Electronic Monitoring: Best Practices for Automation

Volume II • 2024

1.0 Table of Contents

New England Marine Monitoring productops OnDeck fisheries AI

2.0 Introduction

In 2020, the Gulf of Maine Research Institute (GMRI) and CVision AI published "Electronic Monitoring: Best Practices for Automation."[1] Since then, use of Electronic Monitoring (EM) in fisheries management has continued to grow around the world. In addition to pilot studies, a global expansion of operational programs is now well underway. In New England, two EM programs have been federally approved as an alternative to human at-sea monitors (ASM) and are fully operational. New pilots have also have begun exploring how automation can make them more efficient. Globally, as EM expands, significant public and private resources are being dedicated to exploring how automation can reduce the cost of fishery monitoring. The promise of automation and the use of artificial intelligence (AI) has led many to see it as the holy grail of fisheries monitoring, offering low cost, easy to implement solutions that can quickly provide information for science and management. However, the current EM market reflects a different reality where few EM programs are currently experiencing the benefits of automation. This document is intended to provide insight and lessons learned regarding EM automation to help catalyze its broader use.

New England Marine Monitoring (NEMM), productOps, and OnDeck Fisheries AI have collaborated to update the original 2020 report with new learnings and fresh perspectives. It includes content from the previous report with targeted updates and new insights and lessons from applying automation techniques in the field. The update includes a detailed case study from an automation project in New England to help provide real world context. It also provides an overview of relevant technological advancements as well as considerations for regulatory bodies seeking to incorporate automation into EM programs. Successful AI development and implementation requires coordination between multiple disciplines. With that in mind, this report has a broad target audience. Certain sections may not apply to every reader and thus a detailed table of contents is included so each reader can focus on topics that are most relevant to them but also have the benefit of reviewing topics that may be outside of their wheelhouse. The project team hopes this will promote the cross-discipline learning and collaboration that is necessary for success in automation. Readers should be advised to view the technical documentation as a reference specific to automation and not as prescriptive guidance for all EM program implementation.

It's important to note that this document is focused specifically on machine learning-based automation applied to video-based technologies for EM as opposed to the more general use of AI in fisheries management. It is limited to video data and does not explore the use of gear sensors or GPS for automation. This document is also geared towards trawl fisheries although many of the concepts apply to other gear types. The authors acknowledge this limited scope and hope to collaborate with others in the future to expand the report's utility.

2.1 Report Layout

The report is a combination of text from the previous report with targeted updates, new sections, and the real-world application of concepts discussed throughout the report. Text from the original 2020 report was included as it remains relevant and provides the background needed to contextualize the updates and case studies.

1 B. Woodward, M. Hager, H. Cronin, 2020, 'Electronic Monitoring: Best Practices for Automation', Gulf of Maine Research Institute, CVision AI.

2.2 An Introduction to the Case Study

In 2021, New England Marine Monitoring (NEMM) conducted a project titled "Integrating Artificial Intelligence Algorithms to Strengthen Electronic Monitoring" funded by the National Fish and Wildlife Foundation (NFWF). Throughout this report we present lessons learned from this project as case studies to demonstrate the concepts discussed in a real-world setting.

The objective of this NFWF project was to integrate complementary AI developed for two different trawl vessels into a multistage workflow. Both vessels were <19 meters long and targeted groundfish in the Northeast Atlantic. The vessels participated in different EM programs. One vessel was in the New England Audit Project (Audit) and the other was in the Maximized Retention Project (MREM)[2]. Both projects are focused on collecting groundfish discard data. In the Audit project participants measure their undersized groundfish discards at a designated measuring location before discarding them. In the MREM program vessels are not allowed to discard undersized groundfish at sea. MREM vessels retain all undersized groundfish and meet a dockside monitor at the end of their trip.

Audit camera arrays have a dedicated measuring camera to collect information on undersized groundfish. Audit measuring cameras are positioned to identify and record a measurement in centimeters. Measurement of groundfish discards is not required in the MREM program. Camera arrays on MREM vessels are configured to identify any groundfish discarded at sea. The impact the differences in camera setup and fish handling protocols had on algorithm integration and performance are explored in each case study.

Three AI algorithms were developed as part of previous efforts. Activity recognition (AR) and object tracking algorithms had been developed for the MREM vessel. A species recognition algorithm was being trained with data from the Audit vessel. These algorithms were selected for integration because the project team felt they had the highest probability of reducing human video review time.

The project team integrated the 3 algorithms into a multistage workflow. The goal was to create a workflow that could:

1.) Use the AR algorithm to flag to the reviewer when there was activity to review.

2.) Use the object tracking algorithm to detect discards.

3.) Apply the species recognition algorithm to the discards identified by the object tracking tool to determine if the discard was a groundfish.

Throughout the project the team customized video review platforms, developed AI assisted review protocols and conducted AI assisted review.. In the process, the team applied many of the concepts and ideas discussed in this report. The application of these concepts and lessons learned will be shared in case study sections throughout this document.

---

2 "Northeast Groundfish Monitoring Program" Accessed: Feb. 01, 2024,. [Online] Available: https://www.fisheries.noaa.gov/new-england-mid-atlantic/commercial-fishing/northeast-groundfish-monitoring-program

3.0 System Level Design

3.1 Summary

The first step towards understanding the utility of automation in an EM program is to evaluate the program objectives and existing monitoring standards for the fishery. Because EM programs differ in data collection objectives, time and cost burdens and data standards, automation applications will vary across programs. The scope and level of detail captured by footage of tasks intended for automation will influence both EM system installation and algorithm performance. Cameras should be installed in locations which capture the best compromise between the scope of the activity, the necessary level of detail, environmental and structural factors, and camera specifications.

3.2 Broad Fishery and Program Level Considerations

When evaluating a fishery to determine the suitability of introducing automation, it's important to first holistically evaluate the target fishery and the objectives of the EM program. It's critical to take into consideration factors such as the level of fishing effort, the economics of the fishery, the stage of EM program development, the cost structure of the program, and relevant regulatory requirements. Most parties pursuing automation will be looking to achieve a return on investment (ROI) in terms of cost savings, time savings, improved data, program feasibility, or combination of these factors. Identifying the parameters that may impact ROI is a suggested first step in understanding how automation may improve the efficiency of an EM program. Each of these common parameters are discussed below.

3.2.1 Fishery Effort

The number of vessels participating in an EM program is an important factor to consider when identifying opportunities for automation. The number of vessels may be an important indicator of the relative cost benefits successful automation development can provide, given that development costs can be shared across vessels. This is especially true in fisheries with vessels that share a common layout, operating procedures, and catch composition. Larger fleets may be able to better justify the initial investment into automation.

Additionally, understanding how much each vessel fishes or the number of fishing days per vessel is important. This will inform how well any automation will need to generalize, or function under a broad range of circumstances. In fisheries with lower effort per vessel, it is advisable to pursue generalized automation approaches in order to share as much of the costs among the full range of vessels. With fisheries with a high level of effort per vessel, it may be worth engaging in more vessel-specific development as the benefits of more customized automations are worth the higher per vessel development costs. This consideration may impact what type of automation tool is most useful and affordable.

Lastly, vessel-to-vessel structural and operational differences, including gear type and species composition are key factors to consider. Automation is most valuable when it has the potential to scale. However, if the tasks being automated are too different across the fleet, scaling may prove difficult. In some cases, successful automation may require vessel-to-vessel customizations and the associated development costs should be weighed when considering investing in automation.

3.2.2 Developmental stage

The maturity of an EM program should be considered when exploring the potential for automation. Nascent EM programs tend to undergo large-scale changes as they transition from conception, to the pilot stage, to implementation. Data elements may be added or removed, and data collection methodologies may be adapted. It is important to ensure that firm program guidelines are in place prior to making major investments in automation as design shifts can make expensive automation development obsolete. At any stage, when investing in a tool that is focused on a specific data element, it's important to consider if that element could change. The potential evolution of a program can be difficult to predict and having detailed knowledge of the region and program's history can be vital. Developers from outside the target region may want to collaborate with local groups to gain such insight. As demonstrated in Case Study 1, sometimes even changes that are widely accepted as program improvements may render an AI tool obsolete.

3.2.2.1 CASE STUDY: NFWF Case Study 1

A significant challenge the project team encountered in developing AI tools for the Northeast US groundfish project was the requirement to document fish that interacted with fishing gear but were discarded before coming aboard. These were called "drop-offs". An example of a drop-off would be a flounder falling out of the fishing gear as a net came aboard. This was a challenging discard to reliably capture with AI. The project team analyzed video data on drop-offs and noted there was often high occlusion from the fishing gear, very few frames of video to identify drop-offs, and a low pixel density on drop-off fish due to distance from the camera. Instead of putting significant resources into the development of AI tools to identify drop-offs, video analysts manually watched the hauling and setting of gear to document drop-offs. AI development effort was instead focused on fish that came all the way onboard the vessel, presenting an easier case for automation. After the completion of the project the requirement to document drop-offs was eliminated via a program protocol change. If significant resources had been put into developing an algorithm to automate the review of drop-offs, it would have been quickly made obsolete due to protocol changes in the fishery.

3.2.3 Economics of fishery

The economics of a fishery will often dictate the amount of EM automation development it can afford. Ex-vessel value is a good starting point to understanding the dollar value of commercial landing in a fishery. An understanding of net revenues and profitability is ideal. A stepped approach where a traditional EM program based on human video review is created first and then enhanced with future automation can be a successful model. In this stepped approach, managers can gain firsthand experience with the fishery and use the experience to maximize the impact of automation development. With an existing EM program in place, many fisheries have received grant funds to help cover the upfront cost of automation development. In a fishery with poor economics it may be difficult to afford automation development, therefore, it's important to set expectations appropriately and not overpromise the level of cost savings expected or the timeline in which they can be achieved.

3.2.4 Cost structure

Understanding the cost structure of an EM program can help target automation efforts. It's important to examine if the major cost drivers are well suited for automation. Reducing video review cost is often a target, however, it is not always the biggest driver, varying from 2.5 - 60% of overall program costs.[3] Instead, a program may have large data storage costs and thus automation investments could be well spent targeting data volume rather than review time. Understanding the cost structure of a program is helpful to determine an AI tool's potential impact on the overall cost of the program and what tool might provide the most savings and be worth investing in.

3.2.5 Regulatory Requirements

Understanding the path for approval of automated data collection and analysis is key. The pathway to implementing automated review in an experimental pilot project will likely be very different from a project used for fisheries regulation by a government agency. A working tool may not provide benefits due to regulatory factors that don't allow it to be implemented into a program. It's best practice to engage with stakeholders to consider the following items:

- Identify any approvals needed from a government or other entity.
- Is there a process for tool approval? If so, identify the steps, timeline, and costs.
- Identify any regulations in place that may limit your ability to innovate in the EM program.
- Identify any differences between the AI derived data product and the existing data products and confirm the program will be able to utilize the new products.
- Identify the performance metrics for the tool such as required precision and recall requirements.
- Identify how tool performance will be evaluated.
- Understand the likely evolution of the EM program and assess if future regulations are likely to render the tool less impactful.

---

3 J. Pierre et al., 2022, "How much is enough? Review optimization methods to deliver best value from electronic monitoring of commercial fisheries." Accessed: Feb. 01, 2024. [Online] Available: https://iattc.org/GetAttachment/6fa6c7a9-18f2-4aa5-98cb-df036ccd1c4b/WSEMS-04-INF_Pew-Project---How-much-is-enough.pdf

3.2.6 Return on Investment

The most important consideration when assessing the value of a potential EM automation is the return on investment (ROI). The rationale for investing in AI development and implementation is often cost savings. However, those benefits have rarely been realized. In fact, 85% of algorithm projects don't make it into production as they are too costly, too complex, and/or they take too long to develop.[4] It is common to see small AI projects run by academic institutions that target new ideas or programs. These are often promising and attractive projects because they tend to explore valuable automation tasks that may be too risky for private equity parties to heavily invest in. However, these proof-of-concept projects are not usually built for large scale applications and have little viability to operationalize or scale. This can create "pilot fatigue" in industry stakeholders and impact their willingness to invest in automation. Some efforts have less of a focus on financial ROI and are focused on collecting novel data. Even efforts with non-monetary drivers should consider ROI to ensure the effort expended has the opportunity to create the minimum desired outcome.

It's critical to conduct a thorough analysis of the ROI for automation in the specific fishery and EM program in question. Short term grant funding is available for exciting ideas in the automation arena and may not require a full ROI analysis. However, to deliver the desired impact to the target fishery, it's important to assess if the automation approach is practical or sustainable in the long term. A significant hurdle in assessing the ROI for automation is that EM is a relatively young commercial environment that is subject to the extreme variability of the commercial fishing business. Fisheries are regularly open and shut down, vessels enter and exit EM programs, and unforeseen regulations may increase or decrease fishing effort. These variables make estimating ROI more difficult than other markets in the automation space. Given this, practitioners should envision the full life cycle of their automation project and estimate the ROI to the extent possible. Approaches such as Multiple on Invested Capital (MOIC), Net Present Value (NPV) or Internal Rate of Return (IRR), and payback period may be useful in modeling the ROI.

Given the time and resources needed to create, train, assess functionality, and incorporate the tool into the overall program workflow, ROI in the EM automation space can be challenging to achieve. Significant upfront investment often takes many years to recoup. This may be due to development and testing delays, regulatory approval timelines, or limited fishing revenue. For most EM automation projects, cash outlays are large and immediate, and inflows are incremental and come in over a long period of time. Due to the inherent challenges in achieving ROI for these types of projects, it is best practice to put effort into modeling and planning before development gets underway. In addition, it may also be helpful to communicate and collaborate with other stakeholders who have worked on similar automation efforts to ensure the project is drawing upon prior experience. Scoping automation efforts with regulatory bodies in the pre-development stage is also recommended to ensure the tool meets their standards for operational use.

---

4 1st Global Artificial Intelligence in Fisheries Monitoring Summit Report, Pew Charitable Trusts, 2023

One particularly important component of such an ROI analysis should be video review rates. In many EM projects, the percentage of video that is reviewed can change throughout the lifetime of the project. In the early phases of an EM project the review rate may be near 100%[5]. After a project is stable, it is common for the review rate to drop significantly to limit review expense[6]. If an AI tool is only used on this reduced number of reviewed trips, the ROI may be significantly prolonged.

Below we present a non-exhaustive list of items that should be included when assessing automation development costs.

1. Determine the goals and definition of success for the automation project.
2. Identify the hardware and cloud platforms required to run the model.
3. Assess operational variables such as the gear type, vessel types, and number of potential participants in an EM program.
4. See how much training data is available to your team. Decide if the amount possessed is sufficient and plan how to get more if needed.
5. Research existing projects and see if they can contribute to the project. For example, there may be open image libraries that can help supplement training data or existing models that provide a good starting point. Evaluate the cost of licensing any previously developed models or impact of using open-source resources.
6. Estimate the development cost of the model. Set targets for a proof of concept and operational model.
7. Project the cost of evaluating, maintaining, and operating the model.

This is intended as a starting point and each effort should explore other potential cost categories.

3.3 Program Design: Data Collection

Across EM programs, data collection requirements vary according to a variety of factors. Program type, fishery, fishing method, vessel type, and targeted data categories may all affect how automation can be best incorporated into an EM program. Further, the relative effort needed to collect targeted data categories influences how automation is applied and which tasks are designated for automation. Because of the high variance in automation needs between EM programs, we do not intend to describe all case-by-case specifics for program design. Rather, we provide generalizations from a broad range of experiences, with a few descriptive examples.

5 K. Tokunaga et al., "Maximized Retention Electronic Monitoring in the Northeast Multispecies Groundfish Fishery" Accessed: Mar. 15 2024. [Online], Available: https://em4.fish/wp-content/uploads/2022/05/GMRI_MREM_Economic_Report_April2022.pdf

6 "Fishing Year 2023 Electronic Monitoring Program Video Footage Review Rates for the Northeast Multispecies Fishery" Accessed: Feb. 02, 2024. [Online], Available: https://www.fisheries.noaa.gov/bulletin/fishing-year-2023-electronic-monitoring-program-video-footage-review-rates-northeast

3.3.1 Considerations for targeted activities

Depending on program objectives, the EM system will target the capture of certain on-board or dockside activities. Data collection for all or a subset of those activities may be suited for automation. Common activities for which data collection is automated include, but are not limited to: setting of gear, hauling of gear, sorting of catch, discarding of catch, catch processing, catch stowing, measuring of fish, protected species interactions, and crew-specific behavior. These activities may occur in either discreet locations on the vessel or as mobile processes that analyses must track across the vessel or fishing area. Some may require detailed imagery for analysis and data collection, while others only require an overview image. Therefore, when considering automation during EM system design or installation, it is not only important to consider the targeted data categories, but also the manner in which targeted activities are conducted and the level of detail necessary to identify each activity. Installation of the EM system and areas monitored must be appropriate for the location and scope of the targeted activities and data collection categories.

3.3.2 Installation considerations

To implement machine learning across a wide range of activities and conditions, EM system installers should make deliberate decisions regarding the camera field of view, camera distance from the object or scene, angle of incidence, physical mounting, and environmental conditions surrounding the camera. The following paragraphs provide suggestions from our experience of best practices for each of these parameters when designing an EM system for automation.

3.3.2.1 Field of view

The camera field of view should be as narrow as possible to see all relevant information about the targeted activity. For an activity such as sorting or gear retrieval on board a trawl vessel, this may need to be relatively broad, while for gear retrieval on a gillnet vessel, the field of view may be more limited. Finally, if the targeted activity includes detection, counting, identification, or measurement of fish on a measuring board, a camera with a field of view limited to the only measuring board is ideal.

Figure 1

Figure 2

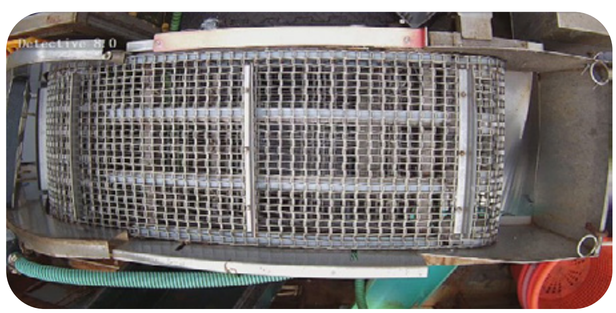

Figure 1: An example of a restricted field of view that captures only the measuring strip and its immediate surroundings. This tight view limits the amount of noise an algorithm needs to accommodate in the area surrounding the activity of interest. Image provided by the NE groundfish audit project.

Figure 2: An example of a restricted view of a groundfish sorting conveyor that is ideal for high quality training data collection. This view also represents an appropriate view to consider applying an automation tool.

3.3.2.2 Angle of Incidence

The best angle for most EM applications is as overhead of the targeted activity as possible. This allows for a consistent size and distance of all objects, regardless of their position within the camera field of view. On many vessels, this is achieved by mounting cameras in the rigging or A frame for broad deck shots. Close shots, such as species ID and length measurements, are best achieved by positioning discard control points or measuring strips near vessel structure such as wheelhouse overhangs. This creates both a restricted field of view and an overhead shot for detailed imagery analysis. In cases where discarding or measuring must occur away from any overhead vessel structure, fabricating a structure (e.g. a mounting arm) can create a view which achieves a top-down image.

Figure 3

Figure 4

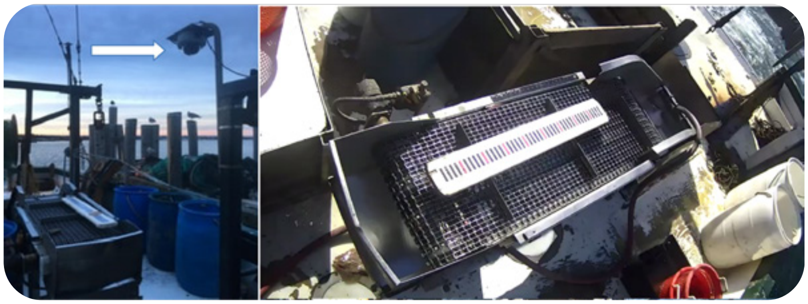

Figure 5

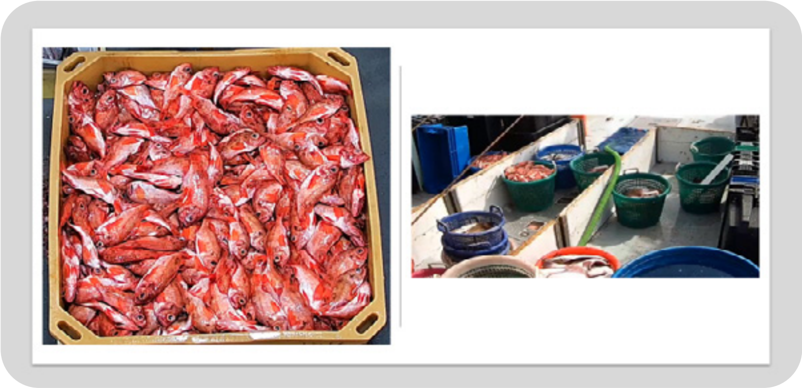

Figure 3: An overhead camera mounting which generates a broad, top-down view of sorting activities. Image provided by the NE groundfish audit project.

Figure 4: A fabricated mounting which generates a close, top down view of species ID and measuring. Image provided by the NE groundfish audit project.

Figure 5: Angle of incidence has been critical in projects where we have tried to train algorithms on what species are in fish containers such as baskets and totes. The image on the left was far better suited for training AI on the species in the container than images on the right.

3.3.2.3 Camera Distance to Target

Once again, the target and objective of the camera footage must be taken into account when thinking about the mounting distance from the target. The combination of the distance from the recorded activity, the resolution of the camera sensor, and the field of view provided by the camera's optics dictate the level of detail provided for a specific object or activity of interest. Most modern object detection algorithms are sized to work on objects that are between 50 and 450 pixels on a side, if one was to draw a square box around the object.

Therefore, the best way to ensure adequate detail for automation is to create tables to calculate the expected object size based on the real size of the object, the distance between the object and the camera, and the camera sensor and lens parameters (Table 1). Calculations in Table 1 show workflows for determining system requirements using the example of a 1-foot length object at a range of 25 feet, that must be 100 pixels in size in the captured image. They solve for the required lens focal length for a given sensor type or the object size, in pixels, given a specific sensor/lens combination.

The first is useful in determining the appropriate components for an EM camera setup, while the second is useful to evaluate the appropriateness of a specific EM setup. A useful lens calculator for making FoV calculations can be found here: https://www.1stvision.com/lens/fov-lens-calculator. For descriptions of object and sensor parameters, see the "Camera Design and Configuration" section on page 18.

Table 1

Object and Sensor Parameters	
Sensor Pixel Pitch	6.4 µm
Sensor Format	1/1.2"
Sensor Horizontal Size	10.67 mm
Sensor Vertical Size	8mm
Minimum Desired Pixels on Object	100
Object Distance from Sensor	25 ft
Object Length	1 ft
Camera + lens focal length	16mm

Computations		
Goal	**Calculation**	**Result**
Object size on sensor	= *Sensor pixel pitch* * *Minimum desired pixels* = 6.4µm * 100	0.64mm
Required focal length	= 56789:: <=>:?@:9D / 6789:; A9@B;C * *Object size on sensor* = (25 / 1) * 0.64	16mm
Resulting Horizontal FoV	= 2 * arctan 5 >96@>60 C60=P6@;?A >=P9 / Q*R9ST=09< U6:?A A9@B;C D = 2 * atan(10.67 / (2 * 16))	36.9°
Resulting vertical FoV	= 2 * arctan 5 V9@>60 W90;=:?A >=P9 / Q*R9ST=09< U6:?A A9@B;C D = 2 * atan (8 / (2 * 16))	28.1°
Resulting diagonal FoV	= 2 * arctan X √V9@>60 C60=P6@;?A >=P9² [V9>@60 W90=;:?A >=P9² / Q*R9ST=09< U6:?A A9@B;C \ = 2 * atan(√ (10.67^2 + 8^2) / (2 * 16))	45.2°
Object size on image	=]789:; >9@>60 >=P9 / V9@>60 ^=_9A `;:C =0.64mm/ 0.64µm	100 pixels

*Note: Computations assume "pinhole" camera model, meaning lens distortion is ignored. Reasonable assumption for lenses with ≥12mm focal length.

Table 1. Calculations for determining expected object size, ideal distance between the object and the camera, and the camera sensor and lens parameters. Example calculations are for an object 1-foot in length.

For broad shots intended to perform activity recognition, object detection and/or object tracking, a camera mounted 10-30 feet from the target is usually suitable when using 1920 X 1080 resolution. For the same resolution we have found that automating the identification of species in the 10-70 cm range requires a closer view of 3-15 feet.

3.3.2.4 Resolution

The common camera resolutions used in EM are 1920X1080 (Full HD) and 1280x720 (HD). In 2023, many marine grade cameras also support higher resolutions like 2560x1440 (QHD) and 3840x2160 (UHD). Higher resolutions make it easier to get an adequate number of pixels on an object of interest. However, as resolution increases so does file size. Automation developers should consider what resolution they need to develop an algorithm and consider the hard and soft costs of higher resolution footage throughout the lifespan of a project. A common approach in AI development is to initially record in as high resolution as possible and retroactively reduce the resolution to see the lowest quality the algorithm can perform on.

3.3.2.5 CASE STUDY: NFWF Case Study Example 2- Consideration for Targeted Activities

In the 2021 NFWF project the activity recognition algorithm demonstrated high interoperability across vessels. This was likely due to the similarity in the image scene and good overlap between objectives. The project team was pleased with how adaptable the activity recognition algorithm worked on a vessel the AR was not trained on with minimal retooling. These results are covered further in Case Study Example 4. However, after initial tests the species recognition algorithm trained on the Audit vessel was not working as well as expected on the MREM vessel. The project team was required to adapt to this challenge and alter our approach.

Due to varying vessel configurations and differences in protocols between the projects, the camera views had different angles of incidence, field of views, and distances to targets. The measuring camera on the Audit vessel was mounted within 3 meters of the fish measuring board, had a tight field of view focused on the measuring strip, and a top-down angle of incidence. The cameras on the MREM vessel were 6 meters away from the deck, but had a top-down angle of incidence, and provided an unobstructed view of potential discard locations.

When we attempted to apply the species identification data collected from the Audit camera to the MREM vessel camera, we did not see satisfactory performance. This reduction in performance was due to a combination of factors. We saw lower pixel density on MREM discards due to the distance to the target (discard location). Due to the placement of the audit measuring camera, the species recognition training data consisted of images that had a high pixel density on the object of interest, a standardized background (e.g. measuring board) and were collected as part of an organized sampling period with the fish stationary.

In contrast to the training data collected on the Audit vessel, the MREM discards we were attempting to identify with AI were stressing each of these categories. Although the MREM vessel had a top-down view ideal for AI development, the discarded fish were further from the camera. The background was complex and often variable. Groundfish discards in the MREM project are accidental or a breach of protocol. Common discards seen on this vessel were fish washing off the stern ramp, falling out of the net, or being accidentally discarded with another individual (for example the crew intends to discard a skate, the fish pick goes through the skate and picks up a flounder the skate was covering). Discards in MREM do not happen in a controlled sampling period. The crew is not presenting the discarded fish in an organized fashion that the EM system is optimized to capture. Obstruction from other fish, crew, or vessel activities reduced the number of video frames the camera had a clear, unobstructed view of the discard when compared to the training data collected on the audit vessel. This, combined with the target individual being in motion lead to a lower chance of an identifiable frame for identification by the algorithm. Most discards were recorded going off the stern of the vessel. The image scene was far more complex than the background in the audit training data. The checker pen of the vessel often had hundreds of fish within camera view, crew moving throughout the pile, the trawl gear in different states (set out vs on the net reel), and views of the constantly variable ocean surface. The frame where a species is discarded in clear view of the camera may occur against any of these backgrounds. This introduced a higher level of complexity than seen in the audit training data.

The combination of these factors required us to pivot. Due to the relative scarcity of groundfish discards in MREM we did not have a robust enough set of training data to adapt the algorithm to the stressors identified. We needed a different approach to create a workflow that would save review time.

We decided to train the species identification algorithm to identify the allowable non-groundfish discards and flag the unknown individuals to the reviewer for identification. We shifted to this approach for three reasons:

1. The majority of review time in both programs is spent watching a vessel sort fish to verify they are not discarding groundfish discards. Having the AI assist with highlighting individuals of interest would reduce the amount of video an analyst would need to watch.

2. Non-groundfish discards are a common occurrence so there was sufficient training data.

3. Groundfish discard monitoring is the primary objective of the Audit and MREM projects. Having a trained analyst make the final determination on positive groundfish species identification is an operational approach regulating bodies may feel more comfortable with implementing.

Using this approach we were able to create a multistage workflow that reduced review time by 20%, had robust available training data, utility across both programs, and considered the path to approval in an operational program.

Modifying our approach and developing a workable solution required a multidisciplinary approach and coordination between EM analysts and developers to best apply each other’s strengths. To successfully adapt our project team needed:

- Knowledge of vessel operations, catch composition, and review protocols
- An understanding of the quantity of training data available for different types of events
- An understanding of what data points take the most review time to collect
- Tight coordination between developers and EM reviewers to understand the level of effort required to automate specific data collection tasks.

Throughout this process we learned firsthand the value of a collaborative automation team with expertise spanning fisheries and AI.

3.3.2.6 Physical Mounting

The point at which the camera connects to the vessel is critical. For all automation applications, the mounting location should experience minimal vibration. However, during normal vessel operation, cameras will experience acute vibration during fishing activity and chronic vibration during steaming. Cameras commonly experience acute vibration when mounted near winches or booms and chronic vibration when mounted to the vessel superstructure. Mounting to high vibration areas such as these is most easily avoided through discussion with the vessel operator. In some cases, it may even be prudent to perform mock gear hauling, setting or hoisting to mimic vibration during fishing activity. While certain cameras may be robust enough to handle both acute and chronic vibration, other models may need a form of vibration absorption. For these, we have found ¼ inch neoprene pads to reduce the effects of vibration. For success with automation both types of vibration are important to avoid or mitigate. Specifically, for objects identified obtaining a single frame associated with an annotation, it is increasingly difficult to obtain such a frame when vibration distorts the imagery.

Another consideration for camera mounting, is minimizing the potential for gear to collide with the camera. Cameras are easily struck and moved off target by gear used during fishing and offload. Therefore, the best places to mount cameras may be on the back of the wheelhouse, which protects cameras from moving gear. If it is essential to mount near moving gear, cameras are most protected by mounting opposite from the moving gear on beams or rigging. A close look at wear and tear on the vessel can help to determine locations of frequent gear contact. Taking time to ensure the crew won't bump the cameras when mounting lower to the deck is also important. While it may be obvious that reduced physical interactions are useful for regular EM practices, such interactions also create incomplete or inconsistent automation training data sets because they change the position of or entirely obscure the target activity.

In some situations, the ¼ inch neoprene pads are not enough to dampen the vibration from some operations. In the cases where moving a camera was not possible, the authors have successfully used cylindrical stud isolation mounts or 7/8 inch rubber isolation pads to maintain image quality. Max load compression should be appropriate for the type of vibration being mitigated. Moving a camera to a safer location is always the best option but may create less than ideal image characteristics for AI development.

It is important to consider impacts to automation tools if a camera is moved either intentionally or unintentionally for a period of time. EM cameras can be accidentally knocked out of the expected field of view by the vessel's operations while at sea. If the view no longer aligns with your training data, view performance may suffer and in some cases the automation tool simply may not function. It's a best practice to work with the vessel to choose locations least likely to be moved, build language into installation contracts to limit disruptive changes, request that crew monitor their cameras throughout a trip for changes, and ultimately anticipate and plan for the fact that some movement will still occur on some vessels.

3.3.2.7 Environmental Conditions

It is important to gather and annotate training footage under the same environmental conditions as are expected in operational footage. This ensures good generalization performance by algorithms, especially for object detection, object tracking, and activity recognition. For example, deck lighting on the vessel used to create algorithm training data sets should be like that used by the rest of the fleet. It may seem advantageous to use training data from a vessel with the best lighting or to add additional lighting to the vessel to improve training data, but building algorithms under ideal conditions can ultimately degrade generalization performance. Therefore, when designing EM systems for the collection of training data, the characteristics and environmental conditions on the training vessel should be representative of the fleet.

Alternatively, algorithms for tasks such as classification of fish within a pre-determined bounding box (e.g. the green box in Figure 6), may be trained using data captured under unrepresentative conditions. For example, if an algorithm can draw bounding boxes around all fish, but not identify species, a second algorithm trained on close-cropped images of fish can identify the sub-images created by the bounding box algorithm. These algorithm pairings are much more robust to different environmental conditions (e.g. pictures of fish in a lab vs on deck).

The following sections expand on specific camera parameters, and how to evaluate them for your application.

4.0 Hardware Specifications

4.1 Summary

EM systems typically consist of an array of machine vision cameras. The camera configuration will determine the automation capacity of footage collected from each camera in the system. Desired camera sensitivity, image blur, and field of view are achieved by balancing the camera configurations to the optimal level for the captured scene or activity.

4.2 Camera Design and Configuration

Unlike cameras meant for consumer applications, cameras used in EM are from a class of cameras called machine vision cameras. Machine vision cameras lack the components meant for human interaction, such as a shutter release or viewfinder, and consist of four main components.

- **Lens:** a system of optical components that focus light onto the imaging sensor.
- **Imaging sensor:** an array of sensors that convert light into electrical signals.
- **Readout electronics:** turn an analog electrical signal into a digital signal, and may apply various transformations such as amplification, filtering, or binning.
- **Interface:** transfers image data to a computer for recording or processing and accepts commands from a software application.

Key requirements that drive design of camera systems include the following:

- **Field of view:** the angular span corresponding to the image, specified as the diagonal, horizontal, and/or vertical field of view.
- **Blur tolerance:** the acceptable amount of blur for objects of interest. For moving objects this is dominated by motion blur.
- **Object size:** the object of interest's dimensions in an angular sense.
- **Frame rate:** the frequency at which the camera acquires consecutive images.
- **Lighting conditions:** the required lighting conditions for proper camera function

The camera's sensor format and lens determine the field of view. The sensor format refers to the physical dimensions of the imaging sensor, which determine aspect ratio. Many machine vision cameras have interchangeable lenses which use C-mount or CS-mount as the interface between the camera case and the lens. For hardware redundancy, it is possible to use the same image sensor for all cameras but achieve different fields of view by using different lenses in each camera. Other cameras are designed with the lens and image sensor packaged together. These often have an optical zoom capability to change the field of view between cameras or dynamically during monitoring activities. However, per the above notes on vibration, optical zoom motors must be sufficiently ruggedized for the operating environment. See Table 1 for an example of how field of view affects automation capabilities.

Image blur is an especially important consideration when designing an EM system for automation. Just as a blurry image can obscure important details for identification by a human, the same is true for an algorithm. Therefore, installations should aim to balance the above parameters against tolerance for expected motion within an image.

Depending on automation goals, tolerance of blur will differ. For simple activity or motion-based actions, a higher level of blur is tolerable because algorithms will use scene context to make decisions. For object detection and classification, it is important to drive motion blur close to zero because algorithms will use fine-grained details such as morphological structure or coloring pattern to make decisions. While simply reducing the amount of time that the camera aperture is open will reduce image blur, it can also create images that are overly dark, and do not have the detail required to make accurate classification decisions. For many human video review scenarios, these challenges can be overcome by tracking the fish through the whole scene and piecing together identifying information from different parts of the scene. While this is technically possible for an algorithm, it requires a lot more training data and a more complicated architecture. Therefore, the recommendation is to have as high quality an image as possible for every frame of interest.

The goal of reducing image blur is in direct tension with higher frame rates (high detail) and low light scenarios. The reason for this is that to overcome low light and achieve higher detail, the camera aperture is left open for as long as possible. Many consumers observe this effect in low light cell phone pictures. The following lists the main parameters that increase the ability of a camera system to take high quality images with minimal motion blur and highlights recommendations for automation.

- **Pixel pitch:** the physical size of each pixel on the sensor.
    - **Recommended:** larger is better.
- **Quantum efficiency:** the percentage of incident photons converted to an electrical signal.
    - **Recommended:** higher is better.
- **Aperture:** the size of the smallest window through which light passes in the lens.
    - **Recommended:** see f-number.
- **F-number:** Aperture is often expressed as a ratio of focal length to diameter of the entrance pupil called f-number. F-number is favored over simply using aperture diameter/area for characterizing lens “brightness” because image illuminance is directly proportional to the square of the f-number.

    - **Recommended:** as a rule of thumb, the lower the f-number, the better the lens for low light or high motion scenarios.

- **Exposure:** the length of time the sensor accepts light for each image.

- **Recommended:** Longer exposure risks more motion blur. Use the lowest possible exposure time to get usable imagery.

- **Lens transmissivity:** the percentage of incoming light that reaches the image sensor.

    - **Recommended:** Higher is better.

5.0 Algorithm Considerations

5.1 Summary

Algorithms used for EM generally fall into one of two categories: object identification or activity recognition. Object identification is used to recognize an object within a scene and possibly perform some task related to that object. Activity recognition is used to identify instances of a particular action or activity occurring in the video footage. Activity recognition and object identification algorithms are able to build off of each other to perform more complex tasks and further reduce the burden to human review, but as the level of automation increases, so does the level of difficulty in creating that automation pathway. Below we describe automation pathways and how they might work together for a number of common EM tasks including: object identification, fish presence recognition, fish species identification, fish counting, feature measurement, and various activity recognition scenarios.

Data collection for AI (referring to deep learning) pathways requires careful EM system design and installation as well as considerations for building training data sets in human-review workflows. For spatial tasks, video reviewers may draw a shape or point on the scene, while for temporal tasks, reviewers may mark a series of timestamps.

5.2 Possible Algorithm Tasks

Broadly, algorithms for EM fit into two overlapping categories: object identification and activity recognition. Algorithms will fall into one or both categories depending on purpose and use, and it is important to identify automation objectives when choosing either category. In object identification, the goal is to recognize the presence of an object within a scene (e.g. fish species of interest), and then localize, identify, or measure that object. An example of this is the Northeast Multispecies Audit model EM program. In this model video reviewers count, identify the species, and obtain a length measurement of each fish in imagery from cameras placed above measuring boards.

Figure 6: A fisherman measures discarded fish, using a measuring strip installed below the EM camera. The fish length, and calculated weight is captured during video review using this image. Image provided by the NE groundfish audit project.

Figure 6

The second algorithm category is (AR). In this case, the goal is to identify when a specific activity is happening within a video scene (e.g discarding of fish, loading or unloading of the hold, etc.). An example of this is the Northeast Multispecies Maximized Retention model EM program. In this model the goal is to identify when fish are brought on board, when they are unloaded, and to identify any discarding of allocated groundfish. Figure 7 shows the outputs of an AR algorithm that has characterized the different activities of a fishing trip. The red bars denote start and stop times of activities as recorded by a human reviewer. The shaded bars denote recognition of specific activities by algorithm. Outputs such as this have the potential to both focus human video review efforts and reduce the amount of stored video footage by differentiating between high-priority and low-priority video. The algorithm is able to have a granular description of the different activities (e.g. Catch Sorting is broken into Sort and Hold Loading), so that even with less than optimal performance, someone reviewing this graph can easily pinpoint where to look for certain activities.

Figure 7

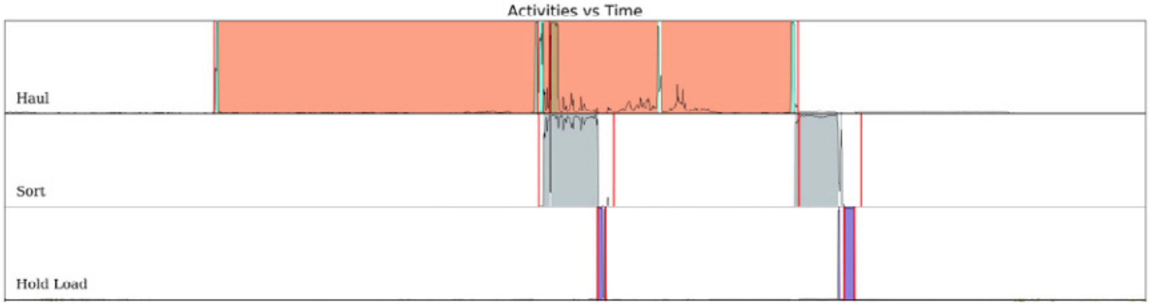

Figure 7: An example of output from activity recognition algorithms. The red bars denote start and stop times of activities as recorded by a human reviewer, while the shaded areas denote activity duration as identified by AR algorithms. Algorithm outputs: Orange segments represent calculated haul time. Green segments denote net handling activity and occur on either side of the haul as the net is set in or hauled out. Gray segments denote when catch handling is occurring. Purple segments denote video in which catch is being loaded into the hold.

Object identification and activity recognition algorithms have even greater potential when used in tandem to perform a specific task. Figure 8. shows how different automation tasks might build off of each other to reduce review time. The difficulty associated with implementing algorithms increases as the level of automation and task-specificity increases. Algorithm implementation must therefore balance the potential reduction in human-review effort and algorithm specificity.

Figure 8

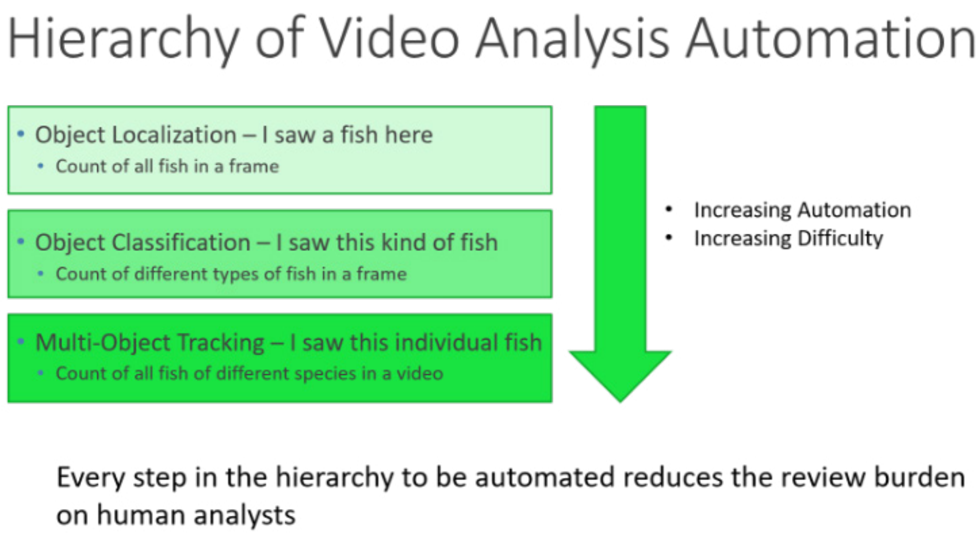

Figure 8. Hierarchy of video analysis. As the automated task becomes more specific, automation difficulty and potential reduction in human-review effort increase.

5.2.1 Object Identification

The task of object identification within a scene is usually associated with image recognition tasks. For this reason, most object identification workflows are applied frame by frame, with additional algorithms layered on top (e.g. tracking algorithms for counting fish). For these workflows it is important to have a training data set that has annotations associated with individual frames. Most modern object detection algorithms use Convolutional Neural Networks (CNNs) as their baseline.

Common implementations for video are YOLO and Faster R-CNN, though there are many others, each with their own balance of accuracy, inference speed and processing requirements.

The following paragraphs describe examples of algorithm workflows, proceeding from least complicated to most complicated. These are intended to provide examples for designing workflows to meet an EM program's needs. There is almost nothing unique about the example of identifying fish in the following, and all of the examples provided would be equally valid for just about any EM application, regardless of species (even humans, which is where most of the technology comes from).

5.2.1.1 Fish Presence Recognition

One of the simplest workflows is to recognize points in a video in which fish are present within the frame. If an algorithm reliably recognizes those points, then a workflow that requires human analysts to work with video containing fish may be drastically reduced by allowing a reviewer to jump from event to event rather than combing through video to locate fish-handling activities. The algorithm for this case does not need to make judgments such as the species of the fish or measure any physical property such as length. The training requirements are much more lenient for such a workflow, and the performance and reliability of the resulting models are generally very high.

With a slight modification this workflow can accomplish fish presence recognition within a Region of Interest (RegOfInt) of an image. Usually the RegOfInt is some subset of the image such as a measuring board or conveyor belt. In some cases, fish presence recognition can even be accomplished without the aid of an additional algorithm. There are many software tools available to interact with images in this way (e.g. OpenCV, Python Imaging Library), and integrating them into review software is not a difficult task. Figure 9 shows an example of a fish recognition (green box) performed inside a defined region of interest (blue box). The region reduces both processing load, as well as the potential for false alarms. In this example, the algorithm is implemented as part of the open source OpenEM algorithm and software library.

Figure 9

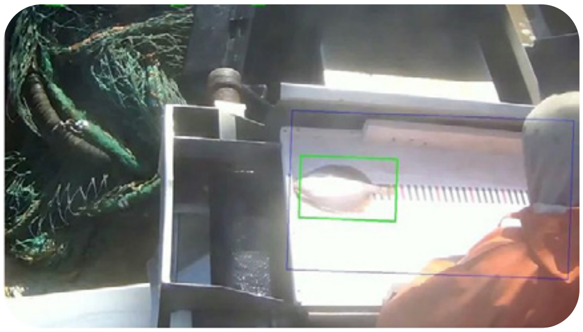

Figure 9. Image with a defined region of interest in blue, as well as a localized fish in green. By restricting an algorithm's field of view to a specific region of interest, performance is improved by excluding potentially confusing scenery (e.g. background, totes with other fish in them, etc.). Image provided by the NE groundfish audit project.

5.2.1.2 Fish Species Identification

A step up from the previous workflow would be to identify the species of a fish within a frame. This usually requires some concept of localization within a scene because in many use cases there are multiple fish of multiple species types within a scene. Performing localization requires the use of more complex algorithms which need increased amounts of training data that is harder to generate.

To build training data sets, human reviewers localize objects within a scene by describing the bounding coordinates for the object. Typically, coordinates are represented as a box, but any polygonal shape or even a pixel mask can be used. This information is used to train a localization algorithm to judge itself when it creates bounding box proposals. By giving it examples of localizations (boxes), it learns to draw the appropriate boxes in new images it is not trained on. For imagery, there are many free tools available to generate these types of localizations. A few of them are Tator (https://github.com/cvisionai/Tator), CVAT (https://github.com/opencv/cvat), LabelMe (https://github.com/wkentaro/labelme), and many others. In the above example, (Figure 9) the green box shows an example of a fish localized within an image.

5.2.1.3 CASE STUDY: NFWF Case Study Example 3-Object tracking/Species Recognition

In the NFWF 2021 project, we developed a multistage algorithm that incorporated object tracking, species classification, and activity recognition. To capture the performance of this workflow we compared review time and data between the AI-assisted and human-only review. The multistage workflow saw a 20% reduction in review time. However, trips reviewed without AI assistance had more groundfish annotated. This discrepancy was traced back to a number of causes including algorithm training, review protocols, and the category of discard the AI review did not capture. We reviewed the discards the AI review missed and saw that many of these discards slid off the stern ramp and were not interacted with by the crew. The object tracking algorithm was trained on discards that were intentionally discarded. See figure 10. This trial was helpful in identifying where we should focus training data and is an example of the revisions and maintenance AI workflows will require to be robust.

We also saw increased savings in review time the more complex a trip was. We concluded that we saw less time savings on AI-assisted reviews for simple trips because they tended to have less complex species composition. In trips with a more homogenous species composition or low catch volume the time spent interacting with the algorithm outputs took more time than watching the trip at high speed. With a simple species composition, a trained video analyst can identify common discards in the fishery such as lobster, dogfish, and skates at high speed.

These results show that a complex algorithm workflow will require more tuning and trial cycles to realize review savings. The effectiveness of the workflow can also be impacted by the operations of the fishing vessel. The vessel in this trial fished with an open stern that allowed some catch to wash over the stern ramp. If the stern ramp was closed or boxed off by a checker pen board, we may have seen a better match in discard capture between the human and AI-assisted review. Understanding the review process and fishing operations is critical to maximizing the impact of any AI implementation effort.

Figure 10

Figure 10. Demonstration of object tracking algorithm identifying an intentionally discarded fish. If a fish slid off the stern on the port side without any human interaction it may not have been identified.

5.2.1.4 Fish Counting

Fish counting is an example of a task in which information needs to be propagated from frame to frame to determine something about the video. Typically, tracking is incorporated to recognize the same fish from frame to frame and when it enters and exits the footage. Tracking requires a concept of localization as in the fish species recognition but it does not require a full species identification algorithm. For cases with one or a few fish with little to no overlapping or long-term occlusion, there are many tracking algorithm options. Cases with more than a handful of fish and high potential for occlusion (e.g. on a conveyor belt) are significantly more difficult. This is known as the multi-object tracking problem, and it is not a well solved problem in most domains. The training data required for multi-object tracking is vastly harder to generate by hand. Most efforts in generating multi-object tracking algorithms are a combination of automation assistance to generate training data as well as model-based tracking algorithms to reduce the amount of required data. With the resources and the patience though, the benefits of generating high quality tracking training data are immense.

5.2.1.5 Feature measurement

Oftentimes there is a desire to not just count and identify, but to also generate a feature measurement (e.g. length) for objects of interest. Depending on the feature and the scene, this task has varying degrees of difficulty. For example, fish that are on a fixed measuring board with a fixed-distance camera view are straightforward to measure accurately. In this instance the algorithm only needs to measure the pixel distance, which is usually well approximated by the bounding box generated by the object detection algorithm around the fish (see Figure 7). That pixel distance is then converted to length units using a fixed conversion factor.

It is significantly more difficult to generate a feature measurement in scenes that do not have a fixed distance or canonical view of the fish (e.g. when fish are hauled over the side of a boat). In this case it is usually required to have a stereo camera set up, which needs much more complex algorithms and training data. An example of work being done for stereo measurement is in the Alaska Fisheries Science Center, in conjunction with the University of Washington (https://www.fisheries.noaa.gov/feature-story/advancing-innovative-technologies-modernize-fishery-monitoring)

5.2.2 Activity Recognition

Activity recognition in video is most often associated with labeling of short video clips (YouTube-8M, Thumos, etc.), but for use in longer, unconstrained video clips, it is usually known as event detection. This is still a very active area of research, and there are not common frameworks to fall back on for this task.

The good news for EM is that the required bounds for event recognition are often much looser than in other domains. For instance, if an algorithm can narrow down to +/- 1 minute for activities of interest, this is usually sufficient for EM purposes. This is especially true if a human will be using those time stamps to jump to parts of video and manually verify events of interest. This level of temporal accuracy is also likely acceptable for use in CPUE or other fishery metric calculations.

Typically, activity recognition algorithms also have less burdensome annotation requirements for building training data sets because they only require start and stop timestamps for events of interest, instead of annotations such as bounding boxes, lines, or other shapes. These are often already recorded in review software and require little modification of review technique to generate appropriate training data for algorithms.

The following paragraphs describe a few common workflows for activity recognition.

5.2.2.1 Fishing Activity

In this use case the goal is to identify portions of video that contain fishing activity of interest. The intended output would be a timestamp, duration, and label. For most activities of interest (e.g. setting or hauling of gear or catch handling) the accuracy can be close to 100%, with relatively tight time bounds. In the case of a 24/7 camera system, this can eliminate vast amounts of the captured video as not relevant, resulting in significant storage and transmission savings, as well as review time savings. When targeting this level of automation, it is best to have clear camera views of all areas where targeted activities can occur. The best views tend to be overhead, with resolution on the order of 720p or better, given the typically available mounting points are higher up on the boat. This is a rule of thumb and depending on available mount points and required fields of view, can be higher or lower (see “Hardware Specifications”, page 18).

5.2.2.2 Hold Loading/Unloading

For non-fishing activity such as hold loading or unloading, activity recognition starts to cross over into object detection or have otherwise more difficult aspects. For instance, EM protocols may aim to determine when a hold is loaded, count the number of totes loaded, and identify the species associated with each tote. This is a compound activity that comprises the recognition of the loading event, and then counting of totes and species identification. The loading event can be bounded with simple time stamps but counting of totes and species identification additionally require object detection and tracking.

Typically, this type of activity recognition needs an unobstructed camera view of the area that contains all objects of interest and their intended destination. Again, this is best achieved through an overhead camera view. For adequate views of individual totes and individual fish within totes, a higher resolution such as 1080p is often required, due to available overhead mounting positions. This is due to the algorithmic specification of an object subtending at least 50 pixels of the field of view.

The data annotations needed to achieve all three of these goals are a little more involved than the time stamp only requirements for fishing activity and are similar in effort to object detection annotation.

5.2.2.3 Anomalous Activity

Oftentimes it is desired to flag anomalous events such as a endangered, threatened, or protected (ETP) species interaction, prohibited discarding, crew in a dangerous position, or other activity that is of special interest to managers or captains. The diversity of possible events in this regime means that detecting these events is usually a noisy problem, with a fair number of false positives. It is possible to construct high performance algorithms with low false positive rates and high recall, but they tend to require more training data than is typically available and cost more to develop than is justified by what they are detecting.

With the above considerations in mind, a useful paradigm for anomalous activity identification is to identify a region of interest where these activities are likely to occur and to focus processing on that area. The requirements are then an unobstructed view of the area, with an overhead view typically desired. In the case of identifying activities such as operations in a dangerous or restricted area, lower resolutions such as 720p can usually be accommodated. However, for the use case of identifying illegal discards or protected species interactions, there is an element of object detection and identification that necessitates a higher resolution. Discriminating between allowable and prohibited discards is usually an easier proposition for an algorithm than full species identification.

Protected species interactions are easier still but usually suffer from a lack of training data.

5.2.2.4 CASE STUDY: NFWF Case Study Example 4-Activity Recognition

Activity recognition (AR) is an extremely promising application of AI in EM. In the 2021 NFWF AI project, we developed AR review protocols that allowed reviewers to use the AR to skip periods of inactivity. This allowed them to save time by highlighting periods of activity to watch. When compared to a fully manual review of the same trips, the AR assisted review showed a 33.5% reduction in review time, a 0.53% difference in groundfish discard entries, and an identical number of hauls documented. These results show all fishing and discarding activity was reviewed with an insignificant impact to groundfish discard data quality. Saving such a large portion of review time while maintaining high data quality was an encouraging outcome.

Another exciting factor was how little retraining the AR algorithm required to work effectively between vessels. There was minimal adaptation required to apply the AR algorithm developed on the MREM vessel to perform well on the Audit vessel. It should be noted that these vessels did have similar deck set ups and operations. Interoperability and flexibility are a key components to realizing ROI and creating a pathway to wide-scale use of AI in EM programs.

In addition to performing well and requiring minimal retooling to work across vessels, AR also has the potential to save more review time on longer trips. The vessels participating in the NFWF 2021 project were primarily fishing single day trips. On day boats, the travel time to the fishing ground is relatively short, and there tends to be activity on deck for a high percentage of the trip. In multi-day operations the travel to and from the fishing ground is often longer, hauls are longer, and there tends to be a lower percentage of activity on deck compared to total trip length. If using the same AR protocols on multiday operations we would expect even greater savings in review time while maintaining good performance.

AR demonstrated tremendous capability to be a cost-effective tool for reviewing video from New England trawl vessels, with the opportunity to be even more effective on multi-day offshore operations.

5.3 Data Collection

For all the use cases described in the previous section, there are some important differences between systems designed for human-only video review instead of algorithm-assisted video review. This section will highlight some of the footage requirements from automated use cases, as well as give examples of considerations that may not seem obvious when designing for human analysis rather than algorithm analysis. In addition to requirements on the EM system for operational functionality, there are also requirements on the data collected to train algorithms to ensure they are robust and perform as expected in application. The following is a list of guidelines for getting the best EM footage for automation.

5.3.1 Camera Frame Rate

Generally, human reviewers are adept at inferring information between frames of video and can make more robust judgements than algorithms regarding an object of interest by using multiple partial views. This allows human reviewers to accommodate slightly lower frame rates than automated analysis.

Camera frame rate is primarily important when the objects of interest are only in the field of view for a short amount of time. For quick moving operations such as a conveyor or a discard measuring operation, we therefore recommend 15 frames per second (fps) as a minimum frame rate to guarantee a good view of the object. For activity recognition, where judgments are made closer to the minute time scale, it is often enough to capture 5 fps, or sometimes even 1 fps footage.

Like resolution, frame rate will have a direct impact on the amount of storage and process required. Higher frame rates will require more storage but will allow better data capture, particularly when objects are in motion. Developers should decide on the preferred frame rate for their application and account for associated costs. Like resolution, developers can retroactively degrade frame rate to assess the minimum frame rate an algorithm can perform at.

EM systems commonly record between 5-15 FPS. The frame rates commonly available on previously installed EM system are acceptable for AR development and fall within the recommended minimum frame rate for more complex tasks like discard measuring.

Many cameras and EM systems can also utilize a variable frame rate. A variable frame rate is a feature that increases or decreases the frame rate of a camera based on an input from the EM system. The input can be done via physical sensors on the vessel or onboard AR. Some common physical EM sensors include pressure sensors and rotation sensors. A more modern approach is to have an AR algorithm running on the EM recording system detect activity and change the frame rate appropriately.

5.3.2 Object Placement and Occlusion

Humans are also particularly adept at tracking objects through occlusions (e.g. hand covering) and using context from separate views to piece together classification clues. For instance, a human reviewer may be able to easily identify a fish that is mostly occluded by a hand when being measured by looking back a few seconds in time to find a better view. Algorithms do best when they are afforded a single, clear view at a known point in space or time. This minimizes the need for complex tracking algorithms, or algorithms that consider every frame in which they see a particular fish. While these types of algorithms are sometimes necessary, a significant amount of development time, and time spent generating training data can be saved if you can provide that “money shot” view.

Occlusions are best avoided through training of vessel crew and a thorough discussion of deck operations during system installation.

5.4 Data Annotation

The primary method of training algorithms is a technique known as supervised learning. This is accomplished by showing the algorithm many positive and negative examples for the concept that it is learning. Algorithms will use various types of reviewer- generated annotations as the cues to understand what is meant by these examples. For spatial tasks, such as species detection and recognition, annotations may be drawn on the scene. For temporal tasks, the most common annotation is either a single time stamp along with a label, or a start and stop time, also with a label.

Data annotation can be a lengthy and expensive process. It can be beneficial to use EM data analysts to assist in the collection of training data, as the accuracy of the training data is very important and not all AI developers will be trained to identify species of fish or relevant fishing events. It is also not an efficient use of resources to have a highly paid AI engineer annotate the large amounts of fishing data required to create a robust algorithm. Their time is better focused on algorithm development and performance. Utilizing the skill set of EM data analysts to create training data is an efficient approach that can provide high quality data. One particularly efficient method is to incorporate the collection of training data into a regular program review flow. Teams should be careful to understand the impact of adding extra annotations may have on the regular program review and scope protocols appropriately.

AI developers should meet with EM data analysts in the early phases of a project and develop detailed documentation to communicate the specific factors required for ideal training data. It is also helpful for AI developers to annotate a small set of data to demonstrate to the EM analyst exactly what characteristics they are looking to capture. For example, when boxing a fish for species identification it can be important to have the box around the individual of interest extremely tight with as little background distraction as possible. However, too tight and a segmentation tool may clip the fish's tail off during processing. Just as EM many vessels have to follow vessel monitoring plans (VMP's) and reviewers follow review protocols, the training data annotator should be following a detailed guide. Having confidence in the way the data is annotated can improve the training process dramatically.

After protocols are communicated to reviewers the project team should process a few small sets of data and meet regularly in the early stages. These steps will ensure the team is in sync and most effectively utilizing the skill sets of the EM data analyst and AI developers. During this type of collaboration, useful domain expertise can be transferred from the subject matter experts to the developer.

The following is a list of spatial annotation types, and where to use them.

5.4.1 Box

This is the most common annotation to use in visual analysis tasks. Box annotations describe the coordinates in a scene that create a rectangle that encompasses the object of interest. For instance, a rectangle may be described by the x and y coordinates of its four corners, or by the x and y coordinates of one of the corners and a width and height. It does not matter which convention is used so long as the same convention is always used, and the algorithm is provided a description of the convention.

5.4.2 Line

Line annotations are most often used to measure an object's length. Usually, they are used in conjunction with box annotations. The best way to describe a line is the x and y coordinates of its endpoints.

5.4.3 Dot

The dot annotation most commonly used to denote an area of interest or to count objects within a scene. It is usually used with more complex algorithms that attempt to infer something about the location marked by the dot by looking at the whole scene.

5.4.4 Pixel mask

Pixel masks are a tool used to generate segmentation masks, which seek to identify all pixels belonging to an object in a scene. With a complete representation of an object, algorithms can try to learn things such as morphometrics or pose of an object. Pixel masks are typically represented as a list of pixel indices within an image, where a defined convention determines how to assign numbers to pixels within the image.

5.4.5 Polygon

Polygons may be used in place of pixel masks, as a lower fidelity form, to accomplish the same types of tasks. It is far easier to draw a polygon than it is to color in every pixel belonging to an object.

Polygons are usually represented as ordered lists of the x and y coordinates of vertices.

5.5 On Vessel Implementations

Automation algorithms may not only have different goals for video analysis but may also have different modes of operation. One extremely useful use case is running algorithms on the vessel on which the video footage is captured. This can drastically reduce both the storage and transmission requirements for EM video data. However, the available computer capability on board a vessel is usually significantly less than what is available at a data review center, and so there are considerations for expected performance and reasonable use cases.

The main considerations for on board vessel algorithm evaluation revolve around Size, Weight, and Power - and Cost (SWAP-C). It is technically possible to put a powerful enough computer on board a vessel to run the types of algorithms discussed in this document, but those computers can range in cost up to several thousands of dollars. Additionally, without optimization, many of the algorithms developed do not run in (near) real time.

There does exist a growing field of low cost, lightweight, ruggedized computers for machine learning inferencing (e.g. NVIDIA's Jetson platform). These systems are sufficient for running the following types of workloads:

- Object detection and classification
- Activity Recognition

These two workflows comprise a large majority of the types of automation that are most useful aboard a vessel, because they can act as trigger points for either prioritizing recordings, or human intervention, via lightweight message outputs. It is much simpler to consider sending a small text file with a summary of the last minute's worth of activity or discards over satellite or some other cellular link than it is to send an entire minute's worth of multi-camera video frames.

6.0 Review Software Interaction

EM video review without any human interaction may be possible as the technology evolves. For the foreseeable future however, there will likely be a human analyst working with AI technology to create the final EM data product. If a project plans to use AI-assisted review, the interaction between analyst and AI needs to be efficient to maximize the time savings potential. Developers should consult with the analyst, understand how the review software works and determine how to display AI outputs and how users will want to interact with them. It may be important to consider business implications if the AI development and review software are from different companies. Understanding the business relationships needed such as access to source code, intellectual property and licensing requirements can prevent roadblocks during implementation.

6.1.1.1 CASE STUDY: Case Study Example 5-Reviewer Software Interaction NFWF 2021

The AR algorithm used on the 2021 NFWF project identified net in/out, catch sorting, hold loading, and offloading. The algorithm classified the activity on a scale of 0-1. The closer to 1 the event classification was, the more likely the algorithm predicted the event was occurring.

Review for this project occurred on two different platforms. Due to differences in user interfaces, the AR results were displayed differently. On the Audit review platform, AR was displayed in a binary way. If any of the categories were above 0.5 a yellow bar was displayed on the trip timeline. On the MREM review platform, each AR output (net in/out, catch sorting, hold loading, and offloading) was displayed as a line graph on the trip timeline.

Both approaches allowed reviewers to skip sections of inactivity. Reviewers reported they felt the binary approach was faster and easier to work with. The reviewers highlighted that they did not need to see which activity the AR was identifying because a value over 0.5 for any category required them to watch the video anyway. We believe the simplicity of the binary AR display combined with how video analysts interacted with it was partly responsible for a greater reduction in review time seen between the Audit and MREM components of this project. While using the binary AR display reviewers annotated the same number of hauls and there was only a 0.53% difference in groundfish discard entries when compared to human review. This demonstrated the AR was properly highlighting periods of activity and reviewers were able to capture virtually identical data. If using a binary approach, it is important to calibrate the threshold correctly to avoid skipping over sections that need further review. This calibration may take some trial cycles to fine tune and should be continuously audited for quality assurance.

How AI is displayed and incorporated is an important consideration for any project. In this project we found the binary approach may be preferable. However, this is not universal. EM projects may want to use AR to identify the number of hauls and then review a subset of those hauls. In a program with these objectives displaying the traces for “net/in out” and “catch sorting” separately would be more useful than a binary approach.

7.0 Technological Advancements In Automation Practices

There are currently several areas of technological innovation having direct impact on automation for EM. It is important to recognize that many of the tools for automation are evolving rapidly, and static lists on the latest tools or trends become out of date within months of publication. Nevertheless, the following are areas of noteworthy advancements in EM and automation that hold particular promise for the future.

7.1 Open Datasets

Managers often want to use EM to identify species (e.g. rare ETP species) or events (e.g. gear abandonment) where there is little available video footage with which to train algorithms. Other industries benefit from high-quality out of the box models and datasets relevant to their application. However, these packaged models do not have training data that includes relevant data, such as fishing footage, which limits their performance. The need to create costly training data sets applicable to EM AI development is currently limiting the development of useful automation tools in EM. Fortunately, there are several efforts focused on creating open datasets which would include images of rare catches as well as more common footage on a variety of vessel configurations from different EM providers. Diversity and scale are some key elements to creating powerful datasets. However, cleanliness/readiness for use, organization, and labeling/annotation are also important for dataset quality and making sure they are effective training data. Datasets are important to training, but a strong community is also vital to the advancement of AI in EM. Additional useful features include: managed registration (not open to anyone), user uploaded data sets (regulated), data set searching, ability for users to add labels (managed crowdsourcing), a strong user community including messaging on updates, ability for users to rate and add comments, masking of key sensitive information, and regular maintenance and updates. These datasets could also include open source models trained on the images and a message/bulletin board on how people are using the datasets and models.

While there are some existing datasets, most of the data available does not include a sufficiently wide range of species and rare events. This may change in the near-future as the market and need for more models increases.

An example is the Fishnet dataset by The Nature Conservancy with more than one hundred forty thousand annotated images of long line tuna fishing activity in the Western and Central Pacific, and has been updated since beta release in 2019 to its primary release in 2022 [https://www.fishnet.ai]. This data set is limited and not applicable to all fisheries or gear types.

7.2 Computer Vision and Deep Learning Methods

Recent advancements in machine learning models relevant to EM include generating plausible imagery given image or textual prompts (useful for synthetic data generation), object detection and identification at high speed, improved self-supervised training, and improved pose estimation. The You-Only-Look-Once (YOLO) models have remained at the forefront of real-time detection, allowing object detection for EM at the same rate as the video runtime (or faster).

Annotated and labeled data remains a bottleneck for many models in use today. However, deep learning methods for self-supervised have been constantly improving, advancing on older ideas such as contrastive learning and auto-encoders. Vision transformers have reached state of the art object detection performance, thanks to these self-supervised designs that have allowed them to scale sample efficiency and performance with less domain specific training data (including in zero or few-shot inference). This makes them really good at detecting and segmenting objects out of images, even very different to their training sets. EM applications, which are bottlenecked by collection of specific fishery/species/camera-angle imagery as well as its annotation, can improve their datasets and performance by incorporating these vision transformer advances.

Software frameworks dedicated to model development, training management, and handling before and during deployment have also advanced significantly. Today, PyTorch[7] remains the scientific community's leading library for model development, thus many open-source models used in EM are (at least initially) built with PyTorch. Previously a leader, TensorFlow[8], is still a mainstay for its training coordination and model compilation capabilities (for efficient deployment). Useful libraries and open standards like ONNX[9] exist to allow for transfer between frameworks and deployment across hardware platforms.

7 A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library." arXiv, Dec. 03, 2019. doi: 10.48550/arXiv.1912.01703.

8 M. Abadi et al., "TensorFlow: A System for Large-Scale Machine Learning," in Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, USENIX Association, 2016, pp. 265–283.

9 "onnx/onnx." Open Neural Network Exchange, 2017. Accessed: Feb. 01, 2024. [Online]. Available: https://github.com/onnx/onnx

Hardware advancements for deep learning training and deployment (inference) have also moved rapidly, due to increasing demand and development of AI integrated solutions[10]. Graphical processing units (GPUs) have had dramatic increases in performance and have been specialized to cater to matrix multiplication and other AI specific operations, across large players like Google's Tensor Processing Units (TPUs)[11] or NVIDIA's edge devices[12] as well as at smaller dedicated chip makers[13]. These are all important to consider for cloud, on premise, or edge components of EM's AI training and deployment.

7.3 Communications

Fisheries is unlike many other industries that have undergone automation, due to its remote operating locations, harsh conditions, and limited access to communications. In some cases, using cell or satellite services to process or transmit data can reduce the costs and complexity. Remote communications also allow for cloud-based AI to be part of the workflow before the data reaches an analyst, potentially saving time for the review. However, this is not true for all programs and careful comparison should be done to ensure utilizing these methods will actually reduce cost or complexity.

Traditional satellites are used by advanced EM systems and fishing vessels mostly for maintenance and operations of EM systems. This may include the transmission of regular heath updates of the EM systems and the ability for EM technicians to remotely log into the EM systems to troubleshoot issues. Traditional satellite plans utility can be limited by high price, data transmission speeds and volume limits as well as other contractual limitations such as mandatory yearly contracts. Often the data plan is not sufficient to send the full volume of video data required within a realistic EM budget. However, if combined with the use of AI, it's possible to send small video clips that are flagged by AI to be of high importance for immediate review (see Edge Computing below).

New Low Earth Orbit (LEO) satellite offerings are now available in some regions by both new and established space groups, with EM providers trialing services such as Starlink from SpaceX. These services offer much higher speeds, bandwidths, and latencies at lower costs. However, this is a nascent range of services and many plans are still capped at low levels of data per month or in limited regions, which means that they still may not be feasible for full video transmissions in all areas and remain in their trial phases. Other groups also launching or have recently launched satellite constellations such as Amazon's Project Kuiper[14] and OneWeb (now Eutelsat OneWeb)[15] and are competing with Starlink. More data on LEO solutions should be available in 2024 as more results from trials become available. There are some concerns regarding relying on LEO satellites and providers as considerations of LEO business stability given government disputes, sanctions and other external factors. This will continue to be a developing technology that has the potential to have a large impact on the EM automation space.

---

10 D. A. Patterson and J. L. Hennessy, Computer Organization and Design RISC-V Edition: The Hardware Software Interface.Elsevier Science, 2020. [Online]. Available: https://books.google.ca/books?id=e8DvDwAAQBAJ

11 N. P. Jouppi et al., "In-Datacenter Performance Analysis of a Tensor Processing Unit." arXiv, Apr. 16, 2017. doi: 10.48550/arXiv.1704.04760.

12 "Meet Jetson, the Platform for AI at the Edge" AS OF 2/14/24] NVIDIA Developer. Accessed: Feb. 01, 2024. [Online]. Available: https://developer.nvidia.com/embedded-computing

13 Z. Jia, B. Tillman, M. Maggioni, and D. P. Scarpazza, "Dissecting the Graphcore IPU Architecture via Microbenchmarking." arXiv, Dec. 07, 2019. doi: 10.48550/arXiv.1912.03413.

14 "Project Kuiper," US - About Amazon. Accessed: Feb. 01, 2024. [Online]. Available: https://www.aboutamazon.com/what-we-do/devices-services/project-kuiper

15 "Our Network," OneWeb. Accessed: Feb. 01, 2024. [Online]. Available: http://oneweb.net/our-network

7.4 Computing

7.4.1 Edge Computing

Edge computing or on-vessel computing is the ability to process information from EM (and other) systems on the vessel while the vessel is still at sea. Edge computing can significantly reduce data transmission costs by only transmitting distilled video or even just summary data reports to managers.

All EM systems have some level of computing power embedded in their systems. This is generally proprietary and not interoperable with other systems. Additionally, the primary function of these systems is the operation of the EM systems themselves. Applications of automation at the edge (onboard) that are being used or trialed include the following:

- Activity detection to determine variable video frame rates and compression algorithms to reduce file sizes when there are periods of no activity
- Face blurring to protect privacy issues of the fishers
- Activity detection for pre-selecting video for review by analysts
- Catch count and species identification while at sea
- Bycatch detection

In some cases, the additional compute power that AI models and automation algorithms may require separate edge devices dedicated to AI inferencing on the edge, such as a Jetson Orin.

7.4.2 Dual Computing

Dual computing is the utilization of processing power from multiple sources to complete a task. Examples of these processing locations include the cloud, on-premise at an onshore location, or the EM system recording the data (on the edge). Dual compute leverages multiple sources of compute power to allow greater storage, processing, and automated review capacity. Cloud resources or on-premise compute power will usually be much more capable of handling complex operations than the EM recording system on a vessel. An example of dual compute in practice would be running a lightweight AR algorithm on the vessel's EM system to selectively record relevant fishing activity, offload the fishing activity to an onshore processing location, and then use on-premise compute power to run more complex algorithms like object tracking and species recognition. The application of dual compute to semi-autonomous systems can be traced back to early computing such as space and military operations (especially for early satellite and lunar activities[16]) due to similar challenges faced today by EM operations, such as limited bandwidth, network speeds, intermittent connectivity, unreliable links, interference, and unsecure channels. Dual compute (edge and onshore) solutions have been applied by certain providers and groups, but insufficient details of their applications and outcomes exist for direct recommendations. As mentioned previously, further publication of positive and negative results for different tasks and fishers/fisheries would be quite useful.

16 "Apollo 11 Mission Overview - NASA." Accessed: Feb. 01, 2024. [Online]. Available: https://www.nasa.gov/history/apollo-11-mission-overview/

7.5 Simplified Automation Approaches-Alternatives to AI

AI has proven to be more costly, slow to develop and difficult to implement than many agencies and program directors planned for. In some cases, it may be beneficial to explore other technologies to automate tasks. Companies and researchers are actively pursuing and, in many instances, already utilizing other methods to automate processes that don't rely on AI. Although the focus of this report is machine learning-based vision automation, the examples below may provide useful alternatives to improve the efficiencies of an EM program.

- Using more sensors on winches and other equipment to trigger cameras and mark significant events. This is already being done but could be expanded upon through use of threshold sensors and movement sensors.
- Using GPS and speed of vessel to trigger events and send notifications.
- Using more advanced review software and UI hardware to allow for hot keys, macros, and accepting/rejecting inputs as opposed to manually entering data.
- Fish count - there are a lot of AI solutions for this - but there are also more standard algorithms that take AI results and then process count.
- eLog integration that can help expedite data processing by highlighting areas of fishing activity or specific species encounters.

8.0 Considerations for Fisheries Regulatory Bodies Interested in Incorporating AI into EM programs

8.1 Summary

As a consumer of AI, regulatory bodies should be aware of the best practices and standards for AI. Regulation for AI is currently quite limited across industries and around the world. Thanks to generalizable language models, there is a rise in demand for governance and policy development. We recommend regulatory bodies review this development to make use of existing efforts for policy development where other industries and governing bodies have helped lead the way. Two examples of such standards are provided below. [https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html, https://www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf]

How regulatory bodies will vet and approve AI for use in an EM program is one of the most important catalysts or barriers to widespread adoption. The development, testing, and maintenance of AI is costly. Without continued funding or a road to operational use at scale, AI can be difficult for stakeholders to invest in However, AI has the potential to unlock enormous power in fisheries monitoring and with careful implementation it can become a reality.

The current lack of a roadmap for the use of AI in operational programs is understandable. Many EM projects started as EFP or pilot projects with the goal of assessing if EM can capture a set of data. Operational adoption of EM as a replacement for human observers is relatively new. AI is an even more nascent technology in fisheries. This and the developing nature of EM worldwide combine to make data goals and standards a moving target.

However, as operational EM data streams are adapted into existing regulatory frameworks the fishing community will have a greater understanding of what the baseline is for EM data quality. When data quality standards are understood, targets can be developed for AI to meet or exceed. Having operational EM programs will not only provide data quality targets but also more stable review protocols that AI development teams can set their sights on. This will mitigate the impacts of moving goalposts.

Investment into EM automation , particularly via grants has historically been and remains high. Tools are becoming available that could have profound impact on the success of EM programs. However, there are still non-technical hurdles to realizing the desired benefits. Regulatory bodies have a large influence over the potential success of automation in EM and in some cases, they may even act as barrier to such progress. For example many programs, especially ones funded by national fishery agencies, have specific rules in place that require human review of the video. The way that many of these programs are defined prohibits significant use of AI for reducing the amount of video that analysts need to review. While currently 100% automation is not feasible or recommended, there are several hybrid plans for AI augmented review that have merit, including requiring more total review but less human review. An example could be to change a program from 20% human review of all fishing sets to 50% AI review and 5% human review.

Until regulations become more defined, flexible and creative to include uses of AI, many EM vendors will be hesitant to invest too much capital in AI research.

Below are a few considerations for regulatory bodies interested in preparing for automation implementation in their region. Additionally, regulatory bodies should delineate between type of automation when considering implementing these technologies. Tools that provide simple algorithmic or rule-based processing of video or imagery should not be bundled with more advanced technologies like deep learning that generally have limited visibility in their decision making and guarantees, thus require empirical evidence of their capabilities, i.e. metrics calculated using test data. Regulatory bodies are encouraged to recognize [and foster inclusive stakeholder discussions on] the full spectrum of automation in standards creation.

8.2 EM Program Specific AI Standards

How regulators define specific AI standards will depend on the goals of the AI used in the program and the AI tool itself. However, at a minimum regulatory bodies should create standards and require reporting on the following metrics:

- Compute Performance: Speed and memory cost, per frame or standard video size.
- Accuracy: Number of correct predictions vs total number of predictions made.
- Confusion matrix scores (true/false positives, true/false negatives) and their ratios, importantly:
- Recall: How many true positives there are in all of the positives (including those not identified).
- Precision: How many true positives are there compared to all of the positives in object/activity detection.
- Task-specific measures of performance.

Determining acceptable values for the metrics above should be part of program design and should include both program directors and technical consultants. This is an important time to include vendors in the discussion. Important questions to ask are:

- What problem is AI solving and how fast are the results needed?
- Does the model need to run on edge hardware, on premises, or in the cloud?
- What decisions will need to be made based on the results of the AI and what scores should metrics reach to support those decisions?
- Are there policies and regulations in place that may have an impact on these metrics?

Defining these metrics may be a challenging task for a regulatory body. However other industries have created tools to assist. One example which has been used recently for large-language models in machine learning is Model Cards, which provides a summary of important factors in the creation of the model as well as its intended use [https://arxiv.org/abs/1810.03993]. A similar description for AI in EM solutions could be a concise summary of EM task capabilities and intended use-cases, along with critical metrics and benchmarking methods. Similar to other industries, AI applications in EM which involve outputs relating to humans must be subject to bias and fairness checks. These should be more stringent than non-human tasks and should be required to provide additional disclosure such as the model's human-related training data distributions.

8.3 Approval Process and Auditing

Companies that use AI should always inform their customers (i.e., fishing agencies and regulatory bodies) how and when AI is being used.

8.3.1 Documenting Processes

Regulatory bodies should have documented processes for using EM data and AI for automation for any AI program that is not a pilot. While a template may not exist for this process, it would be beneficial if one is developed for use in fisheries. This process should be reviewed on an annual basis as the state of AI is changing rapidly. It may be beneficial for a governing agency to have a broad directive for this but allow individual programs or regions to clearly define their goals and expected costs and outcomes. However, these programs should not be so prescriptive that it requires engineers to create tools that meet the letter of the program's definitions but not the spirit of its goals. Programs should be more performance based rather than meeting specific technical requirements.

8.3.2 Auditing

Regulators should have an audit process for any program that uses AI (similar to audit programs for EM analysts). Until models are proven, the audit rates should be higher than standard EM analysts. Any time a model is deployed (beyond pilots or tests) in a new setting (vessel type, fishery, new species, etc.) it should be audited. Any time a model is updated with new training data or other significant updates it should be audited. A program may consider implementing AI alongside human review and gradually decreasing the human review over time as the AI demonstrates sufficient performance metrics.

8.3.3 Flexibility

Programs should be flexible enough to allow for innovation, not just in technology but in the management of fisheries and the collection of data. Since there are few programs that rely heavily on AI automation at this time, the best practice is to allow for flexibility but to verify often. Regulatory effectiveness will be achieved by designing policies and regulations together with industry and allowing informed adjustments according to changes in the field as well as results. Companies should be responsible for building systems that can effectively be audited and have reproducible results.

8.4 Competing Incentives

Stakeholders looking to reduce review rates to save program costs may inadvertently disincentivize groups looking to invest in automation. Programs that are overly focused on reducing review rates can be short sighted and limit the long-term sweeping impact an automated EM program can provide. The challenge is that financial considerations often drive near term decisions to reduce review rates. Roadmaps, stepped approaches, and even short term subsidies for review can go a long way to providing a healthier ecosystem and allow time for automation to demonstrate its value.

9.0 Conclusion

The information provided in this report is intended to provide insight and lessons learned regarding EM automation to help facilitate its broader use. The report provides information for a variety of stakeholders currently engaging or looking to engage in automating EM programs. The authors made a concerted effort to relay lessons learned from their own experiences in hopes to assist others in effectively applying their resources.

Stakeholders should reference this report for technical guidance when planning to automate an EM program. This report highlights that successful automation development is multifaceted, and managers must consider the interaction of systems design, hardware specifications, algorithm considerations, data collection and review software interaction to succeed. In addition, this report calls out the need to examine the return on investment and regulatory considerations of automating an EM program. To date, many projects have suffered from lack of attention in these areas and have not met expectations. Stakeholders are urged not to overlook these less technical considerations. A thorough and honest return on investment analysis should be a cornerstone of any automation effort. With careful approaches, planning, and cross discipline coordination the outlook is bright. Continued advancements in technology and the EM community's innovation are paving the way for huge impact.

In line with increased diligence and planning around EM automation projects there also is a need to think more holistically. There is an opportunity to evolve the way EM automation is being approached. Some EM programs fall short by simply looking to replace human observers instead of recognizing they can provide an entirely new way of observation. EM automation in many cases has repeated those shortcomings by simply trying to make programs cheaper or faster. Many programs turn to automation to reproduce the same results as traditional EM analysis thus constraining the opportunity to the limitations of current EM analysis. Those working in EM, and in particular those working in EM automation have an opportunity to think beyond cheaper and faster to envision totally new fisheries data systems that are only now becoming possible. A shift in thinking of these tools as sweeping opportunities opposed to one-to-one replacements of our current systems will provide an entirely new realm of impact to be driven EM and automation.

10.0 Glossary

Important note on usage of key terminology in this document:

Many words will be used slightly differently to their conventional meanings. They have nuanced meanings and each of their meanings should be understood.

In this document, we will refer to automation, artificial intelligence, deep learning and machine learning interchangeably. In order of most general to most specific we have: automation, artificial intelligence, machine learning, deep learning.

We will also refer to [artificial intelligence, machine learning, deep learning] models and algorithms interchangeably. In order of most general to most specific we have: algorithm, models, AI model, ML model, DL model.

- **Algorithm:** A sequence of steps to accomplish a task. In the context of this document, it is typically software-based.
- **Artificial Intelligence (AI):** All-encompassing terminology for any intelligence that is embedded in technology. Also called machine intelligence.
- **At-Sea Monitor (ASM):** An at-sea scientist whose primary focus is collecting data on groundfish species. ASMs deploy on fishing vessels on groundfish trips in New England.
- **Automation:** Solution for making 1 or numerous actions that previously required some amount of manual, human input to occur without intervention. NOTE: Not a binary term e.g., automation can refer to separate pieces of a workflow which have human/manual intervention in between.
- **Catch Per Unit Effort (CPUE):** The amount of fish caught per unit of fishing effort.
- **Deep Learning:** Subfield of Machine Learning which uses parameters in the structure of a neural network to produce data-informed models. Called deep because neural networks are typically wide and deep in shape.
- **Dockside Monitor (DSM):** Personnel that meet MREM vessels when they return from a trip and offload catch. They collect data on groundfish catch.
- **ETP:** Endangered, threatened, or protected species.
- **Electronic Monitoring:** The use of a combination of computers, cameras, and other sensors to collect data on fishing vessels.
- **Electronic Logbook (eLog):** A report filled out by the fisherman that details their trips. Formats are highly variable but often include the number of fishing efforts, location of effort, and catch composition.
- **Gear Sensors:** Equipment like hydraulic sensors and rotation sensors that can provide non-visual data to an EM system. Spikes in hydraulic pressure or rotation of spools can indicate fishing activity.

- **Groundfish:** In the New England groundfish monitoring projects i.e. ASM, MREM, and Audit, the following species are grouped under groundfish: Atlantic cod, haddock, pollock, white hake, Atlantic halibut, winter flounder, American plaice flounder, yellowtail flounder, Acadian redfish, witch flounder, ocean pout, windowpane flounder, and Atlantic wolffish.

- **Machine Learning (ML):** Field of statistics and computer science where models are produced to try to understand, explain and/or replicate some phenomenon as closely to a ground truth as possible. These are typically capable of generating or predicting outputs using "learned" parameters.

- **Open Datasets/Libraries:** Compilation of data into a collection, sometimes called library. Data will be typically structured according to relationships and usage and ready for statistical analysis or model training. The term "open" refers to the fact that it is provided publicly, possibly with a license that allows for a certain level of free usage. Not to be confused with open-source software libraries.

- **Training Data:** In this document, this refers to data used to train machine learning or deep learning models.

- **Trawl:** A fishing gear type where a net is pulled behind the vessel. The net is hauled in and catch is sorted on the deck of the vessel.

- **Vessel Monitoring Plan (VMP):** A document shared between the vessel, EM provider, and the regulating body. The VMP outlines the EM system, project objectives, and catch handling requirements for an EM Program.

- **Video Analyst/Reviewer:** Interchangeable term for personnel who watch EM footage and record data.