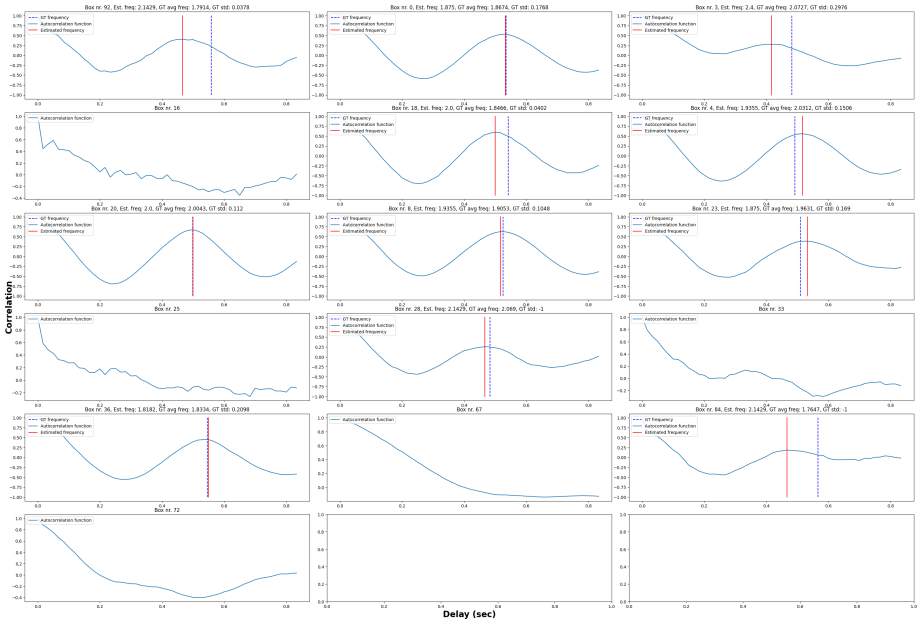
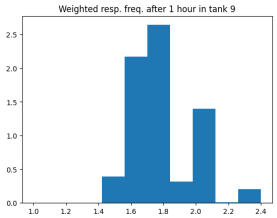
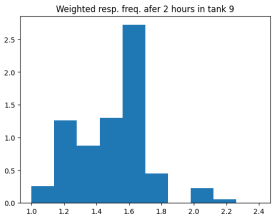


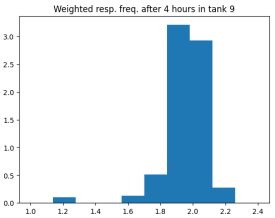
Figure 6.3: Autocorrelations of mouth gape trajectories



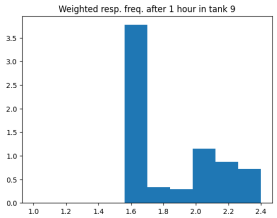
(a) 1 hour, frame $\in [10'000, 11'000]$



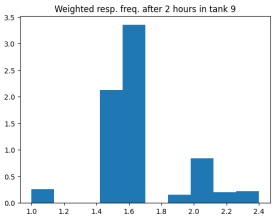
(b) 2 hours, frame $\in [10'000, 11'000]$



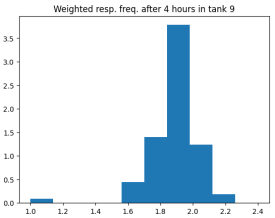
(c) 4 hours, frame $\in [10'000, 11'000]$



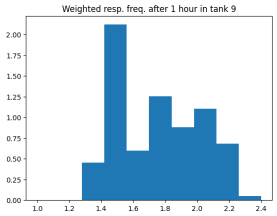
(d) 1 hour, frame $\in [20'000, 21'000]$



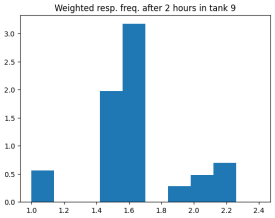
(e) 2 hours, frame $\in [20'000, 21'000]$



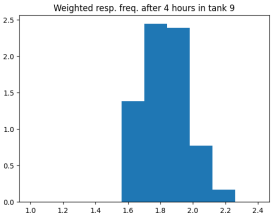
(f) 4 hours, frame $\in [20'000, 21'000]$



(g) 1 hour, frame $\in [30'000, 31'000]$

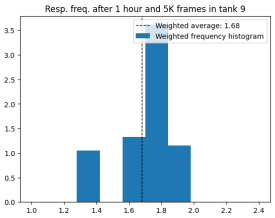


(h) 2 hours, frame $\in [30'000, 31'000]$

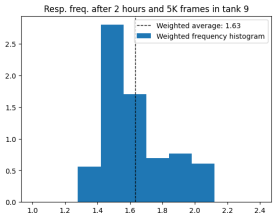


(i) 4 hours, frame $\in [30'000, 31'000]$

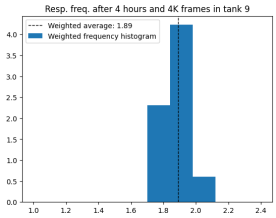
Figure 6.4: Histograms of all ventilation estimates in tank 9



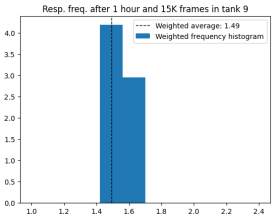
(a) 1 hour, frame $\in [5'000, 6'000]$



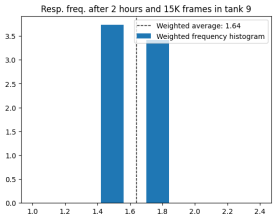
(b) 2 hours, frame $\in [5'000, 6'000]$



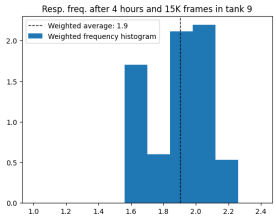
(c) 4 hours, frame $\in [5'000, 6'000]$



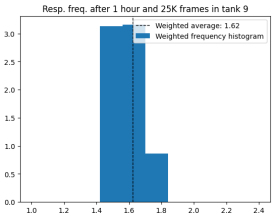
(d) 1 hour, frame $\in [15'000, 16'000]$



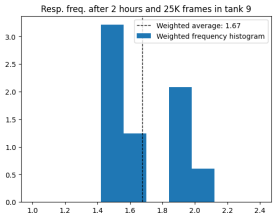
(e) 2 hours, frame $\in [15'000, 16'000]$



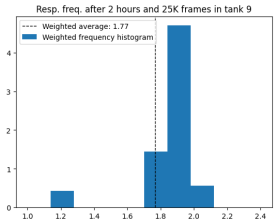
(f) 4 hours, frame $\in [15'000, 16'000]$



(g) 1 hour, frame $\in [25'000, 26'000]$



(h) 2 hours, frame $\in [25'000, 26'000]$



(i) 4 hours, frame $\in [25'000, 26'000]$

Figure 6.5: Histograms of low error ventilation estimates in tank 9

Chapter 7

Discussion

7.1 Methodology

7.1.1 Signal length

Parts of the methodology were constructed by observing preliminary results, and deserves some attention. The reason for removing signals under 50 frames (0,83 seconds), were to guarantee at least one respiration period registered, and the reason for restricting the length to two seconds, was to avoid too much offset and amplitude dynamics. Long sequences were split into multiple signals, which could result in a few well registered fish having a large impact on the estimated shoal frequency. This is not an issue if the shoal frequency distribution is narrow. One way to avoid large influence from a small number of fish is to cut off signals instead of splitting them. This, however, would result in lost information.

7.1.2 Tuning of autocorrelation function

The two parameters in the autocorrelation function were there to ensure only marked peaks corresponding to breathing would be registered. The numerical values were tuned after reviewing the signals in figure 6.3, and must be interpreted with that in mind. The delay threshold is only sensitive to the problem statement, and would not need to be adjusted based on the signal quality. The correlation threshold, however, is highly dependent on noise and drift, and new frame samples will have a different optimal correlation threshold.

7.1.3 Neural network annotation

Furthermore, the annotation must be addressed. The margin added at the snout is inaccurate, however it was necessary to ensure the tip of the dentary and premaxilla were inside the bounding box. The choice to refrain from labelling partly hidden fish was to force the network to only detect individuals with clearly distinguishable keypoints. The consequence of detecting occluded fish

was exemplified by frame 890, with a grave location error of the kype. Another solution to deal with hidden fish, is to train the network to learn not only the position of keypoints, but also whether they are occluded. The keypoint RCNN implementation of PyTorch does not offer this flexibility, so choosing this option would require constructing the network from scratch, or finding some clever way to modify it.

The small amount of manually annotated training data was facilitated by data augmentation and a pretrained backbone. Rather large augmentations were performed, making the network robust to several disturbances. Some of these disturbances, like horizontal flip, are not physically relevant, and hence possibly superfluous. The decision not to freeze the pretrained parameters was to allow a specialization from general animal features in the ImageNet dataset to fish features. This might be too much to expect from our tiny dataset, which should be sufficiently challenged by the flexibility of the network heads.

7.1.4 Validation set

The frequency validation set was constructed with the same data that was used to tune the parameters, and with an output of the model it is supposed to evaluate. Due to this, it is important to state that the set is not supposed to be used for statistical analysis, neither for proving accuracy of the model. Instead, the set demonstrates that the estimated autocorrelation peaks and sinusoid oscillations are close to the respiration frequency observed in the video recordings. This strengthens the evidence that the main signal dynamics is caused by ventilation.

7.1.5 Regularity condition

Also an implicit assumption made by the constructed algorithm is worth addressing, the regularity condition, which can be defined as assuming regularity of the period of max mouth gape. Since the salmon can snap, feed or ram ventilate, this condition does not hold in general, and oscillations not correlated with respiration happens occasionally. This will cause both frequency extraction methods discussed here to break down.

In this project, signals of maximum two seconds were used, so it is likely that most of the trajectories only contain respiration oscillations. If, by improving on the algorithm, longer sequences are extracted, some way of dealing with snapping and feeding must be constructed. This could be done by splitting long signals into short sections, and discard sections with poor fit, assuming they are the result of breakdown of the regularity condition. Another possible solution is to use neural networks to train a salmon behaviour classifier on the time series, and use this to filter out non-breathing oscillations.

7.2 Neural network

Most of the errors observed from the keypoint RCNN were expected. The trade off between double detections and excessive bounding box pruning is inherent in object detection problems, but can be improved by a soft NMS method[148], by changing the IoU threshold, or by changing the network architecture to one that does not require NMS, like the DEtection TRansformer (DETR) network[149]. Furthermore, the bounding box tracking issues were expected, and solutions were addressed in the literature survey chapter.

The false negatives are most likely caused by the network not seeing all salmon poses during training, together with the conservative labeling. The false positive in frame 832 is likely caused by the similarity between salmon and other submerged objects. These kinds of errors should be removable by using more training data.

7.3 Frequency estimation

7.3.1 Outliers

In figure 6.2, the time series after outlier pruning are shown. These are almost outlier free, do not have pronounced processing peculiarities, and have clearly visible oscillations, matching the respiration frequency of salmon. Considering this, the algorithm to remove outliers work well. Some effort were spent trying to improve the outlier removal by using scikit learns[150] implementations of local outlier factor and isolation forest. These algorithms performed significantly worse than the frame jump limiter.

A Kalman filter could be applied to smoothen the mouth gape curves as an alternative to outlier detection and removal. Due to the simple oscillatory movement of the mouth, the system modelling should be straight forward.

7.3.2 Validation set

The ground truth max mouth gape periods were mostly around 30 frames, giving a respiration frequency of 2 Hz. Gape state were often indistinguishable for 3 frames, causing an uncertainty of up to 6 frames each period. This results in a 2 Hz label potentially being as low as 1,7 Hz, or as high as 2,5 Hz.

7.3.3 Autocorrelation function

The frequency estimates extracted from the autocorrelation functions (fig 6.3 are all close to the ground truth. The amount of estimates, however, is rather poor. Even with some parameters tuned on observed signals, the autocorrelation function is unable to extract frequencies in 5 of 16 tracked fish. As another downside, it has no metric to measure quality of estimate.

7.3.4 Sine fitting

For the fish with both sine function fitting and autocorrelation function respiration estimates, the regressed frequencies are very similar. The sine fitting method furthermore provides reasonable estimates for the salmon with flat autocorrelation, a quality metric is included in the estimate, and no parameter tuning, except an initial estimate, is required. Considering this, the sine fitting algorithm is more robust and better generalizable than the autocorrelation function. Not much time were spent on tuning initial parameters, so the function fitting can probably be improved by providing an initial estimate for the phase, and a more accurate initial frequency.

7.3.5 Shoal frequency distribution

It is noteworthy that the frequencies of the salmon in figure 6.2 are reasonably close together, although some spread is noticeable. If we remove the least certain sine fitting estimates, which can be done by requiring the standard deviation of the frequency parameter to be less than 0.05, 10 estimates lands in the range [1,83, 2,100]. This tells us that for our experimental setup, extracting one frequency number for the entire shoal is a justified simplification, but cause some loss of relevant individual fish information.

7.3.6 Small jaw motion

By looking at the time series (fig: 6.2) it does not seem like small jaw motions are an issue. Salmon in the back of the frame with Peak to Peak (P2P) amplitude of about 6 pixels, like box 18 or 28, have at least equally visible oscillations as proximal fish with large gapes, like box 92 or 67. One possible reason is that recordings of distal fish are less noisy, since locomotion has less effects on signal amplitude and offset. Furthermore, the fish are usually filmed from the same direction, and stay longer in the frame. If the experimental setup would have facilitated longer distances and smaller amplitudes, down to 3 pixels, issues with small gapes would likely be more pronounced.

7.4 Complete model

7.4.1 Frequency distribution

Looking at figure 6.4, all histograms show a clear peak around a reasonable frequency, strongly advocating that the method is able to extract average shoal frequencies. There are, however, a significant spread in the estimates, up to 1 Hz. It is tempting to attribute this to random noise, however this is an unreasonable assumption. Remember that the histogram bars represent frequencies, and that the time series plots show very accurate fit for some signals, and very poor for others. Hence, model noise is heavily influenced by outliers, and another

explanation must be put forth to explain the multimodal distributions observed in the figure.

A reasonable hypothesis is to attribute the spread to a physical ventilation distribution among the salmon. This theory can be elucidated by removing functions with poor fit from the histograms, reducing the amount of noise. In figure 6.5 this is done, and a narrower, although still multimodal, distribution is uncovered. When generating these histograms, a visual inspection was performed on the time series plots earlier in the pipeline, which showed that almost all estimates not removed had clear frequency oscillations and good sinusoidal fit. Hence, the refined histograms are almost noise free, and we can conclude that some difference exists in the ventilation frequency of different fish exposed to the same stressor.

7.4.2 Stress

There are some other interesting effects in the histograms that must be mentioned. By looking at figure 5.1, it is clear that the oxygen content after 2 and 4 hours are nearly identical, while the average frequency estimations differ significantly. As explained in the experiment chapter, the link between dissolved oxygen and video recordings are uncertain, however it is likely that the oxygen levels after two and four hours are not too different. If we assume this, we arrive at the conclusion that the oxygen content by itself is not responsible for the increased ventilation, but rather that reduced oxygen 30 minutes before has a delayed effect on ventilation. A tenable hypothesis is to attribute this delayed response to stress. Furthermore, not a lot of difference is observed between respiration after one and two hours, advocating for some threshold over which the fish do not care about reduced DO content. It should also be mentioned that welfare score, although not evaluated, was likely reduced after the downbreathing, since the respiration change suggests a stress response in the fish.

As explored in the literature review, the resting frequency of salmon is around 1 Hz. This means that all the fish in the samples evaluated here are stressed. One possible explanation to this is that the recordings in this report are from the first ten minutes after lowering cameras into position in the tanks. This disturbs the fish for the span of the video. Another explanation could be that the recordings are from one of the tanks keeping eighteen degrees, stressing and increasing oxygen requirements of the salmon ([151], chapter 1).

7.4.3 Statistical analysis

At the time of writing this report, the full data set from the experiment was not received, and only three sample recordings from tank nine could be used. It will be interesting to see how the respiration frequency varies with tanks, and with recordings a while after the camera has been lowered into position. When more data is received, a thorough statistical analysis could be constructed,

making it possible to draw quantitative conclusions with regards to both pipeline performance and stress response of the fish.

7.4.4 Resource requirements

The resource requirements of the algorithm is high, spending over 2 minutes to process one second of video. This is prohibitively large for real time applications, however improvements can be made. The most time consuming part of the algorithm is the neural network, so reducing the depth of this, or use GPUs at evaluation time, would greatly speed up the pipeline.

Chapter 8

Future directions

Through this literature study and baseline development, a large set of opportunities have revealed themselves. An obvious continuation of the work done in this report is to refine the developed algorithm by improving its components. This could be to use RANSAC for frequency estimation, improving outlier rejection, training the neural network with a larger training set or automatically determine some of the parameters.

8.1 Fish identification

Although improvement of current methods are interesting, inclusion of new components seem more important for the performance of the algorithm. In particular, the poor fish tracking solutions are of great concern. As the fish respiration of different individuals in the same tank and at the same time differ by 0.5 to 1 Hz, the specific fish that stays in the camera frame and gives rise to good signals have a large impact on the frequency estimate. It is likely that some of the low error histograms are created from only a couple of fish, skewing the frequency estimate. This bias and randomness is making statistical analysis difficult, and reducing the quality of the results. Motivated by this, the most important addition to the current pipeline is the inclusion of fish identification capabilities. With this, the respiration frequency evolution of single fish can be tracked, making stress and welfare inference a lot easier. Furthermore, shoal analysis is simplified with knowledge of single fish behaviour.

8.2 Statistical analysis

After a sufficiently good pipeline has been constructed, a statistical analysis on the experimental data would allow us to hypothesize about the relationship between DO, temperature and ventilation, and subsequently stress. Interesting statistics would be the evolution of average shoal frequency, spread of shoal frequency or whether temperature and DO has statistically significant effects

on ventilation. By evaluating the welfare score of different salmon, ventilation and stress could be linked to welfare.

8.3 Other pipelines

Beside this, developing new algorithms is also an interesting path to explore. It does seem like the method with keypoint RCNN and mouth gape trajectories works well as a non temporal network, hence new approaches should incorporate temporal relationships. We found in this report that the shoal respiration frequency has a significant distribution among fish, but also that the shoal average contain important information. As such, both a 3D action classification algorithm for shoal frequency classification, or a convolutional LSTM network for single fish ventilation estimation is interesting avenues to pursue. The former requires a lot of computational power to train, and the latter requires a large manually annotated bounding box database. Due to this, neither is ideal for master thesis work.

8.4 Other behaviour indicators

In this project, the focus has been on fish ventilation. As touched upon in the literature survey, however, other features can also infer about fish stress. One such metric is the tail beat frequency, which might be possible to detect with a similar pipeline as the one developed here. The tail oscillations could even be detected in the same neural network as the ventilation frequency, by introducing more keypoints. Then, a combined respiration and tail beat frequency pipeline could be constructed, providing a stronger base for stress elucidation.

8.5 Plan for master thesis

Considering the above, the following list is a preliminary suggestion for work to be performed in the master thesis.

1. Improve robustness of the pipeline constructed in this report, and reduce the number of tunable parameters.
2. Include fish identification into the pipeline. This could be done by adding a new branch of keypoint RCNN used for fish recognition.
3. Perform a thorough statistical analysis on experimental data with the developed pipeline. Use this to link respiration, DO, temperature, stress and welfare.

Chapter 9

conclusion

In this report we have thoroughly examined stress and welfare in salmon, how it relates to breathing, and how ventilation can be inferred by computer vision methods. The validity of ventilation frequency as a stress metric has been demonstrated by the construction of a pipeline capable of coarsely discerning shoals exposed to different levels of Dissolved Oxygen (DO) from video streams. To construct and evaluate the algorithm, an experiment was performed in which salmon was recorded at different temperatures and oxygen levels. This project revealed work that can be performed in the subsequent master thesis of the author, including a complete statistical analysis on the data gathered in the ventilation experiment, as well as improvement on the constructed pipeline by enhancing existing modules, and by adding new ones.

Bibliography

[1] Espen Høgstedt. *Salmon breathing evaluation by deep learning*. 2022. URL: https://github.com/espenbh/salmon_respiration_ml_preproject.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] Kathleen M. Gilmour and Steve F. Perry. "Branchial Chemoreceptor Regulation of Cardiorespiratory Function". In: Fish Physiology 25 (2006), pp. 97–151. ISSN: 1546-5098. DOI: https://doi.org/10.1016/S1546-5098(06)25003-9.

[25] 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.

[26] 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.

[27] 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.

[28] Frank M. Smith and Peter S. Davie. "Effects of sectioning cranial nerves IX and X on the cardiac response to hypoxia in the coho salmon, Oncorhynchus kisutch". In: *Canadian Journal of Zoology* 62.5 (1984), pp. 766–768. DOI: 10.1139/z84-109. eprint: https://doi.org/10.1139/z84-109.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] Selye H. "The evolution of the stress concept". In: *Am Sci.* 61.6 (1973), pp. 692–699.

[35] 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.

[36] Geven EJW and Klaren PHM. "The teleost head kidney: Integrating thyroid and immune signalling". In: *Dev Comp Immunol.* 66 (2017), pp. 73–83. DOI: 10.1016/j.dci.2016.06.025.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] 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.

[61] 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.

[62] 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.

[63] 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.

[64] 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.

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

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] Merriam-Webster. *Merriam-Webster.com Dictionary*. Merriam-Webster, 2022. URL: https://www.merriam-webster.com/dictionary/welfare.

[75] 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.

[76] 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.

[77] Simon Thorpe, Denis Fize, and Catherine Marlot. "Speed of Processing in the Human Visual System". In: *Nature* 381 (July 1996), pp. 520–2. DOI: 10.1038/381520a0.

[78] Simon J. Thorpe. "Image Processing by the Human Visual System". In: *Advances in Computer Graphics*. Ed. by Gérald Garcia and Ivan Herman. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991, pp. 309–341. ISBN: 978-3-642-76286-4.

[79] David R. Bull. "Chapter 2 - The Human Visual System". In: *Communicating Pictures*. Ed. by David R. Bull. Oxford: Academic Press, 2014, pp. 17–61. ISBN: 978-0-12-405906-1. DOI: https://doi.org/10.1016/B978-0-12-405906-1.00002-7. URL: https://www.sciencedirect.com/science/article/pii/B9780124059061000027.

[80] 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.

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

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

[83] Joseph Walsh et al. "Deep Learning vs. Traditional Computer Vision". In: Apr. 2019. ISBN: 978-981-13-6209-5. DOI: 10.1007/978-3-030-17795-9_10.

[84] 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.

[85] *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.

[86] John Canny. "A Computational Approach to Edge Detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6 (1986), pp. 679–698. DOI: 10.1109/TPAMI.1986.4767851.

[87] N. Dalal and B. Triggs. "Histograms of oriented gradients for human detection". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1. 2005, 886–893 vol. 1. DOI: 10.1109/CVPR.2005.177.

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

[89] 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.

[90] Maria Eugenia Paz Reyes et al. “Computer Vision-Based Estimation of Respiration Signals”. In: *VIII Latin American Conference on Biomedical Engineering and XLII National Conference on Biomedical Engineering*. Ed. by César A. González Díaz et al. Cham: Springer International Publishing, 2020, pp. 252–261. ISBN: 978-3-030-30648-9.

[91] 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.

[92] Maria Jorquera et al. “Modelling and Validation of Computer Vision Techniques to Assess Heart Rate, Eye Temperature, Ear-Base Temperature and Respiration Rate in Cattle”. In: *Animals* 9 (Dec. 2019). DOI: 10.3390/ani9121089.

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

[94] 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.

[95] Ekrem Misimi et al. “Robust classification approach for segmentation of blood defects in cod fillets based on deep convolutional neural networks and support vector machines and calculation of gripper vectors for robotic processing”. In: *Computers and Electronics in Agriculture* 139 (June 2017), pp. 138–152. DOI: 10.1016/j.compag.2017.05.021.

[96] Òscar Lorente, Ian Riera, and Aditya Rana. *Image Classification with Classic and Deep Learning Techniques*. May 2021.

[97] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. DOI: 10.48550/ARXIV.1505.04597. URL: https://arxiv.org/abs/1505.04597.

[98] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Commun. ACM* 60.6 (May 2017), pp. 84–90. ISSN: 0001-0782. DOI: 10.1145/3065386. URL: https://doi.org/10.1145/3065386.

[99] Joseph Redmon et al. “You Only Look Once: Unified, Real-Time Object Detection”. In: *CoRR* abs/1506.02640 (2015). arXiv: 1506.02640. URL: http://arxiv.org/abs/1506.02640.

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

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

[102] 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.

[103] Gu Lingyun, Eugene Popov, and Dong Ge. *Fast Fourier Convolution Based Remote Sensor Image Object Detection for Earth Observation*. 2022. DOI: 10.48550/ARXIV.2209.00551. URL: https://arxiv.org/abs/2209.00551.

[104] Neal Wadhwa et al. "Phase-Based Video Motion Processing". In: *ACM Trans. Graph. (Proceedings SIGGRAPH 2013)* 32.4 (2013).

[105] Mohamed A. Elgharib et al. "Video magnification in presence of large motions". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 4119–4127. DOI: 10.1109/CVPR.2015.7299039.

[106] Tae-Hyun Oh et al. "Learning-based Video Motion Magnification". In: *arXiv preprint arXiv:1804.02684* (2018).

[107] 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.

[108] Weihao Yu et al. *MetaFormer Is Actually What You Need for Vision*. 2021. DOI: 10.48550/ARXIV.2111.11418. URL: https://arxiv.org/abs/2111.11418.

[109] 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.

[110] Harold W. Kuhn. "The Hungarian method for the assignment problem". In: *Naval Research Logistics (NRL)* 52 (2010).

[111] Rudolf E. Kálmán. "A new approach to linear filtering and prediction problems" transaction of the asme journal of basic". In: 1960.

[112] 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.

[113] 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.

[114] 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.

[115] M. PARISA BEHAM and S. MOHAMED MANSOOR ROOMI. "A REVIEW OF FACE RECOGNITION METHODS". In: *International Journal of Pattern Recognition and Artificial Intelligence* 27.04 (2013), p. 1356005. DOI: 10.1142/S0218001413560053. eprint: https://doi.org/10.1142/S0218001413560053. URL: https://doi.org/10.1142/S0218001413560053.

[116] Petr Cisar et al. "Computer vision based individual fish identification using skin dot pattern". In: *Scientific Reports* 11 (2021).

[117] 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.

[118] 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.

[119] Ching-Yao Chuang et al. "Debiased Contrastive Learning". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 8765–8775. URL: https://proceedings.neurips.cc/paper/2020/file/63c3ddcc7b23daa1e42dc41f9a44a873-Paper.pdf.

[120] 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.

[121] 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.

[122] 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.

[123] James W Cooley and John W Tukey. "An algorithm for the machine calculation of complex Fourier series". In: *Mathematics of computation* 19.90 (1965), pp. 297–301.

[124] 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.

[125] 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.

[126] Darren R. Myatt et al. "NAPSAC: High Noise, High Dimensional Robust Estimation - it's in the Bag". In: *BMVC*. 2002.

[127] 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.

[128] 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.

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

[130] 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.

[131] 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.

[132] 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.

[133] 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.

[134] 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.

[135] Espen Høgstedt, Hielke Walinga, and Tommaso Tofacchi. *An exploration of fish nets [Unpublished]*. June 2022.

[136] 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.

[137] Ross Girshick. *Fast R-CNN*. 2015. DOI: 10.48550/ARXIV.1504.08083. URL: https://arxiv.org/abs/1504.08083.

[138] Markus M. Breunig et al. "LOF: Identifying Density-Based Local Outliers". In: *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*. SIGMOD '00. Dallas, Texas, USA: Association for Computing Machinery, 2000, pp. 93–104. ISBN: 1581132174. DOI: 10.1145/342009.335388. URL: https://doi.org/10.1145/342009.335388.

[139] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation Forest". In: *2008 Eighth IEEE International Conference on Data Mining*. 2008, pp. 413–422. DOI: 10.1109/ICDM.2008.17.

[140] Ananth Ranganathan. "The levenberg-marquardt algorithm". In: *Tutoral on LM algorithm* 11.1 (2004), pp. 101–110.

[141] 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.

[142] 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.

[143] A. Buslaev et al. "Albumentations: fast and flexible image augmentations". In: *ArXiv e-prints* (2018). eprint: 1809.06839.

[144] J. Deng et al. "ImageNet: A Large-Scale Hierarchical Image Database". In: *CVPR09*. 2009.

[145] Josef Seabold Skipper and Perktold. "statsmodels: Econometric and statistical modeling with python". In: *9th Python in Science Conference*. 2010.

[146] 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.

[147] Precision Measurement Engineering. *miniDOT Logger user's manual*. 2021. URL: https://www.pme.com/products/minidot.

[148] Navaneeth Bodla et al. *Soft-NMS – Improving Object Detection With One Line of Code*. 2017. DOI: 10.48550/ARXIV.1704.04503. URL: https://arxiv.org/abs/1704.04503.

[149] Nicolas Carion et al. “End-to-End Object Detection with Transformers”. In: *CoRR* (2020). arXiv: 2005.12872. URL: https://arxiv.org/abs/2005.12872.

[150] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

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