Automated Salmonid Counting in Sonar Data

Peter Kulits[1], **Angelina Pan**[1], **Sara Beery**[1], **Erik Young**[2], **Pietro Perona**[1], **Grant Van Horn**[3]

Caltech[1], Trout Unlimited[2], Cornell University[3]

Abstract

The prosperity of salmonids is crucial for several ecological and economic functions. Accurately counting spawning salmonids during their seasonal migration is essential in monitoring threatened populations, assessing the efficacy of recovery strategies, guiding fishing season regulations, and supporting the management of commercial and recreational fisheries. While several methods exist for counting river fish, they all rely heavily on human involvement, introducing a hefty financial and time burden. In this paper we present an automated fish-counting method that utilizes data captured from ARIS sonar cameras to detect and track salmonids migrating in rivers. Our fully automated system has an 19.3% per-clip error when compared to human counting performance. There is room to improve, but our system can already decrease the amount of time field biologists and fishery managers need to spend manually watching ARIS clips.[1]

1 Introduction

Salmonids comprise the members of the taxonomic family *Salmonidae* that include salmon, trout, and char. They play a vital ecological role, serving as keystone species and a barometer of the health of our environment [5]. They provide a direct, critical food source for humans as well as for at least 137 other animal species, including orcas, bears, and wolves [5]. Unfortunately, many salmonid populations are threatened due to dams, hatcheries, loss of habitat, excessive fishing, and climate change. The U.S. National Marine Fisheries Service has listed 28 distinct U.S. salmon and steelhead populations as critically endangered or threatened with extinction under the U.S. Endangered Species Act [17]. Nearly all of the salmonid species on the U.S. West Coast listed under the Endangered Species Act are highly vulnerable to expected increases in stream temperatures, sea surface temperatures, and ocean acidification [3]. The southern extent of their range is expected to contract, and their ability to successfully adapt is limited as a result of other anthropogenic changes, such as dams and habitat destruction [3]. Other populations are healthy, and with proper management they are providing an important source of food and supporting the large economies associated with commercial and recreational fishing [5].

To restore and maintain a stable population of fish, ensure the survival of threatened populations, and guide catch limits for non-threatened populations, fishery management practices must be accurate, comprehensive, timely, and cost-efficient. Data-driven estimates of *escapement*, the number of fish that have made it safely upstream to spawn, are used to ensure a sustainable population is maintained. Where populations are threatened with extinction, escapement estimates are used to evaluate the effectiveness of recovery strategies. Accurate estimates are also essential to supporting the successful management of commercial and recreational fisheries. Catch-limit regulations and a fishing season's start and end dates are determined via these estimates. There are currently a handful of different methods used to count returning salmonids, including human observers in fish towers, fish wheels,

[1]Code available at https://github.com/gvanhorn38/fish_eye

Tackling Climate Change with Machine Learning workshop at NeurIPS 2020.

and technicians hand-counting fish in sonar videos. Each of these methods represents a tradeoff between accuracy, cost, and efficiency — our work focuses on sonar-based methods.

This paper presents a system to automatically detect and track fish in ARIS frames, allowing estimates of escapement to scale to 24 hours per day, with near-real-time monitoring across large numbers of sites, a feat that is beyond the capabilities and resources of human-based monitoring efforts. From detected fish tracks our system can produce a total count of fish swimming upstream and downstream, as well as an estimate of the length of each fish. Our method uses the latest advances in computer vision detection methods, and while the individual components of our system are not novel, the system as a whole is a novel application of machine learning to the domain of counting river fish in ARIS data. The review and analysis of this system will be beneficial for future projects with similar goals.

2 Related Works

Sonar Data: For decades, ecologists have used sonar data to monitor fish populations [2, 16]. Automating fish counting in sonar data has been investigated previously in [22, 12, 15, 20]. Sindre Vatnehol et al. [22] studied the problem of boat-mounted acoustic sonar cameras. Liang Liu et al. [12] studied different regularization techniques to combat fish length and fish density challenges. Filip Mandic et al. [15] developed a tracking filter to fuse USBL (ultra-short baseline) acoustic sensors. Our work instead focuses on long-range static sonar in rivers, and it includes both upstream/downstream tracking and fish measurement.

Detection: Localizing objects in images has been well-studied in the computer vision community [6]. The field is constantly releasing more accurate [4] and more efficient [21] neural network architectures that produce object location proposals. In this work we use a Single Shot MultiBox Detector (SSD) architecture [13] with a MobileNet-v2 backbone [19]. Our proposed system is agnostic to the detection architecture and can be easily updated with new computer vision methods.

Tracking: Tracking an object across successive frames of a video clip is another well-studied problem in the computer vision community [14]. We used the recently proposed Simple Online and Realtime Tracking (SORT) algorithm [1] for its simplicity and accuracy.

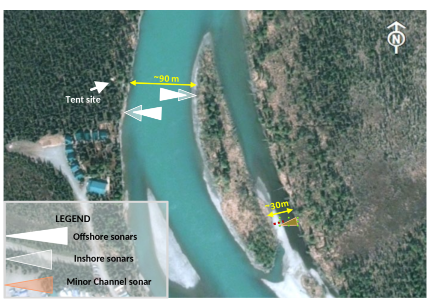


Figure 1: Camera locations on the Kenai River [9]

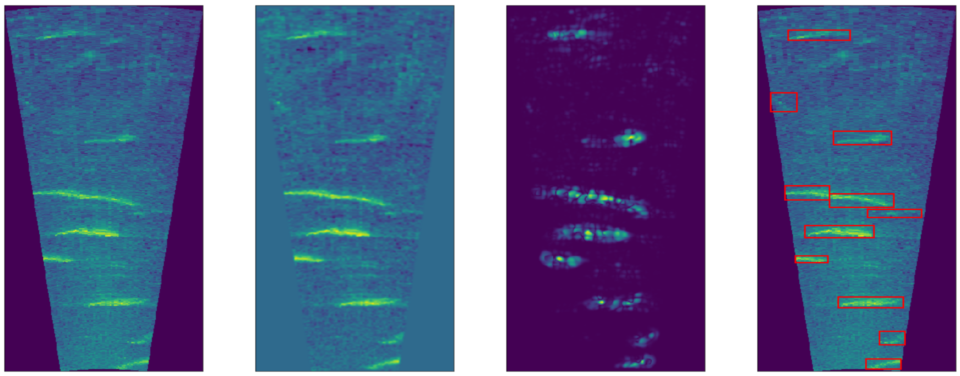
	Train	Test
Total Clips	492	35
Unique ARIS Files	320	29
Biologist Fish Count	785	83
Boxed Fish Count	1,617	95
Total Frames	249,380	18,459
Frames w/ Fish	91,036	3,439

Table 1: Dataset statistics

3 Dataset

We constructed a dataset from five ARIS cameras stationed on the Kenai River in Alaska [18], see Fig. 1. Each camera was configured to monitor a different section of the river. Kenai offshore strata used two ARIS 1200 models with high-resolution lenses (HRL), and nearshore strata used one ARIS 1200 and one ARIS 1800, each with a standard lens. An additional minor river channel was monitored with the ARIS 1200 model with HRL. Kenai data includes both near and far cameras and varied camera placement strategies, resulting in increased image diversity. Field biologists analyzed the footage from these clips using their standard manual-review practices [8], producing timestamps, swimming direction, and fish lengths for Chinook and Sockeye salmon. From these manual annotations we created our dataset in the following steps: (1) We randomly cropped one-minute clips from the ARIS videos that were guaranteed to have fish, as labeled by the field biologist's timestamps. (2) We used an annotation GUI and a crowd workforce to box and track the fish in each clip. (3) We split the clips into train and test sets by grouping them by their camera placement and randomly splitting the placements.

This produced a dataset of 527 clips totalling $267{,}839$ frames, and $1{,}712$ individual fish totalling $169{,}649$ boxes, see Table 1 for details and Fig. 2d for a sample annotated ARIS frame. Note the discrepancy between the number of fish reported by the field biologists and the number of fish tracked by the human crowd annotators (868 vs $1{,}712$). The field biologists typically ignore all non-salmonid species and all fish that are smaller than a particular length (e.g., 40cm). We were unable to communicate these rules to the crowd annotators because of varying image scale and instead had them box all fish.



(a) ARIS Frame (b) Mean-Subtracted (c) Optical Flow (d) Target Boxes

Figure 2: Our model takes three input channels: (a) the original ARIS frame, (b) the blurred and mean-subtracted frame, and (c) the magnitude of the optical flow between the current frame and the next frame. The detector is tasked with predicting the bounding boxes for each fish in the frame (d).

4 Automated Fish Counting

Our automated fish-counting system has three main components: a detector, a tracker, and a direction/length predictor. The input to the system is an ARIS data file, and the output is a list of individual fish with length and direction values.

Detection: We use an SSD architecture [13] for detection with a MobileNet-v2 backbone [19]. The detector takes as input a three channel "image" of size $[640, 320, 3]$ composed of: (1) the original ARIS frame, (2) the Gaussian blurred (kernel size of 5, standard deviation of 1), mean-subtracted frame (the mean is computed across the respective clip), and (3) the magnitude of the optical flow between this frame and the next frame, where the optical flow is computed between the blurred, mean-subtracted versions of the respective frames. See Fig. 3 for an example of each of these channels. We train the model using a focal loss ($\gamma = 2.0$ and $\alpha = 0.75$) [11], a batch size of 64 across four GPUs, and a cosine-decayed learning rate of 0.01 for 100 epochs. Any detected boxes with a confidence score of at least 0.5 are passed to the tracker.

Precision-recall curves for our detector are shown in Fig. 3a, where we assume a correct detection when the intersection-over-union (IoU) of the detected box and the ground truth box exceed 0.5. We plot performance for small boxes (whose area is less than 32^2 pixels), medium boxes (whose area is between 32^2 and 96^2 pixels), and all the boxes in the test set. While small fish are important to localize, bounding boxes for the target salmonid species often fall into the medium-sized range, especially for ARIS cameras that are sampling near-shore strata. We achieve a mean average precision of 0.95 for medium-sized fish, providing a strong set of initial detections for the tracker.

Tracking, Direction Prediction, & Counting: To associate the bounding boxes to tracks, we use a modified version of the SORT algorithm [1]. The algorithm employs the Hungarian [10] and Kalman Filter [7] algorithms. The Hungarian algorithm is used to determine whether an object in one frame is the same object in another frame; the Kalman Filter algorithm is used to predict future position based on current position.

The evaluation of the tracker is directly related to our fish direction prediction. Direction prediction relies only on the first and final boxes of a track. If the center of the first box is on the right and the center of the final box is on the left, then the fish is left-moving, and vice versa for a right-moving fish. If the centers of the first and final boxes are on the same side of the image, the direction is "undefined." These heuristics are based on the instructions given to human salmonid counters at the Alaska Department of Fish and Game (ADFG) and described in [8].

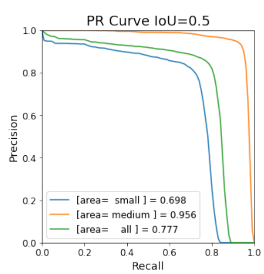
We evaluate our fish detection and tracking performance in relation to ground truth left-traveling and right-traveling fish counts *provided by the field biologists*, which is a non-traditional evaluation setting. We show our evaluation in Table 3b. To capture overall performance across the dataset, we show ground truth counts left and right and compare with our predicted counts in two scenarios: (1) using the crowd sourced human-annotated boxes, which represent our tracking system alone, and (2) using the boxes output by our detector, which represent the performance of our entire pipeline end-to-end. We see that overall performance is close to the expert counts on both scenarios, with overall count differences within five fish in all settings and a maximum absolute difference of two fish across any given clip. The mean absolute difference in count across all clips in the test set is <1 in all scenarios. To capture overall performance over the n test clips, without over- or under-sampling on any set of clips cancelling each other out, we use the following metric:

$$\text{Norm. Sum Abs Count Diff} := \frac{\sum_{i=0}^{n} \left(|\text{left\_pred}(i) - \text{left\_gt}(i)| + |\text{right\_pred}(i) - \text{right\_gt}(i)|\right)}{\sum_{i=0}^{n} \left(\text{left\_gt}(i) + \text{right\_gt}(i)\right)} \tag{1}$$

where left_pred(i) and left_gt(i) are the left predicted and ground truth counts for clip i. This metric captures the ratio between average combined left and right prediction errors, normalized by the average ground truth counts left and right per clip. For our fully automated system, we achieve 19.3% error using this metric, a promising step towards automating salmonid-escapement estimation.

Length Prediction: To estimate the length of a tracked fish, we multiply the 80th-quantile bounding-box width (considering all the bounding boxes in the respective track) by the meters-per-pixel constant (provided in the respective ARIS file) and a constant scalar α that is learned from the training data. We set $\alpha = 0.85$ for our experiments.

Computing statistics for the performance of our length estimator is difficult due to the ambiguity of assigning the field biologist's annotations to a specific fish in the frame (field biologists did not box the fish, they simply provided counts, directions, and length measurements at a particular time stamp). To remove this ambiguity we only consider clips where both the field biologist and the human annotator labeled one fish, which results in 19 matchable examples in the test set. From this subset of clips we compared the performance of our predicted measurements using tracks constructed from either human bounding boxes or detected bounding boxes. When using annotated boxes our mean-absolute length computation error on single-fish examples from the test set is 5.65 ± 7.60 cm. When using detected boxes our mean absolute length computation error is 7.8 ± 6.0 cm.



(a)

	w/ Annotated Boxes		w/ Predicted Boxes	
	Left	Right	Left	Right
Total GT (Expert Counts)	61	22	61	22
Sum Predicted Counts	63	24	60	17
Sum Abs Diff Per Clip	2	6	11	5
Max Abs Diff Across Clips	1	2	2	2
Mean Abs Diff Across Clips	0.06 ± 0.24	0.17 ± 0.45	0.31 ± 0.58	0.14 ± 0.43
Normalized Sum Abs Count Diff	9.6%		19.3%	

(b)

Figure 3: (a) Per-frame detector PR curve. (b) Left and right fish tracking/counting results on the test set, see Section 4 for a discussion. All units are "counts" except the final row, which is unitless. Sum Abs Diff Per Clip is the dividend of Eq. 1. Max Abs Diff Across Clips and Mean Abs Diff Across Clips are the maximum and the mean, respectively, over the per-clip absolute differences between ground truth and predicted count. Normalized Sum Abs Count Diff is defined in Eq. 1.

5 Conclusion

Our proposed system is a first step towards scalable, efficient, and low-cost salmonid escapement estimation. We are excited to work with our collaborators to put our system into practice at the Kenai River, at first in parallel with human counting to build trust in the model, make improvements, and develop a deeper understanding of our failure modes. There are potential ethical concerns with using any automated system to make important conservation and sustainability decisions. We will investigate human-in-the-loop quality control to ensure that any errors or systematic biases in our models do not lead to potentially damaging fisheries-management protocols.

6 Acknowledgements

We would like to thank Amazon Web Services for funding and compute resources for this project; George Pess and Oleksandr Stefankiv from the National Marine Fisheries Service; James Miller, Suzanne Maxwell, Brandon Key, Carl Pfisterer, Gregory Buck, April Faulkner, and Jordan Head from the Alaska Department of Fish and Game; Michael Sparkman and David Kajtaniak from the California Department of Fish and Wildlife; Dean Finnerty from Trout Unlimited; as well as Keith Denton, Mike McHenry, and the Lower Elwha Klallam Tribe.

References

[1] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft. Simple online and realtime tracking. *CoRR*, abs/1602.00763, 2016.

[2] P. Brehmer, T. Lafont, S. Georgakarakos, E. Josse, F. Gerlotto, and C. Collet. Omnidirectional multibeam sonar monitoring: applications in fisheries science. *Fish and Fisheries*, 7(3):165–179, 2006.

[3] L. G. Crozier, M. M. McClure, T. Beechie, S. J. Bograd, D. A. Boughton, M. Carr, T. D. Cooney, J. B. Dunham, C. M. Greene, M. A. Haltuch, E. L. Hazen, D. M. Holzer, D. D. Huff, R. C. Johnson, C. E. Jordan, I. C. Kaplan, S. T. Lindley, N. J. Mantua, P. B. Moyle, J. M. Myers, M. W. Nelson, B. C. Spence, L. A. Weitkamp, T. H. Williams, and E. Willis-Norton. Climate vulnerability assessment for pacific salmon and steelhead in the california current large marine ecosystem. *PloS One*, 14(7):e0217711–e0217711, 2019.

[4] X. Du, T.-Y. Lin, P. Jin, G. Ghiasi, M. Tan, Y. Cui, Q. V. Le, and X. Song. Spinenet: Learning scale-permuted backbone for recognition and localization, 2020.

[5] G. Rahr. Why Protect Salmon.

[6] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy. Speed/accuracy trade-offs for modern convolutional object detectors. *CVPR*, abs/1611.10012, 2016.

[7] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, Mar. 1960.

[8] B. Key, J. Miller, and J. Huang. Operational plan: Kenai river chinook salmon sonar assessment at river mile 13.7, 2020–2022, 2020.

[9] B. H. Key, J. D. Miller, S. J. Fleischman, and J. Huang. Chinook Salmon Passage in the Kenai River at River Mile 13.7 Using Adaptive Resolution Imaging Sonar, 2016. *Fishery Data Series*, 19-07, Dec. 2018.

[10] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.

[11] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[12] L. Liu, H. Lu, Z. Cao, and Y. Xiao. Counting Fish in Sonar Images. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3189–3193, Oct. 2018. ISSN: 2381-8549.

[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. page 21–37, 2016.

[14] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, X. Zhao, and T.-K. Kim. Multiple object tracking: A literature review. *arXiv preprint arXiv:1409.7618*, 2014.

[15] F. Mandic, I. Rendulic, N. Miškovic, and . Nađ. Underwater Object Tracking Using Sonar and USBL Measurements, Aug. 2016.

[16] O. A. Misund. Underwater acoustics in marine fisheries and fisheries research. *Reviews in Fish Biology and Fisheries*, 7(1):1–34, 1997.

[17] N. Oceanic and A. Administration. Endangered and threatened species; initiation of 5-year reviews for 28 listed species of pacific salmon and steelhead, 2019.

[18] A. D. of Fish and Game. Kenai (RM 14) River.

[19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks, 2019.

[20] S. Schneider and A. Zhuang. Counting fish and dolphins in sonar images using deep learning. *arXiv preprint arXiv:2007.12808*, 2020.

[21] M. Tan, R. Pang, and Q. V. Le. Efficientdet: Scalable and efficient object detection, 2020.

[22] S. Vatnehol, H. Peña, and N. O. Handegard. A method to automatically detect fish aggregations using horizontally scanning sonar. *ICES Journal of Marine Science*, 75(5):1803–1812, Oct. 2018.