A novel approach for wild fish monitoring at aquaculture sites: wild fish presence analysis using computer vision

Kana Banno*, Håvard Kaland, Alberto Maximiliano Crescitelli, Stig Atle Tuene, Grete Hansen Aas, Lars Christian Gansel

Department of Biological Sciences Ålesund, Norwegian University of Science and Technology, 6025 Ålesund, Norway

ABSTRACT: Aquaculture in open sea-cages attracts large numbers of wild fish. Such aggregations may have various impacts on farmed and wild fish, the environment, fish farming, and fisheries activities. Therefore, it is important to understand the patterns and amount of wild fish aggregations at aquaculture sites. In recent years, the use of artificial intelligence (AI) for automated detection of fish has seen major advancements, and this technology can be applied to wild fish abundance monitoring. We present a monitoring procedure that uses a combination of multiple cameras and automatic fish detection by AI. Wild fish in images collected around commercial salmon cages in Norway were automatically identified and counted by a system based on the real-time object detector framework YOLOv4, and the results were compared with manual human counts. Overall, the automatic system resulted in higher fish numbers than the manual counts. The performance of the system was satisfactory regarding false negatives (i.e. non-detected fish), while the false positive (i.e. objects wrongly detected as fish) rate was above 7%, which was considered an acceptable limit of error in comparison with the manual counts. The main causes of false positives were confusing backgrounds and mismatches between detection thresholds for automated and manual counts. However, these issues can be overcome by using training images that represent real scenarios (i.e. various backgrounds and fish densities) and setting proper detection thresholds. We present here a procedure with great potential for autonomous monitoring of wild fish abundance at aquaculture sites.

KEY WORDS: Wild fish · Sea-cage aquaculture · Automatic fish detection · Computer vision · Deep neural network · Artificial intelligence

1. INTRODUCTION

Uneaten feed drifting out from aquaculture sea cages is likely to be a significant attractant for wild fish (Tuya et al. 2006). Substantial aggregations of wild fish have been observed around Norwegian salmon farms (Dempster et al. 2009, Uglem et al. 2009, Otterå & Skilbrei 2014) and in other areas such as the Mediterranean Sea (Dempster et al. 2002, 2004), the Canary Islands (Dempster et al. 2005), and the coast of Australia (Dempster et al. 2004).

More than 160 fish species have been observed around open-cage farms worldwide (reviewed in Uglem et al. 2014). In Norway, common species observed at farm sites are saithe *Pollachius virens*, Atlantic mackerel *Scomber scombrus*, and Atlantic cod *Gadus morhua* (Dempster et al. 2009).

The aggregations of wild fish pose numerous concerns. Wild fish may act as vectors for pathogens (e.g. Snow et al. 2002, Uglem et al. 2009). For instance, saithe are known to be preferred hosts for sea lice *Caligus elongatus* and may act as vectors for trans-

*Corresponding author: kana.banno@ntnu.no

ferring sea lice among farms (Bruno & Stone 1990). Migration patterns of wild fish may be altered (e.g. Bjordal & Johnstone 1993), with wild fish becoming resident at farm sites, which may reduce their availability to fisheries, since farms are restricted sites for fishing vessels (Uglem et al. 2014). In addition, access to high-energy feed pellets may benefit wild fish and increase their growth and reproduction (Uglem et al. 2014, Gonzalez-Silvera et al. 2020). Commercial feed can affect the composition of fatty acid components in liver and muscle tissues of saithe (Arechavala-Lopez et al. 2015), and consumers have been found to slightly prefer the taste of fillets from saithe caught near fish farms over that of conspecifics caught away from those sites (Uglem et al. 2017). Some farmed fish species such as Atlantic cod have the ability to create holes in nets (Damsgård et al. 2012, Uglem et al. 2014). Wild cod might also show this behaviour, which could lead to farmed fish escaping.

Methods for wild fish abundance monitoring include visual counts by SCUBA divers (Dempster et al. 2002, 2004, 2005, Boyra et al. 2004, Fernandez-Jover et al. 2008, Oakes & Pondella 2009), underwater video systems combined with manual quantification of fish (Dempster et al. 2009, 2010, Uglem et al. 2009, Bacher et al. 2012), stationary stereo video systems (e.g. AQ1 AM100; AQ1 Systems 2022) (Stagličić et al. 2017), and hydroacoustic tools (Giannoulaki et al. 2005, Goodbrand et al. 2013). These methods have provided valuable knowledge about wild fish abundance around aquaculture farms but are either time-consuming or their precision and repeatability depend strongly on the skill and training of their operators. Hydroacoustic methods can be used for long-term monitoring but do not allow detailed investigations about individual fish such as species determination.

The use of artificial intelligence (AI) in fish recognition is a potential solution for long-term monitoring of wild fish abundance. Much effort has been made to recognize fish species using deep learning techniques based on deep neural networks (DNNs) (Hossain et al. 2016, Mandal et al. 2018, Siddiqui et al. 2018, Rauf et al. 2019, Salman et al. 2020, Zhang et al. 2020). Such methods allow a drastic reduction in human workload by supporting human decision-making and automating some tasks. To perform good quality fish detection, the neural networks need to be trained with large, annotated image data sets. Crescitelli et al. (2021) used computer vision techniques to create a data set containing saithe and salmonid fish images that can be used to train DNNs

to recognize those species. Some preliminary tests have evaluated the performance of the system developed in their study and suggest that the system is promising for the use of automatic detection of wild fish at farm sites.

Given the potential impact of wild fish aggregations on wild and farmed fish, as well as on commercial fisheries, it is important to have thorough knowledge about wild fish abundance patterns around sea cages. In addition, obtaining more information about wild fish around the sea cages can be helpful for the farmers because they can, for instance, inspect the cage nets more carefully when wild cod are present. This process may prevent salmon escaping events that otherwise might have been caused by wild cod biting the nets. Moreover, a good understanding of wild fish abundance is necessary for a proper assessment of aquaculture impacts on the environment. Wild fish have the potential to utilize particulate wastes from the sea cages and mitigate the impacts of sea-cage farming on the surrounding environment (Ballester-Moltó et al. 2017). If it is known that large aggregations of wild fish exist around the sea cages, these mitigating impacts need to be taken into account. The creation and existence of wild fish aggregations is something that aquaculture policy should consider. However, our understanding of wild fish aggregation dynamics is still limited, partly because of a lack of observation methods that can provide continuous and timely information on wild fish abundance patterns.

The aim of the present study was to test a wild fish abundance monitoring procedure using a combination of multiple cameras and automatic fish detection by AI. We evaluated and propose a monitoring system to analyze wild fish abundance around sea cages autonomously which exhibits acceptable accuracy compared to manual analysis. To demonstrate that the presented monitoring method is suitable for the study of wild fish abundance around the sea cages, some examples of the wild fish distribution patterns are presented in this article. However, the focus of this study was on the monitoring method itself. Detailed investigation of wild fish distribution patterns in relation to various abiotic and biotic factors was outside the scope of this study. As a first step, this study aimed at automatic quantification of wild fish in general without species identification.

Considering that automatic recognition models such as salmon individual recognition (e.g. Cermaq 2022) and automatic sea lice counting (e.g. Aquabyte 2022) have already been developed and used in aquaculture, the automatic wild fish detection sys-

tem presented in this study is not the most advanced technology. The novelty of this study is that the available technology (i.e. underwater video monitoring and automatic fish detection techniques) was adjusted for the purpose of monitoring wild fish around coastal sea cages. The setup of multiple cameras around the sea cages combined with the use of an automatic fish detection system is novel and potentially a great help to deepen our understanding of the dynamics of wild fish occurrence and aggregations at farm sites.
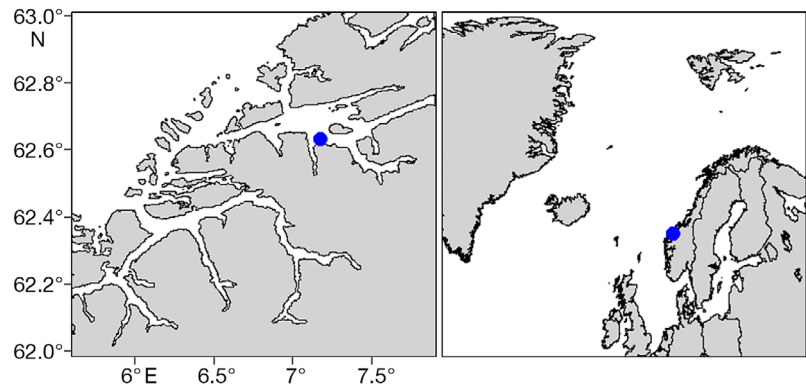

Fig. 1. Farming locality in Gjermundnes (blue dot), Møre and Romsdal county, Norway

2. MATERIALS AND METHODS

2.1. Study site

This study was conducted at the commercial salmon aquaculture site Gjermundnes (62° 63′ N, 7° 19′ E), located in Møre and Romsdal county, Norway (Fig. 1). Wild fish abundance was monitored around 2 sea cages (126 m circumference, 33 m deep, 18 850 m^3 water volume), which were located next to each other in a common 2-row mooring framework (Cage 1 and Cage 2 in Fig. 2). These 2 cages were selected for this study because they were stocked with salmon throughout the entire study period. At the end of the study period, Cage 1 and Cage 2 contained roughly 190 000 (average weight: 1.6 kg) and 140 000 (average weight: 2.5 kg) salmon, respectively. An inner and outer buoy, part of the mooring framework located approximately 100 and 360 m

away from Cage 1, were used as farm-internal control sites (Fig. 2).

2.2. Video recording

Video recordings of wild fish around Cage 1, Cage 2, and the 2 control sites (buoys) were collected using GoPro cameras (GoPro 4 and 5) fixed to a cylindrical plastic pipe on 5 separate days in July and October 2020. Three cameras fixed to the cylinder recorded videos simultaneously, covering approximately 2/3 of a circular horizontal angle (Fig. 3). One camera had approximately 80° horizontal field of view when used underwater. Video recording was conducted at 6 different positions equally spaced around a cage, where 0° was located between 2 feed spreaders, and at a single position at each buoy. The feed spreaders were oriented in a south-west direction from the cen-
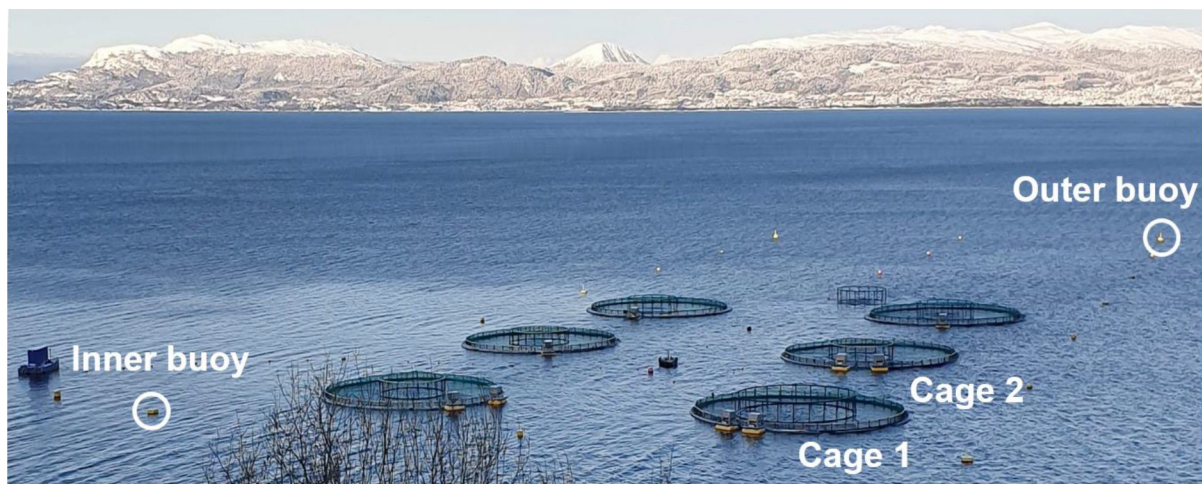

Fig. 2. The 2 sea cages and 2 buoys where video monitoring of wild fish was conducted
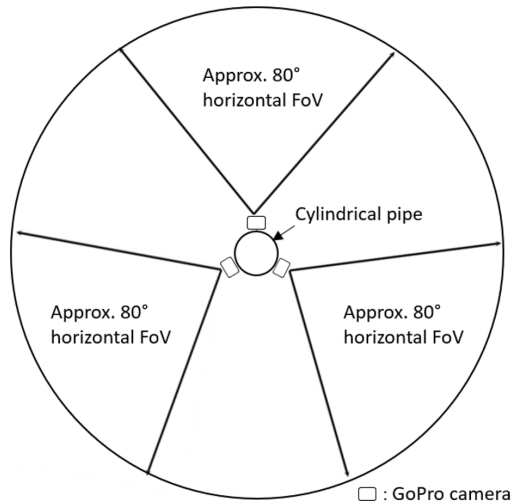
Fig. 3. Simplified sketch of the top view of the video monitoring equipment. Approximately 2/3 of the circular angle was covered by the 3 GoPro cameras. FoV: field of view

ter of the cages. At each position (0, 60, 120, 180, 240, and 300° around the cages and at the 2 buoys), the cameras were lowered down to 2, 7, 13, 20, 30, and 40 m depths and recorded for 2 min at each depth (Fig. 4). Videos were recorded during the morning hours (09:00–12:00 h), when feed was provided from the feed spreaders. Throughout the 5 sampling days, approximately 24 h of video footage was obtained, including the time used for camera transfer between different depths.

2.3. Video quality check, adjustment, and image extraction

The videos from each position around the cages and at each buoy were cut into 6 separate videos (2 min each, corresponding to each depth). The light intensity was low in some of the videos from 30 and
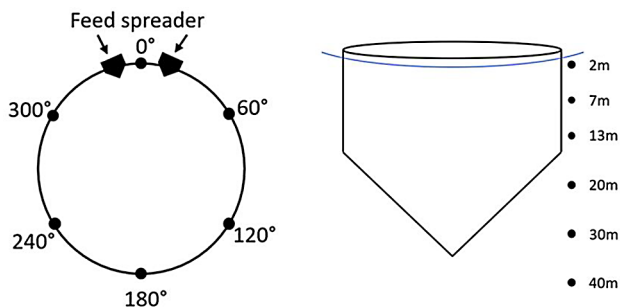

Fig. 4. The 6 positions around a cage (left) and 6 depths (right) where video filming of wild fish was conducted (total: 36 positions)

40 m depths, and it was difficult to detect the fish in the original videos. In such cases, lightness and contrast were manually adjusted using the video processing tool VideoProc 3.9 (Digiarty Software). The original videos were replaced by the adjusted videos when the adjustment enabled the discovery of new fish that were initially hidden in dark areas. In VideoProc, it was possible to adjust lightness and contrast manually while playing videos. This software also allowed us to save new videos with the adjusted settings. The adjusted videos were identical to the originals except for the lightness and contrast settings. The replacement of the original videos with the adjusted videos was done manually, with great caution not to duplicate the video files. In total, 33 out of 120 videos from 30 and 40 m depths were adjusted. From each 2 min video, 8 frames were extracted with an interval of 5 s (starting point: 40 s; ending point: 75 s). The first 40 s and the last 45 s were discarded to avoid extracting images when cameras were transitioning between different depths. In total, 10 080 images were extracted for further analyses.

2.4. Automated counts of wild fish

Wild fish in the extracted images were counted both manually and automatically, the latter by using an automated system based on deep learning. The fish detection systems (hereafter, 'models') were implemented based on the real-time object detector framework YOLOv4 (Bochkovskiy et al. preprint https://arxiv.org/abs/2004.10934). The detector can be trained for multiple types of objects (classes). In this work, the models were trained to detect fish as one class, without specifying species. Three different models were tested, and their performance was compared. Each of the models was trained with a different set of images (Table 1).

Model 1 was trained with a subset of images from Open Images Dataset (OID) v.6. OID is a publicly available data set containing 1.9 million annotated images distributed across 600 classes (https://storage.googleapis.com/openimages/web/download.html; accessed 19 Aug 2021). The 'Fish' class contained 5774 annotated images with a variety of fish species and backgrounds, which were used to train Model 1. Model 2 was trained with a set of 2072 unpublished images taken at salmon farm sites along the Norwegian coast. Those images contained only saithe *Pollachius virens* at varying densities and with similar backgrounds to those in the footage evalu-

Table 1. Data sets used to train each neural network model tested in this study

	Data set	No. of images used for training	Image features
Model 1	Open Image Dataset (OID) v.6, class: fish	5774	Images from public data set Various fish species and backgrounds Presence of wrong labels Presence of fish in the form of food Presence of inaccurate annotations
Model 2	Saithe image data set	2072	Images taken around salmon fish farms in Norway Only saithe images, in various densities
Model 3	Corrected OID plus saithe	6996 (4924 from OID, 2072 from saithe data set)	Combination of images from OID (with corrected annotation) and saithe image data set Undesired images in OID were discarded Undesired annotations in OID were manually corrected

ated in this study. Annotation of fish (saithe) was done automatically by software using the methodology given in Crescitelli et al. (2021). To ensure that these annotations were correct, the images with the annotations were manually checked and corrected in case they were wrong. OID included several types of images that were not suitable for our work (Fig. 5), which was initially not known. Some of the images were of fish processed as food, some were incorrectly labeled, and others had missing annotations (i.e. bounding boxes). After discarding such undesired images and correcting the annotations, a total of 4924 properly annotated OID images were kept and combined with the 2072 saithe images used in the previous model to train Model 3. We initially planned to use Models 1 and 2, and then created Model 3 after it was discovered that not all images from OID were suitable for training the models.

The 3 different models were evaluated using the set of images extracted from the video footage around the fish farm. For each input image, the system generated a new image; this image contained bounding boxes surrounding areas where it predicted the presence of a fish as well as corresponding confidence scores. Any predicted bounding box with confidence scores below 30% was discarded. This threshold was chosen empirically based on pre-tests over a small number of images. The number of detected fish per image was automatically stored in .csv files for further analyses.

2.5. Manual counts of wild fish

Images processed by the models were manually assessed for the correct number of fish and false detections. Counter 1 (hereafter C1) checked the entire set of 10 080 images to avoid potential biases and for a complete comparison with the automated analysis. Objects in the images were identified and counted as fish when they showed visible characteristics of fish e.g. body shape, color, lateral line, pattern (in the case of mackerel), regardless of size and regardless of whether they were part of a fish or the full body (strict threshold). This excluded many objects that were likely fish but did not show any of the aforementioned visible characteristics. Another less strict count method was used to assess the error rates of the automatic detection system (see below). Taxonomical identification to species level was per-
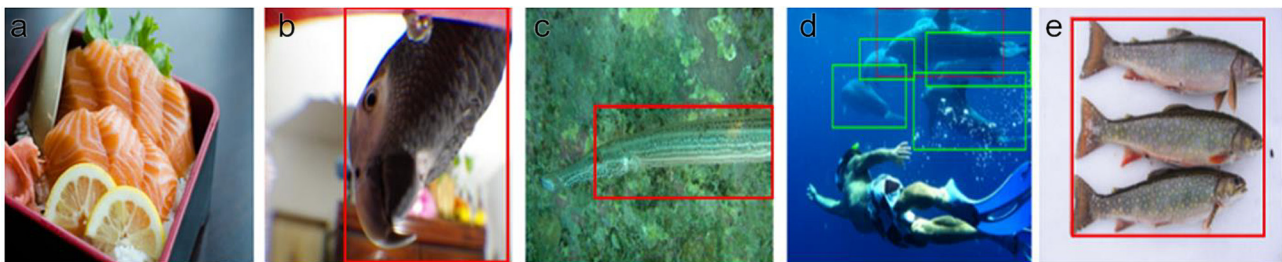

Fig. 5. Types of undesired images and annotations in the Open Image Dataset. (a) Fish processed as food; (b) wrong label (something that is not a fish but labeled as a fish); (c) wrong bounding box alignment; (d) wrong label and missing bounding boxes (original box in black and added bounding boxes in green); (e) individuals grouped as one fish

formed when possible; otherwise, the objects were categorized as 'unidentified fish'. The manual counting categories were as follows: (1) false positive (FP): objects that were wrongly identified as fish by AI; (2) false negative (FN): objects that were clearly fish but were not identified as fish by AI; (3) saithe: objects that were clearly saithe; (4) mackerel: objects that were clearly mackerel; (5) cod: objects that were clearly cod; and (6) unidentified fish: objects that were clearly fish, but doubtful which species it was.

Salmon inside the cages were present in a small number of images because the camera lenses were occasionally directed towards the cage net. In such cases, salmon were excluded from the categories, since our interests were in monitoring wild fish.

2.6. Estimation of error/deviation (manual)

To evaluate the accuracy of the automated detection system, it was necessary to benchmark human accuracy. Potential deviations caused by manual counts needed to be considered in the benchmarking process, and therefore the following types of potential deviations were estimated: (1) within-person deviation; (2) deviation among different people; and (3) effects of the presence of bounding boxes.

Deviation within a person's counts was estimated as follows. C1 counted fish in 10 images 5 times each, randomly in between other images so that the previous count could not be remembered. The test images were selected with the purpose of including different varieties of fish densities. The standard deviation among the counting results of the 5 trials was calculated.

To determine deviation among the counters, 4 different people (C1, C2, C3, and C4) independently counted wild fish in the same set of 864 images taken around Cage 1 on 8 July 2020 (hereafter the 'test cage') under the above-mentioned counting criteria. These were employees at the Department of Biological Sciences Ålesund, Norwegian University of Science and Technology (NTNU), and therefore they had prior experience identifying fish. The standard deviation among the 4 people's total fish counts was calculated.

Lastly, biases caused by the presence of bounding boxes were assessed by counting a set of 50 images twice, with and without the bounding boxes. The images that had the largest number of bounding boxes were used for this test, which was performed by C1. The percentage difference in total number of detected fish based on images with and without the bounding boxes was calculated. In addition, the effect of the presence of bounding boxes on manual counts was assessed by using a Wilcoxon signed-rank test. This analysis was performed using R v.4.0.2 (R Core Team 2020).

2.7. Estimation of error (AI — Models 1, 2, and 3)

To evaluate the accuracy of automated fish counts, 2 types of error were calculated: (1) FP rate: the percentage of incorrect detections of fish by the automated detection system; and (2) FN rate: the percentage of fish that were missed by the automated detection system out of all fish that were detected manually.

For Model 1, FPs and FNs for all 10 080 images were manually assessed, and an overall FP and FN rate including all 5 sampling days was obtained. For Models 2 and 3, FPs and FNs were manually assessed in 200 images with the 100 highest numbers of FPs and FNs, based on the manual FP and FN counts of the Model 1 output. In addition, manual counts of FPs and FNs were conducted for all 864 images from the test cage for all the models to evaluate whether the use of different types of training image data sets could reduce errors.

In addition, the quality of the models was assessed by calculating an F_1 score, which is a harmonic mean of precision and recall. Calculation of F_1 was made based on the results from the 864 images from the test cage, using the formula: $F_1 = 2 \times$ (precision × recall)/(precision + recall). Precision and recall can be calculated as follows (Powers 2007): precision = TP/(TP + FP); recall = TP/(TP + FN); where TP is true positive (the number of positives determined by a model that were actually positive identifications). When we use the F_1 score to compare models, the model with the highest score is considered the best.

An additional assessment of the FPs of Model 3 was performed for the automated counts of wild fish in the 864 images from the test cage. In addition to the initial counts of FPs under the strict threshold, FPs in these images under a relaxed detection threshold were counted. This assessment was made because Model 3 identified more fish than the manual assessments under the strict threshold. Most of the FPs created by Model 3 were, however, not true FPs (i.e. detection of non-fish objects), but rather detection of blurred or small objects that most likely were fish. Under the relaxed threshold, we counted all objects as fish except those objects that were obviously not fish or individual fish that were identified multiple

Table 2. Manual fish count results. Each value is the sum of detected objects under each category. Values for each cage are the sum of detected objects in a total of 864 images per sampling date. Values for each buoy are the sum of detected objects in a total of 144 images per sampling date

Date (yyyy-mm-dd)	Location	Saithe	Mackerel	Cod	Unidentified fish	Total
2020-07-08	Cage 1	791	8	0	861	1660
	Cage 2	310	0	0	679	989
	Inner buoy	0	0	0	0	0
	Outer buoy	0	0	0	0	0
2020-07-15	Cage 1	4	0	0	41	45
	Cage 2	166	3	0	735	904
	Inner buoy	0	0	0	0	0
	Outer buoy	1	0	0	2	3
2020-07-24	Cage 1	65	0	0	506	571
	Cage 2	21	134	0	329	484
	Inner buoy	0	0	0	0	0
	Outer buoy	0	0	0	0	0
2020-07-31	Cage 1	209	7	0	838	1054
	Cage 2	117	44	0	813	974
	Inner buoy	0	0	0	0	0
	Outer buoy	0	0	0	0	0
2020-10-01	Cage 1	180	82	2	2525	2789
	Cage 2	75	3	0	551	629
	Inner buoy	0	0	0	0	0
	Outer buoy	0	0	0	0	0
Total		1939	281	2	7880	10102

times. We also subtracted FPs from the outcome of Model 3 to assess how it affected the overall fit of the manual and automated counting results.

3. RESULTS

3.1. Results of manual counting

Manual counting of 10 080 images resulted in the identification of 10 102 wild fish (Table 2). Throughout the 5 sampling dates, only 3 wild fish were detected at the control sites, whereas an average of 841 wild fish were counted in the same volume of water around the sea cages. In many images, it was not possible to identify species, as the majority of the wild fish were blurred or their orientation made it difficult to recognize their body shape and appearance. This issue was independent of depth and light intensity; rather, it was caused by factors such as water turbidity, distance from the cameras, camera settings, and motion blur. Overall, the most abundant category in the manual counts was 'unidentified fish', which accounted for 78% of the fish detected. In total, 1939 saithe *Pollachius virens*, 281 mackerel

Scomber scombrus, 2 Atlantic cod *Gadus morhua*, and 7880 unidentified fish were detected.

Most of the mackerel (99.6%) were detected at depths above 7 m, while most of the saithe (99.9%) were detected at depths below 20 m. Horizontal distribution patterns of wild fish varied among different sampling dates and cages. Fig. 6 shows the horizontal and vertical distribution of saithe, mackerel, and total fish detected around Cage 1 as an example. We chose to show wild fish distribution around the sea cages and at the control sites to highlight the need for enough sampling stations to detect spatial distribution patterns, as there was clearly variation in fish distribution within a single farm site (Table 2, Fig. 6).

3.2. Potential deviation in manual counting

When C1 counted fish in 10 images 5 times each, the standard deviation among counts in the 5 trials varied between 0 and 1.1, which accounted for 0–14.3% of average fish counts per image. Based on the 10 images, percentage of standard deviations to the average manual fish counts among the 5 trials was 4.8% (Table 3).
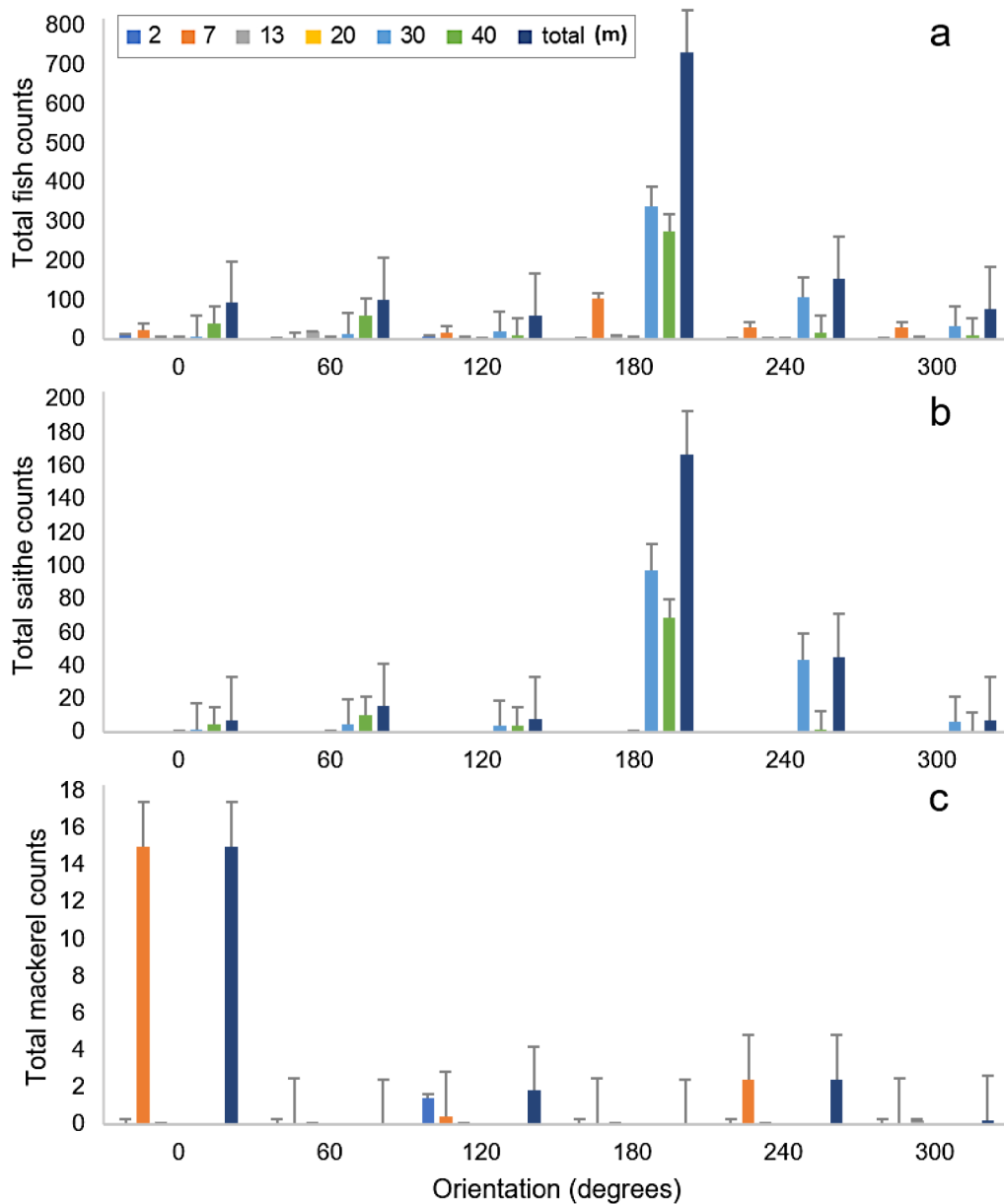
Fig. 6. Average (+SE) number of (a) total manual fish counts, (b) saithe counts, and (c) mackerel counts per day at each filming point around Cage 1 throughout 5 sampling dates

C1, C2, C3, and C4 counted a total of 1660, 1468, 1802, and 1660 fish, respectively, in the same set of 864 images taken around the test cage. The largest difference was between C2 and C3. The latter counted 22.8% more fish than C2, whereas C1 and C4 had exactly the same total fish counts. However, on a single image basis, the number of detected fish also differed between C1 and C4 as well as among all the counters (Fig. 7). The standard deviation among the 4 counters was 112.3. This accounted for 6.9% of the average total fish counts. Two distinct outliers were observed in

C3's counts, and those were excluded from the calculation of the average fish counts and the standard deviation.

In total, 1065 fish were detected when images without bounding boxes were assessed, while 1119 fish were detected when images with the bounding boxes were checked. C1 detected 4.8% more fish when bounding boxes were present on the images. The median number of wild fish detected by manual work was significantly different when images were examined with and without bounding boxes (Wilcoxon signed-rank test, p < 0.001).

Table 3. Standard deviations (SD) in manual fish counts calculated when a single person (Counter 1) counted 10 images 5 times each (in between other images)

Image no.	Average counts among 5 attempts (±SD)	SD/manual fish counts average (%)
1	26.8 (±0.4)	1.5
2	22.4 (±0.8)	3.6
3	5.8 (±0.4)	6.9
4	38.6 (±0.5)	1.3
5	10.0 (±1.1)	11.0
6	4.0 (±0.0)	0.0
7	31.2 (±1.0)	3.1
8	16.6 (±0.5)	3.0
9	26.4 (±0.8)	3.0
10	5.6 (±0.8)	14.3
Average		4.8

3.3. Results of automated counting

Model 1 (fish OID) detected a total of 8350 wild fish (Table 4). Overall, FP and FN rates were 21.3 and 35.0%, respectively. Large numbers of FPs occurred in images that contained objects such as cage structures, lice skirts, and biofouling organisms attached to such structures. FNs often occurred in images where wild fish individuals overlapped substantially or in high-density situa-

tions (Fig. 8). Based on 100 images with the largest numbers of FPs (hereafter 'FP test set'), the FP rate was 37.2%. Similarly, based on 100 images with the largest numbers of FNs (hereafter 'FN test set'), the FN rate was 67.2%. Based on 200 images in the combined test set (combination of FP and FN test sets), FP and FN rates were 25.8 and 49.6%, respectively. Based on 864 images taken around the test cage, FP and FN rates were 7.5 and 26.8%, respectively (Fig. 9).

Model 2 (saithe) detected a total of 14 945 wild fish (a 79.0% increase from Model 1's counts). The FP rate calculated based on the FP test set was 36.3%. The FN rate calculated based on the FN test set was 17.4%. Based on the combined test set, the FP and FN rates were 21.4 and 15.1%, respectively. Based on the 864 images from the test cage, the FP and FN rates were 19.9 and 4.0%, respectively (Fig. 9). Model 2 still detected background objects, and in some cases, the number of FPs was higher than in Model 1. On the other hand, fewer FNs occurred in images with high fish densities and overlapping fish compared to Model 1 (Fig. 8).

Model 3 (corrected fish OID plus saithe) detected a total of 12 980 wild fish (55.4% increase from Model 1 counts). The FP rate calculated based on the FP test set was 22.2%. The FN rate calculated based on the FN test set was 7.2%. Based on the combined
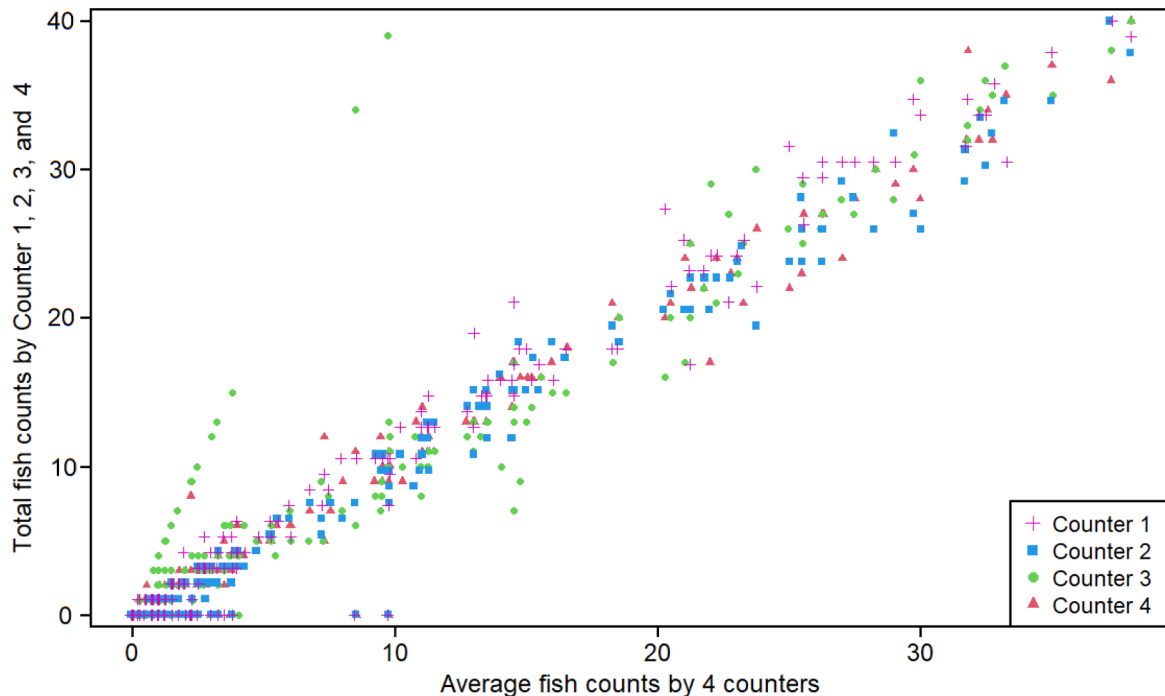

Fig. 7. Manual fish count results by Counters 1, 2, 3, and 4 in the 864 images taken around the test cage (Cage 1 on 8 July). Each mark represents the number of detected fish in each image; *x*-axis is the average count in each image by the 4 counters

Table 4. Automated wild fish count results. Values for each cage are the sum of objects detected as fish in a total of 864 images per sampling date. Values for each buoy are a sum of objects detected as fish in a total of 144 images per sampling date. Model 1: Open Image Dataset (OID) model; Model 2: saithe model; Model 3: corrected OID plus saithe model. FP: false positive; FN: false negative. For Model 1, the number of FPs and FNs in the 10080 images were manually counted

Date (yyyy-mm-dd)	Location	Total, Model 1	Total, Model 2	Total, Model 3	FP, Model 1	FN, Model 1
2020-07-08	Cage 1	1313	1989	1979	98	445
	Cage 2	793	1177	1148	57	253
	Inner buoy	7	21	5	7	0
	Outer buoy	0	1	1	0	0
2020-07-15	Cage 1	65	174	164	35	15
	Cage 2	674	1173	1078	94	324
	Inner buoy	3	0	1	3	0
	Outer buoy	7	19	11	4	0
2020-07-24	Cage 1	693	919	879	198	76
	Cage 2	560	799	736	173	97
	Inner buoy	1	34	4	1	0
	Outer buoy	10	11	5	10	0
2020-07-31	Cage 1	1009	1505	1367	215	260
	Cage 2	1070	1774	1386	342	246
	Inner buoy	2	6	3	2	0
	Outer buoy	4	0	2	4	0
2020-10-01	Cage 1	1610	3546	3317	311	1490
	Cage 2	511	1743	877	208	326
	Inner buoy	7	34	13	7	0
	Outer buoy	11	20	4	11	0
Total		8350	14945	12980	1780	3532

test set, FP and FN rates were 17.8 and 5.8%, respectively. Based on the 864 images from the test cage, the FP and FN rates were 18.5 and 2.8%, respectively (Fig. 9). In some images, Model 3 had a higher number of FPs because the system detected objects that were not identified as fish in the manual counts. Those objects were typically small, black, blurred objects in the background that were most likely fish but were not clear enough to convince C1 during the manual counts (Fig. 8). This phenomenon was also observed when Models 1 or 2 were used, but it was most notable with Model 3.

The F_1 scores of Model 1, Model 2, and Model 3 were 0.82, 0.87, and 0.89, respectively. After subtracting FPs from Model 3's outcome based on the FP, FN, and combined test sets as well as images from the test cage, the percentage difference between Model 3 and the manual counts ranged from 2.5–7.2% (Table 5). Based on the strict fish detection threshold, 366 FPs were identified by Model 3 in the 864 images from the test cage, whereas 53 (14% of 366) were detected with the relaxed detection threshold. The amount of time spent inspecting FPs in the 864 images based on the relaxed detection

Table 5. Total fish counts by manually counting and by artificial intelligence (AI) Model 3 using a false positive (FP) test set (100 images that had the highest number of FPs when Model 1 was used), false negative (FN) test set (100 images that had highest number of FNs when Model 1 was used), the combined test set, and images from the test cage (Cage 1 on 8 July 2020). The percentage of AI overestimates were determined as: [(AI counts − manual counts) / (manual counts)] × 100. The final column shows how subtraction of FPs changed the fit of the AI counts (by Model 3) to the manual counts

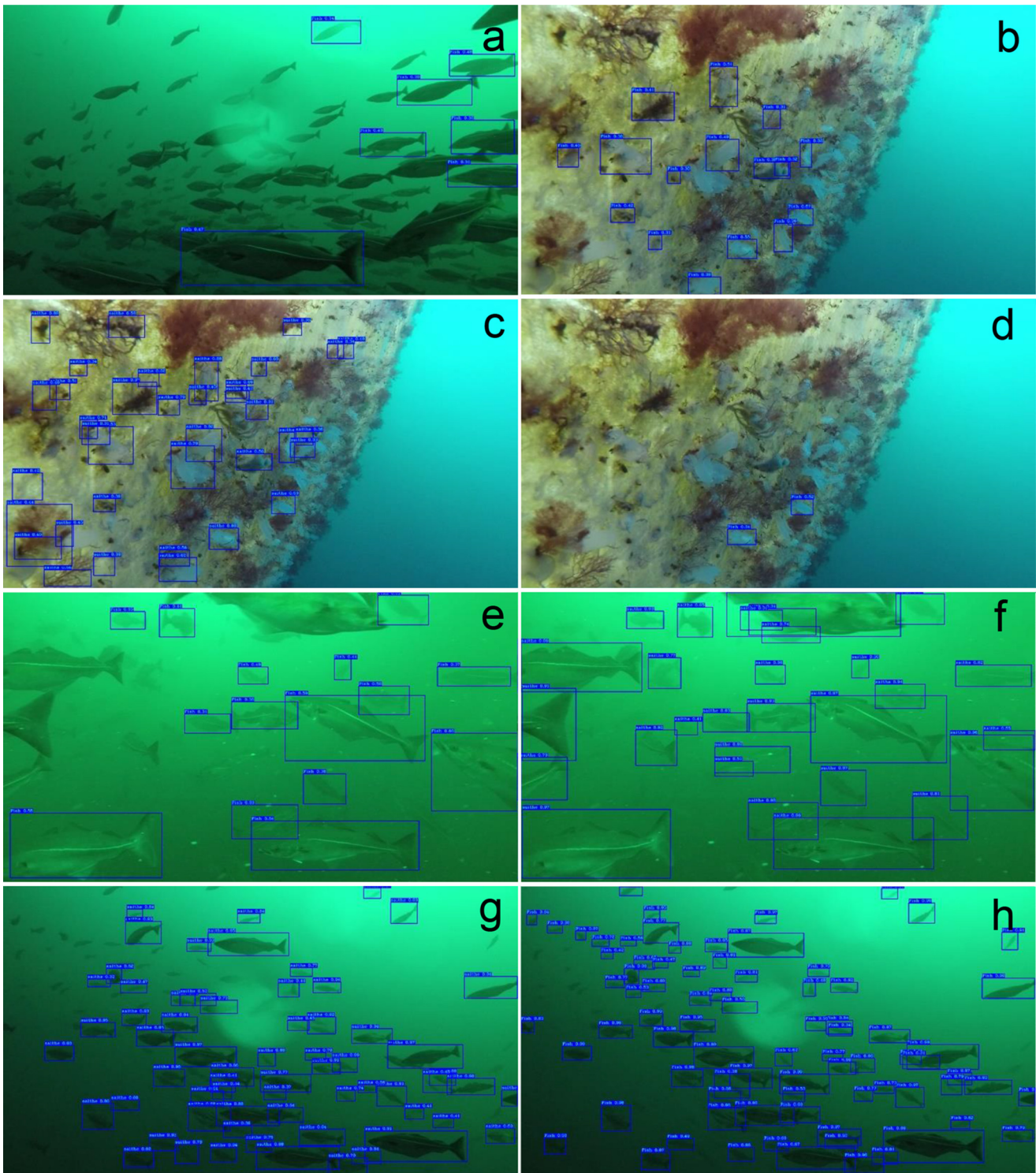| Images analyzed	AI counts, Model 3	Manual counts	FPs	FP rate (%)	AI counts, overestimate (%)	AI counts, overestimate (%) after subtracting FPs
100 images, FP test set (a)	1574	1255	350	22.2	25.4	−2.5
100 images, FN test set (b)	3175	2890	494	15.6	9.9	−7.2
200 images (a) + (b)	4749	4145	844	17.8	14.6	−5.8
864 images from the test cage	1979	1660	366	18.5	19.2	−2.8

Fig. 8. Examples of images processed by the automated counting system (models). Blue boxes were placed on objects which were identified as fish by the models. Some of them were false positives (FPs). Small letters on top of the bounding boxes are confidence scores. Any predicted bounding boxes with confidence scores below 30% were discarded. Large numbers of (a) false negatives (FNs, fish which were not in the blue bounding boxes) and (b) false positives (biofouling organisms in the blue bounding boxes) by Model 1. (c) Model 2 still had large registration of FPs; (d) Model 3 had fewer FPs. (e) Model 1 had some FNs, but these were fewer in (f) Model 2. (g) Model 2 and (h) Model 3 detected objects that were not counted as fish by manual work, which resulted in increased FPs (some of the blue bounding boxes on the small blurred objects)
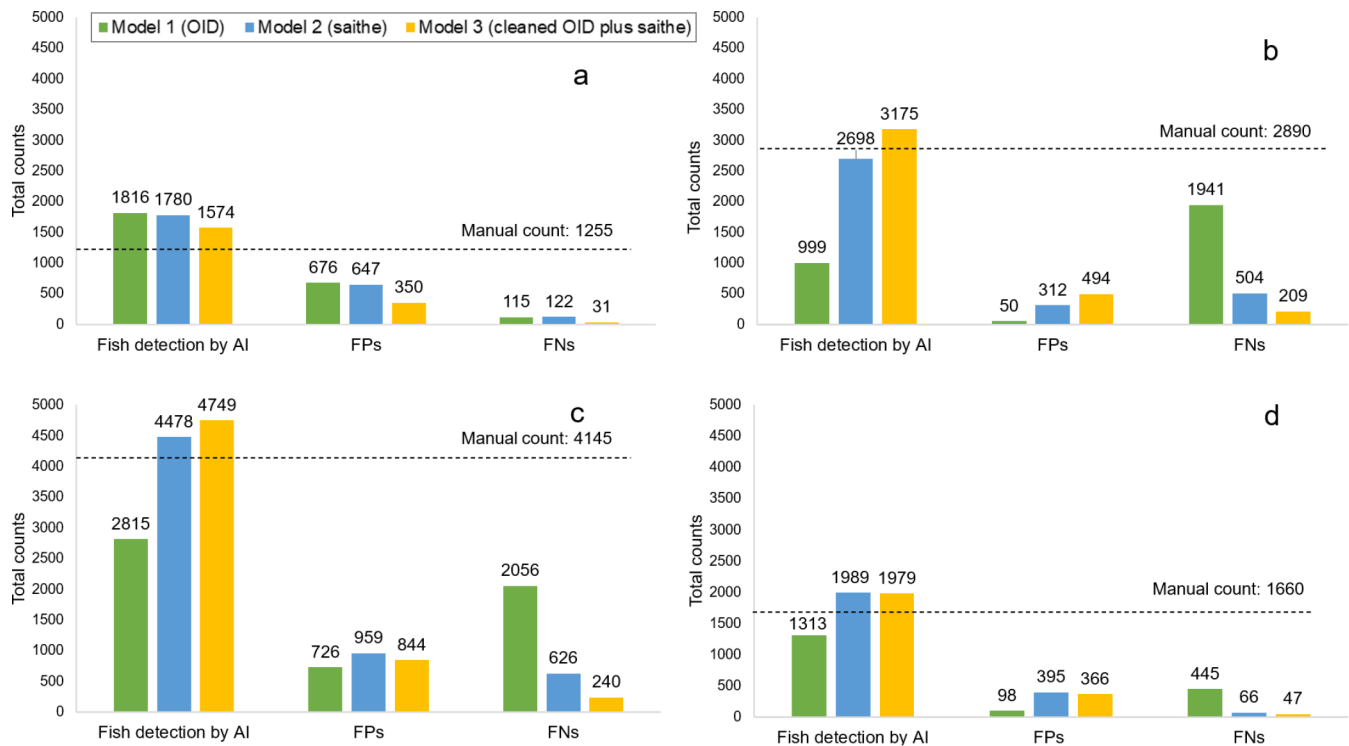
Fig. 9. Comparison of false positives (FPs) and false negatives (FNs) when different models were used in (a) the FP test set (100 images that had the highest number of FPs when Model 1 was used), (b) the FN test set (100 images that had the highest number of FNs when Model 1 was used), (c) the combined test set (combination of FP and FN test sets), and (d) the 864 images taken around the test cage (Cage 1 on 8 July). Dotted line: total manual fish counts by manual work using the same images

threshold was approximately 30 min, which accounted for about 5% of the time spent counting wild fish manually in the same set of images.

4. DISCUSSION

4.1. Filming methods and image quality

Varying light intensity underwater is a common challenge in autonomous fish detection (e.g. Mandal et al. 2018, Marini et al. 2018). Light attenuation in deeper water and higher turbidity makes it difficult to capture objects clearly. Both light sensitivity and dynamic range can vary strongly among different cameras. Using cameras equipped with larger and more light-sensitive sensors can help improve image quality in low-light conditions. Every camera system has inherent physical limitations, and requirements vary with applications (Ulrich & Bonar 2020). One has to consider sensor and system size, resolution, cost, etc., to find the best compromise for a given application and budget (Struthers et al. 2015). In our system, GoPro cameras had the advantage of relatively low cost, but their limited light sensitivity and

dynamic range started to heavily impact image quality at 30–40 m depths. At these depths, there was rarely enough light for the cameras to accurately capture underwater objects. However, giving slight upward angles to the cameras and post-processing helped to improve the visibility of fish in the images. With these improvements, the automated models were able to discern fish in images from 30 and 40 m depths where light intensity was low, but some fish might have not been detectable even after the manual adjustment of video quality. GoPro cameras captured most of the wild fish at depths down to 40 m during bright daylight hours (09:00–12:00 h) in relatively non-turbid water. To ensure visibility of wild fish regardless of depth, cameras with better performance under low light conditions or artificial light could be used as in some previous studies (e.g. Dempster et al. 2009). However, additional light might attract or scare away fish (e.g. Marchesan et al. 2005), which would be an unwanted effect for wild fish monitoring. Therefore, cameras without additional lights are preferable if the image quality is acceptable. Filming needs to encompass multiple depths and positions around the sea cages in order to determine the 3-dimensional distribution patterns of

the wild fish aggregations, as the abundance and species composition of wild fish are not equal in space (e.g. Dempster et al. 2005, 2009).

4.2. Wild fish quantification in the images

To assess whether the current automated system is accurate enough, we first needed to define the acceptable limit for accuracy. Therefore, we assessed deviations in fish counts when a single person was counting identical images multiple times as well as when different people were counting the same set of images independently. There was some potential uncontrollable variability in manual counts by multiple people; in our study, this error was >20% between 2 people. An explanation for the outliers identified by C3 was that salmon was confused with wild fish in turbid water far away from the cameras, even though all counters were given instruction not to count salmon in the sea cages. Within-person variation based on C1's counts was about 5%. Images with low fish numbers tended to cause larger variation in the manual counts due to the higher impacts of each fish on calculation of the variation. Variability among the 4 people was slightly higher, at approximately 7%. Variability in manually counted fish in our tests was mostly due to the presence of objects that could be classified as either fish or non-fish objects, depending on the applied detection thresholds. Such objects were often present in the images due to motion blur, low light conditions, and a long distance between the cameras and the objects. Based on our tests, the presence of bounding boxes had a significant effect on the numbers of detected wild fish by manual counting. However, we believe that the presence of bounding boxes did not considerably affect the evaluation of the models. The presence of bounding boxes increased the detection of wild fish by an average of 4.8% per image, but the same level of variation was observed even when a single person was assessing the images.

Regarding automated wild fish detection, one limitation was FNs in images with high densities of fish or highly overlapped fish, as also observed by Marini et al. (2018). Model 1 had the highest registration of FNs, which can be explained by the inappropriate annotations of fish in the training data set (e.g. original images from OID included a group of fish annotated as one individual). As a result, the neural network was never exposed to high densities of fish during the training stage. This explanation is further supported by the decreased numbers of FNs ob-

served in the outputs of Models 2 and 3, where every fish in each training image was annotated individually. Another challenge was FPs in images that contained background objects such as cage structures, lice skirts, and biofouling organisms attached to the structures. Previous work on autonomous underwater fish detection has also reported the challenge of FPs due to similarities between fish and backgrounds (e.g. Marini et al. 2018, Salman et al. 2020). The error rate was lowest in the output of Model 3, which was trained with images from both OID with corrected annotations and those from the saithe data set; thus, one explanation for Model 3 having the lowest error was the improved representation of training images to the images that were analyzed. The F_1 score for Model 3 was also the highest among the 3 models, indicating that this model performed best.

4.3. Comparison of results — manual vs. automated

The results from C1 were chosen as representative for manual counts, as they were in the middle range of the 4 people's counts, and C1 counted the entire data set. In this study, the acceptable limit for accuracy was set based on between-person variation, which was about 7%, since we assume that multiple people would contribute to manual-based wild fish quantification work.

In this study, when the FP and FN rates were <7%, the errors generated by the system were equivalent to those made by the human counters; therefore, we considered the model to be as good as manual counts. None of the models fulfilled the criteria of both rates being lower than 7%. However, the FN rate in Model 3 was close to or lower than this limit in all test image sets used to assess the error rate. The FP rate fluctuated depending on the fish detection threshold for manual counts. Considering that only 14% of FP detections by Model 3 were obvious 'wrong detections' and the rest of the detected objects were in ambiguous zones, the FP rate by Model 3 would have been lower than 7% if a more realistic threshold had been selected for manual detection.

4.4. How can we improve the automated detection of wild fish?

In this study, the training of each model was done using the same parameters and settings — only the training images were different. The error rates of the

different models indicated that the training images had a direct influence on the accuracy of the detection system. Training images need to be chosen carefully, as the models can only learn based on the input, and their performance is directly dependent on the training image quality.

A few thousand images were used for training the neural network in this study, which was a relatively small data set compared to some of the studies in which researchers developed automated counting systems for use in aquaculture (reviwed in Li et al. 2020). Increasing the number of training images may improve the performance of the system; however, focusing on image quality and representativeness of the target scenario may be equally or more important. One way of increasing the system's accuracy is to re-train the models with new images (e.g. Casado-García et al. 2020). As an example, with fixed cameras and a continuously operating system, one could take a few hundred images with a defined frequency, check and correct annotations, and incorporate those images into the data set. The neural network can be re-trained with the new images, thus being constantly exposed to the environment it is inspecting and therefore decreasing systematic FPs.

In this study, some FPs occurred due to background objects regardless of which model was used. Such FPs could be reduced not only by improving the performance of the system itself but also by modifying filming methods. At shallower depths, our cameras were located close to lice skirts and occasionally captured only the surface of those structures. This led to high numbers of FPs because the system identified biofouling organisms attached to the lice skirt as fish. Filming only the area of interest and not directing the camera lenses towards cage structures could remove those sources of FPs. Reduction of FPs could also be achieved by manually correcting the output of the automatic detection system. Automatic counts followed by manual correction could bring the FP rate to within acceptable thresholds with little time and effort. Manual correction of FPs could also save time compared to a fully manual analysis of the images (approximately 95% of time was saved in this study). Lastly, as mentioned above, proper detection thresholds must be identified in addition to using suitable images for training the neural network. It is therefore of utmost importance to decide on suitable detection criteria for any study that investigates the number of fish or any objects using an automatic detection system.

4.5. Conclusions and future perspectives

The performance of our automatic wild fish detection system is already satisfactory when it comes to not missing fish; i.e. the FN rate of our model was negligible (below the variation among multiple manual counters in this study). The FP rate, however, could be further improved. We believe this improvement could be achieved to a large extent by reducing any mismatches between detection thresholds for manual and automated counts and by improving the training image data sets and filming methods. In addition, FPs could be eliminated manually with little time and effort after running the automated detection, thus also obtaining quality assurance for the outcome of the automated system. Eliminating FNs manually is more time-consuming and needs to be reduced to a minimum in the automated system, which has already been achieved in this study. In case there is a trade-off between FNs and FPs, for example by thresholding or by choice of available training images, one should always prioritize the reduction of FNs. Model 3 seems suitable for the investigation of wild fish abundance in scenarios similar to the case presented in this study. In different scenarios, such as more turbid waters or with different fish species, training data sets would need to be adjusted.

Through further improvements in the quality of videos and images, it will become more feasible to perform automated fish species identification using computer vision. In this study, only 22% of the manually detected fish were identified to species level, which indicates that better quality of videos/images is needed to achieve automatic species identification. It is thus worthwhile to explore different camera models and filming methods that obtain brighter and sharper images regardless of depths.

Frequent or continuous monitoring of wild fish at more farm sites is needed, given their potential impacts on farmed and wild fish (e.g. Snow et al. 2002, Uglem et al. 2009), the environment (e.g. Ballester-Moltó et al. 2017), and commercial fisheries (e.g. Uglem et al. 2014). We have presented here a novel procedure with great potential for autonomous monitoring of wild fish abundance at aquaculture sites through further development.

LITERATURE CITED

AQ1 Systems (2022) AM100. www.aq1systems.com/uploaded/271/13510020_78am-100broc.pdf

Aquabyte (2022) Automatic lice counting. https://aquabyte.ai/produkt/welfare/ (accessed 18 May 2022)

Arechavala-Lopez P, Sæther BS, Marhuenda-Egea F, Sanchez-Jerez P, Uglem I (2015) Assessing the influence of salmon farming through total lipids, fatty acids, and trace elements in the liver and muscle of wild saithe *Pollachius virens.* Mar Coast Fish 7:59–67

Bacher K, Gordoa A, Sagué O (2012) Spatial and temporal extension of wild fish aggregations at *Sparus aurata* and *Thunnus thynnus* farms in the north-western Mediterranean. Aquacult Environ Interact 2:239–252

Ballester-Moltó M, Sanchez-Jerez P, Aguado-Giménez F (2017) Consumption of particulate wastes derived from cage fish farming by aggregated wild fish. An experimental approach. Mar Environ Res 130:166–173

Bjordal Å, Johnstone ADF (1993) Local movements of saithe (*Pollachius virens* L.) in the vicinity of fish farm cages. ICES Mar Sci Symp 196:143–146

Boyra A, Sanchez-Jerez P, Tuya F, Espino F, Haroun R (2004) Attraction of wild coastal fishes to an Atlantic subtropical cage fish farms, Gran Canaria, Canary Islands. Environ Biol Fishes 70:393–401

Bruno DW, Stone J (1990) The role of saithe, *Pollachius virens* L., as host for the sea lice, *Lepeophtheirus salmonis* Krøyer and *Caligus elongatus* Nordmann. Aquaculture 89:201–207

Casado-García A, del-Canto A, Sanz-Saez A, Pérez-López U and others (2020) LabelStoma: a tool for stomata detection based on the YOLO algorithm. Comput Electron Agric 178:105751

Cermaq (2022) iFarm. https://www.cermaq.no/ifarm (accessed 21 Jan 2022)

Crescitelli AM, Gansel LC, Zhang H (2021) NorFisk: fish image dataset from Norwegian fish farms for species recognition using deep neural networks. Model Identif Control 42:1–16

Damsgård B, Høy E, Uglem I, Hedger RD, Izquierdo-Gomez D, Bjørn PA (2012) Net-biting and escape behaviour in farmed Atlantic cod *Gadus morhua*: effects of feed stimulants and net traits. Aquacult Environ Interact 3:1–9

Dempster T, Sanchez-Jerez P, Bayle-Sempere JT, Giménez-Casalduero F, Valle C (2002) Attraction of wild fish to sea-cage fish farms in the south-western Mediterranean Sea: spatial and short-term temporal variability. Mar Ecol Prog Ser 242:237–252

Dempster T, Sanchez-Jerez P, Bayle-Sempere J, Kingsford M (2004) Extensive aggregations of wild fish at coastal sea-cage fish farms. Hydrobiologia 525:245–248

Dempster T, Fernandez-Jover D, Sanchez-Jerez P, Tuya F, Bayle-Sempere J, Boyra A, Haroun RJ (2005) Vertical variability of wild fish assemblages around sea-cage fish farms: implications for management. Mar Ecol Prog Ser 304:15–29

Dempster T, Uglem I, Sanchez-Jerez P, Fernandez-Jover D, Bayle-Sempere J, Nilsen R, Bjørn PA (2009) Coastal salmon farms attract large and persistent aggregations of wild fish: an ecosystem effect. Mar Ecol Prog Ser 385:1–14

Dempster T, Sanchez-Jerez P, Uglem I, Bjørn PA (2010) Species-specific patterns of aggregation of wild fish around fish farms. Estuar Coast Shelf Sci 86:271–275

Fernandez-Jover D, Sanchez-Jerez P, Bayle-Sempere JT, Valle C, Dempster T (2008) Seasonal patterns and diets of wild fish assemblages associated with Mediterranean coastal fish farms. ICES J Mar Sci 65:1153–1160

Giannoulaki M, Machias A, Somarakis S, Karakassis I (2005) Wild fish spatial structure in response to presence of fish farms. J Mar Biol Assoc UK 85:1271–1277

Gonzalez-Silvera D, Izquierdo-Gomez D, Sanchez-Jerez P, Elbal MT, López-Jiménez JA, Martínez-López FJ (2020) Influence of aquaculture waste on fatty acid profiles and gonad maturation of wild fish aggregations at fish farms. Mar Environ Res 156:104902

Goodbrand L, Abrahams MV, Rose GA (2013) Sea cage aquaculture affects distribution of wild fish at large spatial scales. Can J Fish Aquat Sci 70:1289–1295

Hossain E, Alam SMS, Ali AA, Amin MA (2016) Fish activity tracking and species identification in underwater video. In: Proc 5th Int Conf Informatics, Electronics and Vision (ICIEV), p 62–66

Li D, Miao Z, Peng F, Wang L and others (2020) Automatic counting methods in aquaculture: a review. J World Aquacult Soc 52:269–283

Mandal R, Connolly RM, Schlacher TA, Stantic B (2018) Assessing fish abundance from underwater video using deep neural networks. In: Proc Int Joint Conf Neural Networks (IJCNN), p 1–6.

Marchesan M, Spoto M, Verginella L, Ferrero EA (2005) Behavioural effects of artificial light on fish species of commercial interest. Fish Res 73:171–185

Marini S, Fanelli E, Sbragaglia V, Azzurro E, Del Rio Fernandez J, Aguzzi J (2018) Tracking fish abundance by underwater image recognition. Sci Rep 8:13748

Oakes CT, Pondella DJ (2009) The value of a net-cage as a fish aggregating device in Southern California. J World Aquacult Soc 40:1–21

Otterå H, Skilbrei OT (2014) Possible influence of salmon farming on long-term resident behaviour of wild saithe (*Pollachius virens* L.). ICES J Mar Sci 71:2484–2493

Powers DMW (2007) Evaluation: from precision, recall and *F*-factor to ROC, informedness, markedness & correlation. Technical report SIE-07-001 presented at the 2003 International Cognitive Science Conference and the 2007 Human Communication Science SummerFest. http://david.wardpowers.info/BM/Evaluation_SIETR_2up.pdf

R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rauf HT, Lali MIU, Zahoor S, Shah SZH, Rehman AU, Bukhari SAC (2019) Visual features based automated identification of fish species using deep convolutional neural networks. Comput Electron Agric 167:105075

Salman A, Siddiqui SA, Shafait F, Mian A and others (2020) Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system. ICES J Mar Sci 77:1295–1307

Siddiqui SA, Salman A, Malik MI, Shafait F, Mian A, Shortis MR, Harvey ES (2018) Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. ICES J Mar Sci 75:374–389

Snow M, Raynard R, Bruno DW, van Nieuwstadt AP, Olesen NJ, Løvold T, Wallace C (2002) Investigation into the susceptibility of saithe *Pollachius virens* to infectious salmon anaemia virus (ISAV) and their potential role as a vector for viral transmission. Dis Aquat Org 50:13–18

Stagličić N, Šegvić-Bubić T, Ugarković P, Talijančić I, Žužul I, Tičina V, Grubišić L (2017) Ecological role of bluefin tuna (*Thunnus thynnus*) fish farms for associated wild fish assemblages in the Mediterranean Sea. Mar Environ Res 132:79–93

Struthers DP, Danylchuk AJ, Wilson ADM, Cooke SJ (2015) Action cameras: bringing aquatic and fisheries research into view. Fisheries (Bethesda, Md) 40:504–512

Tuya F, Sanchez-Jerez P, Dempster T, Boyra A, Haroun RJ (2006) Changes in demersal wild fish aggregations beneath a sea-cage fish farm after the cessation of farming. J Fish Biol 69:682–697

Uglem I, Dempster T, Bjørn PA, Sanchez-Jerez P, Økland F (2009) High connectivity of salmon farms revealed by aggregation, residence and repeated movements of wild fish among farms. Mar Ecol Prog Ser 384:251–260

Uglem I, Karlsen Ø, Sanchez-Jerez P, Sæther BS (2014) Impacts of wild fishes attracted to open-cage salmonid farms in Norway. Aquacult Environ Interact 6:91–103

Uglem I, Ulvan EM, Toledo-Guedes K, Hegstad E, Blakstad S, Buserud B, Sæther BS (2017) Does a diet including pellets from salmon cage farms affect the taste of wild saithe (*Pollachius virens* Linnaeus, 1758)? J Appl Ichthyol 33:374–376

Ulrich TL, Bonar SA (2020) Inexpensive, underwater filming of rare fishes in high definition. Fisheries (Bethesda, Md) 45:121–130

Zhang S, Yang X, Wang Y, Zhao Z and others (2020) Automatic fish population counting by machine vision and a hybrid deep neural network model. Animals (Basel) 10: 364