VARS Annotation Assistance (VAA)

Danelle Cline

dcline@mbari.org

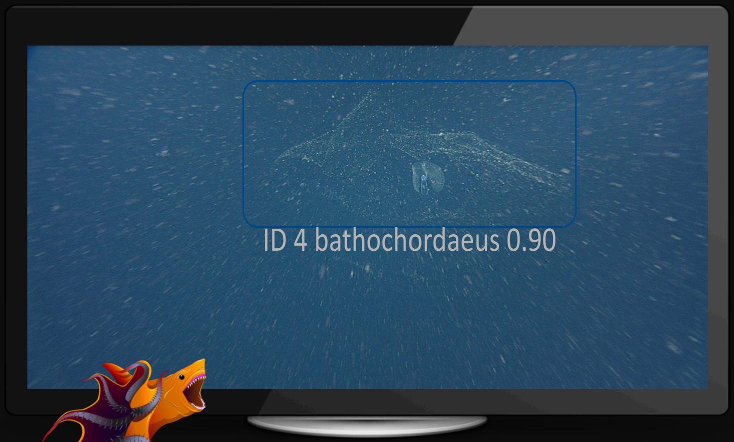
Goals and reflections

A quick reflection of high-level goals, where we have been and what the future might bring

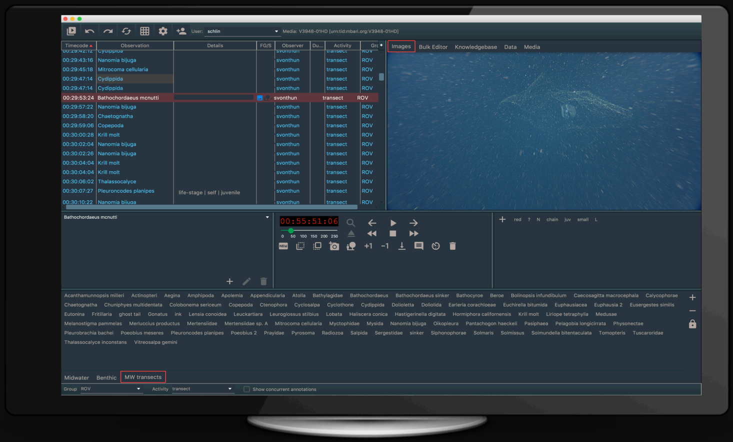
Goals

Integrated Assistance

ML Aware Sharktopoda

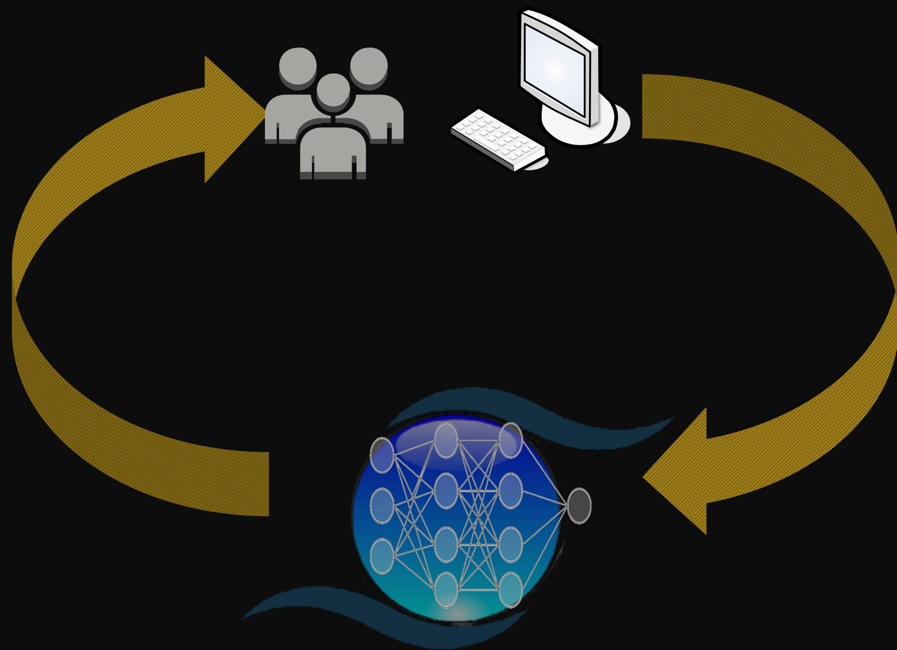


VARS



Focus on typical *quantitative* transect process for benthic and midwater dives

Science Annotators



Drive New Discoveries

Life cycle

Behavior

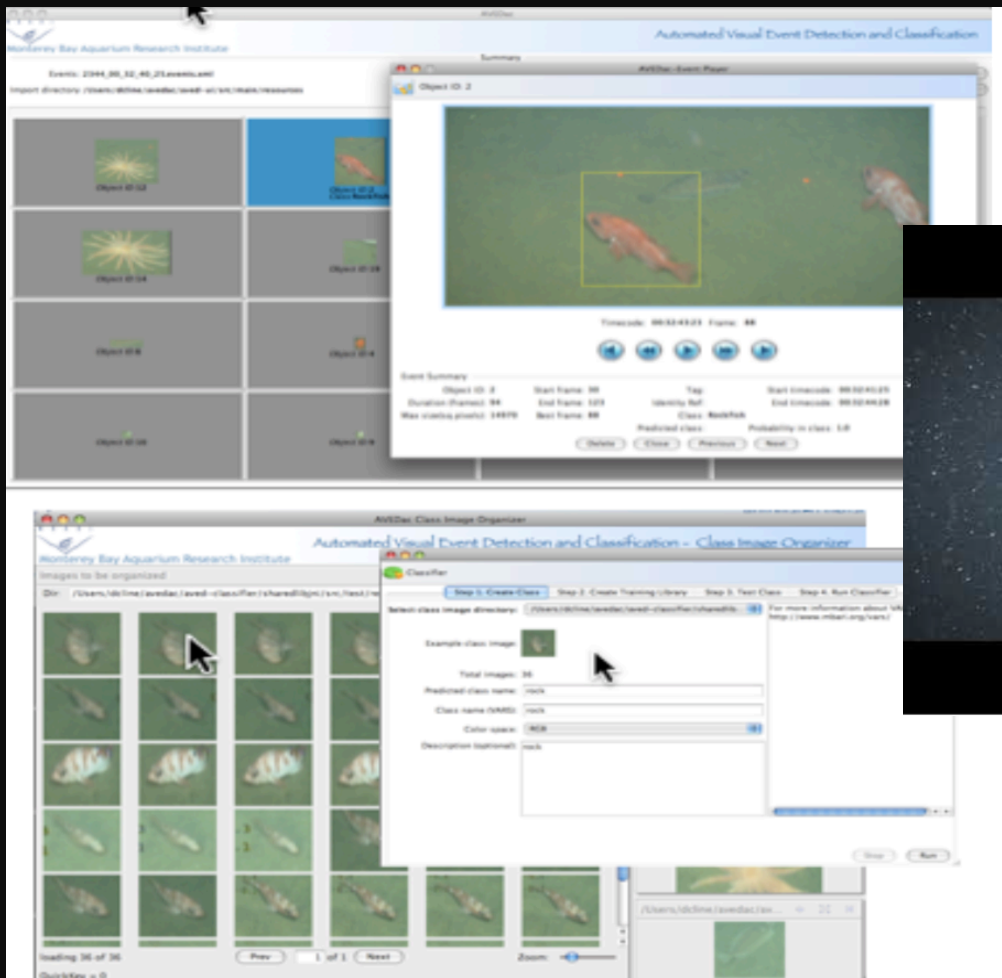
Less Workload

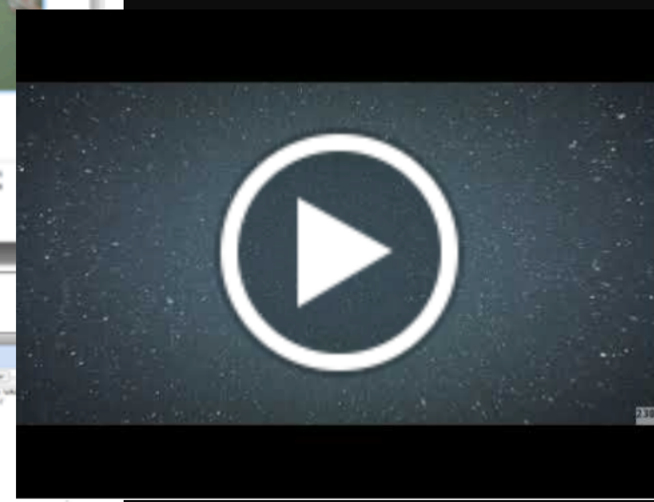
Machine Learning (ML)

Share with the community

Early visions of annotation assistance

Automated Visual Event Detection





Projects

- Station M
- Eye-in-the-Sea
- Army Corps

High-level requirements

Assist annotation process using machine learning (ML)

Achieve TBD speed-up over current annotation process

Share results with the community

Simple enough design to be supported by a small engineering team

VAA

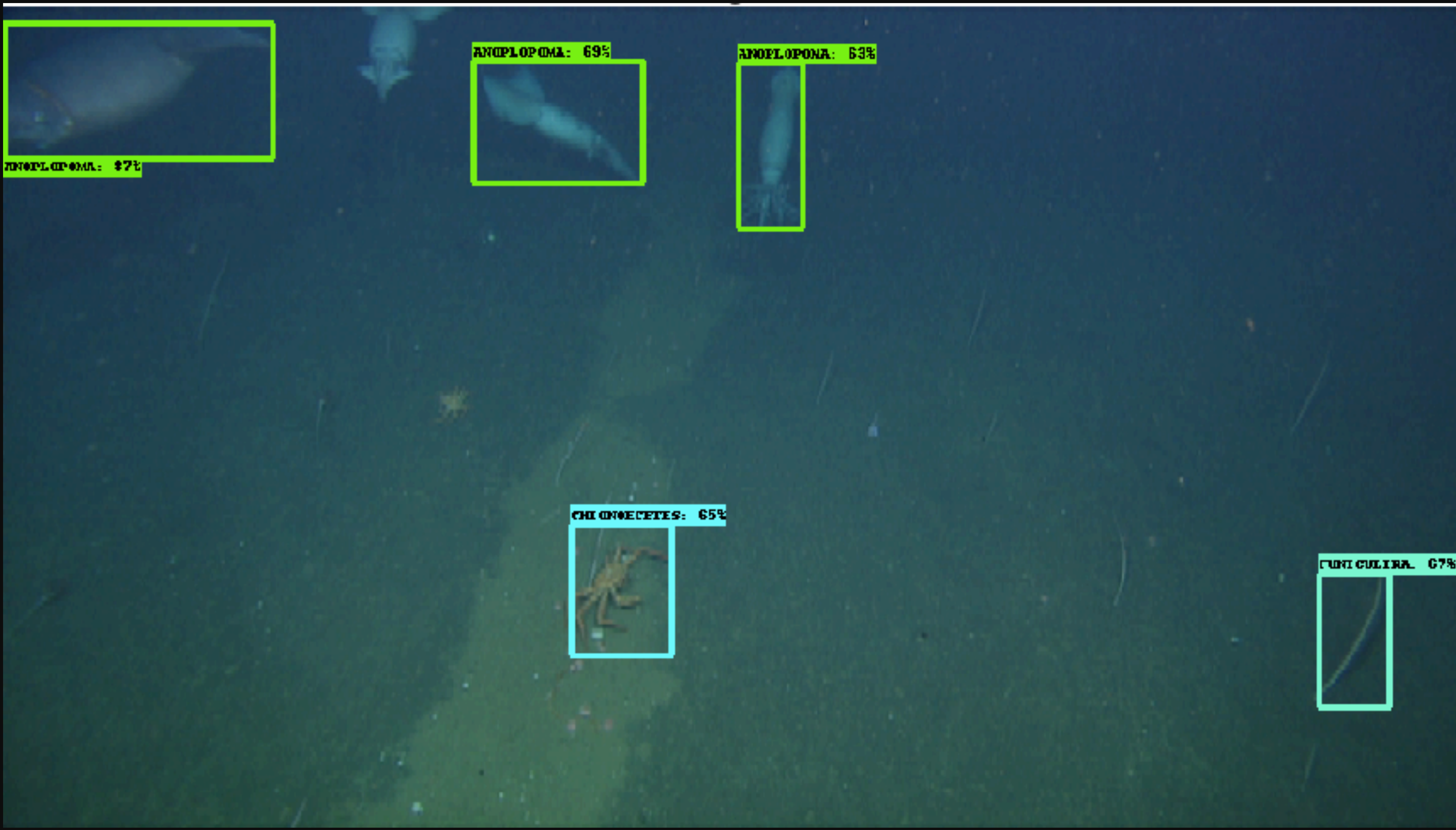
CAUTION

Expect few bumps in the road

Improving ML performance is a highly iterative process . The tools for production level ML are in the nascent stages and the state of the art needs to address important shortcomings.

Misclassifications

Misclassifications



Administrivia

Schedule, Budget , Roles
and Responsibilities

Oct 29, 2019 3:10:11 PM

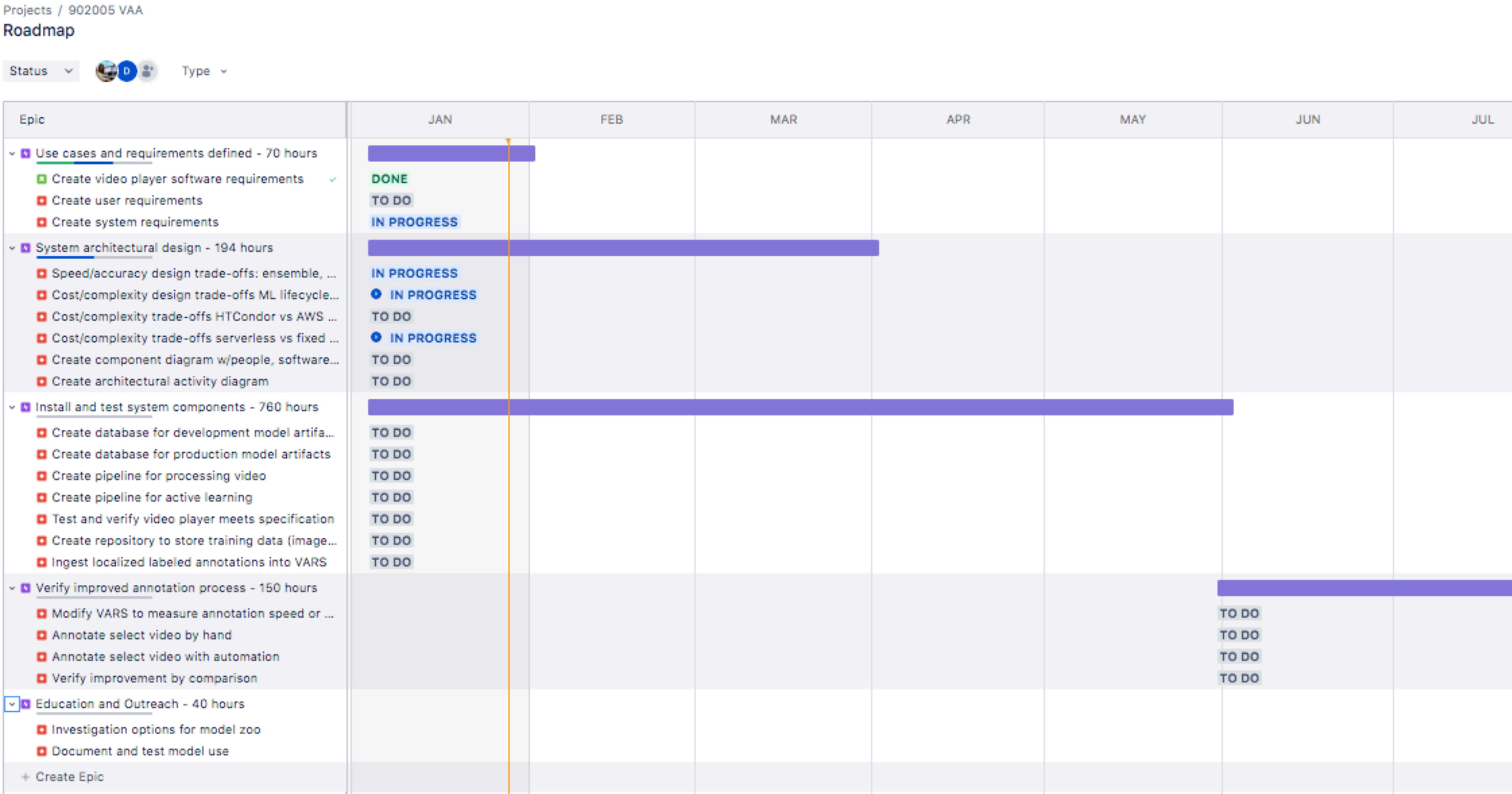
Project: **902005 - VARS Annotation Assistance** Project Manager: **Lundsten, Lonny**

Expenses:

Project	Account	Description	Amount
902005	5110-000 - OS-Sftwr Design Program	External cloud compute service to run model training	3,000
	5110-000 - OS-Sftwr Design Program	VARS video player modifications	0
	5500-000 - Travel - Domestic	Strata Data Conference, March 2020 San Jose	1,500
	5500-000 - Travel - Domestic	GTC Conference March 2020, San Jose, CA	3,000
902005 - Total			**7,500**
Total Expenses			**7,500**

Labor Hours:

Project	Labor Category	Employee Name	Hours
902005			
	ENG-Software Eng-Info Eng	Cline, Danelle	350
	ENG-Software Eng-Info Eng	Edgington, Duane	384
	ENG-Software Eng-Info Eng	Schlining, Brian	480
	ENG-Software Tech	Salamy, Karen	40
	ITD-Video Support	TBD (ITD-Video Support)	200
902005 - Total			**1,454**
Total Hours			**1,454**


Projects / 902005 VAA
Roadmap
Status
Type
Epic
JAN
FEB
MAR
APR
MAY
JUN
JUL
Use cases and requirements defined - 70 hours
Create video player software requirements
DONE
Create user requirements
TO DO
Create system requirements
IN PROGRESS
System architectural design - 194 hours
Speed/accuracy design trade-offs: ensemble, ...
IN PROGRESS
Cost/complexity design trade-offs ML lifecycle...
IN PROGRESS
Cost/complexity trade-offs HTCondor vs AWS ...
TO DO
Cost/complexity trade-offs serverless vs fixed ...
IN PROGRESS
Create component diagram w/people, software...
TO DO
Create architectural activity diagram
TO DO
Install and test system components - 760 hours
Create database for development model artifa...
TO DO
Create database for production model artifacts
TO DO
Create pipeline for processing video
TO DO
Create pipeline for active learning
TO DO
Test and verify video player meets specification
TO DO
Create repository to store training data (image...
TO DO
Ingest localized labeled annotations into VARS
TO DO
Verify improved annotation process - 150 hours
Modify VARS to measure annotation speed or ...
TO DO
Annotate select video by hand
TO DO
Annotate select video with automation
TO DO
Verify improvement by comparison
TO DO
Education and Outreach - 40 hours
Investigation options for model zoo
Document and test model use
Create Epic

Project management: *Lundsten*

Training data requirements, integration and test: Engineering team *(Cline, Edgington, Schlining)*

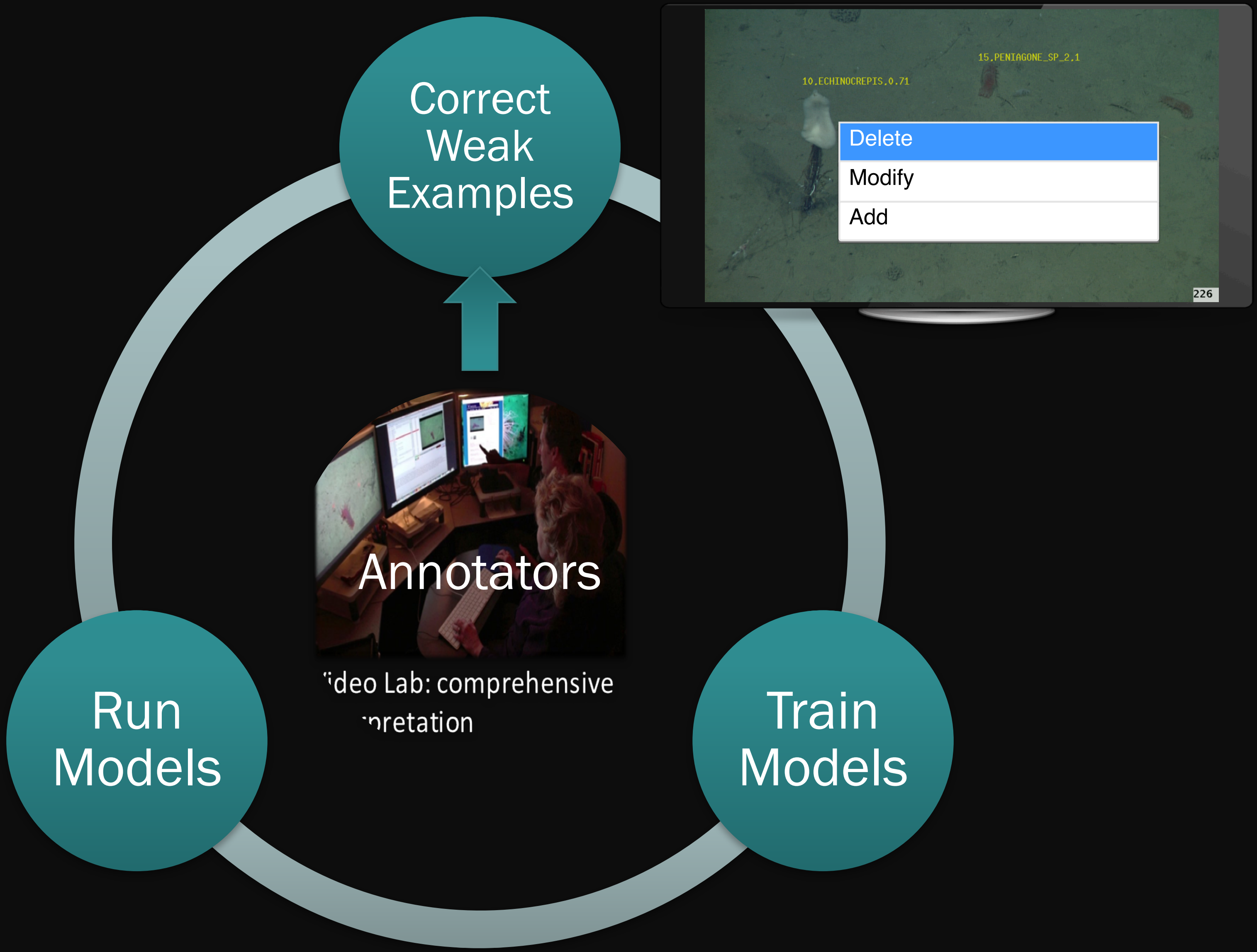
Training data: *FathomNet team*

Quality and validation: ***Video Lab***

Workflow

Learning cycle and proposed workflow

Learning Cycle



Workflow

3-stage pipeline for video

VisualEvents are voted across frames to find the best classification. The "best" classification is then assigned.

KalmanHough Filter Tracking

Anatomy of a Visual Event

Frame 0

Mitrocoma, 0.9

Frame 1

Mitrocoma, 0.92

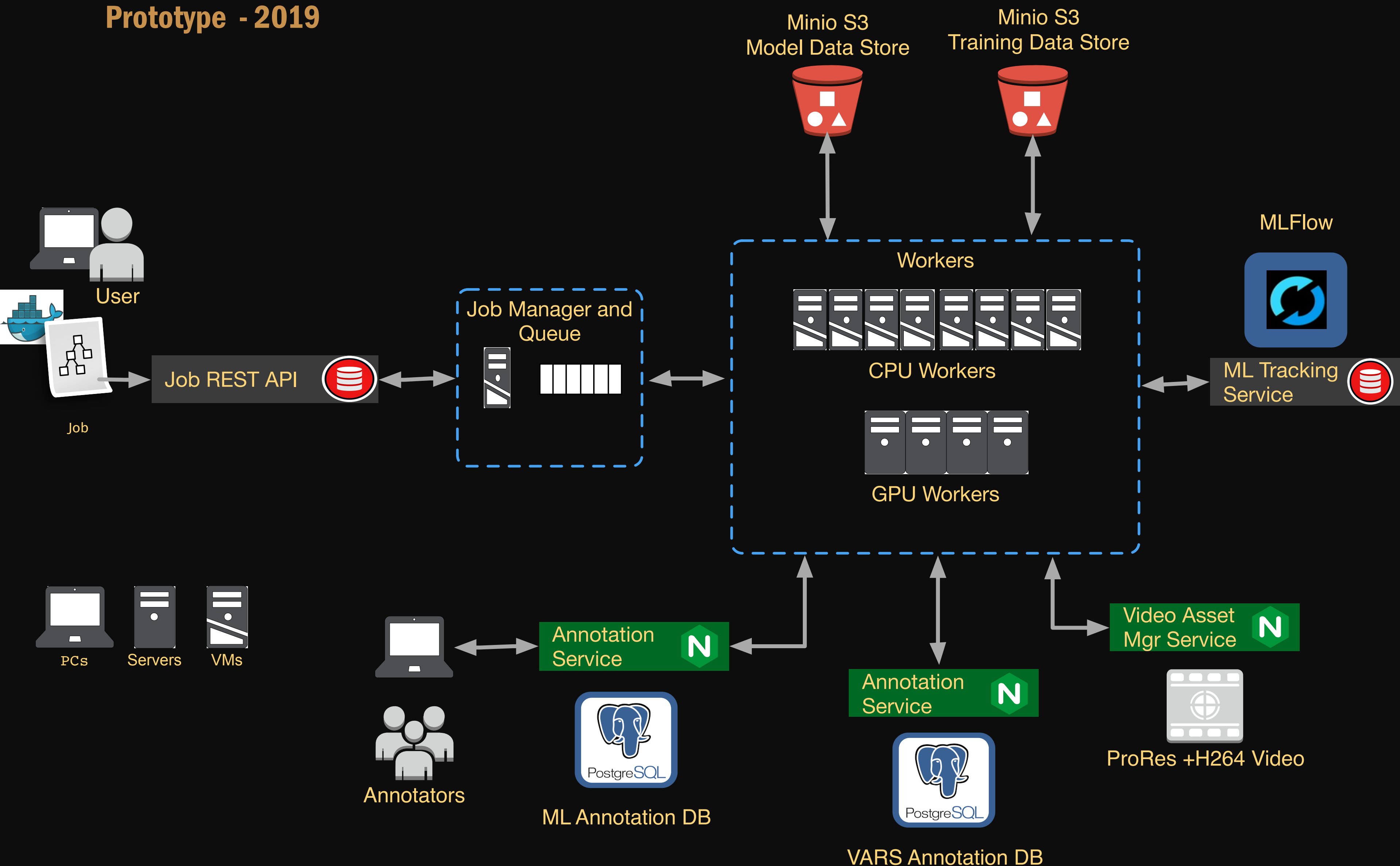
Frame 2

Mitrocoma, 0.99

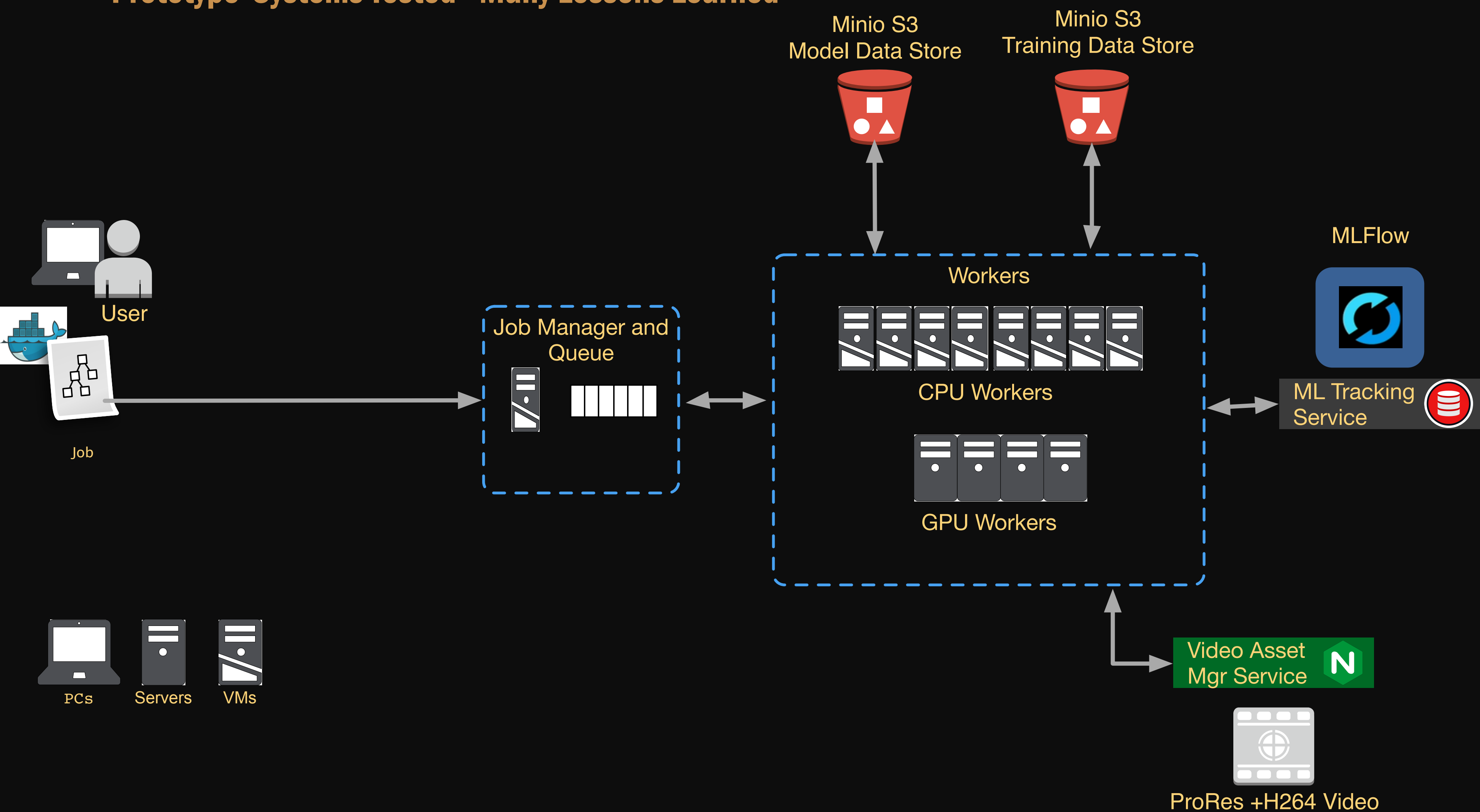
Attributes

- Number of occluded pixels
- Eccentricity, centroid, major/minor axis length, etc.
- Bounding box
- (optional) Shape mask

Prototype - 2019



Prototype Systems Tested - Many Lessons Learned



Key Technologies

Object Storage

MLFlow

Java

Titan Access

AWS Cloud Services

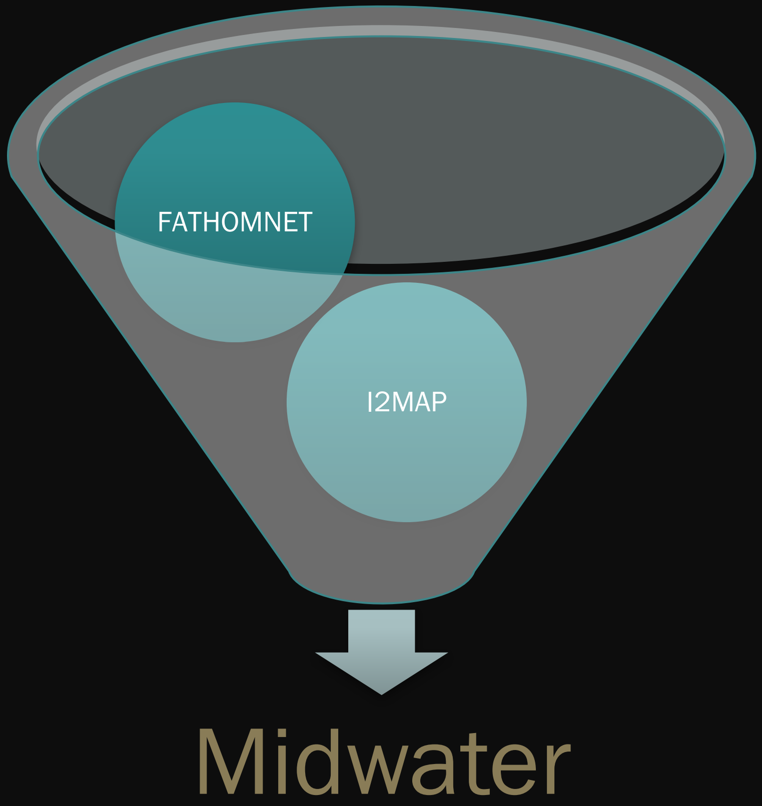
Data

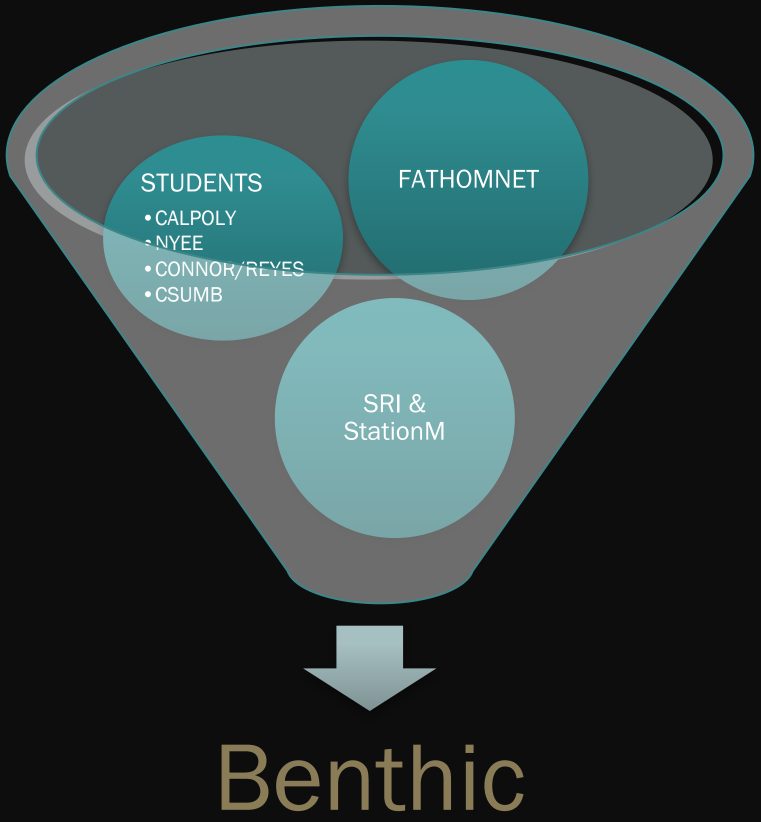
A detour into data

How much data do we have, and what is a good training, testing, and development set for the CNN methods we are using?

Data

Data pool





Data pool

- FathomNet ~10K frames, ~34K midwater labels, ~32K benthic labels, 200+ concepts
- Students
  - Cal Poly ~*16.5K* frame, ~67K human labels
  - ~6.5 million human+track annotations, 40 concepts
  - CONNOR/REYES/NYEE – 6K mostly benthic labels, 16 concepts
- i2MAP– 184 frames 663 midwater labels, 19 concepts
- SRI – mostly rockfish?

How much data is enough?

Classification

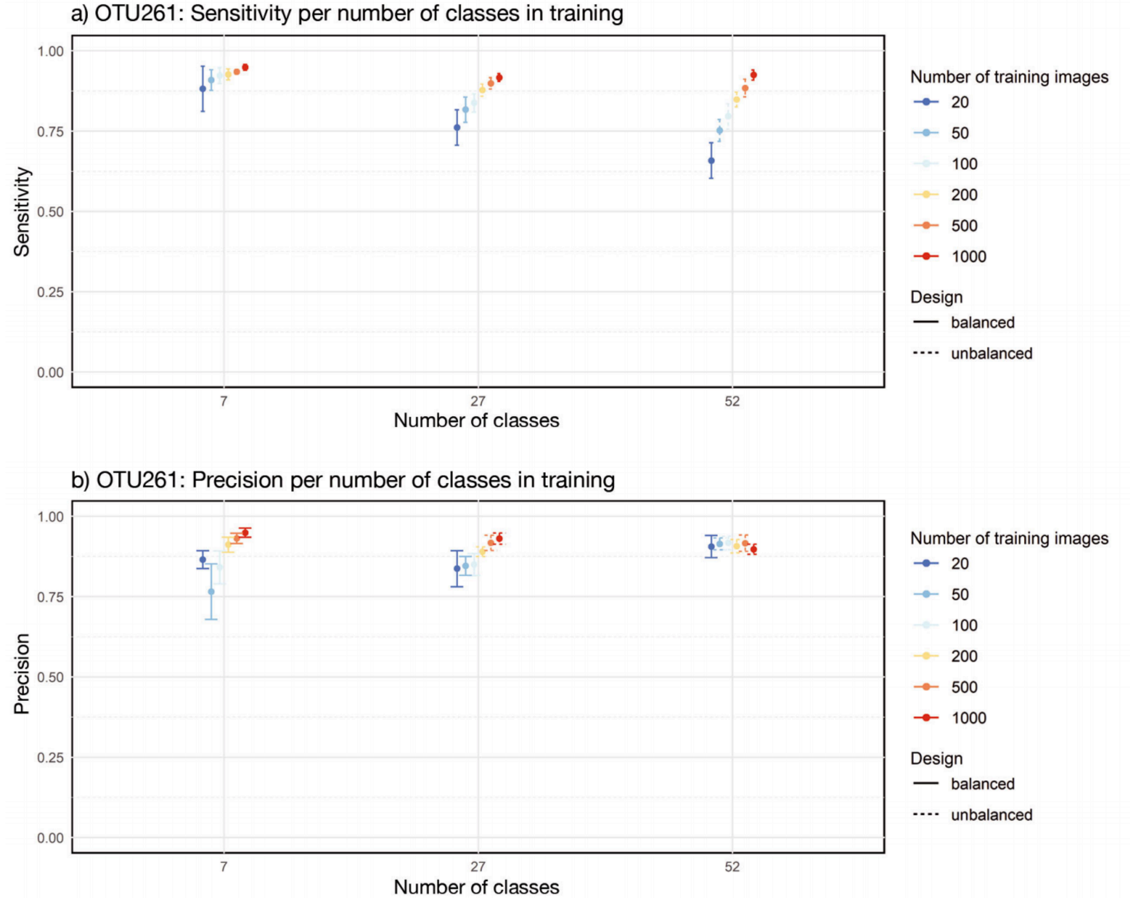


Fig. 5. Differences in (a) sensitivity and (b) precision for operational taxonomic unit (OTU) 261 in classifier trained with different number of classes and images (7 for group A, 27 for group B, 52 for group C). Unbalanced designs (not equal number of images in every class) are displayed with dashed lines. Error bars: standard deviation of the 10 random splits

Piechaud, N., Hunt, C., Culverhouse, P.F., Foster, N.L., & Howell, K.L. (2019). Automated identification of benthic epifauna with computer vision.

Classification

How much data is enough?

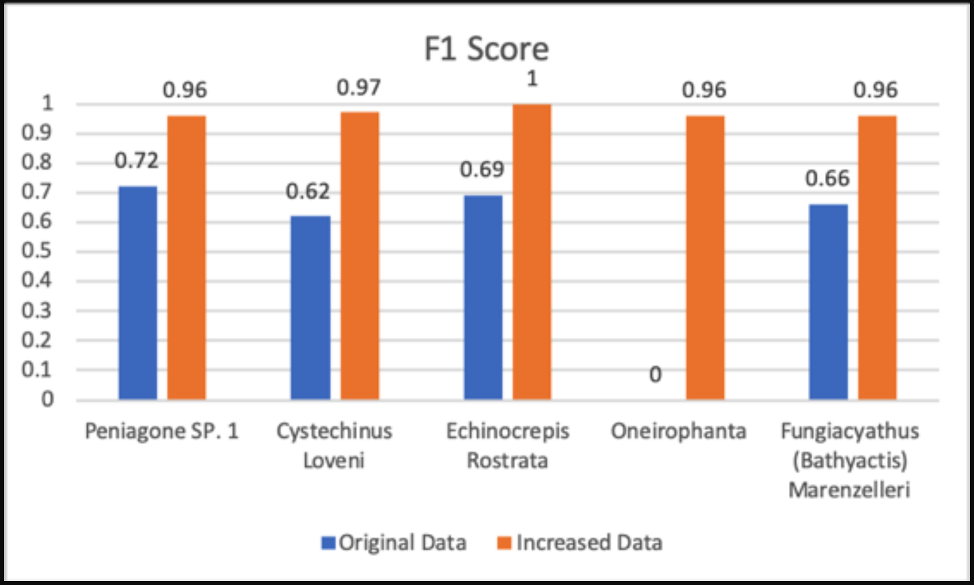
Peniagone SP.1 increased from **113** to **158**

Cystechinus increased from **212** to **266**

Echinocrepis rostrata increased from **108** to **192**

Oneirophant increased from **9** to **142**

Fungiacyathus increased from **74** to **141**

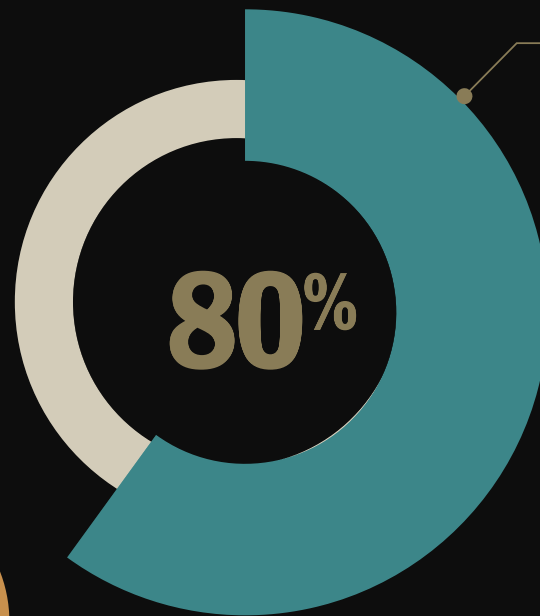


*Hale, Reyes MBARI 2019 Summer Intern results

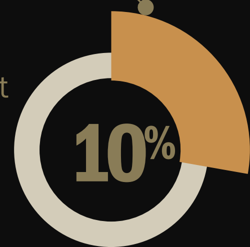
Data

Data breakdown for ML

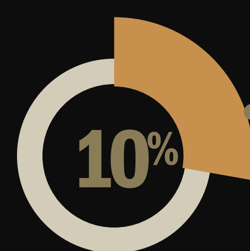
Training Data
Images and labels the models are trained on



Testing Data
Images and labels used to test the model performance but **not** make decisions about fine-tuning



Development/ cross-validation Data
Images and labels used to make decisions about the model. This should be what you want to perform well on.



Distribution Matters

All these should be from the same distribution you want to perform well on for best results.

Labeling Requirements

Images

- **Object detection** (label all the concepts in a frame)
- **Image classification** (beauty shots or crop images from object detection labels)
- **Attributes** per each label or for taxa groups to enable zero-shot classification models. This is useful to identify species not previously seen before. (stretch goal)

Video

- **Tracking** sequences (60 - 600 frames) Can be generated from the image labels extracted from video

QA

- Multiple annotators?

Metrics

A discussion about the metrics for evaluation and a criteria for success

How do we know if have enough training data, or if more focus is needed on models? What is the desired performance?

"A **single** metric speeds up iterations with a well-chosen test and development set" - Andrew Ng, Machine Learning Yearning

Technical Strategy for AI Engineers in the Era of Deep Learning

Metrics are task dependent

***Detection**
Precision

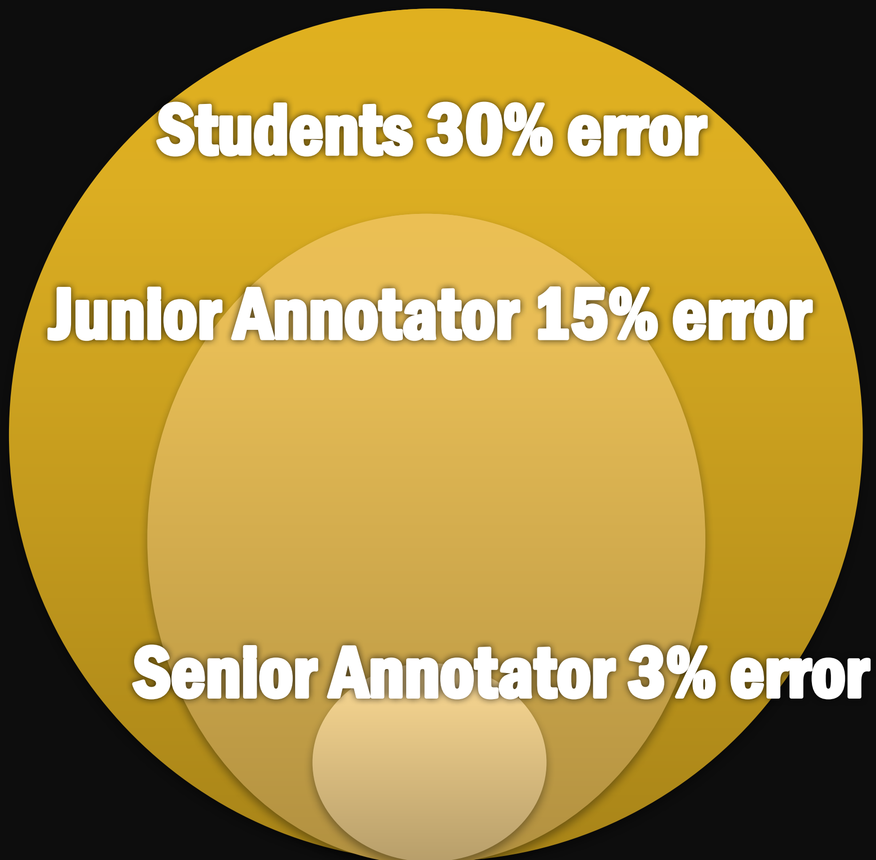
Classification
Accuracy/Error
Precision
F1 Score

**** Tracking**
Accuracy/Error
Precision

* mean Average Precision (mAP) at 50% Intersection Over Union (IOU) small, medium, large
** Multiple Object Tracking Accuracy and Precision

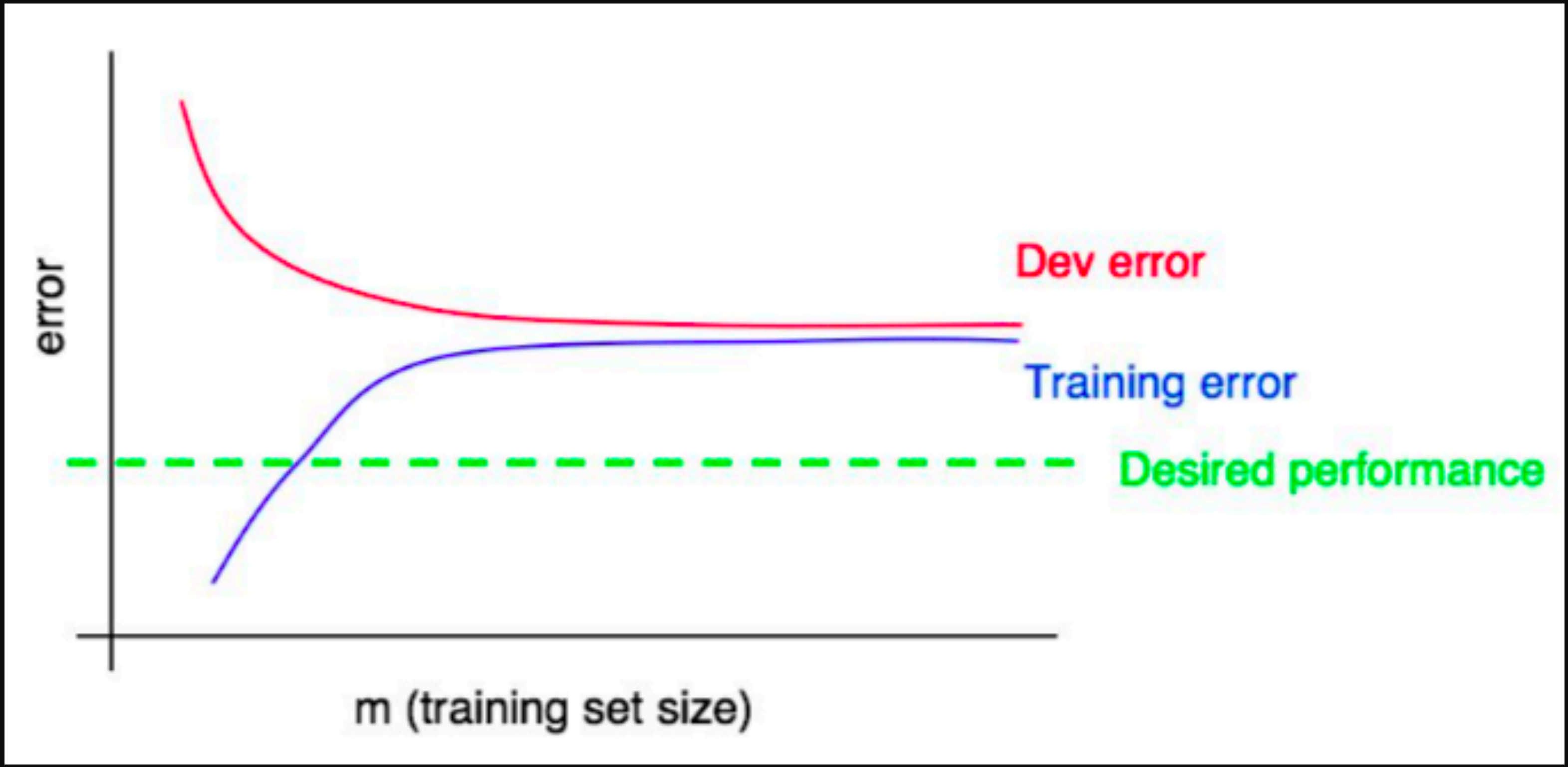
Metrics

What is the desired performance?



The value of learning curves

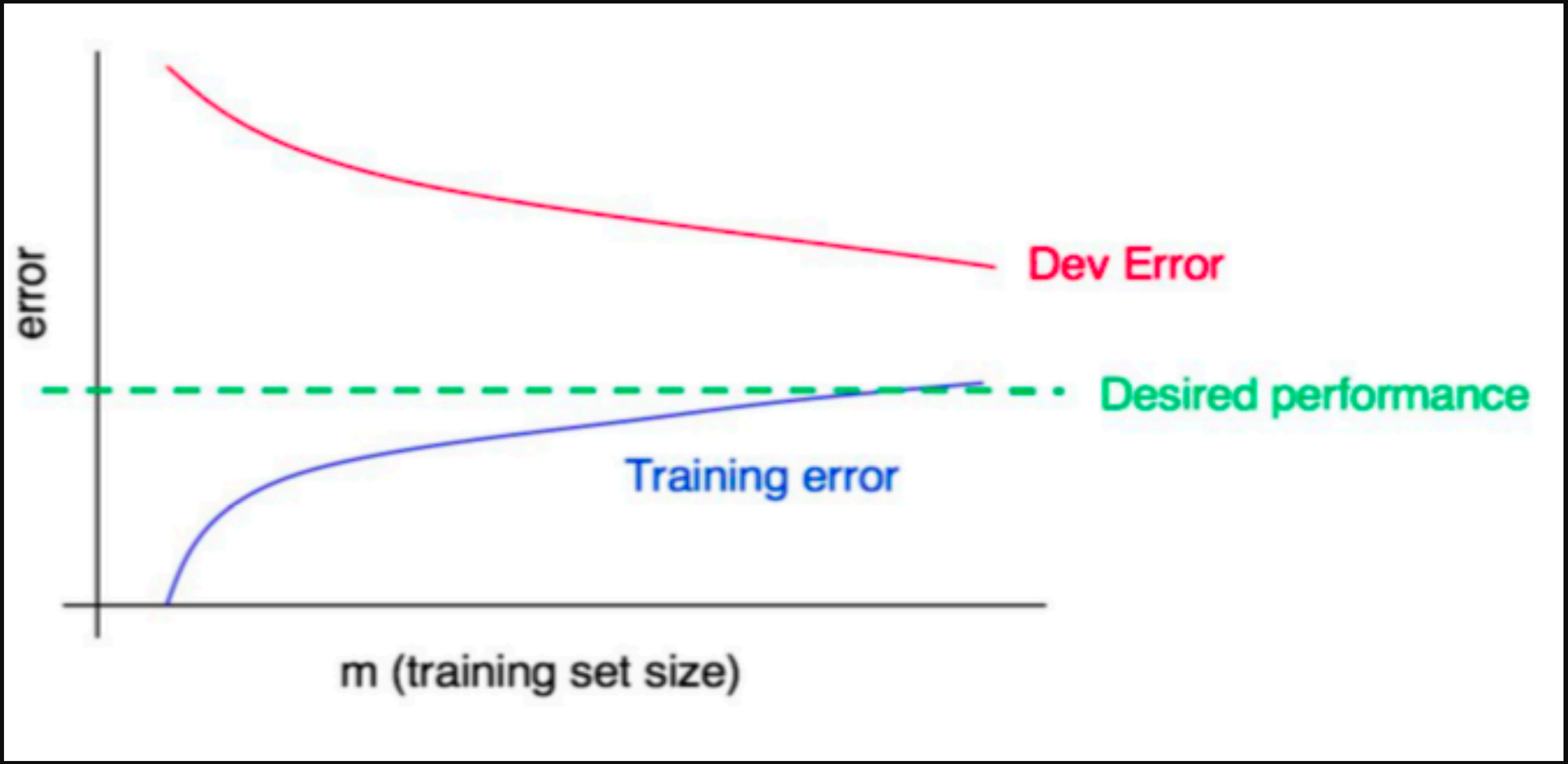
Adding more training won't help. New model? data



Source: Andrew Ng, Machine Learning Yearning Technical Strategy for AI Engineers in the Era of Deep Learning

The value of learning curves

Add training data will help close the gap



Source: Andrew Ng, Machine Learning Yearning Technical Strategy for AI Engineers in the Era of Deep Learning

thank you

Danelle Cline
dcline@mbari.org
831-775-1947