# Optimizing video sampling for juvenile fish surveys: Using deep learning and evaluation of assumptions to produce critical fisheries parameters

Sheaves M, Bradley M, Herrera C, Mattone C, Sheaves J, Konovalov D

Target journal: Fish and Fisheries

## Abstract

The limitations imposed by traditional sampling methods have restricted the acquisition of data on key fisheries parameters. This is particularly the case for juveniles because most traditional gear explicitly avoids the capture of juveniles, and the juveniles of many species use habitats in which traditional gear is ineffective. The increasing availability and sophistication of Remote Underwater Video Techniques(RUVs) such as Baited Remote Underwater Video, Unbaited Remote Underwater Video and Remotely Operated Underwater Vehicles offer the opportunity of over-coming some of the key limitations of more traditional approaches. However, RUV techniques come with their own set of limitations that need to be addressed before they can fully realize their potential to shed new light on the early life history of fish. We evaluate key strengths and limitations of RUV techniques, and how these can be overcome, in particular by employing bespoke computer vision Artificial Intelligence approaches, such as Deep Learning in its Convolutional Neural Networks instantiation. In addition, we investigate residual issues that remain to be solved despite the advances made possible by new technology, and the role of explicitly identifying and evaluating key residual assumptions

## Introduction

Fisheries research and management suffer from a lack of funda -mental data needed to underpin some of the most critically important fisheries parameters. Catch records of various types are the basis for a substantial part of available fisheries data, and these data are limited in the information they can provide. While they supply much of the information needed on the structure of exploited components of stocks, they usually provide little information about juveniles and other non-exploited elements. This is because most catch data, even fisheries-independent data, come from gears that target fish in harvestable size ranges (Aymaet al., 2016; Stoner, Laurel, & Hurst, 2008), and so exclude early life-history stages. Consequently, for many, and perhaps most species, there is a dearth of information on the spatial distribution, habitat requirements and temporal dynamics of juveniles between settlement and entry into the adult population. The mis -match between egg production (or spawning potential) of a stock and subsequent recruitment strength is a key barrier to fisheries management, as it obscures predictions of harvestable stock size(Magnuson, 1991). For many species, mortality during the juvenile phase, between settlement from the pelagic environment and entry into the fished population, is critical in determining recruitment strength (Bradford & Cabana, 1997; Peterman, Bradford, Lo,& Methot, 1988; Rice, Crowder, & Marschall, 1997), and habitat factors can be critical in regulating this mortality (Gibson, 1994).Information about this phase is key to identify factors that con -strain population success, such as potentially vital juvenile habitats, and to understand variation in recruitment strength (Levin &Stunz, 2005). Such information is critical to both fisheries and eco-system management and conservation (Bradford & Cabana, 1997;Levin & Stunz, 2005; Musick et al., 2000) because it provides (a)knowledge of juvenile habitats that need to be protected, (b) understanding of the extent and direction of change of populations,(c) the ability to predict the size of future harvestable stocks and(d) understanding of the impact of habitat/environmental change on recruitment and survival through early life-history stages. To understand the factors that constrain population success during the juvenile phase for any fishery, key information includes the following: (a) where juveniles are found throughout the sea -scape, and which of these locations contribute individuals to fished stocks (Gillanders, Able, Brown, Eggleston, & Sheridan, 2003),(b) whether there are particular habitats where predation is reduced and survival is higher (Beukers & Jones, 1998), and if so, what habitat qualities are important, (c) what food resources are important, and whether there are particular habitats where food resources are more abundant and growth is higher (Gibson, 1994).This information is likely to differ throughout ontogeny (Bradford& Cabana, 1997), and so is needed for multiple life-history stages, at the very least at the settlement/metamorphosis stage, ear-ly-juvenile stage and late-juvenile stage (Eggleston, 1995). This information can underpin an understanding of what ontogenetic stages present the most significant

population bottlenecks and which resources are limiting factors during these bottlenecks. To inform an understanding of yearly recruitment strength and proj -ect future harvestable stock size, key information includes variation in the abundance of juveniles in habitats and locations that are known to contribute to the fished population, which requires much of the above information along with age-structured estimates or indices of juvenile abundance (Deyle, Schueller, Ye, Pao,& Sugihara, 2018; Zhang, Reid, & Nudds, 2017).

## The Role of a Basic Juvenile Census

In developing a comprehensive knowledge of juvenile dynamics ina fishery where there is limited data on the juvenile phase, there isa logical order in which this information can be produced. Juvenile habitats are often imperfectly known (Adams, Wolfe, Kellison, &Victor, 2006; Bradley, Baker, Nagelkerken, & Sheaves, 2019; Bradley,Baker, & Sheaves, 2017; Rooper, Boldt, & Zimmermann, 2007). First, in order to begin gathering data on any aspect of juvenile ecology, a basic census must be conducted to determine the locations where juveniles are present and the abundance or density at which they occur. Ideally, this knowledge should be comprehensive, giving abroad understanding of distribution in the seascape, necessitating broad-scale exploratory surveys. With broad-scale surveys, very large areas need to be sampled, with most samples likely to only yield information that the juveniles are absent. This results in expensive sampling effort for potentially little gain. Even where some juvenile habitats are known, to provide valid detail needed for a basic census of the spatial distribution of early life-history stages, sapling an area intensively, and spreading that sampling extensively, is required. Once basic patterns of distribution and abundance havebeen determined, various approaches must be employed to gain deeper understanding. Measurement of fish condition and diet will require the capture of fish for tissue and gut content analyses, often achieved through beach seining and trawl sampling (Duffy,Beauchamp, Sweeting, Beamish, & Brennan, 2010). The capture of live individuals will also be necessary for experimental studies of growth and survival during the juvenile stage (Poletto et al., 2018).However, none of this information can be gathered until the spatial distribution of early life-history stages is understood. In this article, we focus specifically on current techniques and technologies that can enable the basic censuses needed to develop a comprehensive understanding of the spatial distribution of early life-history stages of fish and allow the widespread implementation of reliable, repeat-able surveys of juvenile recruitment. Various approaches have been used to provide basic census in-formation on juvenile fish. Each has its advantages and limitations(see Table 1), so each varies in its applicability to particular questions.For instance, many netting or trapping techniques can be used toprovide indices of juvenile recruitment strength over time. However, most of these have limited practical applicability for broad-scale exploratory surveys, because they can only be employed in habitats for which their capture methods

are designed (Paradis, Mingelbier,Brodeur, & Magnan, 2008; Sheaves, 1996a, 1998; Sheaves, Johnston,& Abrantes, 2007) and therefore are only effective on species that use those habitats extensively, or are highly selective in the species and life stages they can effectively sample. Moreover, their habitat specificity means that potentially unknown habitats remain un-censored. Surveys aimed at identifying juvenile habitats of species where nurseries are unknown are even more challenging because, by their nature, such investigations need to assess the diversity of habitats available over large areas. In fact, to be effective as a large-scale technique to provide suitable data for fisheries management, surveys of juvenile fish need to meet a diverse set of challenging criteria (Figure 1). Clearly, no currently available gear type addresses all the criteria perfectly (see Table 1) meaning that compromises are unavoidable, and in turn, the assumptions underpinning these compromises need to be identified and their impacts well understood. Inevitably, some limitations will render a sampling approach untenable for certain applications.

## The Use of RUVs for Basic Juvenile Census

Visual techniques provide a sampling approach that is able to sample structured and unstructured habitats effectively. The most obvious limitation of most traditional netting techniques is their unsuitability for sampling structurally complex habitats (seeTable 1); probably a large part of the reason why the juvenile habitats of many species are poorly resolved. While the use of visual techniques across different habitats comes with its own set of limitations and assumptions (see Table 1), the development of these approaches provides the potential to remedy the lack of information on fish early life-history stages, particularly for species that use structured habitats. Of the visual techniques, Diver Conducted Underwater Visual Census (DUVC) is a relatively well-developed technique that has allowed a massive expansion of knowledge of the fish fauna of many shallow water habitats. Remote Underwater Video Techniques (RUVs) such as Baited Remote Underwater Video (BRUV), Unbaited Remote Underwater Video (UBRUV) and Remotely Operated Underwater Vehicles (ROVs) offer the opportunity of overcoming some of the key limitations of DUVC, principally their limitation to use in shallow water, the high cost of conduct-ing diver surveys, the disturbance of a human entering a habitat(see Mallet & Pelletier, 2014 for more detail factors) and the riskto divers in habitats where dangerous creatures such as crocodiles, sharks or seals are present (Sheaves, Johnston, & Baker, 2016). In addition, they bring with them the advantage of providing a permanent record of the raw data collected, allowing for later reanalysis (Cappo, Harvey, Malcolm, & Speare, 2003). However, Remote Underwater Video Techniques (RUVs) come with their own set of limitations (see Table 1) that need to be overcome before they can fully realize their potential to shed new light

on the early life history of fish and indeed on all fish life-history stages. The impact of these limitations can be minimized by understanding the assumptions needed to accommodate them, articulating those assumptions explicitly and carefully interpreting data in light of the assumptions. This process is actually necessary to make valid use of data of any kind. However, as with every other technique, the confidence we can place in the output of RUVs will increase as we limit the constraining impact of the assumptions required. As a step towards minimizing the number and potential impact of assumptions when employing RUVs, we evaluate key strengths and limitations of RUV techniques, and how these can be overcome, in particular by employing bespoke Deep Learning (DL) (LeCun, Bengio, &Hinton, 2015) approaches, such as Convolutional Neural Networks(CNN).

Table 1: Sampling options for juvenile fish census, their advantages, limitations and successes

| gear | advantages | limitations | successes | References |
|---|---|---|---|---|
| Trawl nets | • Cover large areas efficiently<br>• Can be operated over a range of depths<br>• Can be deployed at night<br>• Collect spatially extensive samples<br>• Do not require the operator to enter the water<br>• Sampling area can be calculated | • Limited to unstructured habitats<br>• Mesh sizes maybe inappropriate<br>• Destructive<br>• Usually used to collect samples over extensive area so time intensive per sample so usually provide low spatial replication<br>• Representation uncertain due to potential for gear avoidance | • Oozeki et al. (2004) developed an effective frame trawl for sampling pelagic larval and juvenile fish.<br>• Jůza and Kubečka (2007) used fixed-frame trawls to estimate juvenile fish densities in the offshore zone of reservoirs in the Czech republic. The trawls worked best at night and for pikerperch (*S. lucioperca*), bream (*A. brama*) and bleak (*A. alburnus*). | (Jůza & Kubečka, 2007)<br>(Oozeki, Hu, Kubota, Sugisaki, & Kimura, 2004)<br>(Oozeki, Hu, Tomatsu, & Kubota, 2012)<br>(Rotherham, Johnson, Kesby, & Gray, 2012) |
| Seine nets | • Can target particular areas<br>• Can be operated to minimise destructive potential<br>• Can be deployed at night<br>• Collect moderately spatially extensive samples<br>• Can be managed so operators do not need to enter the water<br>• Sampling area can be calculated | • Limited to unstructured habitats<br>• Most approaches require shallow water and a bank to haul on to<br>• Usually used to collect samples over moderately extensive area so time intensive per sample so usually provide low spatial replication<br>• Representation uncertain due to potential for gear avoidance | • Paradis et al. (2008) found seine nets most effective at sampling juvenile yellow perch. They received better measures of abundance, precision and occurrence of juveniles than with pop-nets. The nets worked well in sparse and densely vegetated areas in fresh water environments.<br>• Carassou et al. (2009) captured juvenile reef fish with underwater seines. In seagrass or macro-algae habitats seine nets were most effective with juveniles within the first meter from the seafloor.<br>• Beach seines are more effective at catching juvenile silver carp in river-floodplain systems and beach seines better able to capture smaller individuals. Beach seines also more cost-effective in terms of output per | (Carassou et al., 2009)<br>(Collins, Diana, Butler, & Wahl, 2017)<br>(Espino, González, Haroun, & Tuya, 2015)<br>(Kanou, Sano, & Kohno, 2004)<br>(Paradis et al., 2008) |

| gear | advantages | limitations | successes | References |
|---|---|---|---|---|
| | | | unit effort of labour (Collins et al. 2017).<br>• Espino et al. (2015) demonstrated seine nets were effective at assessing juvenile parrotfish in seagrass habitats and best for accurate size measurements. | |
| Cast nets | • Can be used across many habitats<br>• Can be deployed at night<br>• Do not require the operator to enter the water<br>• Able to collect many small samples per unit time so lends itself to high replication | • Difficult to control sampling area precisely<br>• Limited to unstructured habitats (but can be deployed closely adjacent to structured habitats)<br>• Representation uncertain due to potential for gear avoidance | • Wang et al. (2009) used sast nests used in non-vegetated estuaries and mangrove sites/ mudflats. Cast nets caught a comparable amount of fish but differed in the community composition relative to other methods (gill net and centipede net) | (Stevens, 2006)<br>(Stein III, Smith, & Smith, 2014)<br>(Wang, Huang, Shi, & Wang, 2009) |
| Fyke nets | • Can provide a comprehensive sample for particular habitat types<br>• Provide non-destructive samples | • Limited to particular sampling situations | • Collins et al. (2017) demonstrated mini-fyke nets were most effective at capturing high densities of juvenile silver carp in river floodplains. Mini-fyke nets also performed well in comparison to other netting methods.<br>• Fyke netting was used to monitor a population of golden galaxias (*Galaxias auratus*) in a man-made dam in They asmania by Hardie et al. (2006). They found fyke netting was most effective at capturing juveniles during the day and night in the littoral zone. | (Bonvechio, Sawyers, Bitz, & Crawford, 2014)<br>(Collins et al., 2017)<br>(Hardie, Barmuta, & White, 2006)<br>(Van Der Veer et al., 1992) |
| Baited fish traps | • Can be used across many habitats | • Uncertain and variable extent of bait plume makes precise control | • Merilä (2015) assessed the success of baited and unbaited minnow traps to sample nine-spine sticklebacks | (Bosch et al., 2017)<br>(Harvey et al., 2012) |

| gear | advantages | limitations | successes | References |
|---|---|---|---|---|
| | • Can be used in structurally complex habitats<br>• Do not require the operator to enter the water<br>• Can be deployed at night<br>• Able to collect many small samples per unit time so lends itself to high replication | and knowledge of sampling area uncertain<br>• Representation uncertain due to potential for gear avoidance | (*Pungitius pungitius*) in lakes in Finland. Baited traps had a similar catch-per unit effort for juveniles, but baited traps included more adults in the sample.<br>• Sheaves (1996) used baited traps to sample juveniles in structurally complex habitats. | (Merilä, 2015)<br>(M Sheaves, 1996)<br>(Sheaves, Johnston, Johnson, Baker, & Connolly, 2013) |
| Drop samplers | • Provide comprehensive, quantitatively precise samples from particular habitat types<br>• Provide non-destructive samples | • Limited to unstructured, shallow water habitats<br>• Slow to deploy and involves considerable boating activity to deploy so likely to result in avoidance<br>• Produces relatively few samples per day so usually provides low spatial replication | • Baltz et al. (1993) utilised drop samplers to assess microhabitat use by fish in an extensive saltmarsh in Louisiana. Over three years drop samplers were deployed along transects and resulted in 57 different fish species all with a high count of larval and juvenile stages. | (Baltz, Rakocinski, & Fleeger, 1993)<br>(Rozas & Minello, 2015) |
| DUVC | • Is not restricted by habitat type<br>• Provides a real-time visual sample so allows the collection of additional information (e.g. behaviour) in response to observation | • Limited to depths that divers can access<br>• Daytime and night-time samples require different techniques so have limited comparability<br>• Disturbance of having a human intruder in a habitat likely to lead to avoidance by many species<br>• Limits on safe dive-time means replication per day is limited, reducing cost-effectiveness and spatio-temporal replication<br>• Involve personnel to enter the water leading to unacceptable risks such as crocodile attack<br>• Observer bias potentially high | • Espino et al. (2015) used DUVC to assess habitat use and abundance of parrotfish in seagrass meadows. They recorded a similar abundance measure of target fish to seine nets but because of difficultly detecting juveniles, the distribution of juvenile sizes was underestimated relative to their seine net data. | (Espino et al., 2015)<br>(Lindfield, Harvey, McIlwain, & Halford, 2014)<br>(Mallet & Pelletier, 2014)<br>(Warnock, Harvey, & Newman, 2016) |

| gear | advantages | limitations | successes | References |
|---|---|---|---|---|
| ROV | • Is not restricted by habitat type<br>• Provides a real-time visual sample so allows the collection of additional information (e.g. behaviour) in response to observation<br>• No depth limitation | • Daytime and night-time samples require different techniques so have limited comparability<br>• Disturbance of having a large moving device in a habitat likely to lead to avoidance by many species<br>• Limits on complexity of operation means replication per day is limited, reducing spatio-temporal replication<br>• Observer bias potentially high | • ROV system with a mounted 'tickler chain' (to encourage fish movement immediately in front of camera) was an effective means to locate juvenile flatfish and assess their density. This technique worked well in various habitat types from fine mud to rocky bottoms with kelp and coral (Norcross and Mueter 1999). | (Ayma et al., 2016)<br>(Norcross & Mueter, 1999)<br>(Mallet & Pelletier, 2014)<br>(Struthers, Danylchuk, Wilson, & Cooke, 2015)<br>(Warnock et al., 2016) |
| BRUVS | • Can be used across all habitat types<br>• No depth limitation<br>• Medium sized camera units minimise and localise disturbance and therefore avoidance<br>• Many units can be deployed at one time allowing high spatial replication<br>• Techniques available to extend to size measures | • Uncertain and variable extent of bait plume makes precise control and knowledge of sampling area uncertain and limits value for habitat-specific studies<br>• Daytime and night-time samples require different techniques so have limited comparability<br>• Limited to relatively high water clarity situations and detection varies with water clarity<br>• Limits on field-of-view, and structure in front of camera can limit detection<br>• Identification of target species and life-stages subjective<br>• Generates large bodies of data making data analysis time consuming<br>• Observer bias potentially high | • Stoner et al. (2008) successfully deployed BRUVS to assess the relative abundance of juvenile Pacific cod in Alaska. This technique performed well in seagrass, kelp, and open habitats both in shallow and deeper waters.<br>• Although not specifically sampling for juveniles Hardinge et al. (2013) successfully recorded the juveniles of many fish species in a temperate reef system in Western Australia. In their comparison between 3 different bait treatments (200g, 1000g & 2000g) juveniles were present at all baited cameras. | (Bosch et al., 2017)<br>(Hardinge et al., 2013)<br>(Harvey et al., 2012)<br>(Mallet & Pelletier, 2014)<br>(Stoner et al., 2008)<br>(Struthers et al., 2015) |
| UBRUVS | • Can be used across all habitat types<br>• No depth limitation | • Daytime and night-time samples require different techniques so have limited comparability | • Using UBRUVS Bradley, Baker, and Sheaves (2017) were able to sample fish assembles in previously | (Bradley et al., 2017)<br>(Cullen & Stevens, 2017) |

| gear | advantages | limitations | successes | References |
|---|---|---|---|---|
| | • Small camera units minimise and localise disturbance and therefore avoidance<br>• No attractant used therefore suitable for habitat-specific data collection<br>• Many units can be deployed at one time allowing high spatial replication<br>• Techniques available to extend to size measures | • Limited to relatively high water clarity situations and detection varies with water clarity<br>• Limits on field-of-view, and structure in front of camera can limit detection<br>• Identification of target species and life-stages subjective<br>• Generates large bodies of data making data analysis time consuming<br>• Observer bias potentially high | unstudied deep waters from an estuary system in Northern Australia. They revealed that 22% of all species surveyed were only present as juveniles, demonstrating UBRUVS as a successful tool for juvenile detection.<br>• Similarly, Sheaves et al. (2016) deployed UBRUVs within mangrove forests to produce new insights into the use of mangroves by fish | (Hardinge et al., 2013)<br>(Sheaves et al., 2016)<br>(Struthers et al., 2015) |

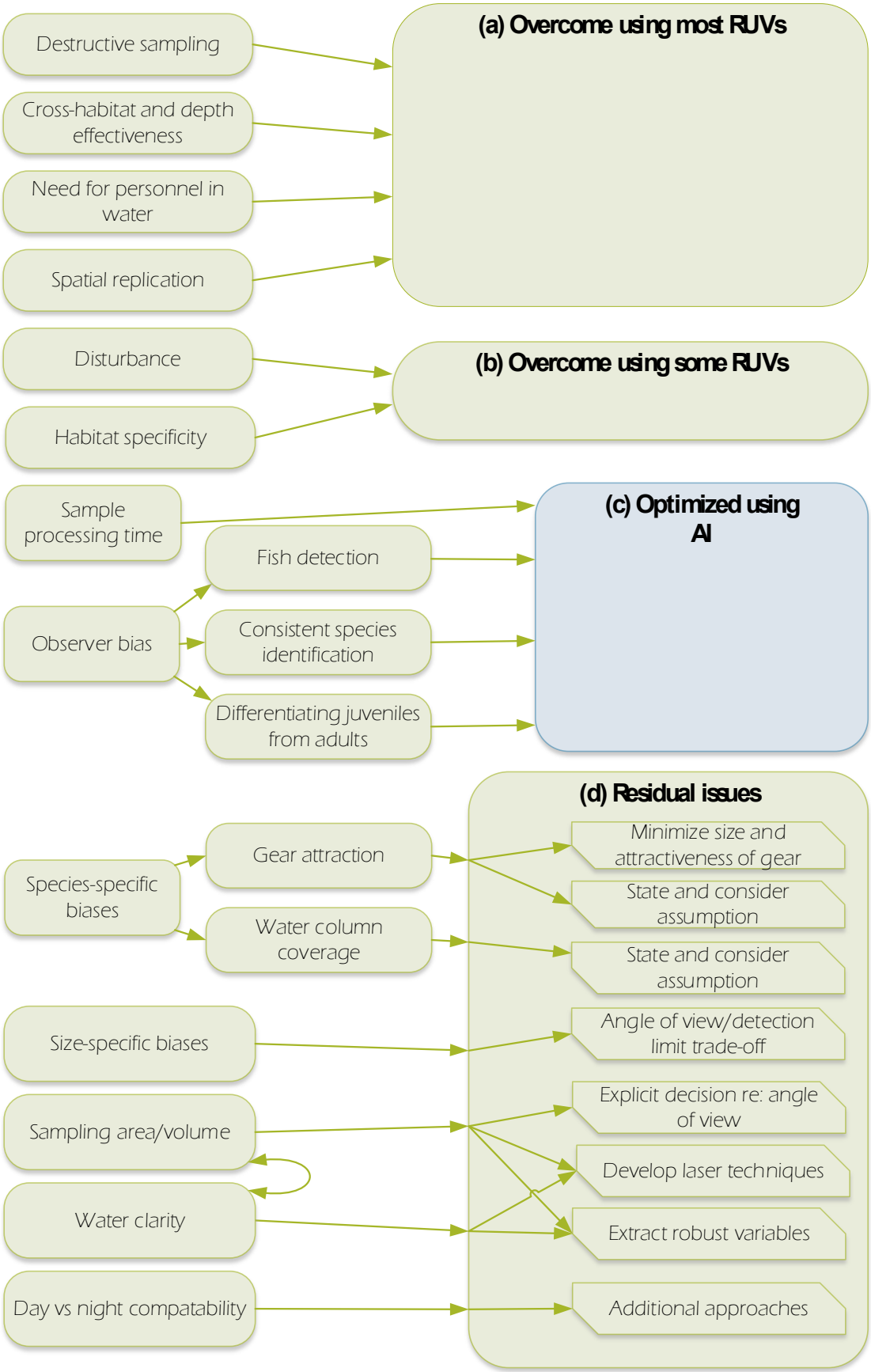
Destructive sampling
Cross-habitat and depth effectiveness
Need for personnel in water
Spatial replication
(a) Overcome using most RUVs
Disturbance
Habitat specificity
(b) Overcome using some RUVs
Sample processing time
Fish detection
Observer bias
Consistent species identification
Differentiating juveniles from adults
(c) Optimized using AI
(d) Residual issues
Species-specific biases
Gear attraction
Water column coverage
Minimize size and attractiveness of gear
State and consider assumption
State and consider assumption
Size-specific biases
Angle of view/detection limit trade-off
Sampling area/volume
Explicit decision re: angle of view
Develop laser techniques
Water clarity
Extract robust variables
Day vs night compatability
Additional approaches

*Figure 1: Issues and solutions relating to effective juvenile fish surveys. (c) indicates approaches where Deep Learning provides substantial opportunities for optimization*

## RUV Techniques; strengths, limitations and solutions

### *Techniques- specific strengths*

Current RUV techniques offer solutions to many of the issues preventing comprehensive understanding of the spatial distribution of early life-history stages of fish and limiting the widespread implementation of reliable, repeatable surveys of juvenile recruitment (Figure 1a). One obvious feature of RUVs is that they are non-destructive and among the least invasive of sampling approaches(Mallet & Pelletier, 2014); overcoming the disadvantages shared by many other approaches that makes them unsuitable for routine ap-plication to sampling delicate juvenile stages. A second advantage is that RUVs can be deployed effectively in almost any habitat type and at all depths (Ayma et al., 2016; Mallet & Pelletier, 2014), al-though modifications may be required for use in very deep waters(Norcross & Mueter, 1999).Because of their different operational characteristics, the various RUVs differ in the roles they can play and to biases that come with their deployment. This makes their values situation-specific. For instance, while ROVs provide the possibilities for large-scale transect-based surveys, the data produced may be biased compared to the minimal disturbance caused by stationary approaches such as UBRUVs. While it is possible to operate ROVs without personnel in the water, the size and motion of most ROVs means they are still have the potential for disturbance leading to avoidance by some species (Ayma et al., 2016). This means that the value of ROVs will often need to be established on a use-by-use basis, relative to particular situations and goals. Small stationary BRUVs and UBRUVs largely overcome the problem of fish avoidance due to movement, and the ability to deploy them without continual operator control means they can provide high spatial replication (Mallet & Pelletier, 2014).The aggregating effect of their bait plume makes BRUVs particularly useful in situations where fish densities are low (Mallet &Pelletier, 2014) but the presence of bait and variation in the area of attraction means they produce estimates with biases that are difficult to assess (Hardinge, Harvey, Saunders, & Newman, 2013). Consequently, where habitat-specific information is required UBRUVs provide an obvious advantage over BRUVs because, in not relying on a bait plume to attract fish to the camera, the fish they record can be assumed to be normal inhabitants of the habitat in which they are deployed.

## Current Limitations that could be addressed by emerging DL technologies

The list of potential limitations for RUVs is extensive (Figure 1).Perhaps most crucially, while the stream of video information provided by RUVs is one of their strengths it also brings with it one of their greatest limitations, that they inevitably produce massive volumes of data (Mallet & Pelletier,

2014; Matabos et al., 2017).Combined with the high replication that can be achieved, particularly with BRUVs and UBRUVs, the amount of data that needs to be pro -cessed can quickly overwhelm the resources of human video view -ers. This often renders video analysis prohibitively costly, militating against consistent high-quality assessment or resulting in reliance on rudimentary summary information from massive data collections. Recent development of computer vision Deep Learning approaches could potentially overcome this limitation. Deep Learning (DL), asub-discipline within Artificial Intelligence research, has proven to be useful in detecting and classifying objects from images, in particular through the use of Convolutional Neural Networks (CNN) variations(Konovalov et al., 2019; LeCun et al., 2015). CNN and relevant programming packages (e.g., Keras, Tensorflow and PyTorch) have made it possible to train and implement fish image detectors and classifiers for research (Konovalov et al., 2019). This substantially reduce the amount of video that needs to be assessed by human viewers. Although this seems a modest ability, over time as training data and network models improve, this would free research staff from watching and labelling videos, removing a major cost impediment and al-lowing staff to devote that time to the research component of the work. Observer bias is a problem for traditional video analysis. Becauseit relies entirely on humans to detect and identify fish, consistency of detection and identification varies complexly over time, influenced by both the viewer's level of proficiency and experience, and their fatigue levels (Rattray, Ierodiaconou, Monk, Laurenson, &Kennedy, 2014). This problem is magnified when multiple viewers are involved because different viewers are likely to have different levels of detection and identification skills. Moreover, even rudimentary quality control requires multiple viewing of at least a proportion of videos. The use of multiple viewers brings with it the possibility of a type of cognitive bias (Keil, Depledge, & Rai, 2007), where a doubtful identification is converted into a confident one due to the reinforcement of multiple subjective opinions. Deep Learning (DL)can help minimize cognitive bias; by acting as an empirically deter-mined statistical model, DL can provide detection and identification with a consistent, defined level of error. Additionally, because it uses explicit criteria, DL can help overcome the related problem of human observers making positive identification based on “experience” rather than on empirical characteristics in situations where difficult conditions (e.g., poor water clarity) make recognition of defining characteristics uncertain. However, DL introduces its own biases (i.e., data bias and/or algorithm bias). Although these will usually be easy to identify, justify and control, it is important that they are explicitly assessed and accounted for.The problem of observer bias is even more problematic in studying juveniles because of the need to consistently differentiate juveniles from later stages. Juveniles often have marking patterns or morphological characteristics that are different from those of sub-adults or adults. In some cases, these differences may be distinct enough for relatively easy identification by human observer. However, even then

there is the problem of subjectivity involved in deciding when a shape or pattern is distinctive enough to trigger positive differentiation into either juvenile or non-juvenile categories. The extent of the problem is non-trivial, with different human assessors producing different estimates of even simple parameters when viewing underwater video (Mataboset al., 2017). Added to this is the problem of drift in stage-identification as observers come to recognize additional characteristics, and so subtly (and usually unconsciously) change the criteria they are using. Because of its explicit criteria, trained CNN can provide a solution by enabling consistent differentiation of individuals into categories based on characteristic sets using specific cut-off points that do not vary over time and have a known error associated with the cut-offs. Because CNN models involve explicit parameter definitions (categories, weights and bias) and training over many samples, they can provide consistent differentiation in situations where human observers' decisions would diverge or vary over time. Because DL techniques are continually evolving, new cut-off points can be developed as prediction models improve and can be applied to video that has been assessed previously, if  that is necessary

## Limitations and Challenges Associated with the Application of DL to RUV

Some of the challenges that have limited CNN use in ecology areas follows: (a) complex and dynamic backgrounds in natural settings that hinder object detection and classification; (b) object deformation, where the variant scale, orientation and flexible shape of an entity might reduce CNN accuracy; and (c) analysis time. The suc-cess of CNN largely depends on the quality of data, optimized hard-ware required to handle and process data (e.g. large memory CPUs and GPUs) and software (Abadi et al., 2016; Campbell, Salisbury, First and foremost, the strategy used for collecting data should be determined by (a) the study question and aim, (b) infrastructure capabilities and (c) the expertise of the research/organization team. Thus, as with any other enterprise, developing DL solutions in fisheries leans on securing appropriate budget and yielding the required domain knowledge (e.g., fish biology, computer and data science).Based on this premise, the following section constitutes a set of guidelines, recommendations and considerations that researchers and managers should consider when planning and executing data collection with the purpose of training detectors and classifiers of juvenile fish.

## Understanding Computer Vision Tasks

There are three foundational tasks in computer vision that under-pin, and would enable, more sophisticated fisheries solutions: fish detection, fish segmentation and fish classification. These are the computing ability to detect and localize fish in an image, the computing ability to isolate fish from

background and the computing ability to correctly classify detected fish with one or multiple "labels,"respectively. There are several algorithms and combinations of algorithms capable of achieving detection, segmentation and classification of fish (Kaur & Kaur, 2014; Matai, Kastner, Cutter, & Demer,2012; Shafait et al., 2016; Villon et al., 2016), see historical over-view by Salman et al. (2016). However, in recent years Convolutional Neural Networks (CNN, Fukushima, 1980; LeCun, Bottou, Bengio,& Haffner, 1998) have become the most common and success-ful computational approach used in fish detection and classification (Konovalov et al., 2019; Li, Shang, Qin, & Chen, 2015; Mandal,Connolly, Schlacher, & Stantic, 2018; Salman et al., 2016; Storbeck& Daan, 2001; Villon et al., 2016). It has been demonstrated with-out a doubt that a modern CNN, for example the state-of-the-art EfficientNet (Tan & Le, 2019), could achieve human-expert-level accuracy if a sufficiently large and diverse training set of labelled images is available. When a novel class of objects does not have a large corresponding data set of labelled images, the main practical (and project-cost) challenge is currently associated with creating such a CNN training data set. A typical overall labelling cost is $1–$10 per labelled image, where the per-image-cost naturally decreases with larger volumes.

## Constructing a training library for effective Deep Learning

The effective number of images/videos required to develop CNNs for detection and classification of juvenile fish is not a trivial matter. The image set must capture the variability of conditions expected during CNN implementation. In many situations, this would mean capturing images for the target fish across all underwater habitats and conditions the species occupy. In addition, the image set would ideally contain all possible representations of the target species. Of particular importance is variation in fish shape during swimming (Shafait et al., 2016), variation in phenotype, which occurs both at a population level and in response to environmental cues(Meuthen, Baldauf, Bakker, & Thünken, 2018) and throughout development, which can also differ in response to environmental con -text (Nyboer, Gray, & Chapman, 2014). Failing to account for those variability sources would limit the generalization capabilities of the CNN model, compromising DL success over large spatial scales and broader applications. Gathering a data set with the characteristics described above can be expensive. There are costs associated with sampling, data curation and labelling. With the current trend of improvement and cost reduction of necessary DL equipment, as well as the increasing diversity of DL open source and freeware software, costs associated with data collection and curation are likely to become the most important challenges to tackle. However, three strategies can be used to overcome this challenge. Firstly, researchers and managers can take advantage of historical images and video recordings which can be repurposed to train CNN. Secondly, using data augmentation or transferring learning (see Table 2) small data sets can be trained with moderate generalization capabilities.

Thirdly, curation and ground truth labelling can be partially automated using already available fish detectors and other computer vision algorithms (motion detection and object segmentation).

Table 2: Table of considerations for the application of DL to RUVs

| Component | Consideration |
|---|---|
| *Infrastructure for data storage and processing:* Footage should be stored in a medium that protects data and allows for easy retrieval, therefore minimizing the cost of sharing and reusing it. Data protection implies physically safeguarding data from damage, but also ensuring it lasts into the future in mediums and formats that can be later consulted. Processing of footage is faster if run on appropriate hardware such as Graphic Processor Units (GPUs) | Meticulously curated metadata improves processing data fetch into the AI algorithms and facilitates searching and finding data of interest With the advent of high-quality image sensors with high resolution(e.g., 4K), storage and processing are something that should be thoughtfully considered. When used for large spatio-temporal scale monitoring, the universally accessible nature of video technology is challenged by the expense of storing and processing big volumes ofraw data |
| *Labelling:* A data set of training images/videos in which the objects of interest are identified (and preferably verified), is required for training supervised DL/CNN. Labelling is the most time-consuming task on the analysis cycle and requires prior expert knowledge. There are three main types of labelling: image/video-level, bounding box and mask. As their names indicate, image-level label specifies the target fish present or not anywhere in the image (or video for the video level), a bounding box label enclose the target fish inside abox or region of interest, while a mask label segments the target fish(crop around fish contour) from the background. CNN models and architectures are designed to work with either of these labels' | It is important to remember that generally supervised DL/CNN will learn from the training/validation data set, and then when presented with new data will make predictions based on what it has learnt.For this reason, and depending on the specific DL/CNN task, data collection should include samples of most of the potential cases expected. For detecting different life stages of fish, obtaining highly accurate training images is crucial. To ground truth images of different life stages/size-classes, ideally fish should be captured, imaged throughout their growth from early-juvenile to adult, then euthanized and aged to determine an accurate age estimate for each set of images throughout the life cycle. Otherwise, the DL/CNN will |

| | |
|---|---|
| types.Lastly, there is an increasing use and demand of mask label data. The type of labelling would ultimately depend on the research purpose and CNN type used, where the image-level labelling is the easiest/fastest (and hence least costly) to curate, followed the bounding boxes and then by the most time-consuming segmentation masks | inherent any inconsistencies and biases from subjective human identification |
| *Augmentation:*<br>Data augmentation refers to the process of applying a series of random changes in the appearance of the images set. Changes include rotation, warping, scaling, contrast and many more. Thus, a given image set can be augmented by randomly applying these transformations during training (or validation or testing). Artificially creating variation is useful to increase the CNN generalization capability | Depending on the classification task, researchers can select which modifications can be appropriate to generate variation and improve model generalization, and which ones should be avoided. The magnitude of change can also be restricted. Augmentation must be severe enough to give flexibility to the network, but not so great that it leads to misidentification |
| *Network selection:*<br>Depending on the topology of the neural layers used, Deep Learning algorithms are classified according to their different known CNN architectures (e.g., ResNet-50 He, Zhang, Ren, & Sun, 2016), which balance accuracy against speed in different ways. Networks can also be individually customized to improve performance | The advantages of customization may be outweighed by the reproducibility and transparency in using publicly available networks "off the shelf" (discussed above) |

| | |
|---|---|
| *Training, validation and testing—Learning:* The network trains on a set of training images, then it can be tuned using validation images, re-trains based on these results, and tests itself on a separate set of testing images | The behaviour of the network, including levels of accuracy required, can be set by researchers. Here, important decisions must be made about the relative acceptability of false positives versus false negatives. In general, either false positives or false negatives could be easily reduce but not both (Fawcett, 2006) |
| *Observing the prediction model:* Once the prediction model has been developed, the criteria used by the AI to distinguish between species and life stages can be extracted, summarized and presented. | Here, researchers can determine whether the criteria developed by theAI are biologically/ecologically relevant |
| *Applying the DL/CNN model—automated processing of data:* If the DL/CNN model achieved acceptable performance on test and validation images/videos, it can now be used to automatically process unlabelled video | Here, a balance must be reached between comprehensive processing of the video sample and acceptable processing time. Analysis speed is needed to be at least at the video's fps or faster (i.e., 30 frames of30 fps footage then would take <1 s to process). For the identification of fish, it may be acceptable not to process every frame, as fish usually inhabit the frame for >1 s (i.e. could be detected in at least 30 frames) |

## Pilot case-study: detecting early-juveniles in RUV surveys

To illustrate the feasibility of DL/CNN-based processing of RUVs, we present a pilot case-study of detecting early-juvenile mangrove snap-per (*Lutjanus argentimaculatus*, Lutjanidae) in videos. Mangrove snap-per are an important commercial and recreational fisheries species, distributed widely throughout the Indo-Pacific. With spatially distinct juvenile and adult populations, knowledge of recruitment variability and early-juvenile habitat requirements is critical for a complete un-understanding of the fishery. Initially, Konovalov et al. (2019) developed a human-labour efficient

labelling approach for detecting fish in under-water videos, where habitat-specific video clips with and without fish were manually cropped and then converted to individual images (one image per video frame). Xception CNN (François Chollet, 2017) was used within the Keras deep learning framework running via Tensorflow as the backend. For 20 different habitat types considered, four thou-sand labelled images of target fish species were sufficient to train the Xception CNN to achieve 0.17% false positives and 0.61% false negatives on the project's 36,000 test images. For our pilot study focusing on the early-juvenile phase of mangrove snapper (see Figure 2), we only used a small initial set of training images (140 positives and 1,398negatives). Following the labelling approach of Konovalov et al. (2019),we trained a PyTorch-based Xception CNN using the training pipe-line from Konovalov et al. (2020). To achieve a high level of accuracy, we developed a novel technique of training a CNN on mixed resolution images. Each high-resolution image was cut into four quarters and combined with a downsized version of the original image. Hence, each training image yielded five distinct images (the four quarters and the downsized version). When processing, each frame was converted to five images in a similar fashion. Figure 3 presents an example of a CNN prediction where the target juvenile fish was detected (blue highlights in Figure 3) in both the original high-definition resolution and the downsized version. However, it was equally common to see detections only in the high resolution or only in the lower resolution.Typically, very small (on the image scale) juveniles where only detected in high resolution and very large visual instances were detected only in lower resolution. For full details of the training and testing procedure, see Supplementary Materials. The CNN we created was able to identify early-juvenile man-grove snapper from a new set of videos with a high level of accuracy(95%) compared to a human viewer. These videos contained a range of different fish species, habitats and conditions. Importantly, these videos also contained late-juvenile mangrove snapper, and the CNN was successfully able to distinguish between them and our target—the early-juvenile stage. The CNN usually detected target fish at the same point in the video as the human viewer and in some cases detected them before the human viewer. The results of this pilot study(see Table 3) demonstrate that human-level detection accuracy of particular juvenile stages of a fish species can be achieved with a modest set of training images.

Figure 2: Sample image of early-juvenile mangrove snapper(*Lutjanus argentimaculatus*, Lutjanidae)

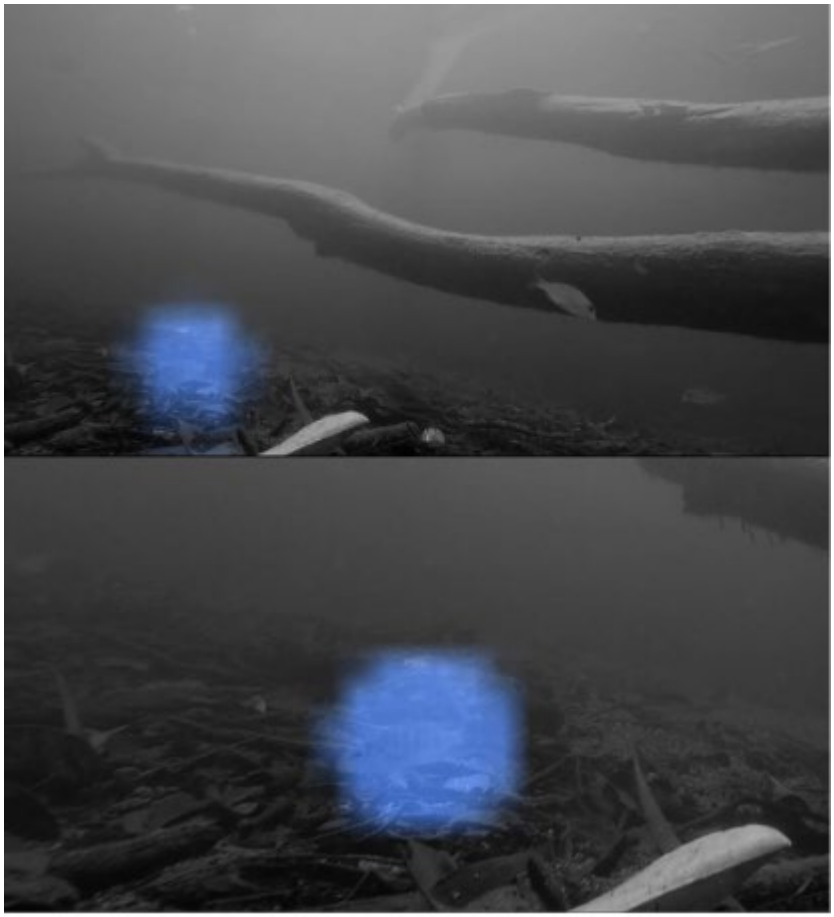

Figure 3: Sample of correct CNN detection and localization in the 1,920 × 1,080 resolution image (bottom subfigure) and the downsized 1,024 × 512 resolution image (top subfigure)

Tabel 3: Summary of comparison between trained network andhuman viewer on a set of video deployments not used in training.

| | |
|---|---|
| Number of video deployments | 49 |
| Total footage (min) | 735 min |
| Number of targets detected by human viewer | 42 |
| Number of targets detected by network | 40 |
| Network accuracy compared to human viewer | 95% |
| Instances where network detected target before human viewer | 17% |
| Instances where network detected target at a similar point in video (±3 s) | 43% |
| Instances where network detected target after human viewer | 38% |

## Residual issues

Although DL-enhanced RUV techniques offer solutions to some of the fundamental issues that have inhibited understanding of the ecology of juveniles of many species, a number of residual issues remain (Figure 1d). The problem of species-specific bias is common across sampling methodologies (Baker & Minello, 2011; Rotherham,Johnson, Kesby, & Gray, 2012; Sheaves, 1995). Most obviously is the problem of attraction or repulsion caused by imposing a foreign object in a habitat, the disturbance associated with deployment of the equipment or the specific attraction that is part of baited sampling techniques. The problem of differential attraction is obviously a concern for BRUVs, where different species are likely to be attracted(or repulsed) by the bait plume (Hardinge et al., 2013), and where the attractive effect varies with the extent and direction of water movement. Moreover, what is attracted also depends on the mix of habitat types intersected by the bait plume (Logan, Young, Harvey,Schimel, & Ierodiaconou, 2017). The utility of BRUVs for the question at hand should be carefully considered, and at the very least, an understanding of species/community bait response, plume size and surrounding seascape should be acquired before results are interpreted (Ghazilou, Shokri, & Gladstone, 2016; Heagney, Lynch,Babcock, & Suthers, 2007; Klages, Broad, Kelaher, & Davis, 2014;Taylor, Baker, & Suthers, 2013). A more localized issue relates to attraction to the flashing lights enabled on many videos that indicate they are recording. Impacts of attraction or repulsion are likely tobe minimized by the use of small, stationary, unbaited video units with neutral (e.g., black) coloured casings and no flashing light, and by careful deployment that minimizes disturbance. While this mix of characteristics and operating procedures is likely to minimize impacts, they are unlikely to be completely eliminated. Consequently, the assumption of no impact needs to be kept in mind when analysing video data and should be considered as a matter of course. A second source of species-specific bias relates to the extent to which a video sample represents the range of habitats available. In part, this is solved by careful spatial sampling design, and however, some issues

remain. One particular problem relates to the extent of the water column sampled by video methods. For example, vid-eos placed on the substratum with only a few centimetres of water over them, may sample the whole water column and therefore include pelagic species, while those deployed in only slightly deeper water will often miss those pelagic even if they are present. The only way around this is to use stacked cameras to cover the whole water column, something that quickly becomes infeasible as water depth increases. Consequently, where only a single camera is used(either bottom-set or floating), there is inherent assumption that the recorded video data only represent the layer of the water column covered. Although this assumption is rarely explicitly stated, if not considered, it has the potential to produce anomalous interpretations. As a result, it is advisable that this water column coverage assumption is routinely explicitly addressed in RUV studies. As with all gears (Gwinn, Allen, & Rogers, 2010; Kubečkaet al., 2012; Sheaves, Johnston, & Connolly, 2010), size-specific biases place limits on interpretation. With RUVs, the sizes of fish that can be detected and identified varies with the angle of view, the characteristics of the lens and water clarity (Mallet & Pelletier, 2014),leading to a trade-off between such things as the area that can be viewed and the ability to detect and identify small fish. Not only does this mean that an explicit decision is needed about the lens, angle of view, etc., based on the main purpose of the study, but that videos collected with different lenses, such as wide angle versus narrow, will be implicitly incomparable for some purposes. Added to this, large fish distant from the video are easier to detect and identify than small fish at the same distance, adding to size-specific bias. This problem is intertwined with the questions of defining the area sampled (Kubečka et al., 2012) and accounting for water clarity (Boland& Lewbel, 1986). Stereo video or laser techniques to measure distance and area sampled, and standardize sampling accordingly, offer possible solutions. The question of defining the area/volume sampled is one of the most vexed for underwater video (Mallet & Pelletier, 2014). Although this is in part controllable by maintaining a constant angle of view, the issue of increasing field of view with distance from the camera remains. Moreover, the distance from the camera in which fish can be detected and identified varies with both habitat complexity and water clarity, and in shallow water is further complicated by effects such as backscattering and lens flare. Complex habitat features, for example macroalgal fronds, mangrove roots or seagrass blades, can block visibility of fish. While the most severe of these effects can be overcome by accepting or rejecting video samples based on degree of obstruction of the field of view, the effects of detectability of fish between structured and unstructured habitats remain. For species that actively move through their surroundings, it is an issue of allow-ing adequate sampling time to encounter individuals, whereas for species that are relatively stationary, this presents a serious limitation. Water clarity is also complex, affected by both dissolved (e.g.,tannins) and particulate (inorganic or organic) matter. The various components have different

impacts on the optical properties of water (Mobley, 2001) and combine to influence light transmission and scattering in complex ways. Dissolved organic components principally impact light transmission while organic and inorganic particles effect backscatter in complex ways so their influence varies substantially in responses to factors like the angle of the sun. It is common practice to measure water clarity and only deploy RUVs when clarity is above a pre-determined threshold. This is fraught with difficulties, however. Firstly, different approaches (e.g., using a nephelometer vs. a Secchi disk) measure water clarity in different ways and provide inconsistent information depending on the source of turbidity (dissolved materials, different types of particles; Davies-Colley & Smith, 2001). It is also often difficult to take measurements at the actual depth where the camera is deployed. In attempting to account for water clarity, it is common to employ a marker at a set distance from the video (e.g., a card with field of dots at different densities) as a way of determining water clarity. In practice, it can be difficult to link this to objective criteria, although it does provide an opportunity for further development, especially when these measurements can be objectively automated with DL technology.Although emerging technologies can help, the issue of variable sampling area/volume remains problematic for two-dimensional video recordings where water clarity, and visibility in general, is subject to complex variability. Quite often there will be no simple solution beyond limiting interpretation to videos with some identifiable level of water clarity or obstruction. This is not only subjective but does not provide consistency of sample area/volume. Consequently, in most situations it is advisable to focus on robust measures such as probability of encounter (Sheaves, Johnston, & Connolly, 2012) and clearly address the assumption involved in even that interpretation. Other more complex measures such as NMax can provide unreliable estimates (Schobernd, Bacheler, & Conn, 2013) so should be used with caution, again with the assumption clearly stated. Finally, there is the question of the differential effectiveness of current video technology between low versus high light situations, particularly at night. For instance, although specialized systems have been developed for specific applications (Hung et al., 2016), current technology requires lighting to be used in most low light and night situations. There are obvious issues of differential attraction suchas those described above. In effect, this places limits on many ap-plications of RUV techniques. In particular, at the moment it is only practical to utilize daytime video data when investigating habitat utilization, leaving a gap in understanding of changes in habitat occupancy at night. The use of artificial illumination and ultrasensitive low-light cameras have the potential to fill this gap (Fitzpatrick,McLean, & Harvey, 2013) but the extent of the advantages they provide still needs to be quantified and the specific assumptions as -sociated with their use understood. Another strategy that can help overcome some of the limitations resulting from low light limitations is to link video collected at times of high light (e.g., daytime) with information from

complimentary sources, such acoustic telemetry(Cooke et al., 2005) that can provide metre resolution of the location of a fish over time. Given the potential of bias, if day–night changes in habitat utilization are not considered, linking video surveys to such complementary techniques should be employed, if at all possible, in studies where understanding fish–habitat relationships is the object.

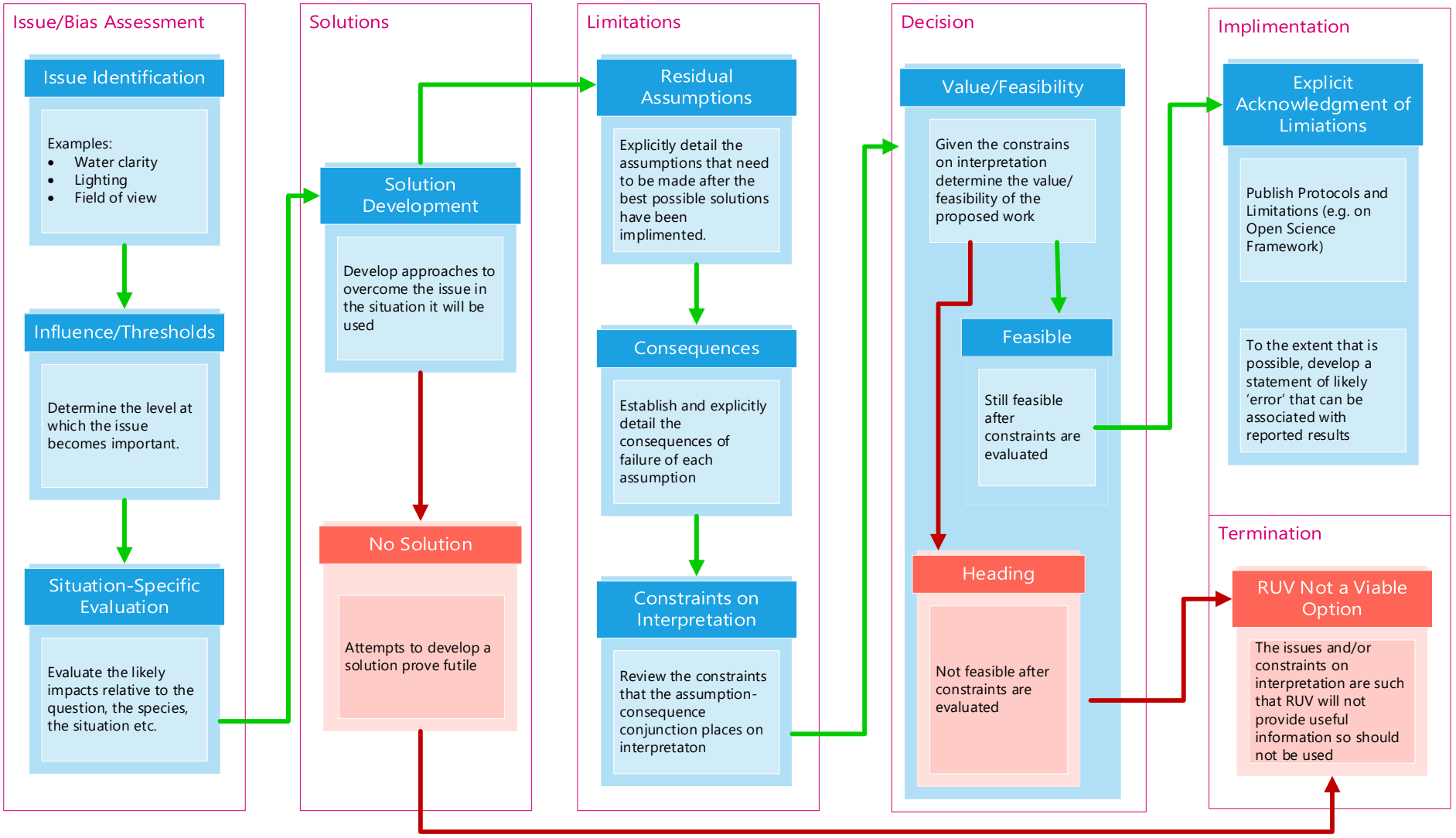


*Figure 4: A framework for assessing issues and biases in RUV projects, understanding limitations and determining reasonable, defensible solutions.*

## A Solution Framework

The range of residual issues (Figure 1d) suggests that RUVs may have limited value in many situations. However, similar issues accompany most other data collection approaches. To move forward, it is necessary to consider; (a) how to overcome or minimize limitations, (b) how limitations constrain reasonable interpretation, (c) whether RUV approaches are suitable to address specific questions in specific circumstances and (iv) how to ensure that the use and interpretation of RUVs are valid and useful. These questions are best addressed by developing a standard framework for assessing and resolving issues and biases (Figure 4).The first step (Figure 4a) involves determining when a specific issue is likely to constrain interpretation and evaluating its likely impact on the investigation (e.g. variability of the bait plume in BRUV sampling). Once the impact of the issue is clear, it may be possible to develop a solution (Figure 4b), often procedural or mechanical, to partially or completely overcome the issue for the particular situation (e.g., standardizing BRUV data based on relative plume size,Taylor et al., 2013). However, there are few situations where all issues can be overcome, and as a result, interpretation of RUV data will be constrained by residual assumptions. The extent to which these assumptions are justified will limit interpretation of data, the generality of findings and, ultimately, the value of using RUV to address the question at hand (e.g., the use of BRUVS for fish community studies where there is differential attraction to bait). Consequently, it is important that these residual assumptions and their consequences are clearly understood and explicitly stated (Figure 4c). If the use of RUV techniques is feasible under this framework, all protocols, assumptions, limitations and associated estimates of error should be made explicit. This is best done using a transparent process, for instance, by using the Open Science Framework (https://osf.io/). This would enable others to be confident in the interpretation, share solutions and produce complementary data. Faced with the innate uncertainty of ecological systems (Harris & Heathwaite, 2012), recognizing key assumptions, and the constraints they place on interpretation, is critical, and often overlooked. An assessment of feasibility allows for either the informed implementation of a RUV solution (Figure 4e) or recognition that RUV is not a viable option (Figure 4f). In sum, RUV techniques, combined with deep learning technology and the explicit evaluation of assumptions, provide a promising tool for performing basic juvenile surveys to fill critical knowledge gaps in fisheries ecology. While not relevant to all species, the carefully considered application of RUV techniques would be suitable for a wide variety of species, particularly demersal species. Once the location of juveniles and key habitats are understood, more intensive, destructive expensive sampling techniques can be employed in a targeted and informed way to fully understand requirements throughout ontogeny and factors limiting population success. While traditional gears will need to be employed into the future,

the ease of deployment of RUVs make them an attractive tool for use by fisheries management organizations as repeatable surveys that could directly inform recruitment strength as part of annual stock assessments.

## The Future of RUV Optimization for Ecology

In a general sense, the optimization of RUVs with emerging techno -logical solutions (e.g. marrying RUVs outputs with DL and Big Data solutions) is a fertile area of methodological development. While we have evaluated these techniques in the context of juvenile fisher-ies surveys, DL and other computer vision technologies hold much promise for enabling previously unavailable data streams, new areas of enquiry, new ways of looking at the world (e.g., a 360° field of view, Campbell et al., 2018) and greatly enhanced knowledge in many areas of ecology. Not only could this boost of knowledge on composition, abundance, size and habitat utilization of countless fauna, but there is the potential to extract entirely new lines of information relating to phenology, environment, behaviour and activity, from both existing and future footage. For example, Herrera, Baker, Sheaves, &Sheaves (2020) demonstrated the capabilities of image-based sampling and computer vision analysis to obtain ecological data on small invertebrate behaviour and bioturbation rate. There are many challenges associated with going from what is theoretically possible with DL to being able to use the technology in these new ways (e.g., using Long Short Term Memory and/or Generative Adversarial Networks methods to model and understand fish movement and swimming de-formations). However, by making sampling and analysis workflows available, and clearly explaining sampling and analysis assumptions(Herrera, Sheaves, Baker, & Sheaves, 2020), DL and computer vision approaches will likely be adopted by diverse researchers, catalysing new developments and improvements. Close partnerships between marine/freshwater ecologists, fisheries scientists, engineers and machine learning researchers will become increasingly important, as well as mutually beneficial (Weinstein, 2018). Furthermore, by marrying optimized ROVs with Big Data, environmental sensors and the Internet of Things, a range of new information streams for ecol -ogy, conservation and fisheries management will become possible, with collaboration in this space becoming fundamental. However, there will always be merit in at least some human viewing of footage. Underwater footage is a rich source of observational information. For any resulting data to be understood properly, domain experts, in this case ecologists, must engage in considered observation of footage, in order to understand the context of the data, make new observations and generate new hypotheses. Now that video data and computer vision are becoming common-place, researchers and practitioners are faced with various pathways for the organization of these new technologies (Wu, Hou, Zhu, Zhang,& Peha,

2001). In terms of the DL architecture itself, it is possible to use off-the-shelf solutions or to develop bespoke solutions optimized for particular situations. This flexibility has encouraged the creation of many in-house solutions for fish detection and identification. Different research domains will clearly require different DL tools, and however, there are advantages in standardizing practises and DL approaches across the community. Advancements can be shared between groups more easily, and common and robust DL system for fish detection and identification developed and tuned to particular domains. These can be made freely available and constantly updated. The extension of this logic calls for the use of standardized methods for the assessment and quality control of DL solutions and outputs, or the use of a single DL by multiple groups working on the same species. An important resource that should be shared across the community are curated video data-bases (Myers, Trevathan, & Atkinson, 2012), particularly labelled training image or video sets. This will allow for the rapid development of powerful DL and maintain consensus and agreement in identification between them. Additionally, the set of procedures used in augmentation during network training, many of which require prior knowledge, can be published for peer review and permanently attached to publications emanating from DL efforts. Unlike human processing, all prior "ex-pert knowledge" used to train DL in classification of different species and life stages can be made available through the publication of training image/video sets for peer review. Similarly, the criteria developed by DL to differentiate between species and life stages based on these im-ages can be extracted during post-processing and published as freely available CNN architecture. This makes all steps in video processing transparent and repeatable in a way that the "black-box" of human pro-cessing is not, eliminating much of the subjectivity that plagues ecological survey data. More broadly, one of the most important benefits of image and video sampling is the capacity to revisit raw data, make new observations or substantiate old ones, and perform new analyses under new paradigms, and this should be fostered throughout the community.

## References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Devin,M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

Adams, A. J., Wolfe, R. K., Kellison, G. T., & Victor, B. C. (2006). Patterns of juvenile habitat use and seasonality of settlement by permit,Trachinotus falcatus. Environmental Biology of Fishes, 75(2), 209–217.https://doi.org/10.1007/s10641-006-0013-5

Ayma, A., Aguzzi, J., Canals, M., Lastras, G., Bahamon, N., Mechó, A.,& Company, J. (2016). Comparison between ROV video and Agassiz trawl methods for sampling deep water fauna of submarine canyons in the Northwestern Mediterranean Sea with observations on behavioural reactions of target species. Deep Sea ResearchPart I: Oceanographic Research Papers, 114, 149–159. https://doi.org/10.1016/j.dsr.2016.05.013

Baker, R., & Minello, T. J. (2011). Trade-offs between gear selectivity and logistics when sampling nekton from shallow open water habitats: A gear comparison study. Gulf and Caribbean Research, 23(1), 37–48.https://doi.org/10.18785/gcr.2301.04

Baltz, D. M., Rakocinski, C., & Fleeger, J. W. (1993). Microhabitat use by marsh edge fishes in a Louisiana estuary. Environmental Biology of Fishes, 36(2), 109–126. https://doi.org/10.1007/BF00002790

Beukers, J. S., & Jones, G. P. (1998). Habitat complexity modifies the im -pact of piscivores on a coral reef fish population. Oecologia, 114(1),50–59. https://doi.org/10.1007/s004420050419

Boland, G., & Lewbel, G. (1986). The estimation of demersal fish densitiesin biological surveys using underwater television systems. Paper pre-sented at the OCEANS'86.

Bonvechio, K., Sawyers, R., Bitz, R., & Crawford, S. (2014). Use of mini-fyke nets for sampling shallow-water fish communities in Floridalakes. North American Journal of Fisheries Management, 34(4), 693–701. https://doi.org/10.1080/02755947.2014.901261

Bosch, N. E., Gonçalves, J. M., Tuya, F., & Erzini, K. (2017). Marinasas habitats for nearshore fish assemblages: Comparative analysis of underwater visual census, baited cameras and fishtraps. Scientia Marina, 81(2), 159–169. https://doi.org/10.3989/scimar.04540.20A

Bradford, M. J., & Cabana, G. (1997). Interannual variability in stage-specific survival rates and the causes of recruitment variation. R. C.Chambers & E. A. Trippel (Eds.), In Early life history and recruitment in fish populations (pp. 469–493). Berlin, Germany: Springer.

Bradley, M., Baker, R., Nagelkerken, I., & Sheaves, M. (2019). Context ismore important than habitat type in determining use by juvenile fish.Landscape Ecology, 34(2), 427–442. https://doi.org/10.1007/s10980-019-00781-3

Bradley, M., Baker, R., & Sheaves, M. (2017). Hidden components in tropical seascapes: Deep-estuary habitats support unique fish assemblages. Estuaries and Coasts, 40(4), 1195–1206.

Campbell, M. D., Salisbury, J., Caillouet, R., Driggers, W. B., & Kilfoil, J.(2018). Camera field-of-view and fish abundance estimation: A comparison of individual-based model output and empirical data. Journalof Experimental Marine Biology and Ecology, 501, 46–53. https://doi.org/10.1016/j.jembe.2018.01.004

Cappo, M., Harvey, E., Malcolm, H., & Speare, P. (2003). Potential ofvideo techniques to monitor diversity, abundance and size of fishing studies of marine protected areas. Aquatic Protected Areas-what Works Best and how do We know, 455–464.

Carassou, L., Mellin, C., & Ponton, D. (2009). Assessing the diversity and abundances of larvae and juveniles of coral reef fish: A synthesis ofsix sampling techniques. Biodiversity and Conservation, 18(2), 355.https://doi.org/10.1007/s10531-008-9492-3

Chollet, F. (2015). Keras. Retreived from https://keras.ioChollet, F. (2017). Xception: Deep learning with depthwise separable convolution. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251–1258).

Collins, S. F., Diana, M. J., Butler, S. E., & Wahl, D. H. (2017). A Comparison of sampling gears for capturing juvenile Silver Carp in river–flood-plain ecosystems. North American Journal of Fisheries Management,37(1), 94–100. https://doi.org/10.1080/02755947.2016.1240121

Collobert, R., Bengio, S., & Mariéthoz, J. (2002). Torch: a modular ma-chine learning software library.

Cooke, S. J., Niezgoda, G. H., Hanson, K. C., Suski, C. D., Phelan, F. J.,Tinline, R., & Philipp, D. P. (2005). Use of CDMA acoustic telemetry to document 3-D positions of fish: Relevance to the design and monitoring of aquatic protected areas. Marine Technology Society Journal,39(1), 31–41. https://doi.org/10.4031/002533205787521659

Cullen, D. W., & Stevens, B. G. (2017). Use of an underwater video system to record observations of black sea bass (Centropristis striata) in waters off the coast of Maryland. Fishery Bulletin, 115(3), 408–419.https://doi.org/10.7755/FB.115.3.10

Davies-Colley, R., & Smith, D. (2001). Turbidity, suspended sediment,and water clarity: A review 1. JAWRA Journal of the American Water Resources Association, 37(5), 1085–1101. https://doi.org/10.1111/j.1752-1688.2001.tb03624.x

Deyle, E., Schueller, A. M., Ye, H., Pao, G. M., & Sugihara, G. (2018).Ecosystem based forecasts of recruitment in two menhaden species.Fish and Fisheries, 19(5), 769–781. https://doi.org/10.1111/faf.12287

Duffy, E. J., Beauchamp, D. A., Sweeting, R. M., Beamish, R. J., & Brennan,J. S. (2010). Ontogenetic diet shifts of juvenile Chinook salmon in nearshore and offshore habitats of Puget Sound. Transactions of the American Fisheries Society, 139(3), 803–823. https://doi.org/10.1577/T08-244.1

Eggleston, D. (1995). Recruitment in Nassau grouper Epinephelus stria-tus: Post-settlement abundance, microhabitat features, and ontogenetic habitat shifts. Marine Ecology Progress Series. Oldendorf, 124(1),9–22. https://doi.org/10.3354/meps124009

Espino, F., González, J. A., Haroun, R., & Tuya, F. (2015). Abundance and biomass of the parrotfish Sparisoma cretense in seagrass meadows:Temporal and spatial differences between seagrass interiors and sea-grass adjacent to reefs. Environmental Biology of Fishes, 98(1), 121–133. https://doi.org/10.1007/s10641-014-0241-z

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fitzpatrick, C., McLean, D., & Harvey, E. S. (2013). Using artificial illumination to survey nocturnal reef fish. Fisheries Research, 146, 41–50.https://doi.org/10.1016/j.fishres.2013.03.016

Fukushima, K. (1980). Neocognitron: A self-organizing neural net-work model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36(4), 193–202. https://doi.org/10.1007/BF003 44251

Ghazilou, A., Shokri, M., & Gladstone, W. (2016). Animal v. plant based bait: Does the bait type affect census of fish assemblages and trophic groups by baited remote underwater video (BRUV) systems? Journal of Fish Biology, 88(5), 1731–174

Gibson, R. (1994). Impact of habitat quality and quantity on the recruit-ment of juvenile flatfishes. Netherlands Journal of Sea Research, 32(2),191–206. https://doi.org/10.1016/0077-7579(94)90040-X

Gillanders, B., Able, K. W., Brown, J. A., Eggleston, D. B., & Sheridan,P. F. (2003). Evidence of connectivity between juvenile and adult habitats for mobile marine fauna: An important component of nurseries. Marine Ecology-Progress Series, 247, 281–295. https://doi.org/10.3354/meps247281

Gwinn, D. C., Allen, M. S., & Rogers, M. W. (2010). Evaluation of procedures to reduce bias in fish growth parameter estimates resulting from size-selective sampling. Fisheries Research, 105(2), 75–79.https://doi.org/10.1016/j.fishres.2010.03.005

Hardie, S. A., Barmuta, L. A., & White, R. W. (2006). Comparison of dayand night fyke netting, electrofishing and snorkelling for monitor-ing a population of the threatened golden galaxias (Galaxias aura-tus). Hydrobiologia, 560(1), 145–158. https://doi.org/10.1007/s10750-005-9509-9

Hardinge, J., Harvey, E. S., Saunders, B. J., & Newman, S. J. (2013). A little bait goes a long way: The influence of bait quantity on a temperate fish assemblage sampled using stereo-BRUVs. Journal of ExperimentalMarine Biology and Ecology, 449, 250–260. https://doi.org/10.1016/j.jembe.2013.09.018

Harris, G. P., & Heathwaite, A. L. (2012). Why is achieving good ecological outcomes in rivers so difficult? Freshwater Biology, 57, 91–107.https://doi.org/10.1111/j.1365-2427.2011.02640.x

Harvey, E. S., Newman, S. J., McLean, D. L., Cappo, M., Meeuwig, J. J., & Skepper, C. L. (2012). Comparison of the relative efficiencies of ste-reo-BRUVs and traps for sampling tropical continental shelf demersal fishes. Fisheries Research, 125, 108–120. https://doi.org/10.1016/j.fishres.2012.01.026

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. (pp. 770–778). In Proceedings of the IEEE conference on computer vision and pattern recognition.

Heagney, E. C., Lynch, T. P., Babcock, R. C., & Suthers, I. M. (2007).Pelagic fish assemblages assessed using mid-water baited video:Standardising fish counts using bait plume size. Marine Ecology Progress Series, 350, 255–266. https://doi.org/10.3354/meps07193

Herrera, C., Sheaves, J., Baker, R., & Sheaves, M. (2020). A computer vision approach for studying fossorial and cryptic crabs. bioRxiv.https://doi.org/10.1101/2020.05.11.085803

Hung, C.-C., Tsao, S.-C., Huang, K.-H., Jang, J.-P., Chang, H.-K., & Dobbs,F. C. (2016). A highly sensitive underwater video system for use inturbid aquaculture ponds. Scientific Reports, 6, 31810. https://doi.org/10.1038/srep31810

Jůza, T., & Kubečka, J. (2007). The efficiency of three fry trawls for sampling the freshwater pelagic fry community. Fisheries Research, 85(3),285–290. https://doi.org/10.1038/srep31810

Kanou, K., Sano, M., & Kohno, H. (2004). Catch efficiency of a small seine for benthic juveniles of the yellowfin goby Acanthogobius flavimanuson a tidal mudflat. Ichthyological Research, 51(4), 374–376. https://doi.org/10.1007/s10228-004-0231-9

Kaur, D., & Kaur, Y. (2014). Various image segmentation techniques: A re-view. International Journal of Computer Science and Mobile Computing,3(5), 809–814.

Keil, M., Depledge, G., & Rai, A. (2007). Escalation: The role of problem recognition and cognitive bias. Decision Sciences, 38(3), 391–421.https://doi.org/10.1111/j.1540-5915.2007.00164.x

Klages, J., Broad, A., Kelaher, B. P., & Davis, A. (2014). The influence of gummy sharks, *Mustelus antarcticus*, on observed fish assemblage structure. Environmental Biology of Fishes, 97(2), 215–222. https://doi.org/10.1007/s10641-013-0138-2

Konovalov, D. A., Saleh, A., Bradley, M., Sankupellay, M., Marini,S., Sheaves, M. (2019). Underwater Fish Detection with Weak Multi-Domain Supervision. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1–8). Budapest, Hungary: IEEE.

Konovalov, D. A., Swinhoe, N., Efremova, D. B., Birtles, R. A., Kusetic,M., Hillcoat, S., … Sheaves, M. (2020). Automatic sorting of dwarf Minke whale underwater images. Information, 11(4), 200. https://doi.org/10.3390/info11040200

Kubečka, J., Godø, O. R., Hickley, P., Prchalová, M., Říha, M., Rudstam, L.,& Welcomme, R. (2012). Fish sampling with active methods. FisheriesResearch, 123, 1–3. https://doi.org/10.1016/j.fishres.2011.11.013

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature,521(7553), 436. https://doi.org/10.1038/nature14539

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE,86(11), 2278–2324. https://doi.org/10.1109/5.726791

Levin, P. S., & Stunz, G. W. (2005). Habitat triage for exploited fishes:Can we identify essential "Essential Fish Habitat?" Estuarine,Coastal and Shelf Science, 64(1), 70–78. https://doi.org/10.1016/j.ecss.2005.02.007

Li, X., Shang, M., Qin, H., & Chen, L. (2015). Fast accurate fish detection and recognition of underwater images with fast r-cnn. In OCEANS2015-MTS/IEEE Washington (pp. 1–5). Washington , D.C.: IEEE.

Lindfield, S. J., Harvey, E. S., McIlwain, J. L., & Halford, A. R. (2014).Silent fish surveys: Bubble-free diving highlights inaccuracies associated with SCUBA-based surveys in heavily fished areas. Methods in Ecology and Evolution, 5(10), 1061–1069. https://doi.org/10.1111/2041-210X.12262

Logan, J. M., Young, M. A., Harvey, E. S., Schimel, A. C., & Ierodiaconou,D. (2017). Combining underwater video methods improves effectiveness of demersal fish assemblage surveys across habitats. Marine Ecology Progress Series, 582, 181–200. https://doi.org/10.3354/meps12326

Magnuson, J. J. (1991). Fish and fisheries ecology. Ecological Applications,1(1), 13–26. https://doi.org/10.2307/1941844

Mallet, D., & Pelletier, D. (2014). Underwater video techniques for ob-serving coastal marine biodiversity: A review of sixty years of publications (1952–2012). Fisheries Research, 154, 44–62. https://doi.org/10.1016/j.fishres.2014.01.019

Mandal, R., Connolly, R. M., Schlacher, T. A., & Stantic, B. (2018).Assessing fish abundance from underwater video using deep neural networks. In Paper presented at the 2018 International Joint Conference on Neural Networks (IJCNN) (pp. 1–6). IEEE. https://doi.org/10.1109/IJCNN.2018.8489482

Matabos, M., Hoeberechts, M., Doya, C., Aguzzi, J., Nephin, J., Reimchen,T. E., … Juniper, S. K. (2017). Expert, Crowd, Students or Algorithm:Who holds the key to deep-sea imagery 'big data'processing Methods in Ecology and Evolution, 8(8), 996–1004. https://doi.org/10.1111/2041-210X.12746

Matai, J., Kastner, R., Cutter, G. R., & Demer, D. A. (2012). Automated techniques for detection and recognition of fishes using computervision algorithms. In Report of the National Marine Fisheries Service Automated Image Processing Workshop (pp. 35–37).

Merilä, J. (2015). Baiting improves CPUE in nine-spined stickleback(Pungitius pungitius) minnow trap fishery. Ecology and Evolution, 5(17),3737–3742.

Meuthen, D., Baldauf, S. A., Bakker, T. C., & Thünken, T. (2018). Neglected patterns of variation in phenotypic plasticity: Age-and sex-specific antipredator plasticity in a cichlid fish. The American Naturalist,191(4), 475–490. https://doi.org/10.1086/696264

Mobley, C. D. (2001). Radiative transfer in the ocean. Encyclopedia of OceanSciences, 2321–2330. https://doi.org/10.1006/rwos.2001.0469

Musick, J. A., Harbin, M. M., Berkeley, S. A., Burgess, G. H., Eklund, A. M.,Findley, L., … Wright, S. G. (2000). Marine, estuarine, and diadromousfish stocks at risk of extinction in North America (exclusive of Pacificsalmonids). Fisheries, 25(11), 6–30. https://doi.org/10.1577/1548-8446(2000)025<0006:MEADFS>2.0.CO;2

Myers, T., Trevathan, J., & Atkinson, I. (2012). The tropical data hub: A virtual research environment for tropical science knowledge and discovery. International Journal of Sustainability Education, 8, 11–27.

Norcross, B. L., & Mueter, F.-J. (1999). The use of an ROV in the study of juvenile flatfish. Fisheries Research, 39(3), 241–251. https://doi.org/10.1016/S0165-7836(98)00200 -8

Nyboer, E. A., Gray, S. M., & Chapman, L. J. (2014). A colourful youth: Ontogenetic colour change is habitat specific in the invasive Nileperch. Hydrobiologia, 738(1), 221–234. https://doi.org/10.1007/s10750-014-1961-y

Oozeki, Y., Hu, F., Kubota, H., Sugisaki, H., & Kimura, R. (2004). Newlydesigned quantitative frame trawl for sampling larval and juvenile pelagic fish. Fisheries Science, 70(2), 223–232. https://doi.org/10.1111/j.1444-2906.2003.00795.x

Oozeki, Y., Hu, F., Tomatsu, C., & Kubota, H. (2012). Development of anew multiple sampling trawl with autonomous opening/closing net control system for sampling juvenile pelagic fish. Deep Sea ResearchPart I: Oceanographic Research Papers, 61, 100–108. https://doi.org/10.1016/j.dsr.2011.12.001

OpenCV. (2015). OpenCV. Open Source Computer Vision Library.

Paradis, Y., Mingelbier, M., Brodeur, P., & Magnan, P. (2008). Comparisons of catch and precision of pop nets, push nets, and seines for samplinglarval and juvenile yellow perch. North American Journal of FisheriesManagement, 28(5), 1554–1562. https://doi.org/10.1577/M07-122.1

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel,O., … Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python.Journal of Machine Learning Research, 12, 2825–2830.

Peterman, R. M., Bradford, M. J., Lo, N. C., & Methot, R. D. (1988).Contribution of early life stages to interannual variability in recruitment of northern anchovy (Engraulis mordax). Canadian Journal of Fisheries and Aquatic Sciences, 45(1), 8–16. https://doi.org/10.1139/f88-002

Poletto, J. B., Martin, B., Danner, E., Baird, S. E., Cocherell, D. E., Hamda,N., … Fangue, N. A. (2018). Assessment of multiple stressors on the growth of larval green sturgeon Acipenser medirostris: Implications for recruitment of early life-history stages. Journal of Fish Biology,93(5), 952–960. https://doi.org/10.1111/jfb.13805

Rattray, A., Ierodiaconou, D., Monk, J., Laurenson, L., & Kennedy, P.(2014). Quantification of spatial and thematic uncertainty in the application of underwater video for benthic habitat mapping. Marine Geodesy, 37(3), 315–336. https://doi.org/10.1080/01490419.2013.877105

Rice, J. A., Crowder, L. B., & Marschall, E. A. (1997). Predation on juvenile fishes: Dynamic interactions between size-structured predators and prey. R. C. Chambers & E. A. Trippel (Eds.), In Early life history and recruitment in fish populations (pp. 333–356). Berlin, Germany: Springer.

Rooper, C. N., Boldt, J. L., & Zimmermann, M. (2007). An assessment of juvenile Pacific ocean perch (Sebastes alutus) habitat use in a deep-water nursery. Estuarine, Coastal and Shelf Science, 75(3), 371–380.https://doi.org/10.1016/j.ecss.2007.05.006

Rotherham, D., Johnson, D. D., Kesby, C. L., & Gray, C. A. (2012). Samplingestuarine fish and invertebrates with a beam trawl provides a different picture of populations and assemblages than multi-mesh gill-nets. Fisheries Research, 123, 49–55. https://doi.org/10.1016/j.fishres.2011.11.019

Rozas, L. P., & Minello, T. J. (2015). Small-scale nekton density and growthpatterns across a saltmarsh landscape in Barataria Bay, Louisiana.Estuaries and Coasts, 38(6), 2000–2018. https://doi.org/10.1007/s12237-015-9945-3

Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., & Harvey,E. (2016). Fish species classification in unconstrained underwater environments based on deep learning. Limnology and Oceanography:Methods, 14(9), 570–585. https://doi.org/10.1002/lom3.10113

Schobernd, Z. H., Bacheler, N. M., & Conn, P. B. (2013). Examining theutility of alternative video monitoring metrics for indexing reef fish abundance. Canadian Journal of Fisheries and Aquatic Sciences, 71(3),464–471. https://doi.org/10.1139/cjfas-2013-0086

Shafait, F., Mian, A., Shortis, M., Ghanem, B., Culverhouse, P. F.,Edgington, D., … Harvey, E. S. (2016). Fish identification from videos captured in uncontrolled underwater environments. ICES Journal of Marine Science, 73(10), 2737–2746. https://doi.org/10.1093/icesjms/fsw106

Sheaves, M. J. (1995). Effect of design modifications and soak time variations on Antillean-Z fish trap performance in a tropical estuary. Bulletin of Marine Science, 56, 475–489.

Sheaves, M. (1996a). Do spatial differences in the abundance of two serranid fishes in estuaries of tropical Australia reflect long term salinity patterns  Marine Ecology-Progress Series, 137(1–3), 39–49. https://doi.org/10.3354/meps137039

Sheaves, M. (1996b). Habitat-specific distributions of some fishes in a tropical estuary. Marine and Freshwater Research, 47(6), 827–830.https://doi.org/10.1071/MF9960827

Sheaves, M. (1998). Spatial patterns in estuarine fish faunas in tropical Queensland: A reflection of interaction between long-term physical and biological processes? Marine and Freshwater Research, 49(1),31–40. https://doi.org/10.1071/mf97019

Sheaves, M., Johnston, R., & Abrantes, K. (2007). Fish fauna of dry tropical and subtropical estuarine floodplain wetlands. Marine and Freshwater Research, 58(10), 931–943. https://doi.org/10.1071/mf06246

Sheaves, M., Johnston, R., & Baker, R. (2016). Use of mangroves by fish:New insights from in-forest videos. Marine Ecology Progress Series,549, 167–182. https://doi.org/10.3354/meps11690

Sheaves, M., Johnston, R., & Connolly, R. M. (2010). Temporal dynamics of fish assemblages of natural and artificial tropical estuaries. Marine Ecology Progress Series, 410, 143–156. https://doi.org/10.3354/meps0 8655

Sheaves, M., Johnston, R., & Connolly, R. (2012). Fish assemblages as indicators of estuary ecosystem health. Wetlands Ecology and Management, 20, 477–490. https://doi.org/10.1007/s11273-012-9270-6

Sheaves, M., Johnston, R., Johnson, A., Baker, R., & Connolly, R. (2013).Nursery function drives temporal patterns in fish assemblage structure in four tropical estuaries. Estuaries and Coasts, 36, 893–905.https://doi.org/10.1007/s12237-013-9610-7

Stein, W. III, Smith, P. W., & Smith, G. (2014). The cast net: An overlooked sampling gear. Marine and Coastal Fisheries, 6(1), 12–19. https://doi.org/10.1080/19425120.2013.864737

Stevens, P. W. (2006). Sampling fish communities in saltmarsh impoundments in the northern Indian River Lagoon, Florida: Cast net and cul-vert trap gear testing. Florida Scientist, 135–147.

Stoner, A., Laurel, B., & Hurst, T. (2008). Using a baited camera to assess relative abundance of juvenile Pacific cod: Field and laboratory trials. Journal of Experimental Marine Biology and Ecology, 354(2), 202–211.https://doi.org/10.1016/j.jembe.2007.11.008

Storbeck, F., & Daan, B. (2001). Fish species recognition using computer vision and a neural network. Fisheries Research, 51(1), 11–15. https://doi.org/10.1016/S0165 -7836(00)00254-X

Struthers, D. P., Danylchuk, A. J., Wilson, A. D., & Cooke, S. J. (2015).Action cameras: Bringing aquatic and fisheries research into view. Fisheries, 40(10), 502–512. https://doi.org/10.1080/03632415.2015.1082472

Tan, M., & Le, Q. V. (2019). Efficient net: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946.

Taylor, M. D., Baker, J., & Suthers, I. M. (2013). Tidal currents, sampling effort and baited remote underwater video (BRUV) surveys: Are we drawing the right conclusions? Fisheries Research, 140, 96–104.https://doi.org/10.1016/j.fishres.2012.12.013

Theano Development Team. (2016). Theano: A Python framework for fast computation of mathematical expressions. arXiv. http://arxiv.org/abs/1605.02688

Van Der Veer, H. W., Witte, J. I., Beumkes, H. A., Dapper, R., Jongejan, W.P., & Van Der Meer, J. (1992). Intertidal fish traps as a tool to study long-term trends in juvenile flatfish populations. Netherlands Journal of Sea Research, 29(1–3), 119–126. https://doi.org/10.1016/0077-7579(92)90013-5

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F.,Warner, J. D., Yager, N., … Yu, T. (2014). scikit-image: Image process -ing in Python. PeerJ, 2, e453. https://doi.org/10.7717/peerj.453

Villon, S., Chaumont, M., Subsol, G., Villéger, S., Claverie, T., & Mouillot,D. (2016). Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between Deep Learning and HOG+ SVM methods. In International Conference on Advanced Concepts for Intelligent Vision Systems (pp. 160–171). Cham, Switzerland: Springer.

Wang, M., Huang, Z., Shi, F., & Wang, W. (2009). Are vegetated areas of mangroves attractive to juvenile and small fish? The case of Dongzhaigang Bay, Hainan Island, China. Estuarine, Coastal and Shelf Science, 85(2), 208–216. https://doi.org/10.1016/j.ecss.2009.08.022

Warnock, B., Harvey, E. S., & Newman, S. J. (2016). Remote drifted and diver operated stereo–video systems: A comparison from tropical and temperate reef fish assemblages. Journal of Experimental Marine Biology and Ecology, 478, 45–53. https://doi.org/10.1016/j.jembe.2016.02.002

Weinstein, B. G. (2018). A computer vision for animal ecology. Journal of Animal Ecology, 87(3), 533–545. https://doi.org/10.1111/1365-2656.12780

Wu, D., Hou, Y. T., Zhu, W., Zhang, Y.-Q., & Peha, J. M. (2001). Streaming video over the Internet: Approaches and directions. IEEE Transactions on Circuits and Systems for Video Technology, 11(3), 282–300. https://doi.org/10.1109/76.911156

Zhang, F., Reid, K. B., & Nudds, T. D. (2017). Relative effects of biotic and abiotic factors during early life history on recruitment dynamics: A case study. Canadian Journal of Fisheries and Aquatic Sciences, 74(7),1125–1134

TA B L E 2 Table of considerations for the application of DL to RUVsComponent ConsiderationInfrastructure for data storage and processing:Footage should be stored in a medium that protects data and allowsfor easy retrieval, therefore minimizing the cost of sharing andreusing it. Data protection implies physically safeguarding data fromdamage, but also ensuring it lasts into the future in mediums andformats that can be later consulted. Processing of footage is faster ifrun on appropriate hardware such as Graphic Processor Units (GPUs)Meticulously curated metadata improves processing data fetch into theAI algorithms and facilitates searching and finding data of interestWith the advent of high-quality image sensors with high resolution(e.g., 4K), storage and processing are something that should bethoughtfully considered. When used for large spatio-temporal scalemonitoring, the universally accessible nature of video technology ischallenged by the expense of storing and processing big volumes ofraw dataLabelling:A data set of training images/videos in which the objects of interestare identified (and preferably verified), is required for trainingsupervised DL/CNN. Labelling is the most time-consuming task onthe analysis cycle and requires prior expert knowledge.There are three main types of labelling: image/video-level, boundingbox and mask. As their names indicate, image-level label specifiesif the target fish present or not anywhere in the image (or video forthe video-level), a bounding box label enclose the target fish inside abox or region of interest, while a mask label segments the target fish(crop around fish contour) from the background. CNN models andarchitectures are designed to work with either of these labels' types.Lastly, there is an increasing use and demand of mask label data. Thetype of labelling would ultimately depend on the research purposeand CNN type used, where the image-level labelling is the easiest/fastest (and hence least costly) to curate, followed the boundingboxes and then by the most time-consuming segmentation masksIt is important to remember that generally supervised DL/CNN willlearn from the training/validation data set, and then when presentedwith new data will make predictions based on what it has learnt.For this reason, and depending on the specific DL/CNN task, datacollection should include samples of most of the potential casesexpected. For detecting different life stages of fish, obtaining highlyaccurate training images is crucial. To ground truth images of differentlife stages/size-classes, ideally fish should be captured, imagedthroughout their growth from early-juvenile to adult, then euthanizedand aged to determine an accurate age estimate for each set of imagesthroughout the life cycle. Otherwise, the DL/CNN will inherent anyinconsistencies and biases from subjective human identificationAugmentation:Data augmentation refers to the process of applying a series ofrandom changes in the appearance of the images set. Changesinclude rotation, warping, scaling, contrast and many more. Thus,a given image set can be augmented by randomly applying thesetransformations during training (or validation or testing). Artificiallycreating variation is useful to increase the CNN generalizationcapabilityDepending on the

classification task, researchers can select whichmodifications can be appropriate to generate variation and improvemodel generalization, and which ones should be avoided. Themagnitude of change can also be restricted. Augmentation must besevere enough to give flexibility to the network, but not so great thatit leads to misidentificationNetwork selection:Depending on the topology of the neural layers used, Deep Learningalgorithms are classified according to their different known CNNarchitectures (e.g., ResNet-50 He, Zhang, Ren, & Sun, 2016), whichbalance accuracy against speed in different ways. Networks can alsobe individually customized to improve performanceThe advantages of customization may be outweighed by thereproducibility and transparency in using publicly available networks“off the shelf” (discussed above)Training, validation and testing—Learning:The network trains on a set of training images, then it can be tunedusing validation images, re-trains based on these results, and testsitself on a separate set of testing imagesThe behaviour of the network, including levels of accuracy required,can be set by researchers. Here, important decisions must bemade about the relative acceptability of false positives versus falsenegatives. In general, either false positives or false negatives could beeasily reduce but not both (Fawcett, 2006)Observing the prediction model:Once the prediction model has been developed, the criteria usedby the AI to distinguish between species and life stages can beextracted, summarized and presented.Here, researchers can determine whether the criteria developed by theAI are biologically/ecologically relevantApplying the DL/CNN model—automated processing of data:If the DL/CNN model achieved acceptable performance on test andvalidation images/videos, it can now be used to automatically processunlabelled videoHere, a balance must be reached between comprehensive processingof the video sample and acceptable processing time. Analysis speedis needed to be at least at the video's fps or faster (i.e., 30 frames of30 fps footage then would take <1 s to process). For the identificationof fish, it may be acceptable not to process every frame, as fish usuallyinhabit the frame for >1 s (i.e. could be detected in at least 30 frames)