Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)
Special Track on AI and Social Good

Exploring Multimodal Foundation AI and Expert-in-the-Loop for Sustainable Management of Wild Salmon Fisheries in Indigenous Rivers

Chi Xu[1] , **Yili Jin**[1,2] , **Sami Ma**[1] , **Rongsheng Qian**[1] , **Hao Fang**[1] , **Jiangchuan Liu**[1] , **Xue Liu**[2] , **Edith C.H. Ngai**[3] , **William I. Atlas**[4] , **Katrina M. Connors**[5] , **Mark A. Spoljaric**[6]
[1]Simon Fraser University, Vancouver, Canada
[2]McGill University, Montréal, Canada
[3]The University of Hong Kong, Hong Kong, China
[4]Wild Salmon Center, Portland, USA
[5]Pacific Salmon Foundation, Vancouver, Canada
[6]Haida Fisheries Program, Skidegate, Canada
chix@sfu.ca, yili.jin@mail.mcgill.ca, {masamim, rqa4, fanghaof, jcliu}@sfu.ca,
xueliu@cs.mcgill.ca, chngai@eee.hku.hk, watlas@wildsalmoncenter.org,
kconnors@psf.ca, mark.spoljaric@haidanation.com

Abstract

Wild salmon are essential to the ecological, economic, and cultural sustainability of the North Pacific Rim. Sustaining their populations requires effective fisheries management, which remains challenged by climate variability, habitat loss, and limitations in data collection and analytics, particularly in remote ecosystems with limited infrastructure. This research project explores the integration of multimodal foundation AI and expert-in-the-loop frameworks to enhance wild salmon monitoring and sustainable fisheries management in Indigenous rivers across Pacific Northwest. With video- and sonar-based monitoring, we develop AI-powered tools for automated species identification, counting, and length measurement, reducing manual effort, expediting result delivery, and improving decision-making accuracy. Expert validation and active learning frameworks ensure ecological relevance while reducing annotation burdens. To address unique technical and societal challenges, we bring together a cross-domain, interdisciplinary team of university researchers, fisheries biologists, Indigenous stewardship practitioners, government agencies, and conservation organizations. Through these collaborations, our research fosters ethical AI co-development, responsible data sharing, and culturally informed fisheries management.

1 Problem Statement

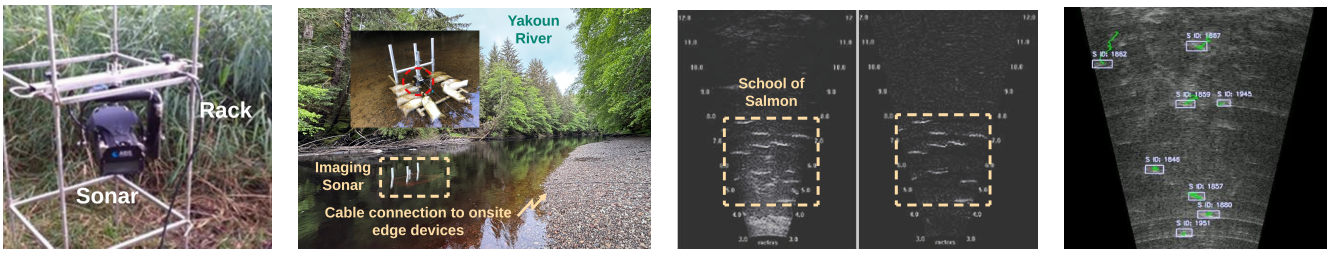
Wild salmon are integral to the social-ecological systems of the North Pacific Rim. For over 10,000 years, they have supported thriving fisheries [Yoshiyama, 1999; Carothers *et al.*, 2021], sustained local economies, enriched cultures, and maintained ecological balance [Walsh *et al.*, 2020; Earth Economies, 2021]. Yet rapid environmental changes driven by climate variability are threatening the resilience of salmon ecosystems [Waples *et al.*, 2008; Di Lorenzo and Mantua, 2016; Frölicher and Laufkötter, 2018]. Across their range, wild salmon populations have declined significantly, with increasingly unpredictable returns [Kilduff *et al.*, 2015; Dorner *et al.*, 2018]. These declines pose significant threats to their long-term sustainability and the communities, such as Indigenous people, that depend on them [Atlas *et al.*, 2021].

Sustaining salmon fisheries is further complicated by mixed-stock fisheries, which indiscriminately harvest co-migrating populations [Walters *et al.*, 2008; Moore *et al.*, 2021], and by the high costs and logistical challenges of monitoring in remote, roadless areas of the Pacific Northwest [Price *et al.*, 2017]. These challenges have created the need for adaptive AI models and systems that support in-season management and selective terminal fisheries targeting healthy populations [Atlas *et al.*, 2021]. Such AI models and systems can bolster ecosystem resilience and maintain productivity through cycles of salmon abundance, even amidst climate change [Schindler and Hilborn, 2015].

The initial effort to integrate computer vision and artificial intelligence into salmon monitoring focused on video-based weir systems aimed to expedite in-season fish counting [Atlas *et al.*, 2023], a process traditionally requiring extensive manual effort. Early approaches leveraged underwater RGB cameras; however, the limitations of visual clarity and environmental variability necessitated alternative sensing modalities. Sonar-based monitoring emerged as a viable solution, exemplified by efforts such as Caltech's Fish Counting (CFC) work [Kay *et al.*, 2022], which introduced fish detection and tracking in sonar videos. Unlike conventional Multi-Object Tracking (MOT) datasets focused on urban environments, CFC highlights the challenges of domain generalization in low signal-to-noise underwater settings [Kay *et al.*, 2024]. SALINA [Xu *et al.*, 2024] further extended these efforts by enabling real-time sonar analytics and energy-efficient deployment, supporting sustainable fisheries management in re-

(a) (b) (c)

Figure 1: **(a)** A salmon counting weir at Koeye River (in Heiltsuk First Nation's traditional territory, northern British Columbia) with salmon swimming passing the fish channel, **(b)** sample underwater video frames with salmon appearances, **(c)** object segmentation with species identification.



(a) (b) (c) (d)

Figure 2: **(a)** A mounted ARIS sonar camera, **(b)** sonar deployment in the Yakoun River, Haida Nation's traditional territory, **(c)** sample frames from ARIS sonar, **(d)** salmon detection and tracking in sonar frames.

mote ecosystems within Indigenous territories.

Despite these advancements, new challenges and opportunities remain. Moving beyond basic fish object detection and tracking, there is a growing need to integrate fishery management models and forecasting with motion-based counting, length measurement, and AI-assisted decision-making. Incorporating the newly developed vision foundation model [Achiam *et al.*, 2023; Liu *et al.*, 2023] also helps improve performance and generate timely, accurate insights.

For effective AI system deployment, ensuring reliable data capture, sufficient energy availability, and stable long-term operation is essential. However, the lack of basic infrastructure in the remote forests of the Pacific Northwest makes deployments particularly challenging [Ma *et al.*, 2024]. Therefore, integrating the expertise of Indigenous stewardship practitioners and fisheries biologists becomes even more critical to enhancing the system's resilience and applicability. Furthermore, expert-in-the-loop frameworks and cross-referencing across multiple sensing modalities hold promise for improving accuracy, robustness, and actionable insights for sustainable fisheries management [Wu *et al.*, 2022].

In response to this need, we formed an interdisciplinary collaborative team to apply multimodal foundation AI and co-develop expert-in-the-loop frameworks for: (1) automated species identification and counting of salmon from video generated at salmon counting weirs, and (2) automated tracking, counting, and length measurement from in-river sonar camera units. As shown in Figures 1 and 2, these two domains are complementary: video-based analysis can provide accurate estimates of species abundance passing through a fixed location, while sonar-based techniques can monitor salmon across the entire river width without the need to build special-purpose weirs and fish channels.

A key objective of this research project is to strengthen collaboration among university researchers, conservation practitioners, and Indigenous communities who lead the stewardship of wild salmon within their territories. Insights from multimodal data can enhance our shared capacity to detect, interpret, and analyze salmon activity, deepening our understanding of species behavior such as run timing, abundance, and year-to-year variation. These insights support broader efforts to protect wild salmon and to better understand the ecological and social factors shaping their health. Moreover, data-informed improvements to fish passage and habitat conditions can help address ongoing challenges related to biodiversity and ecosystem change.

2 Alignment with Sustainable Development Goals and the LNOB principle

The United Nations' 2030 Agenda for Sustainable Development [United Nations, 2015] outlines 17 goals to promote global prosperity, equality, and sustainability. A core principle, Leaving No One Behind (LNOB) [United Nations, 2021], upholds the commitment that all communities, especially vulnerable groups, benefit from development efforts. Wildlife conservation plays a crucial role in sustainability, affecting both ecosystems and human livelihoods [Liu *et al.*, 2024; Gordon *et al.*, 2023; Kshitiz *et al.*, 2023]. This research project, emphasizing wild salmon monitoring, management, and conservation, aligns with several Sustainable Development Goals (SDGs) as follows.

SDG 14: Life below Water. The multimodal foundation AI for monitoring and conserving wild salmon aligns with SDG 14.2 ("Protect and restore ecosystems") by supporting habitat preservation and ecosystem resilience. Meanwhile, it

Prompt: Identify the fish in this video frame and explain what features led you to that conclusion.

Vison Language Model: This is a spawning-phase sockeye salmon. You can tell by the deep red body and grayish/greenish head (rather than bold vertical stripes or heavy spotting), which is characteristic of sockeye as they mature and enter fresh water to spawn.

Prompt: why it is not a Chinook?

Vison Language Model: Because Chinook salmon (especially in spawning phase) typically show black spots on the back and both lobes of the tail, as well as a black gum line, none of which are evident here. Instead, this fish shows the uniform deep-red body and pale head typical of spawning-phase sockeye..

Prompt: Identify the fish in this video frame and explain what features led you to that conclusion.

Vison Language Model: Likely a **Chinook** (king) salmon. Its large, robust body; silvery-bronze coloration; and dark jaw area are characteristic of Chinook. The early-summer timing also aligns with typical Chinook runs.

Expert Comments: It is a sockeye salmon, not a Chinook. Sockeye have no black spots on their back or tail and have lighter (white) gums, whereas Chinook usually show small black spots (especially on the tail) and have black gums.

Figure 3: An off-the-shelf vision language model (e.g., OpenAI o1) can identify some cases but also makes errors, thus requiring further refinement.

contributes to SDG 14.4 ("Regulate harvesting and end overfishing") by enabling sustainable fisheries management and aiding in the conservation of marine biodiversity.

SDG 15: Life on Land. Wild salmon sustain both aquatic and terrestrial ecosystems. Their migrations transfer nutrients to forests and wildlife. Declining populations disrupt these cycles, affecting species such as bears and eagles. The management and conservation efforts in this research project, such as data-driven fish passage improvements and habitat restoration, address biodiversity loss and degradation of terrestrial and inland freshwater ecosystems (aligning with SDGs 15.1 and 15.4).

SDG 17: Partnerships for the Goals. Sustainable salmon conservation requires collaboration with government agencies, Indigenous rights holders, and diverse stakeholders. In this project, our interdisciplinary team includes Indigenous communities, university researchers, conservation organizations, and industry partners. Integrating Indigenous knowledge with modern technologies, such as AI-powered monitoring, enhances conservation strategies and culturally informed approaches. These efforts align with SDG 17.16 ("Strengthen partnerships through knowledge and resource sharing") and SDG 17.17 ("Encourage multi-stakeholder collaborations").

Alignment with LNOB. Wild salmon are essential to Indigenous culture, economy, and food security, yet declining stocks exacerbate food insecurity and economic hardship. LNOB promotes equitable conservation efforts by recognizing Indigenous communities as key decision-makers in managing local salmon populations, strengthening co-governance and long-term fishery access. AI-powered monitoring enhances data analytics and fisheries management, contributing to ecosystem conservation. When integrated with traditional Indigenous knowledge, AI supports more culturally grounded policy development, strengthens sustainability initiatives, and reinforces Indigenous sovereignty in conservation.

3 Strategies and Methods

To address key challenges in salmon monitoring and management, we integrate multimodal foundation AI, scalable user-centric system design, and collaborative governance. The following subsections detail our methods for enhancing perception across video and sonar data, supporting real-time deployment in remote environments, and enabling responsible, community-aligned data sharing and model development.

3.1 Multimodal Foundation AI for Monitoring

Wild salmon monitoring requires accurate species identification, counting, and length measurement. Video-based and sonar-based approaches offer complementary advantages, yet each faces unique challenges. In this research project, we explore multimodal foundation AI to improve model accuracy, reduce annotation burden, and enhance AI explainability.

Video-based monitoring at salmon counting weirs uses underwater RGB cameras to capture detailed visual features, yet occlusion, environmental variability, and data imbalance affect accuracy [Khan *et al.*, 2023]. Salmon frequently overlap in dense aggregations, making single-camera detection unreliable. To address this, we leverage multi-view fusion, where synchronized cameras or optical mirrors at different angles provide complementary perspectives. This approach reveals occluded fish objects and trajectories. Species identification also suffers from data skew, particularly when rare species are underrepresented. We apply data augmentation techniques such as synthetic image generation and class-balanced sampling [Cui *et al.*, 2019] to mitigate dataset imbalance.

For low-confidence salmon detections and classifications in videos, we leverage pre-trained vision language models (e.g., LLaVA [Liu *et al.*, 2023], GPT-4o [Zhong *et al.*, 2024], and OpenAI o1 [OpenAI, 2024]) that generate natural language descriptions and species identifications. As shown in Figure 3, while an off-the-shelf vision language model can identify some cases, it also produces errors compared to expert feedback from fisheries biologists. Therefore, expert validation is necessary to refine model predictions further. Another issue we identified during annotation for this domain-specific task is that inexperienced annotators further introduce labelling errors, thus reducing model reliability. To enhance annotation quality and model interpretability, we also integrate vision language models into the annotation phase, helping to minimize errors.

Sonar-based monitoring in rivers enables salmon detection, tracking, and counting in turbid environments, but presents challenges in noise reduction, spatial-temporal modeling, and cross-modal integration [Xu *et al.*, 2024]. Sonar data con-

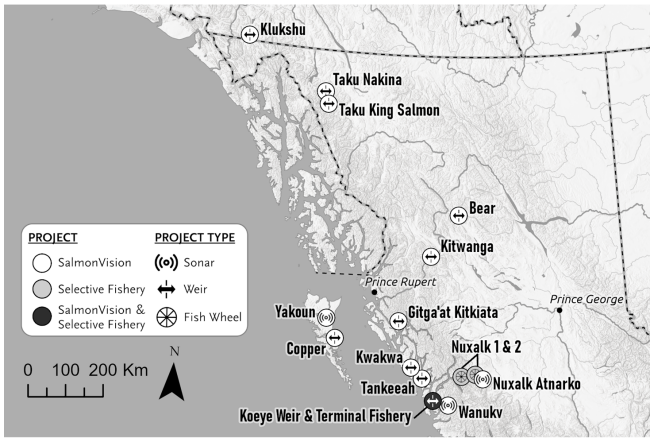


Figure 4: This research project supports SalmonVision & Selective Fishery in multiple Indigenous rivers in British Columbia, Canada.

tains substantial background noise due to environmental factors such as water turbulence and reflections. In this project, we employ deep learning-based denoising models [Garber and Tirer, 2024; Chihaoui and Favaro, 2024] trained on sonar datasets to enhance signal clarity. Existing sonar-based tracking systems struggle with false positives and temporal inconsistencies. To improve salmon tracking performance across frames, we incorporate spatial-temporal features using Transformer-based architectures, which will be further introduced in our implementation plan (Section 4.3). Notably, sonar frames alone lack species-level resolution, limiting classification accuracy. To address this, we synchronize sonar and video data when they are both available, integrating features through early fusion techniques. By aligning spatial and temporal cues with multimodal inputs, our approach enhances tracking and counting performance.

3.2 Scalable and User-Centric Framework Design

Scalability and real-time performance are critical for deploying AI-powered monitoring across different Indigenous rivers to generate in-season insights for fisheries management. Both video and sonar data streams generate large volumes of high-dimensional data, necessitating efficient processing pipelines. Limited power, computation resources, and network connectivity in the remote forests of the Pacific Northwest further exacerbate challenges [Xu *et al.*, 2024]. To this end, we have implemented a hybrid edge-cloud architecture where lightweight models perform initial detection on edge devices, reducing computational demand and transmission costs. More complex tasks, such as fine-grained species classification and anomaly detection, are offloaded to cloud servers. Such task offloading maintains computational efficiency without sacrificing accuracy. To further improve model adaptability, we incorporate continual learning mechanisms that update model parameters based on multi-year data while preserving previously learned patterns.

In summary, our framework design enables collaborative edge–cloud operation while maintaining remote access for diverse stakeholders. Edge-based computing allows immediate fish detection at monitoring sites, even in locations with limited internet connectivity. Meanwhile, cloud integration supports large-scale data storage, remote model updates, and collaborative access to processed data. We also explore federated learning [Liu *et al.*, 2020], which enhances privacy and Indigenous data sovereignty by enabling model improvements without direct data transfer between monitoring sites.

Beyond system architecture, the adoption of AI-powered monitoring tools also depends on usability. We design a user-friendly application interface that enables Indigenous fisheries biologists to access real-time AI-generated insights with minimal technical expertise. Standardized hardware and software integration supports seamless deployment across different monitoring locations within Indigenous territories, as shown in Figure 4. To further promote accessibility, we provide open-source documentation and training resources, allowing fisheries practitioners to deploy and maintain the system without specialized AI knowledge. By integrating scalable computing solutions with intuitive design, we create an adaptive and inclusive monitoring framework that incorporates fisheries experts in the loop.

3.3 Collaboration and Responsible Data Sharing for Effective Fisheries Management

Effective fisheries management relies on transparent processes, cross-sector collaboration, and adaptive decision-making. In this project, we work closely with government agencies, Indigenous communities, and conservation organizations to develop monitoring strategies that are both scientifically rigorous and culturally informed. To support this collaboration, we share carefully selected non-sensitive datasets and AI models for non-commercial use, in line with data governance agreements and in consultation with partner Indigenous communities. These curated resources encourage innovation in AI and fisheries science, while upholding ethical standards and respecting data sovereignty.

By integrating multimodal foundation AI, real-time monitoring systems, and expert-in-the-loop frameworks, we aim to support a shift from data-limited, forecast-based fisheries management to more adaptive, in-season decision-making. This transition enables fisheries experts and managers to respond dynamically to changing environmental conditions, improving conservation outcomes and enabling sustainable harvest opportunities when appropriate. Project findings will be disseminated through peer-reviewed publications in both AI and fisheries research communities. To support reproducibility and technical collaboration, we plan to maintain open-access repositories of code and selected non-sensitive data[1]. All data releases will comply with established agreements and be subject to review by partner communities. Through responsible technology development and inclusive partnerships, we contribute to a more resilient, data-informed framework for wild salmon stewardship.

4 Implementation Plan

We implement our strategies and methods through a combination of expert-in-the-loop workflows, multimodal model integration, and deployment strategies tailored for video and sonar-based salmon monitoring in remote environments.

[1]https://github.com/Salmon-Computer-Vision

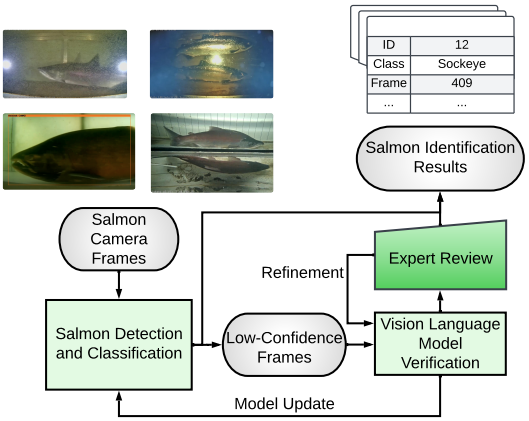


Figure 5: Vision language model verification and refinement.

4.1 Expert-in-the-Loop for Salmon Monitoring

Building on our previous work, we have developed the SalmonVision web app[2], which enables user-led data review and annotation of salmon detection and classification results generated on the edge. This research project extends that groundwork by exploring multimodal foundation AI to transform video and sonar data into actionable insights for salmon monitoring, also emphasizing the critical role of human expertise in the loop. We are currently enhancing SalmonVision web app to include the following features: 1) data collected from different monitoring sites within Indigenous territories is processed using advanced AI models capable of integrating multiple modalities, 2) AI-generated outputs (detections, counts, and species classifications) are further refined through experts' multimodal input, including dot annotations, bounding boxes, and text prompts.

The AI-human collaborative workflow ensures that expert knowledge informs every stage of the process. Fisheries experts contribute their domain expertise by validating and enhancing annotations and AI-generated outputs, creating a rich dataset of labeled frames for AI model refinement. This iterative process strengthens AI models' ability to generalize and perform reliably in real-world conditions, enabling them to better adapt to the unique environmental characteristics of each monitoring site within Indigenous territories. Once refined, the AI models are updated on edge-computing systems installed at monitoring sites. These systems operate autonomously to analyze incoming data in real time, but the process remains firmly anchored by human oversight. Fisheries practitioners provide ongoing feedback and technical support to support continuous system operation and data accuracy throughout the monitoring season.

4.2 Video-based Salmon Detection, Classification, and Counting

The existing system at our salmon counting weir employs single-modality models such as YOLO [Wang *et al.*, 2024] and RT-DETR [Zhao *et al.*, 2024] for salmon detection, classification, and counting. While these models achieve reasonable performance, they struggle with cases involving occlusions, poor lighting, or rare species due to their reliance on visual features alone. Misclassifications and low-confidence detections introduce errors that require extensive manual verification. To address these limitations, we incorporate a vision language model (VLM) that enhances explainability and integrates additional modalities, improving both detection and classification accuracy and expert review efficiency.

[2]https://salmonvision.org/

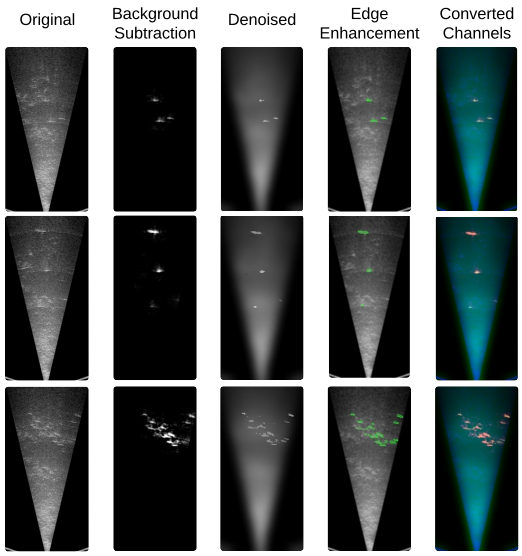


Figure 6: Traditional preprocessing of sonar frames.

As shown in Figure 5, our implementation leverages a VLM to refine low-confidence cases where the confidence scores of the base detection and classification model are low. Instead of relying solely on pixel-based features, the VLM generates descriptive textual explanations of its predictions, providing interpretable insights into classification decisions. This process involves prompt engineering to guide the model in handling specific challenges, such as distinguishing between visually similar species. When uncertainty remains high, both the text-based explanations and the corresponding video frames are flagged for expert review. By combining human expertise with model-driven reasoning, we facilitate the correction of misclassifications and their incorporation into the model's continuous learning process.

As expert-reviewed frames accumulate, the refined VLM progressively improves its performance, reducing reliance on manual verification over time. The overall system transitions from a semi-automated workflow to a fully automated one powered by multimodal foundation model, capable of replacing single-modality approaches. This shift enhances scalability by enabling high-accuracy fish monitoring with minimal human intervention.

4.3 Sonar-based Monitoring

The current sonar-based monitoring system in Indigenous rivers relies on traditional preprocessing techniques to enhance image quality for expert review and AI inference, as shown in Figure 6. However, these preprocessing methods often introduce frame distortion and feature loss, degrading the performance of downstream tasks such as salmon detection, tracking, counting, and length measurement. To address these

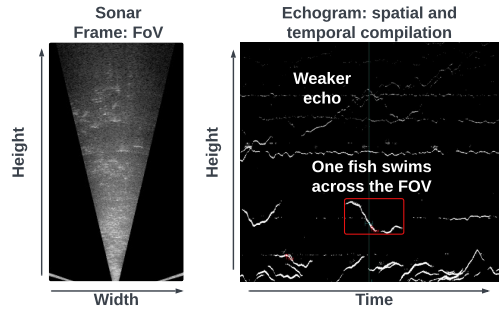


Figure 7: Sonar frames and echogram, as two different modalities.

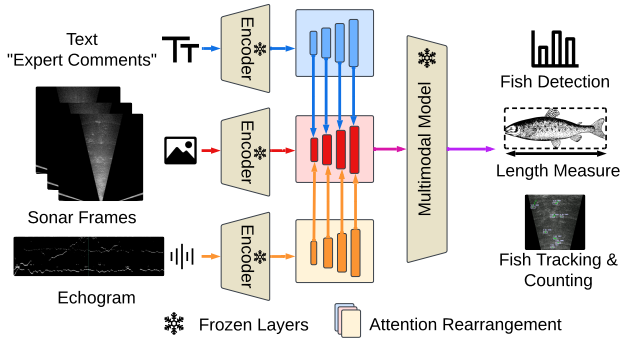


Figure 8: Multimodal foundation model for sonar domain.

challenges, we propose adapting SAM2 model [Ravi *et al.*, 2024], a recently published foundation model for these downstream tasks with multimodal inputs, including sonar frames and echograms. As shown in Figure 7, a sonar echogram is a time-series visualization of sonar returns, representing how acoustic signals interact with underwater objects and the riverbed over time. The echogram serves as a key multimodal input, providing both spatial and temporal information to the adapted foundation model. By integrating echograms with sonar frames, we improve robustness against noise and enhance the performance of downstream tasks.

The first phase of our implementation focuses on integrating multimodal data to improve representation learning. Sonar frames and echograms are fused to create a more comprehensive input representation. To suppress noise while preserving critical information, we employ a lightweight foundation model such as CLIP [Radford *et al.*, 2021] to encode sonar frames and echograms into token representations. Unlike traditional denoising techniques, this approach reduces artifacts and prevents loss of essential details.

The second phase focuses on adapting the multimodal SAM2 model to sonar-specific tasks and improving performance through fine-tuning. A labeled dataset of sonar frames and echograms, annotated with salmon appearances, positions, and length estimates, is prepared for training.

We apply transfer learning to the adapted SAM2 model, freezing early layers while fine-tuning later layers for sonar data. As shown in Figure 8, sonar frames and echograms are processed through separate encoders, facilitating effective multimodal integration. The extracted features are combined using attention-based fusion, incorporating expert comments as additional inputs when available. This structured fusion approach enhances the model's ability to generalize under varying noise and environmental conditions, improving its adaptability to real-world sonar data.

For salmon detection, the adapted SAM2 model generates segmentation masks or bounding boxes, refined using non-maximum suppression (NMS) to remove redundant predictions. Salmon tracking and counting are further performed using DeepSORT [Wojke *et al.*, 2017], which combines motion and appearance features to ensure stable object association across frames. Centerline extraction is conducted using skeletonization algorithms [Hao *et al.*, 2015], which refine fish contours for improved structural analysis. Length measurement is achieved through attention-based feature rearrangement, where extracted features are mapped to real-world metrics using known scaling factors. This approach ensures robust and accurate analysis of salmon instances in challenging underwater environments by leveraging both sonar frames and echogram signals.

5 Evaluation Criteria

5.1 Video-Based Salmon Species Identification and Counting

For species identification, we evaluate multi-class classification performance using mean Average Precision at IoU 0.5 (mAP@50) and F1 score [Goodfellow, 2016]. These metrics assess precision and recall, with robustness across different species. We compare our VLM-enhanced identification results against baseline performance of single-modality models such as YOLO [Wang *et al.*, 2024] and RT-DETR [Zhao *et al.*, 2024], measuring improvements in classification performance. For salmon counting, we assess the counting accuracy using Mean Absolute Percentage Error (MAPE) and F1 score, which represents reliable enumeration in dense and occluded scenarios. Our approach is benchmarked against standard detection-based counting models, with improvements evaluated in terms of both precision and computational efficiency. We also compare VLM-enhanced species counts to expert-reviewed species counts to provide insight into VLM-enhanced salmon counting in real-world deployments.

Preliminary Results. Initial experiments were conducted on a subset of annotated video data from two Indigenous monitoring sites. The VLM-enhanced species identification model achieved an mAP@50 of 81.2%, compared to 72.5% from the YOLO baseline. The average F1 score improved from 0.76 to 0.83, particularly benefiting rare species with limited training examples. For salmon counting, our method reduced the Mean Absolute Percentage Error (MAPE) from 18.7% (YOLO baseline) to 11.4%, with consistent improvements across both high-density and occluded scenes. These preliminary results demonstrate the potential of our approach in improving both accuracy and reliability in real-world salmon monitoring deployments.

5.2 Sonar-Based Salmon Monitoring

For sonar-based salmon detection performance, we also report mAP@50 and mAP@50:75, with the latter further capturing precision-recall trade-offs across different IoU thresholds. These metrics quantify the effectiveness of our multi-

modal approach in handling sonar-specific challenges such as substantial noise and low contrast. We compare against traditional contour-based methods and state-of-the-art baselines, including CFC-YOLO [Kay *et al.*, 2022], RT-DETR [Zhao *et al.*, 2024], and STSVT [Xu *et al.*, 2024].

For tracking evaluation, we employ Multiple Object Tracking Accuracy (MOTA [Bernardin and Stiefelhagen, 2008]), Higher Order Tracking Accuracy (HOTA [Luiten *et al.*, 2021]), and IDF1 [Ristani *et al.*, 2016] as benchmarks. MOTA quantifies the trade-off between missed detections, false positives, and identity switches, while HOTA incorporates temporal consistency and object association. IDF1 evaluates the accuracy of maintaining consistent object identities over time. For salmon counting, we evaluate numerical accuracy using Mean Average Error (MAE) and Root Mean Squared Error (RMSE), comparing the model with expert counts. For salmon length estimation, we apply similar MAE and RMSE metrics to measure the deviation between model-predicted and manually measured salmon lengths. Evaluations are also conducted against expert-reviewed sonar data to validate its reliability in real-world deployment. These criteria support reliable assessment and practical deployment of our adapted model in Indigenous monitoring settings.

6 Expected Results and Impacts

Diverse foundation models with multimodal inputs are transforming society at an unprecedented rate; however, these AI models have rarely been co-developed with local or Indigenous communities. In this research project, we co-develop AI models in collaboration with Indigenous communities, government agencies, and conservation practitioners across the North Pacific Rim. Our work aims to create lasting benefits for fisheries management and conservation while supporting equitable co-governance, empowering communities as decision-makers and stewards of local salmon populations.

Across the Pacific Northwest, thousands of locally adapted wild salmon populations remain unmonitored, despite being actively targeted in ongoing fisheries across marine and freshwater ecosystems. In an era of rapid climate change with no historical precedent, our work advances fisheries management by integrating multimodal foundation AI, real-time monitoring, and expert validation to enable adaptive, data-driven decision-making. Enhanced accuracy in fish population assessments will strengthen conservation strategies, establish management benchmarks for previously data-limited populations, and mitigate overfishing risks while supporting sustainable harvest opportunities.

Sharing our datasets and models will accelerate research and innovation, fostering collaboration across AI, fisheries science, and conservation communities. Our cross-domain, interdisciplinary team will ensure that monitoring strategies are both scientifically rigorous and culturally relevant, enabling the translation of research into actionable fisheries management outcomes. By shifting from preseason forecasting to adaptive in-season management, this research project will provide resilient, responsive tools for the sustainable management of wild salmon fisheries in an increasingly dynamic environment.

7 Assumptions and Risks

Applying multimodal foundation AI to wild salmon monitoring and fisheries management is an emerging area with inherent risks. In particular, model performance for automated detections, tracking, counting, and length measurement across different sites remains uncertain, requiring multiple iterations of training and testing to meet fisheries management standards. However, our preliminary results and deployment experiences suggest that automated model analysis is not only feasible but also critical for efficiently reviewing multimodal data. To mitigate these risks, we incorporate multiple rounds of data annotation, model training, testing, expert verification, and refinement to iteratively improve performance.

In addition, careful attention to data ownership and AI-related risks is essential to the ethical co-development of multimodal foundation AI with Indigenous communities. To address these concerns, we follow an iterative co-development process in which partner First Nations receive regular updates, contribute input on research outcomes and tool development, and have their priorities and concerns addressed throughout. Data-sharing agreements are established between our team and each partner Indigenous community. These agreements safeguard community ownership of raw data outputs while allowing access to labeled, non-sensitive data for model training. Communities retain the right to withdraw from these agreements at any time and request the removal of their data from shared repositories. To date, no community has exercised this option.

Ethical Statement

This research is grounded in ethical, inclusive, and culturally respectful practices. We actively collaborate with Indigenous communities to ensure that fisheries monitoring strategies incorporate traditional ecological knowledge and uphold Indigenous data sovereignty. All data collection and analysis will follow established ethical guidelines, including consent-based participation and transparent data use. We prioritize fair representation by ensuring that all stakeholders—including Indigenous communities, government agencies, and conservation organizations—are equitably involved in decision-making and share in the benefits of the research.

In addition, we adhere to responsible AI principles, ensuring that models are interpretable, unbiased, and aligned with conservation goals. By promoting responsible collaboration and ethical data governance, this research supports sustainable fisheries management while upholding scientific integrity and social responsibility.

Acknowledgements

This research is supported by an NSERC Discovery Grant, a British Columbia Salmon Recovery and Innovation Fund (BCSRIF_2022_401), and a MITACS Accelerate Cluster Grant. It also received additional funding from experiment.com. We are grateful for the trust and collaboration of the Heiltsuk, Haida, Kitasoo Xai'xais, Taku River Tlingit, and Gitga'at First Nations, as well as the Skeena Fishery Commission and Gitanyow Fisheries Authority for their ongoing partnership in this work.

References

[Achiam *et al.*, 2023] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[Atlas *et al.*, 2021] William I Atlas, Natalie C Ban, Jonathan W Moore, Adrian M Tuohy, Spencer Greening, Andrea J Reid, Nicole Morven, Elroy White, William G Housty, Jess A Housty, et al. Indigenous systems of management for culturally and ecologically resilient pacific salmon (oncorhynchus spp.) fisheries. *BioScience*, 71(2):186–204, 2021.

[Atlas *et al.*, 2023] William I Atlas, Sami Ma, Yi Ching Chou, Katrina Connors, Daniel Scurfield, Brandon Nam, Xiaoqiang Ma, Mark Cleveland, Janvier Doire, Jonathan W Moore, et al. Wild salmon enumeration and monitoring using deep learning empowered detection and tracking. *Frontiers in Marine Science*, 10:1200408, 2023.

[Bernardin and Stiefelhagen, 2008] Keni Bernardin and Rainer Stiefelhagen. Evaluating multiple object tracking performance: The clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008:1–10, 2008.

[Carothers *et al.*, 2021] Courtney Carothers, Jessica Black, Stephen J Langdon, Rachel Donkersloot, Danielle Ringer, Jesse Coleman, Erika R Gavenus, Wilson Justin, Mike Williams, et al. Indigenous peoples and salmon stewardship: a critical relationship. 26(1), 2021.

[Chihaoui and Favaro, 2024] Hamadi Chihaoui and Paolo Favaro. Masked and shuffled blind spot denoising for real-world images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3025–3034, 2024.

[Cui *et al.*, 2019] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9268–9277, 2019.

[Di Lorenzo and Mantua, 2016] Emanuele Di Lorenzo and Nathan Mantua. Multi-year persistence of the 2014/15 north pacific marine heatwave. *Nature Climate Change*, 6(11):1042–1047, 2016.

[Dorner *et al.*, 2018] Brigitte Dorner, Matthew J Catalano, and Randall M Peterman. Spatial and temporal patterns of covariation in productivity of chinook salmon populations of the northeastern pacific ocean. *Canadian Journal of Fisheries and Aquatic Sciences*, 75(7):1082–1095, 2018.

[Earth Economies, 2021] Earth Economies. The sociocultural significance of pacific salmon to tribes and first nations. *Special Report to the Pacific Salmon Commission*, 2021.

[Frölicher and Laufkötter, 2018] Thomas L Frölicher and Charlotte Laufkötter. Emerging risks from marine heat waves. *Nature communications*, 9(1):650, 2018.

[Garber and Tirer, 2024] Tomer Garber and Tom Tirer. Image restoration by denoising diffusion models with iteratively preconditioned guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25245–25254, 2024.

[Goodfellow, 2016] Ian Goodfellow. *Deep learning*, volume 196. MIT press, 2016.

[Gordon *et al.*, 2023] Lucia Gordon, Nikhil Behari, Samuel Collier, Elizabeth Bondi-Kelly, Jackson A. Killian, Catherine Ressijac, Peter Boucher, Andrew Davies, and Milind Tambe. Find rhinos without finding rhinos: Active learning with multimodal imagery of south african rhino habitats. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI*, pages 5977–5985, 2023.

[Hao *et al.*, 2015] Mingming Hao, Helong Yu, and Daoliang Li. The measurement of fish size by machine vision-a review. In *International Conference on Computer and Computing Technologies in Agriculture*, pages 15–32. Springer, 2015.

[Kay *et al.*, 2022] Justin Kay, Peter Kulits, Suzanne Stathatos, Siqi Deng, Erik Young, Sara Beery, Grant Van Horn, and Pietro Perona. The caltech fish counting dataset: a benchmark for multiple-object tracking and counting. In *European Conference on Computer Vision*, pages 290–311. Springer, 2022.

[Kay *et al.*, 2024] Justin Kay, Timm Haucke, Suzanne Stathatos, Siqi Deng, Erik Young, Pietro Perona, Sara Beery, and Grant Van Horn. Align and distill: Unifying and improving domain adaptive object detection. *arXiv preprint arXiv:2403.12029*, 2024.

[Khan *et al.*, 2023] Faizan Farooq Khan, Xiang Li, Andrew J Temple, and Mohamed Elhoseiny. Fishnet: A large-scale dataset and benchmark for fish recognition, detection, and functional trait prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20496–20506, 2023.

[Kilduff *et al.*, 2015] D Patrick Kilduff, Emanuele Di Lorenzo, Louis W Botsford, and Steven LH Teo. Changing central pacific el niños reduce stability of north american salmon survival rates. *Proceedings of the National Academy of Sciences*, 112(35):10962–10966, 2015.

[Kshitiz *et al.*, 2023] Kshitiz, Sonu Shreshtha, Ramy Mounir, Mayank Vatsa, Richa Singh, Saket Anand, Sudeep Sarkar, and Sevaram Mali Parihar. Long-term monitoring of bird flocks in the wild. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI*, pages 6344–6352, 2023.

[Liu *et al.*, 2020] Yang Liu, Anbu Huang, Yun Luo, He Huang, Youzhi Liu, Yuanyuan Chen, Lican Feng, Tianjian Chen, Han Yu, and Qiang Yang. Fedvision: An online visual object detection platform powered by federated learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 13172–13179, 2020.

[Liu *et al.*, 2023] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916, 2023.

[Liu *et al.*, 2024] Weizhen Liu, Jiayu Tan, Guangyu Lan, Ao Li, Dongye Li, Le Zhao, Xiaohui Yuan, and Nanqing Dong. Benchmarking fish dataset and evaluation metric in keypoint detection - towards precise fish morphological assessment in aquaculture breeding. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI*, pages 7376–7384, 2024.

[Luiten *et al.*, 2021] Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip HS Torr, Laura Leal-Taixé, and Bastian Leibe. Hota: A higher order metric for evaluating multi-object tracking. *International Journal of Computer Vision*, 129(2):548–578, 2021.

[Ma *et al.*, 2024] Sami Ma, Yi Ching Chou, Miao Zhang, Hao Fang, Haoyuan Zhao, Jiangchuan Liu, and William I Atlas. Leo satellite network access in the wild: Potentials, experiences, and challenges. *IEEE Network*, 2024.

[Moore *et al.*, 2021] Jonathan W Moore, Brendan M Connors, and Emma E Hodgson. Conservation risks and portfolio effects in mixed-stock fisheries. *Fish and Fisheries*, 22(5):1024–1040, 2021.

[OpenAI, 2024] OpenAI. Learning to reason with llms. https://openai.com/index/learning-to-reason-with-llms/, 2024. [Accessed: 2024-10-04].

[Price *et al.*, 2017] Michael HH Price, Karl K English, Andrew G Rosenberger, Misty MacDuffee, and John D Reynolds. Canada's wild salmon policy: an assessment of conservation progress in british columbia. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(10):1507–1518, 2017.

[Radford *et al.*, 2021] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[Ravi *et al.*, 2024] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024.

[Ristani *et al.*, 2016] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision (ECCV)*, pages 17–35. Springer, 2016.

[Schindler and Hilborn, 2015] Daniel E Schindler and Ray Hilborn. Prediction, precaution, and policy under global change. *Science*, 347(6225):953–954, 2015.

[United Nations, 2015] United Nations. Transforming our world: The 2030 agenda for sustainable development. https://sdgs.un.org/2030agenda, 2015. Adopted by the United Nations General Assembly on 25 September 2015. [Accessed: 2025-05-21].

[United Nations, 2021] United Nations. Leave no one behind. https://unsdg.un.org/2030-agenda/universal-values/leave-no-one-behind, 2021. [Accessed: 2025-05-21].

[Walsh *et al.*, 2020] Jessica C Walsh, Jane E Pendray, Sean C Godwin, Kyle A Artelle, Holly K Kindsvater, Rachel D Field, Jennifer N Harding, Noel R Swain, and John D Reynolds. Relationships between pacific salmon and aquatic and terrestrial ecosystems: implications for ecosystem-based management. *Ecology*, 101(9):e03060, 2020.

[Walters *et al.*, 2008] Carl J Walters, JA Lichatowich, Randall M Peterman, and John D Reynolds. Report of the skeena independent science review panel. *A report to the Canadian Department of Fisheries and Oceans and the British Columbia Ministry of the Environment*, 15, 2008.

[Wang *et al.*, 2024] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, and Guiguang Ding. Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*, 2024.

[Waples *et al.*, 2008] Robin S Waples, George R Pess, and Tim Beechie. Evolutionary history of pacific salmon in dynamic environments. *Evolutionary Applications*, 1(2):189–206, 2008.

[Wojke *et al.*, 2017] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In *International Conference on Image Processing (ICIP)*, pages 3645–3649. IEEE, 2017.

[Wu *et al.*, 2022] Xingjiao Wu, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He. A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, 135:364–381, 2022.

[Xu *et al.*, 2024] Chi Xu, Rongsheng Qian, Hao Fang, Xiaoqiang Ma, William I Atlas, Jiangchuan Liu, and Mark A Spoljaric. Salina: Towards sustainable live sonar analytics in wild ecosystems. In *Proceedings of the 22nd ACM Conference on Embedded Networked Sensor Systems*, pages 68–81, 2024.

[Yoshiyama, 1999] Ronald M Yoshiyama. A history of salmon and people in the central valley region of california. *Reviews in Fisheries Science*, 7(3-4):197–239, 1999.

[Zhao *et al.*, 2024] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, and Jie Chen. Detrs beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16965–16974, 2024.

[Zhong *et al.*, 2024] Tianyang Zhong, Zhengliang Liu, Yi Pan, Yutong Zhang, Yifan Zhou, Shizhe Liang, Zihao Wu, Yanjun Lyu, Peng Shu, Xiaowei Yu, et al. Evaluation of openai o1: Opportunities and challenges of agi. *arXiv preprint arXiv:2409.18486*, 2024.