The Caltech Fish Counting Dataset:
A Benchmark for Multiple-Object Tracking and Counting

Justin Kay, Peter Kulits, Suzanne Stathatos, Siqi Deng, Erik Young, Sara Beery, Grant Van Horn, and Pietro Perona

Caltech

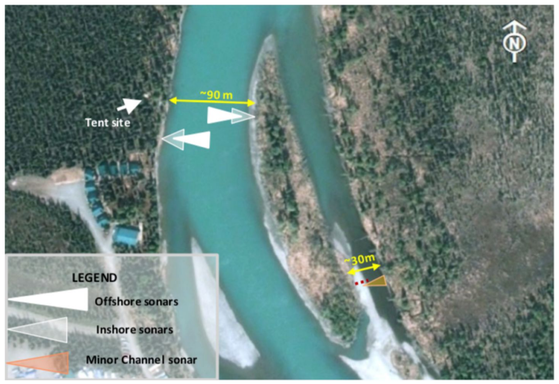
Counting Salmon in Sonar Video

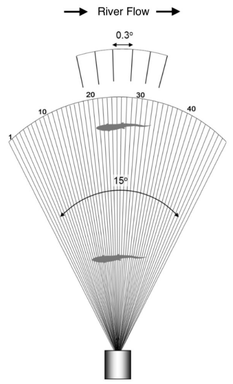
Important application in conservation ecology: **how many salmon migrate upstream each year to spawn?**

Sonar video cameras are deployed in rivers as a **non-invasive and accurate** way to monitor salmon migration.

Counting is currently performed manually by technicians who watch video.

→ Automating counting is a **high-impact challenge** for computer vision.

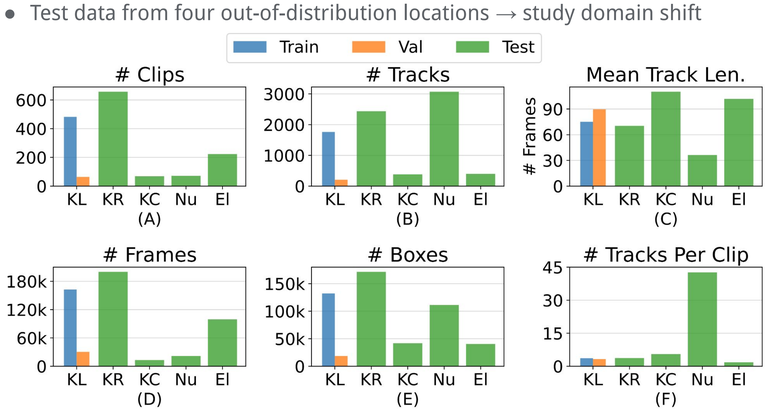




The Caltech Fish Counting Dataset

A large-scale dataset for **video object detection**, **multiple-object tracking**, and **video-based counting.**

- 1,567 video sequences (16.7 hours of video)
- 527,000 image frames
- 516,000 bounding boxes
- 8,254 object instances
- Test data from four out-of-distribution locations → study domain shift



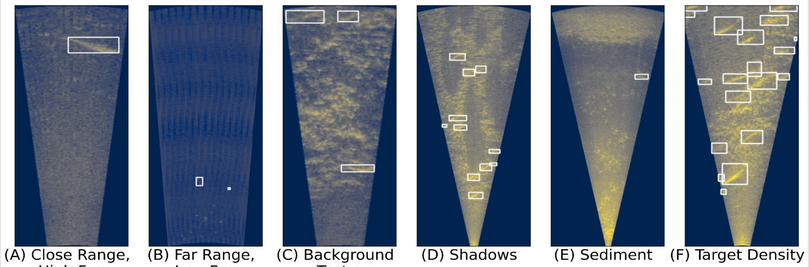
Five source locations: **KL**: Kenai River Left Bank; **KR**: Kenai River Right Bank; **KC**: Kenai River Minor Channel; **Nu**: Nushagak River; **El**: Elwha River

Data Challenges

Domain shift

Each camera deployment presents different challenges to detection, tracking, and counting due to **location-specific environmental conditions**.

Example frames and common challenges



(A) Close Range, High Freq. (B) Far Range, Low Freq. (C) Background Texture (D) Shadows (E) Sediment (F) Target Density

Performance degrades at OOD locations

	Loc	Baseline AP↑ @IoU=0.5	MOTA↑	IDF1↑	HOTA↑	nMAE↓
ID	KL *Val*	66.4	44.9	66.7	49.2	4.9%
OOD	KR	57.7	-28.5	45.4	33.5	11.8%
OOD	KC	32.0	-60.8	35.6	30.9	53.0%
OOD	NU	70.6	30.2	60.8	44.4	14.0%
OOD	EL	39.9	-376.7	18.8	21.3	32.3%

Low signal-to-noise data

- **Individuals look similar to background**: Detection not always possible in a single frame (need to incorporate temporal information).
- **Individuals look similar to each other:** Visual features ineffective for target re-identification.
- **Artifacts from sonar**: speckle noise, acoustic shadows, deterioration of signal at long-range, "ghost fish" (echoes that reflect off water surface)

Comparison with other tracking datasets

In Caltech Fish Counting, trackers cannot rely heavily upon visual association

MOT Challenge UA-DETRAC TAO / LaSOT Ours

Baseline Results

Evaluation

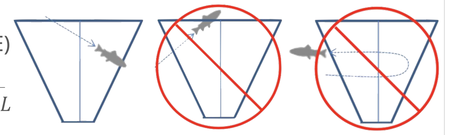
- Detection: AP@IoU=0.5
- Tracking: MOTA, IDF1, HOTA
- Counting: normalized MAE (nMAE)

$$nMAE = \frac{Abs.\ count\ err.\ at\ location\ L}{Ground\ truth\ count\ at\ location\ L}$$

→Target: < 10% nMAE

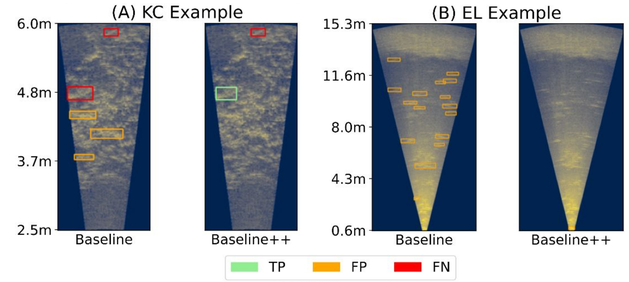
Crossing-line based counting

Matches protocol used in the target application

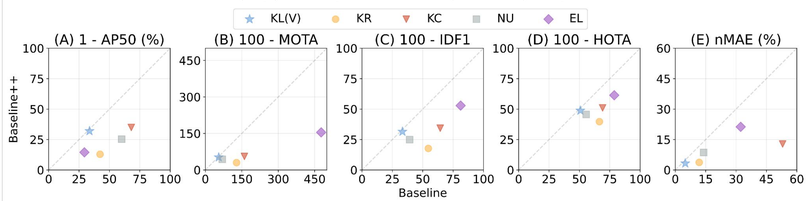


Baseline and Baseline++

- YOLOv5m + SORT (Kalman filter + IoU-based association)
- "Baseline++" **incorporates temporal information** by augmenting input format w/ background subtraction and frame differencing

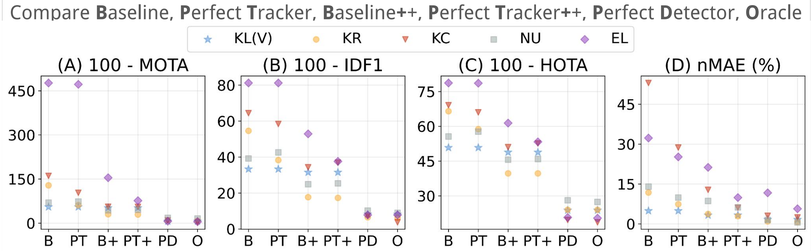


Qualitative (top) and quantitative (bottom) improvements



Upper bounds analysis

Compare **B**aseline, **P**erfect **T**racker, **B**aseline++, **P**erfect **T**racker++, **P**erfect **D**etector, **O**racle



- Still **significant room for improvement** (13% nMAE at KC, 21% nMAE at EL)
- **Most effective path forward is detector improvements**

Dataset Release

visipedia/caltech-fish-counting

Video example from dataset