Out of the shadows: automatic fish detection from acoustic cameras

Aquat Ecol
https://doi.org/10.1007/s10452-022-09967-5

Out of the shadows: automatic fish detection from acoustic cameras

R. M. Connolly · K. I. Jinks · A. Shand · M. D. Taylor · T. F. Gaston · A. Becker · E. L. Jinks

Received: 28 November 2021 / Accepted: 4 May 2022

Abstract Efficacious monitoring of fish stocks is critical for efficient management. Multibeam acoustic cameras, that use sound-reflectance to generate moving pictures, provide an important alternative to traditional video-based methods that are inoperable in turbid waters. However, acoustic cameras, like standard video monitoring methods, produce large volumes of imagery from which it is time consuming and costly to extract data manually. Deep learning, a form of machine learning, can be used to automate the processing and analysis of acoustic data. We used convolutional neural networks (CNNs) to detect and count fish in a publicly available dual-frequency identification sonar (DIDSON) dataset. We compared three types of detections, direct acoustic, acoustic shadows, and a combination of direct and shadows. The deep learning model was highly reliable at detecting fish to obtain abundance data using acoustic data. Model accuracy for counts-per-image was improved by the inclusion of shadows (F1 scores, a measure of the model accuracy: direct 0.79, shadow 0.88, combined 0.90). Model accuracy for MaxN per video was high for all three types of detections (F1 scores: direct 0.90, shadow 0.90, combined 0.91). Our results demonstrate that CNNs are a powerful tool for automating underwater acoustic data analysis. Given this promise, we suggest broadening the scope of testing to include a wider range of fish shapes, sizes, and abundances, with a view to automating species (or 'morphospecies') identification and counts.

Keywords Acoustic camera · Deep learning · DIDSON · Estuary · Fish · Sonar

Handling Editor: Sébastien Villeger.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10452-022-09967-5.

R. M. Connolly (✉) · K. I. Jinks · A. Shand · E. L. Jinks
Coastal and Marine Research Centre, School of Environment and Science, Australian Rivers Institute, Griffith University, Gold Coast, QLD 4222, Australia
e-mail: r.connolly@griffith.edu.au

M. D. Taylor · A. Becker
Port Stephens Fisheries Institute, New South Wales Department of Primary Industries, Taylors Beach, NSW 2315, Australia

T. F. Gaston
School of Environment and Life Sciences, University of Newcastle, Ourimbah, NSW 2258, Australia

Introduction

Monitoring of fish stocks across a wide range of environments is a critical task for effective management. Fisheries scientists and managers monitor fish stocks by collecting data on population abundance, biomass and densities (Egerton et al. 2018; Smith et al. 2021), schooling behaviours (Trenkel et al. 2011), predator–prey relationships (Becker and Suthers

Published online: 26 May 2022

2014; Boswell et al. 2019), and movement via key passageways (Bennett et al. 2020). Common direct sampling methods including nets (e.g. seine, gill, fyke or trawl), traps, and line fishing can be invasive and introduce sampling bias (Kuriyama et al. 2019; French et al. 2021). Indirect sampling techniques such as visual census and baited or unbaited remote underwater video (BRUV/RUV) are alternatives to direct, invasive methods, but are ineffective when visibility is poor [e.g. in turbid waters or deep offshore habitats, and at night (Becker et al. 2011; Benoit-Bird and Lawson 2016; Sheaves et al. 2020; Kimball et al. 2021)]. Acoustic camera monitoring (which uses sound reflectance, instead of light reflectance) offers a non-invasive survey method in underwater environments to overcome the obstacle of sampling where standard video imagery or visual census is unfeasible (Horne 2000).

Fisheries scientists have used acoustic cameras to monitor fish (and other animals) by detecting their direct acoustic image and/or their acoustic shadow (Horne 2000; Trenkel et al. 2011; Martignac et al. 2015). For example, acoustic cameras have been used in saltmarsh habitats to analyse predator–prey interactions (Boswell et al. 2019) and fish movement in tidal passageways (Kimball et al. 2010; Bennett et al. 2020); in areas of high turbidity caused by sedimentation to estimate size and abundance of key demersal fish (Artero et al. 2021); and, in intermittently closed estuaries to determine the abundance and the direction of fish movement, and the distribution of different sized fish (Becker et al. 2016, 2017). Coupling of direct acoustic images and acoustic shadows has enabled identification of different species (Able et al. 2014; Artero et al. 2021). Furthermore, different size classes of fish have been determined with high accuracy through the direct analysis of acoustic shadows (Langkau et al. 2012). While sampling fish using direct acoustic images or shadows is helpful when visibility is poor, as for normal video imaging, acoustic sampling produces vast amounts of data that require laborious and costly processing and analysis.

Automation techniques to overcome the challenges and costs of manually processing video footage are revolutionising monitoring in aquatic environments. For instance, a type of machine learning called deep learning (DL) uses convolutional neural networks (CNN) to analyse standard video footage to detect and classify fish (Mandal et al. 2018; Villon et al. 2018; Salman et al. 2020). Automatic detection and classification of fish increases the efficiency of monitoring the abundance of fish populations (Marini et al. 2018; Ditria et al. 2020a, 2020b), tracking movement of fish (Lopez-Marcano et al. 2021), measuring fish sizes (Álvarez-Ellacuría et al. 2020; Coro and Walsh 2021), and monitoring behaviour patterns (Saberioon et al. 2017; Ditria et al. 2021). Similar approaches to automate the processing and analysis of acoustic data have been used to detect fish aggregations (Shahrestani et al. 2017; Vatnehol et al. 2018; Tarling et al. 2021), track the speed and direction of fish in trawls (Handegard and Williams 2008), track the direction, abundance and size of salmonids (Kulits et al. 2020), detect the presence/absence of tuna (Uranga et al. 2017), and to identify and track marine mammals such as seals (Hastie et al. 2019). Although useful, these studies rely solely on the direct acoustic detection of the species of interest. Deep learning algorithms that simultaneously evaluate the direct and shadow detections might improve the accuracy of automation and, ultimately, provide a valuable tool for automated analysis of acoustic data.

Our goal was to use CNNs to train and detect fish from direct acoustic images, acoustic shadows, and the combination of direct images and shadows. This is a step towards being able to use automatic detection of direct and shadow detections from acoustic data for continuous monitoring of a wide range of metrics. We predicted that the model would be enhanced by training on both the direct and shadow detections rather than the direct detections alone.

Materials and methods

To achieve our aim of using acoustic camera data to automatically detect fish using deep learning (DL) models, we firstly acquired a dataset of labelled fish species sampled using DIDSON (dual-frequency identification sonar). DIDSON is a multi-beam high-frequency (1.1 or 1.8 MHz) sonar device that transmits acoustic pulses through the water to detect objects. Acoustic sounds are reflected when the sound wave meets an object of a different density to the medium in which the sound wave is propagated. DIDSON displays video-like images of the reflected acoustic echoes on an echogram, using colours and

colour intensity to represent the objects and strength of the signal (Martignac et al. 2015).

We annotated and trained a model to detect both the fish and shadow of the fish. To obtain abundance data the model detected fish and shadows separately in each image. We then used an automated post-processing step selecting the maximum count of *either* fish *or* shadow detections (not both) in each image, generating the "combined" count. We then analysed the accuracy of the model using common performance metrics (Fig. 1). We focussed on the most widely used measure of abundance in videos, MaxN, the maximum number of fish visible in a video in any one frame (Ellis and DeMartini 1995; Whitmarsh et al. 2016; Langlois et al. 2020).

Dataset

The data used for the DL model were sourced from a professionally labelled acoustic dataset that is publicly available under a creative commons licence that permits use with acknowledgements (McCann et al. 2018). This dataset contains acoustic observation data sampled using DIDSON in the Ocqueoc River

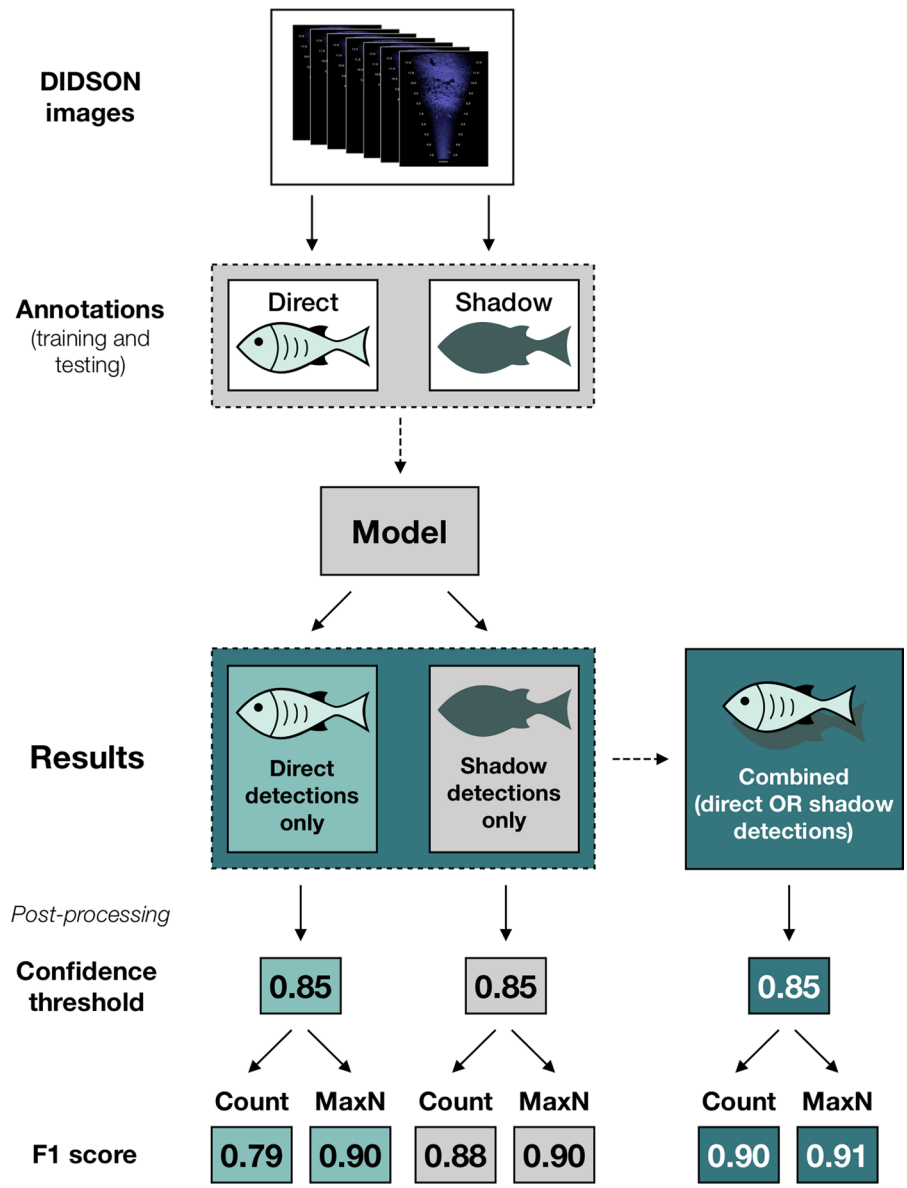

Fig. 1 Flow diagram of deep learning models trained on acoustic imagery to detect fish and fish shadows. During post-processing, count and MaxN was calculated for direct and shadow detections only and then merged to calculate the combined count and MaxN of direct or shadow detection

in northern Michigan, USA, between 2013 and 2016. Each video is sorted and labelled with the fish species known to be present. All videos used a window length of 2.5–12.5 m. To obtain enough training data, videos containing the two most common species, walleye (*Sander vitreus*) and common carp (*Cyprinus carpio*), were selected for training and testing of the DL model.

Although two species were selected from the dataset for training and testing, all data were pooled and labelled as 'fish' so the model could be trained and tested against different backgrounds. The two species grow to a similar size, ranging up to 80 cm in length. The carp to walleye ratio in this pooled dataset was approximately 2:1 both for direct and shadow. Because the different species did not occur in the same frame, we felt there would be bias if we trained and tested on 'species' because each species had a different background the model may have learned to associate with species identification. Therefore, the goal was to test automatic identification of 'fish', rather than to identify the different species. From the total pool of walleye and carp videos, 157 segments of video from 21 days of DIDSON deployments (stratified by day) were allocated to one of two datasets: training (115 segments) and testing (42 segments). Raw video resolution varied between 1002×564 (47 segments) and 1920×1080 (110 segments), but in any case all videos were standardised to a scale of 1002×564 prior to processing. Each segment was unique to either training or testing, and segments from each day were randomly allocated to each of the datasets.

Manual annotation of imagery provided the 'ground-truth' fish counts both to train the model and to evaluate performance. Images were extracted at 5 frames per second. Using bounding boxes around the fish and shadows, both the direct and shadow detections of the fish were manually annotated in each of the extracted images (Table 1). To assist with identification, the annotator was able to play videos back and forth to increase confidence that the object in the video was moving and could be correctly identified as either 'fish' or 'shadow' (see animation in Online Resource 1 for an example of how movement was used to annotate our dataset).

Table 1 Numbers of annotations of acoustic imagery for the direct detections of fish and shadows of fish used in deep learning models

Annotations	Direct 	Shadow 
Training	7566	7764
Testing	1469	2189

Object detection model and performance metrics

We used a convolutional neural network (CNN) for object detection. Specifically, our model was trained using Faster-RCNN with a ResNet50 configuration, pre-trained using the ImageNet1k dataset (Massa and Girshick 2018). Model training, testing, and prediction tasks were conducted on a Microsoft Azure Data Science Virtual Machine powered by an NVIDIA V100 GPU. Overfitting was mitigated by using the early-stopping technique (Prechelt 2012).

We tested the performance of our model using two key metrics of fish abundance: count-per-image and MaxN per video (for direct, shadow and combined detections). Count-per-image was calculated over a total of 1287 images, and was used to assess performance on an image-by-image basis. MaxN was calculated for 42 video segments, and used to assess performance in an application context of providing a typical metric of abundance. For each of these metrics two performance criteria, precision (P) and recall (R), were determined for confidence thresholds between 5 and 95% in 5% increments. The confidence threshold is the level of prediction certainty required to state a detection. Precision measures the fraction of fish detections that were correct, and recall measures the fraction of fish actually present that were detected.

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

Overall performance for count-per-image and MaxN was determined by the $F1$ score, which represents the balance between precision and recall. $F1$ is calculated as follows:

$$F1 = 2 \times \frac{P \times R}{P + R}$$

We checked for any systematic biases in false detections by examining the size and distance from camera of objects, and comparing these for false positives and false negatives against true positives. Distances from camera were extracted directly via DIDSON software, and sizes were calculated as the area of bounding boxes as a percentage of the total image size, from predicted detections for true positives and false positives, and from manually annotated boxes for false negatives. For both distances and sizes, the frequency distributions of the three categories were compared using Kolmogorov–Smirnov two-sample tests on non-standardised frequency data, pairwise among the three detection types.

Results

The model was successful in automatically counting fish in acoustic imagery using either the direct detection, shadows, or a combination of both (Fig. 1). At a confidence threshold of 85%, shadows improved the direct $F1$-score from 0.79 to 0.90 for counts, and from 0.90 to 0.91 for MaxN. Performance of the model increased because shadow detections sometimes occurred when a direct detection was missed (see the example in Fig. 2).

For the count-per-image results, at both a lower (70%) and higher (85%) confidence threshold, our model performed best for the shadow detections alone and combined detections (Table 2; Fig. 3a). F1 scores were lowest for direct detections alone (Table 2).

For the MaxN per video results, at a lower (70%) confidence threshold, the model performed slightly better for shadow detections alone and combined detections than for direct detections alone (Table 3; Fig. 3b). However, at a higher confidence threshold (85%), the model performed nearly as well for all three methods of detecting fish, with the combined detections only marginally higher than the direct or shadow detections alone (Table 3; Fig. 3b).

In comparing the distances from camera and sizes of false detections against true positives, we found no pattern for distance from camera, but object detection size varied significantly among these detection types (Kolmogorov–Smirnov tests, all p values < 0.01; Fig. 4, noting that for visual interpretation, frequencies are displayed standardised by total counts, whereas KS tests were on non-standardised data). Both types of false detections had a higher proportion of very small detections than for true positives, with marginally smaller images for false negatives than false positives. Most false detections were around 10% or less of the total image area. These are small images as observable on screen, and do not necessarily reflect fish sizes, which vary with distance from cameras.

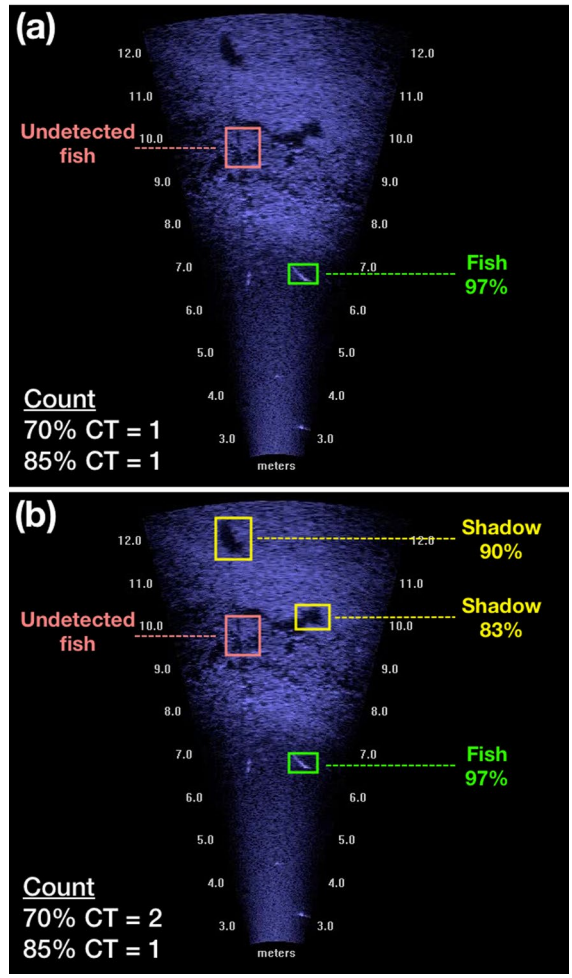

Fig. 2 Example DIDSON image with ground-truthed fish count = 2. Detections shown in green (direct) and yellow (shadow) with probabilities, and counts given for 70% and 85% confidence thresholds (CT). Panel **a** direct detection only, which underestimates fish count, and Panel **b** direct and shadow detections combined, which correctly estimates fish count at 70% CT, and underestimates by one at 85% CT

Table 2 Count per image results of a deep learning model trained on acoustic imagery of direct and shadow detections of fish, at confidence thresholds of 70% and 85%

Count was determined for direct and shadow detections separately, and results were then merged to give the combined count. True positives as proportion of Ground Truths are same as Recall and are not shown again

Detections	Direct		Shadow		Combined	
Confidence threshold	0.70	0.85	0.70	0.85	0.70	0.85
F1 score	0.81	0.79	0.91	0.88	0.90	0.90
Precision	0.81	0.92	0.89	0.96	0.88	0.95
Recall	0.81	0.70	0.93	0.82	0.93	0.85
Ground-truths (GT)	1469	1469	2189	2189	2273	2273
True positives (TP)	1192	1027	2027	1788	2121	1934
False negatives (FN)	277	442	162	401	152	339
FN proportion of GT	0.19	0.30	0.07	0.18	0.07	0.15
True negatives (TN)	174	221	6	10	0	0
False positives (FP)	280	90	238	80	297	104

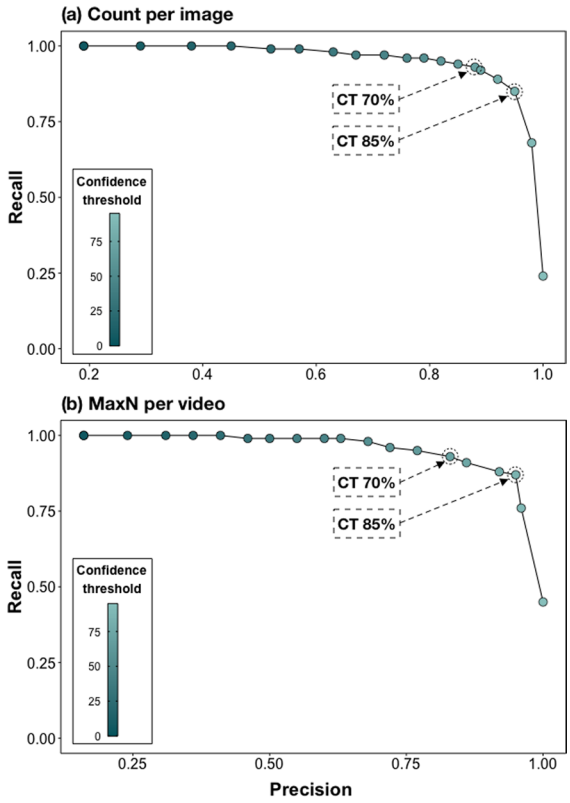

Fig. 3 Precision and recall scores for the combined detection of fish for **a** Count per image, and **b** MaxN per video. Confidence intervals are in 5% increments, and confidence thresholds of 70% and 85% are indicated for comparison of performance (these are the two CTs reported in Tables 2 and 3)

Discussion

We have presented a successful method for automatically detecting fish from acoustic imagery. The CNN reliably detected fish using either direct or shadow detections, or in combination, achieving high *F*1 scores for all three methods of detection. This automated method has the potential to reduce the time and cost of manually counting fish using acoustic data, and particularly so when MaxN is the desired measure of fish abundance. The level of accuracy achieved is equal to or above that reported previously for CNNs on sonar imagery. Using a CNN model to detect eels swimming through a weir in Canada, Zang et al. (2021) reported high accuracy (0.89), although on a relatively small number of videos. These authors had previously achieved higher accuracy (0.99) using a similar model in a controlled laboratory environment (Zang et al. 2021), but found that the model performed poorly on field data (0.5). Automated detection of salmon in sonar imagery using a CNN in conjunction with optical flow to detect pixel changes between sequential frames yielded accuracy of 0.8 (Kulits et al. 2020). The presence of large schools of mullet swimming along the coast have also been detected with accuracy of 0.89 (Tarling et al. 2021). All of these methods used only direct detection, not shadows. The current paper adds substantially to the view that CNNs will be very useful for automatically detecting fish versus no-fish in sonar imagery. Reliable detection of fish using DL techniques such as CNNs is clearly possible, and as the field develops,

Table 3 MaxN per video results of a deep learning model trained on acoustic imagery of direct and shadow detections of fish, at confidence thresholds of 70% and 85%

MaxN was determined for direct and shadow detections separately, and results were then merged to provide the combined MaxN. True positives as proportion of Ground Truths are same as Recall and are not shown again

Detections	Direct		Shadow		Combined	
Confidence threshold	0.70	0.85	0.70	0.85	0.70	0.85
F1 score	0.83	0.90	0.89	0.90	0.88	0.91
Precision	0.76	0.96	0.88	0.94	0.83	0.95
Recall	0.91	0.85	0.91	0.87	0.93	0.87
Ground-truths (GT)	80	80	106	106	108	108
True positives (TP)	73	68	96	92	100	94
False negatives (FN)	7	12	10	14	8	14
FN proportion of GT	0.09	0.15	0.09	0.13	0.07	0.13
True negatives (TN)	0	0	0	0	0	0
False positives (FP)	23	3	13	6	20	5

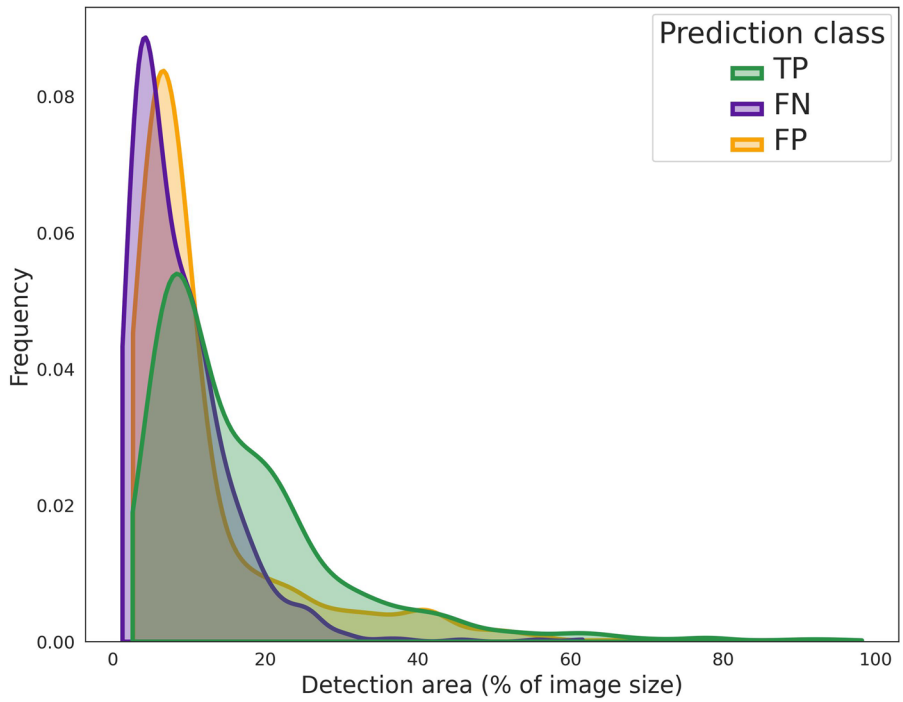

Fig. 4 Frequency distribution of detection area for different detection types. Frequencies for true positive (TP), false positive (FP) and false negative (FN) detections are shown standardised as a proportion of the total number of detections for that detection type. Standardisation simplifies visual analysis since the total counts for true positives were much greater than either false category. Detection area is reported as area of detection box as a percentage of total image area, a proxy for the size of the fish or shadow as actually observed in the frame

we encourage others to consider the inclusion of shadow detections.

Our results indicate that shadows can be a useful addition to include in model training and predicting when using CNNs, and probably for any other automation technique where shadows are present in the acoustic data being analysed. Previous studies using semi-autonomous fish counting methods have suggested that acoustic shadows are an impediment that reduced the accuracy of software solutions (Eggleston et al. 2020; Perivolioti et al. 2021). We have shown convincingly that if shadow information is included in training of detection algorithms, shadows detection can improve performance. We suspect that previous reports of difficulties with shadows adversely affecting fish counts might have resulted from a lack of shadow input in model training, or perhaps from shadows being unusable. Although we have demonstrated

the usefulness of shadows in the imagery analysed, further investigation will be required to test how generalisable this finding is to DIDSON imagery more broadly. Shadow formation is affected by factors such as the angle of the acoustic camera to the substrate, substrate complexity, and fish orientation and behaviour. We largely used imagery in which fish were migrating up and down stream, swimming perpendicularly to cameras, throwing relatively large and easily detected shadows. When fish milled around, perhaps for foraging, orientation and direction changed frequently and shadows were often small or thin as a smaller body profile was exposed to the sonar, with poorer detectability. The usefulness of shadows for identification of different species (or morphospecies) in manual analysis of DIDSON imagery has been pointed out by Langkau et al. (2012), although they suggest that accuracy is poor for smaller sized individuals. Further experimentation into the usefulness of shadows is warranted, to distinguish the roles of camera position and orientation relative to fish, and substrate type. Where the morphology of background substrate is known, the relationships between camera position and the distance between fish and shadow detection can potentially provide a metric of fish position within the water column.

Automatic detection of species (or morphospecies) using multi-class models will be an important future step in improving the value of automating acoustic monitoring. At this stage, however, both manual and automatic species identification has proven problematic due to the nature of acoustic data (Martignac et al. 2015). High accuracy of manual species identification can occur when species have distinct morphological features (Martignac et al. 2015; Jones et al. 2021), and automation should also be successful for species that show clear morphological differences. Automatic species identification has been partially successful for eels (Zang et al. 2021), but attempts for other types of fish have had limited success (Rogers et al. 2004; Jones et al. 2021). Automatic species identification could be improved by analysing behavioural characteristics, such as tailbeat frequencies, which have been used successfully in manual analysis of acoustic data for species identification (Kang 2011; Helminen et al. 2021). Other behaviours such as swimming speed and feeding activities could also be investigated to improve automation and combined with length data where only particular species are known to obtain sizes above certain limits. Sequential non-maximum suppression (SeqNMS), an object tracking method where the model examines neighbouring images in a video to improve the accuracy of detection, has been used to automatically detect the direction and speed of fish in underwater videos (Lopez-Marcano et al. 2021). SeqNMS could also prove useful for acoustic data for species identification. The unique grazing behaviour exhibited by fish in seagrass has also been automatically tracked (Ditria et al. 2021), and with some refinement to the model, this may also be a useful method to detect feeding behaviours to differentiate among species in acoustic data, so long as seagrass does not adversely affect quality of shadows or the acoustic imagery overall.

Apart from DL methods, automating and semi-automating the analysis of acoustic data has been performed using classic machine learning techniques. Commercially available software called Echoview (https://echoview.com) allows users to semi-automate the acoustic data analysis through training of predefined algorithms (Boswell et al. 2008). Some applications that demonstrated a reduction in analysis time or successful semi-automation of the process using Echoview include the tracking of migrating fish (Kang 2011; Helminen and Linnansaari 2021), counting fish (Kang 2011; Eggleston et al. 2020), and monitoring behaviour such as tailbeat frequencies (Mueller et al. 2010; Kang 2011; Helminen et al. 2021). Other studies have demonstrated the usefulness of more 'traditional' ML techniques using statistics and/or advanced algorithms for classifying, counting, and sizing of fish using acoustic data (Han et al. 2009; Bothmann et al. 2016; Jing et al. 2017; Lawson et al. 2019). We suggest that because all of these methods require specialised statistical skills for each new application, a DL model, once evaluated and performing reliably, will be an easier method for scientists to apply. Even the problematic detections of very small objects in the current study might in future be overcome as resolution of imagery from acoustic cameras continues to improve. The DIDSON is already being superseded, for example, by imaging sonar that can operate at 3 MHz, improving resolution at shorter ranges. In terms of the amount of effort required for training, in the order of several thousand annotations typically will be required to achieve suitable model performance (Sheaves et al. 2020). While the accuracy required for applied automation varies

with study objectives, generally $F1$ scores above 0.8 are considered useful, and above 0.9 very good.

Our model performed well using direct, shadow, and the combination of these detections; however, we acknowledge that our study used a limited database with low densities of fish present in each image (typically 3 or fewer). Even for manual counting of acoustic imagery, higher densities can render counts unreliable (Horne 2000), and dense schooling behaviour makes automated tracking of individual fish difficult (Handegard and Williams 2008; Lopez-Marcano et al. 2021). Despite the challenges, high densities of fish are common in any form of video imagery and higher densities of fish in acoustic images should be included in training and testing to improve the usability of the model. Post-processing steps such as varying confidence thresholds, the use of SeqNMS, and statistical adjustment equations could assist in overcoming the issue of individuals obscuring other individuals in acoustic imagery, as has been demonstrated in underwater videos (Connolly et al. 2021).

We have shown that using a DL technique such as CNN can automatically detect fish in acoustic data and has the potential to substantially improve the efficiency of acoustic data analysis. For the short videos analysed here, with relatively low fish abundances, manual extraction of MaxN data took a fish expert on average 1.8 min per min of video (SE 0.16). Computer estimates of MaxN took about half the time in the current study, at 0.95 min per min of video (no variation, so no SE). Much faster computer speeds are possible, however, using parallelisation; e.g. using two servers in parallel doubles the processing speed. We have also highlighted the usefulness of acoustic shadow detections in DIDSON data to improve model accuracy for counting of fish. This method is suitable for fisheries-independent monitoring of exploited species to inform fisheries stock assessments, and quantifying use of fish passageways when fish densities are low. The approach needs further investigation at higher fish densities and for species identification.

Acknowledgements RMC, KIJ, AS, ELJ acknowledge funding support from the Australian Research Data Commons, and from the Global Wetlands Project (GLOW), supported by a charitable organization which neither seeks nor permits publicity for its efforts.

Author contributions All authors conceptualised the study and interpreted results. ELJ led model development, AS and KIJ annotated imagery, MDT, TFG and AB provided access to and knowledge of DIDSON software. Manuscript writing was led by RMC and KIJ with contributions from all authors.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data availability Sonar imagery used in this paper is available via a publicly available dataset (McCann et al. 2018) from DIDSON surveys in the Ocqueoc River, Michigan, USA. Annotations and code will be released on the author Github site at the time of publication. Analysis was solely of existing acoustic data via a publicly available dataset

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Able K, Grothues T, Rackovan J, Buderman F (2014) Application of mobile dual-frequency identification sonar (DIDSON) to fish in estuarine habitats. Northeast Nat 21:192–209. https://doi.org/10.1656/045.021.0207

Álvarez-Ellacuría A, Palmer M, Catalán IA, Lisani J-L, Beyan C (2020) Image-based, unsupervised estimation of fish size from commercial landings using deep learning. ICES J Mar Sci 77:1330–1339. https://doi.org/10.1093/icesjms/fsz216

Artero C, Marchetti S, Bauer E, Viala C, Noël C, Koenig C, Berzins R, Lampert L (2021) High-resolution acoustic cameras provide direct and efficient assessments of large demersal fish populations in extremely turbid waters. Appl Sci 11:1899. https://doi.org/10.3390/app11041899

Becker A, Suthers IM (2014) Predator driven diel variation in abundance and behaviour of fish in deep and shallow habitats of an estuary. Estuar Coast Shelf Sci 144:82–88. https://doi.org/10.1016/j.ecss.2014.04.012

Becker A, Cowley PD, Whitfield AK, Järnegren J, Næsje TF (2011) Diel fish movements in the littoral zone of a temporarily closed South African estuary. J Exp Mar Biol

Ecol 406:63–70. https://doi.org/10.1016/j.jembe.2011.06.014

Becker A, Whitfield AK, Cowley PD, Cole VJ, Taylor MD (2016) Tidal amplitude and fish abundance in the mouth region of a small estuary. J Fish Biol 89:1851–1856. https://doi.org/10.1111/jfb.13056

Becker A, Whitfield AK, Cowley PD, Cole VJ (2017) Does water depth influence size composition of estuary-associated fish? Distributions revealed using mobile acoustic-camera transects along the channel of a small shallow estuary. Mar Fresh Res 68:2163–2169. https://doi.org/10.1071/mf16230

Bennett MA, Becker A, Gaston T, Taylor MD (2020) Connectivity of large-bodied fish with a recovering estuarine tidal marsh, revealed using an imaging sonar. Estuar Coast 44:1579–1587. https://doi.org/10.1007/s12237-020-00822-0

Benoit-Bird KJ, Lawson GL (2016) Ecological insights from pelagic habitats acquired using active acoustic techniques. Ann Rev Mar Sci 8:463–490. https://doi.org/10.1146/annurev-marine-122414-034001

Boswell KM, Wilson MP, Cowan JH (2008) A semiautomated approach to estimating fish size, abundance, and behavior from dual-frequency identification sonar (DIDSON) data. N Am J Fish Manag 28:799–807. https://doi.org/10.1577/m07-116.1

Boswell KM, Kimball ME, Rieucau G, Martin JGA, Jacques DA, Correa D, Allen DM (2019) Tidal stage mediates periodic asynchrony between predator and prey nekton in salt marsh creeks. Estuar Coast 42:1342–1352. https://doi.org/10.1007/s12237-019-00553-x

Bothmann L, Windmann M, Kauermann G (2016) Realtime classification of fish in underwater sonar videos. J R Stat Soc Ser C Appl Stat 65:565–584

Connolly RM, Fairclough DV, Jinks EL, Ditria EM, Jackson G, Lopez-Marcano S, Olds AD, Jinks KI (2021) Improved accuracy for automated counting of a fish in baited underwater videos for stock assessment. Front Mar Sci 8:658135. https://doi.org/10.3389/fmars.2021.658135

Coro G, Walsh MB (2021) An intelligent and cost-effective remote underwater video device for fish size monitoring. Ecol Inform 63:101311. https://doi.org/10.1016/j.ecoinf.2021.101311

Ditria EM, Lopez-Marcano S, Sievers M, Jinks EL, Brown CJ, Connolly RM (2020a) Automating the analysis of fish abundance using object detection: optimizing animal ecology with deep learning. Front Mar Sci 7:429. https://doi.org/10.3389/fmars.2020.00429

Ditria EM, Sievers M, Lopez-Marcano S, Jinks EL, Connolly RM (2020b) Deep learning for automated analysis of fish abundance: the benefits of training across multiple habitats. Environ Monit Assess 192:698. https://doi.org/10.1007/s10661-020-08653-z

Ditria EM, Jinks EL, Connolly RM (2021) Automating the analysis of fish grazing behaviour from videos using image classification and optical flow. Anim Behav 177:31–37. https://doi.org/10.1016/j.anbehav.2021.04.018

Egerton JP, Johnson AF, Turner J, LeVay L, Mascarenas-Osorio I, Aburto-Oropeza O (2018) Hydroacoustics as a tool to examine the effects of marine protected areas and habitat type on marine fish communities. Sci Rep 8:47. https://doi.org/10.1038/s41598-017-18353-3

Eggleston MR, Milne SW, Ramsay M, Kowalski KP (2020) Improved fish counting method accurately quantifies high-density fish movement in dual-frequency identification sonar data files from a coastal wetland environment. N Am J Fish Man 40:883–892. https://doi.org/10.1002/nafm.10451

Ellis D, DeMartini E (1995) Evaluation of a video camera technique for indexing abundances of juvenile pink snapper, *Pristipomoides filamentosus*, and other Hawaiian insular shelf fishes. Fish Bull 93:67–77

French B, Wilson S, Holmes T, Kendrick A, Rule M, Ryan N (2021) Comparing five methods for quantifying abundance and diversity of fish assemblages in seagrass habitat. Ecol Indic 124:107415. https://doi.org/10.1016/j.ecolind.2021.107415

Han J, Honda N, Asada A, Shibata K (2009) Automated acoustic method for counting and sizing farmed fish during transfer using DIDSON. Fish Sci 75:1359–1367. https://doi.org/10.1007/s12562-009-0162-5

Handegard NO, Williams K (2008) Automated tracking of fish in trawls using the DIDSON (Dual frequency IDentification SONar). ICES J Mar Sci 65:636–644. https://doi.org/10.1093/icesjms/fsn029

Hastie GD, Wu GM, Moss S, Jepp P, MacAulay J, Lee A, Sparling CE, Evers C, Gillespie D (2019) Automated detection and tracking of marine mammals: a novel sonar tool for monitoring effects of marine industry. Aquat Conserv Mar Freshw Ecosyst 29:119–130. https://doi.org/10.1002/aqc.3103

Helminen J, Linnansaari T (2021) Object and behavior differentiation for improved automated counts of migrating river fish using imaging sonar data. Fish Res 237:105883. https://doi.org/10.1016/j.fishres.2021.105883

Helminen J, O'Sullivan AM, Linnansaari T (2021) Measuring tailbeat frequencies of three fish species from adaptive resolution imaging sonar data. Trans Am Fish Soc 150:627–636. https://doi.org/10.1002/tafs.10318

Horne J (2000) Acoustic approaches to remote species identification: a review. Fish Oceanogr 9:356–371. https://doi.org/10.1046/j.1365-2419.2000.00143.x

Jing D, Han J, Wang X, Wang G, Tong J, Shen W, Zhang J (2017) A method to estimate the abundance of fish based on dual-frequency identification sonar (DIDSON) imaging. Fish Sci 83:685–697. https://doi.org/10.1007/s12562-017-1111-3

Jones RE, Griffin RA, Unsworth RKF (2021) Adaptive resolution imaging sonar (ARIS) as a tool for marine fish identification. Fish Res 243:106092. https://doi.org/10.1016/j.fishres.2021.106092

Kang M-H (2011) Semiautomated analysis of data from an imaging sonar for fish counting, sizing, and tracking in a post-processing application. Fish Aquat Sci 14:218–225. https://doi.org/10.5657/fas.2011.0218

Kimball ME, Rozas LP, Boswell KM, Cowan JH (2010) Evaluating the effect of slot size and environmental variables on the passage of estuarine nekton through a water control structure. J Exp Mar Biol Ecol 395:181–190. https://doi.org/10.1016/j.jembe.2010.09.003

Kimball ME, Connolly RM, Alford SB, Colombano DD, James WR, Kenworthy MD, Norris GS, Ollerhead J, Ramsden S, Rehage JS, Sparks EL, Waltham NJ, Worthington TA, Taylor MD (2021) Novel applications of technology for advancing tidal marsh ecology. Estuar Coast 44:1568–1578. https://doi.org/10.1007/s12237-021-00939-w

Kulits P, Pan A, Beery SM, Young E, Perona P, Van Horn G (2020) Automated salmonid counting in sonar data. In: NeurIPS 2020 Workshop on tackling climate change with machine learning

Kuriyama PT, Branch TA, Hicks AC, Harms JH, Hamel OS (2019) Investigating three sources of bias in hook-and-line surveys: survey design, gear saturation, and multispecies interactions. Can J Fish Aquat Sci 76:192–207. https://doi.org/10.1139/cjfas-2017-0286

Langkau MC, Balk H, Schmidt MB, Borcherding J (2012) Can acoustic shadows identify fish species? A novel application of imaging sonar data. Fish Manag Ecol 19:313–322. https://doi.org/10.1111/j.1365-2400.2011.00843.x

Langlois T, Goetze J, Bond T, Monk J, Abesamis RA, Asher J, Barrett N, Bernard ATF, Bouchet PJ, Birt MJ, Cappo M, Currey-Randall LM, Driessen D, Fairclough DV, Fullwood LAF, Gibbons BA, Harasti D, Heupel MR, Hicks J, Holmes TH, Huveneers C, Ierodiaconou D, Jordan A, Knott NA, Lindfield S, Malcolm HA, McLean D, Meekan M, Miller D, Mitchell PJ, Newman SJ, Radford B, Rolim FA, Saunders BJ, Stowar M, Smith ANH, Travers MJ, Wakefield CB, Whitmarsh SK, Williams J, Harvey ES, Codling E (2020) A field and video annotation guide for baited remote underwater stereo-video surveys of demersal fish assemblages. Methds Ecol Evol 11:1401–1409. https://doi.org/10.1111/2041-210x.13470

Lawson KM, Ridgway JL, Mueller AT, Faulkner JDA, Calfee RD (2019) Semiautomated process for enumeration of fishes from recreational-grade side-scan sonar imagery. N Am J Fish Man 40:75–83. https://doi.org/10.1002/nafm.10373

Lopez-Marcano S, Jinks EL, Buelow CA, Brown CJ, Wang D, Kusy B, Ditria EM, Connolly RM (2021) Automatic detection of fish and tracking of movement for ecology. Ecol Evol 11:8254–8263. https://doi.org/10.1002/ece3.7656

Mandal R, Connolly RM, Schlacher TA, Stantic B (2018) Assessing fish abundance from underwater video using deep neural networks. In: 2018 International joint conference on neural networks (IJCNN). IEEE, pp 1–6

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. https://doi.org/10.1038/s41598-018-32089-8

Martignac F, Daroux A, Bagliniere J-L, Ombredane D, Guillard J (2015) The use of acoustic cameras in shallow waters: new hydroacoustic tools for monitoring migratory fish population. A review of DIDSON technology. Fish Fish 16:486–510. https://doi.org/10.1111/faf.12071

Massa F, Girshick R (2018) Maskrcnn-benchmark: fast, modular reference implementation of instance segmentation and object detection algorithms in PyTorch. https://github.com/facebookresearch/maskrcnn-benchmark

McCann E, Li L, Pangle K, Johnson N, Eickholt J (2018) An underwater observation dataset for fish classification and fishery assessment. Sci Data 5:180190. https://doi.org/10.1038/sdata.2018.190

Mueller A-M, Burwen DL, Boswell KM, Mulligan T (2010) Tail-beat patterns in dual-frequency identification sonar echograms and their potential use for species identification and bioenergetics studies. Trans Am Fish Soc 139:900–910. https://doi.org/10.1577/t09-089.1

Perivolioti T-M, Tušer M, Terzopoulos D, Sgardelis SP, Antoniou I (2021) Optimising the workflow for fish detection in DIDSON (Dual-Frequency IDentification SONar) data with the use of optical flow and a genetic algorithm. Water 13:1304. https://doi.org/10.3390/w13091304

Prechelt L (2012) Early stopping-but when? In: Montavon G, Orr G, Müller K (eds) Neural networks: tricks of the trade, Book 7700. Springer, Berlin, Heidelberg, pp 53–67

Rogers EO, Fleischer GW, Simpson PK, Denny GF (2004) Broadband fish identification of Laurentian Great Lakes fishes. In: IEEE international geoscience and remote sensing symposium, 2004. IGARSS '04. Proceedings. 2004, pp 1430–1434

Saberioon M, Gholizadeh A, Cisar P, Pautsina A, Urban J (2017) Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues. Rev Aquacult 9:369–387. https://doi.org/10.1111/raq.12143

Salman A, Siddiqui SA, Shafait F, Mian A, Shortis MR, Khurshid K, Ulges A, Schwanecke U (2020) Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system. ICES J Mar Sci 77:1295–1307. https://doi.org/10.1093/icesjms/fsz025

Shahrestani S, Bi H, Lyubchich V, Boswell KM (2017) Detecting a nearshore fish parade using the adaptive resolution imaging sonar (ARIS): an automated procedure for data analysis. Fish Res 191:190–199. https://doi.org/10.1016/j.fishres.2017.03.013

Sheaves M, Bradley M, Herrera C, Mattone C, Lennard C, Sheaves J, Konovalov DA (2020) Optimizing video sampling for juvenile fish surveys: using deep learning and evaluation of assumptions to produce critical fisheries parameters. Fish Fish 21:1259–1276. https://doi.org/10.1111/faf.12501

Smith CS, Paxton AB, Donaher SE, Kochan DP, Neylan IP, Pfeifer T, Van Hoeck RV, Taylor JC (2021) Acoustic camera and net surveys reveal that nursery enhancement at living shorelines may be restricted to the marsh platform. Ecol Eng 166:106232. https://doi.org/10.1016/j.ecoleng.2021.106232

Tarling P, Cantor M, Clapés A, Escalera S (2021) Deep learning with self-supervision and uncertainty regularization to count fish in underwater images. http://arxiv.org/abs/2104.14964

Trenkel VM, Ressler PH, Jech M, Giannoulaki M, Taylor C (2011) Underwater acoustics for ecosystem-based management: state of the science and proposals for ecosystem indicators. Mar Ecol Prog Ser 442:285–301. https://doi.org/10.3354/meps09425

Uranga J, Arrizabalaga H, Boyra G, Hernandez MC, Goni N, Arregui I, Fernandes JA, Yurramendi Y, Santiago J (2017) Detecting the presence-absence of bluefin tuna by automated analysis of medium-range sonars on fishing vessels.

PLoS ONE 12:e0171382. https://doi.org/10.1371/journal.pone.0171382

Vatnehol S, Peña H, Handegard NO, Demer D (2018) A method to automatically detect fish aggregations using horizontally scanning sonar. ICES J Mar Sci 75:1803–1812. https://doi.org/10.1093/icesjms/fsy029

Villon S, Mouillot D, Chaumont M, Darling ES, Subsol G, Claverie T, Villéger S (2018) A deep learning method for accurate and fast identification of coral reef fishes in underwater images. Ecol Inform 48:238–244. https://doi.org/10.1016/j.ecoinf.2018.09.007

Whitmarsh SK, Fairweather PG, Huveneers C (2016) What is Big BRUVver up to? Methods and uses of baited underwater video. Rev Fish Biol Fish 27:53–73. https://doi.org/10.1007/s11160-016-9450-1

Zang X, Yin T, Hou Z, Mueller RP, Deng ZD, Jacobson PT (2021) Deep learning for automated detection and identification of migrating American eel *Anguilla rostrata* from imaging sonar data. Remote Sens 13:2671. https://doi.org/10.3390/rs13142671