
Assisted Annotation using Machine Learning

Danelle Cline

Monterey Bay Aquarium Research Institute (MBARI)
Senior Software Engineer

Past, Present and Future

A quick reflection of where
we are, have been and what
the future might bring.

Past, Present and Future

VARs – 25+ years

Science Insight

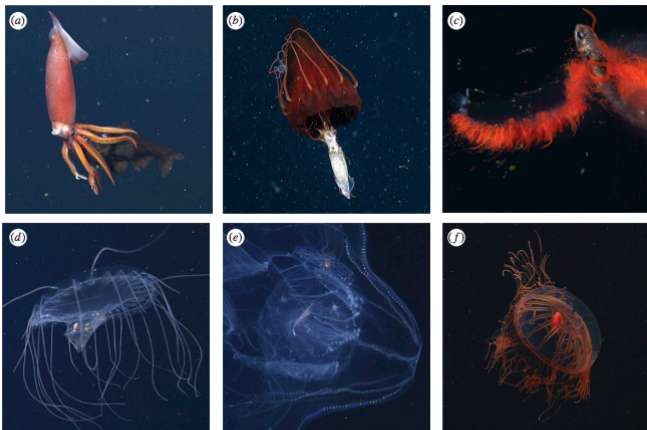
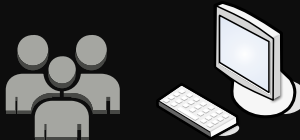


Figure 2. A suite of six illustrative ROV frame grabs of pelagic predators and their prey included in this study. No scale bars for size are available from these sequences. From left to right, top to bottom: (a) *Gonatus* sp. (squid) feeding on a bathylagid fish (Bathylagidae); (b) *Periphylla periphylla*, the helmet jellyfish, feeding on a gonatid squid (Gonatidae), with a small narcomedusa (*Aequorea* sp.) also captured; (c) an undescribed physonekt siphonophore known as 'the galaxy siphonophore' feeding on a lanternfish of the family Myctophidae; (d) a narcomedusa, *Solmissus*, ingesting a salp chain (Salpida); (e) the ctenophore *Thalassocalyce inconstans*, with a euphausiid (Euphausiacea) in its gut; and (f) the tachymedusa, *Halltrephes moosi*, with a large, red mysid (Mysidae) in its gut.

Science Annotators



Unexpected Discovery

June 5, 2013

MBARI research shows where trash accumulates in the deep sea

Surprisingly large amounts of discarded trash end up in the ocean. Plastic bags, aluminum cans, and fishing debris not only clutter our beaches, but accumulate in open-ocean areas such as the "Great Pacific Garbage Patch." Now, a paper by researchers at the Monterey Bay Aquarium Research Institute (MBARI) shows that trash is also accumulating in the deep sea, particularly in Monterey Canyon.



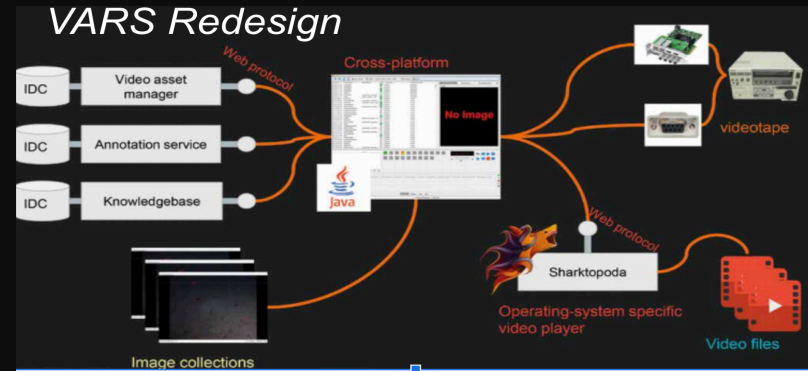
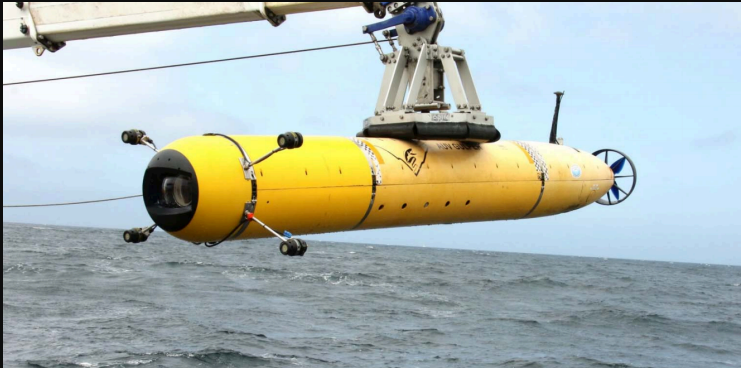
A discarded tire sits on a ledge 868 meters (2,850 feet) below the ocean surface in Monterey Canyon. Image: ©2009 MBARI

Kyra Schlining, lead author on this study, said, "We were inspired by a fisheries study off Southern

California that looked at seafloor trash down to 365 meters. We were able to continue this search in deeper water—down to 4,000 meters. Our study also covered a longer

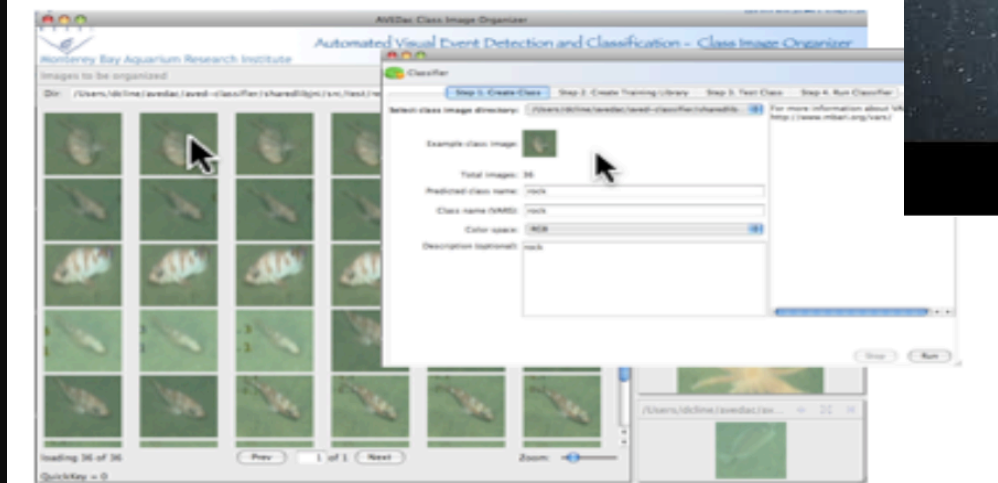
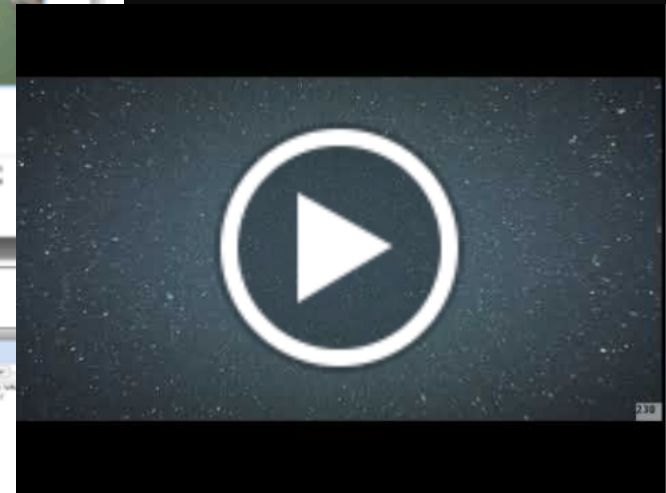
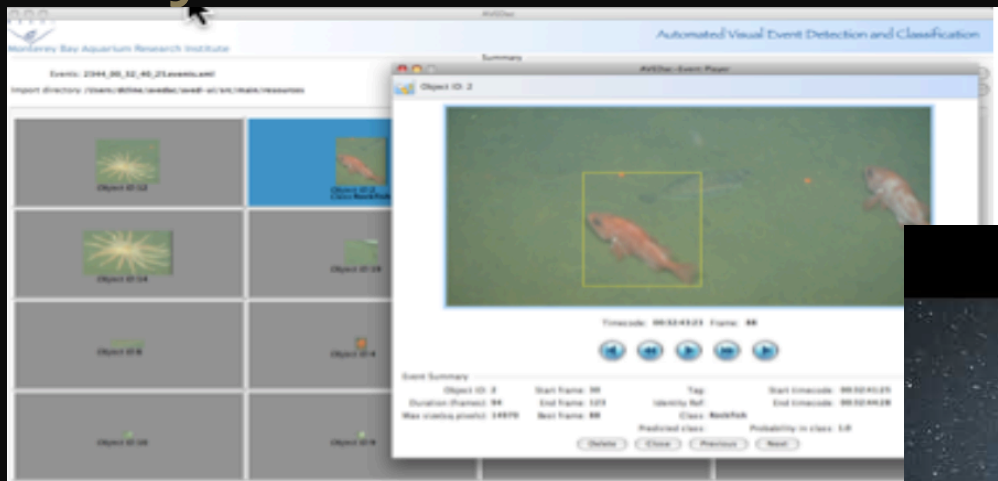
Past, Present and Future

New platforms and we've gone digital



Past, Present and Future

Early visions of annotation assistance

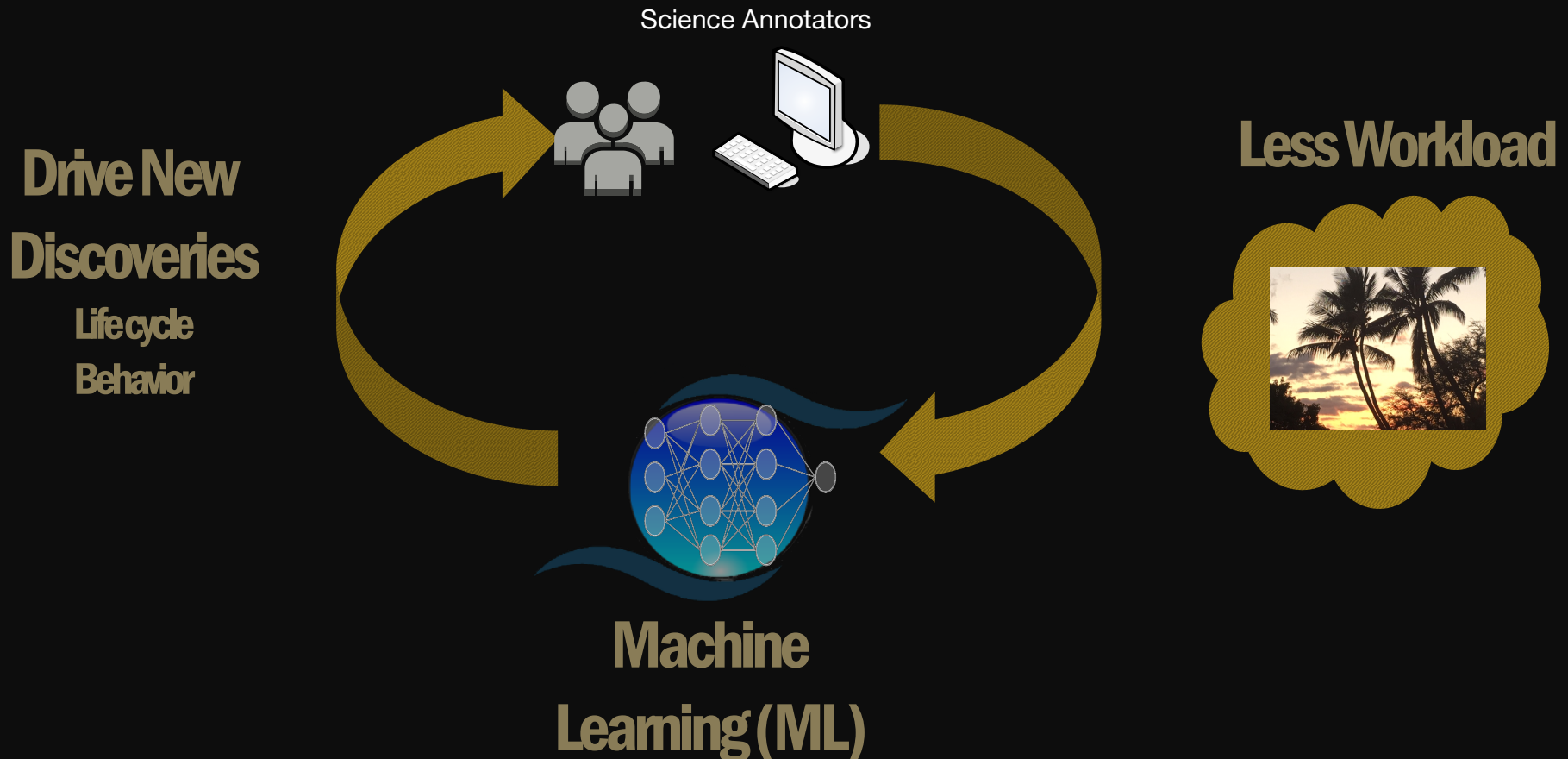


Projects

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

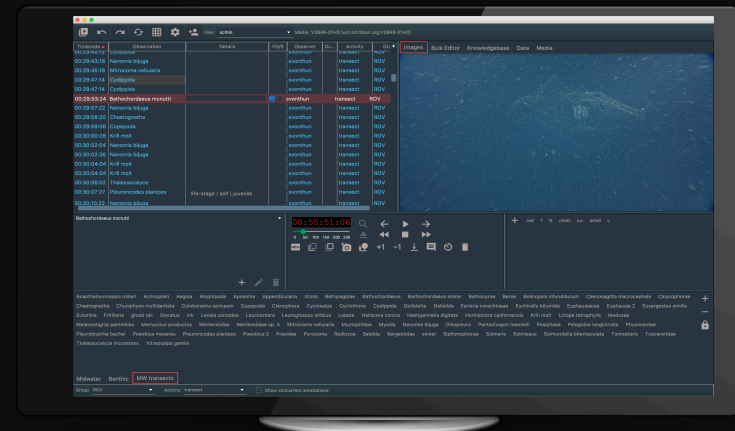
Past, Present and Future

A new direction – Annotation Assistance using Machine Learning



Assistance using the tools we know

VARs

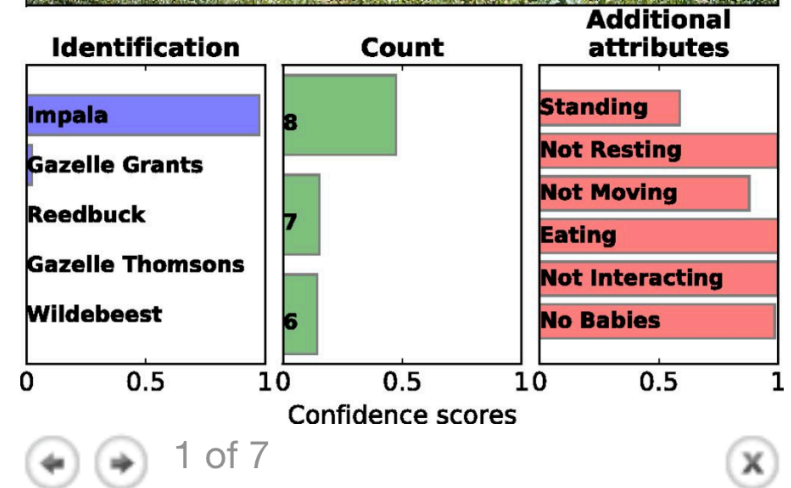


Past, Present and Future

Success using ML in animal camera-trap images

Human answer: 8 Impala (Standing, Eating)

Model answer: 8 Impala (Standing, Eating)



Source: Norouzzadeh, Mohammad Sadegh & Nguyen, Anh & Kosmala, Margaret & Swanson, Ali & Packer, Craig & Clune, Jeff. (2017). Automatically identifying wild animals in camera trap images with deep learning. Proceedings of the National Academy of Sciences. 115. 10.1073/pnas.1719367115.

Annotation Tasks

**A quick look at essential
tasks**

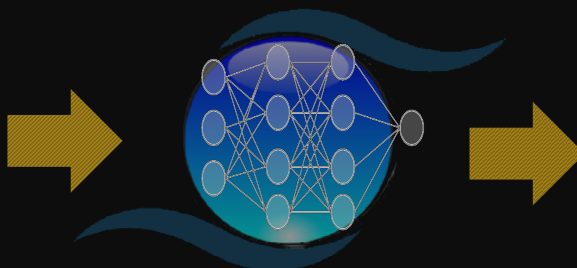
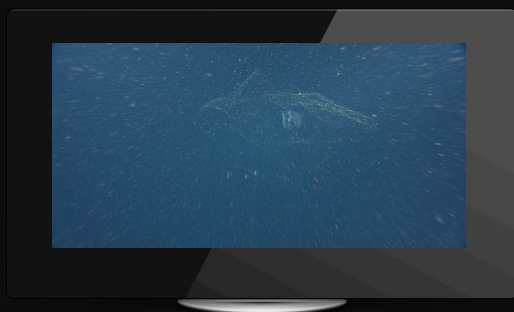
What is the equivalent task
in ML?

Annotation Tasks

Where is a concept and what is it?

ML

Object Detection and
Classification

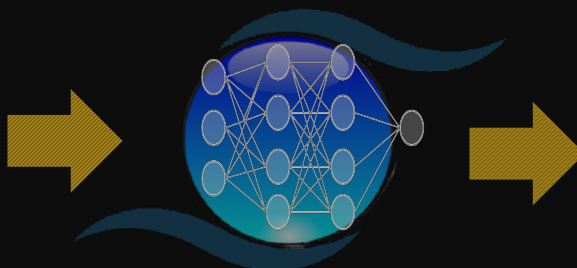
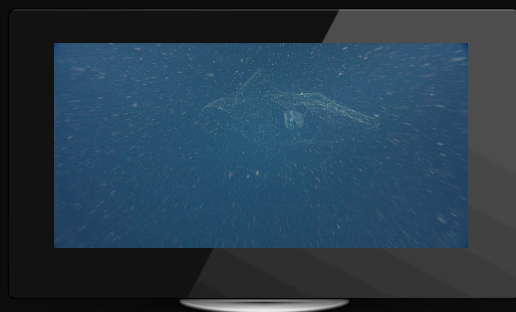


Annotation Tasks

Track the concept

ML

Visual Event ID4

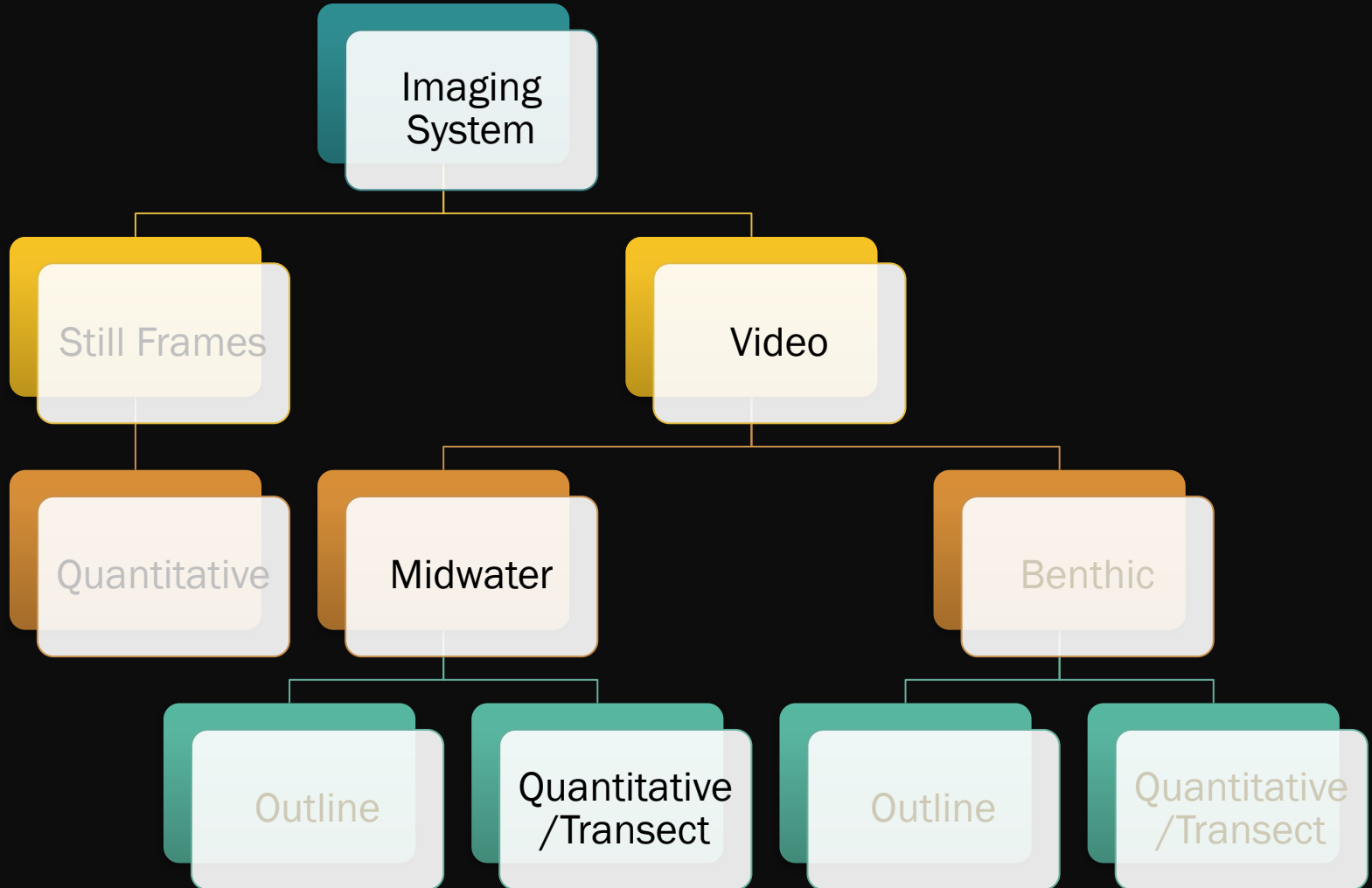


Use Cases and Workflow

A discussion about use
cases and a workflow

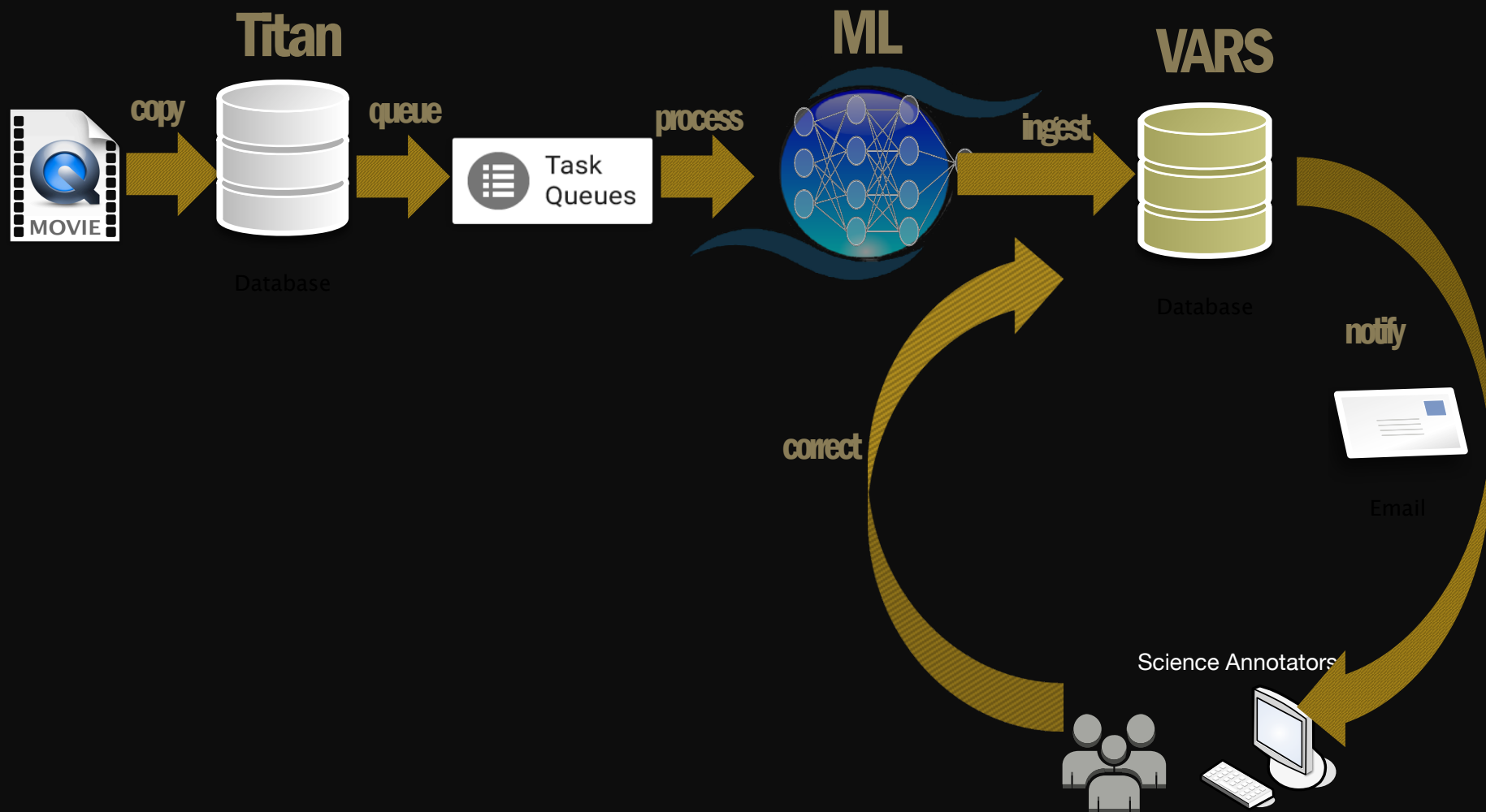
Use Cases and Workflow

Initial focus



Use Cases and Workflow

Final Workflow



Use Cases and Workflow

3-stage pipeline for video



Use Cases and Workflow

2-stage pipeline for stills



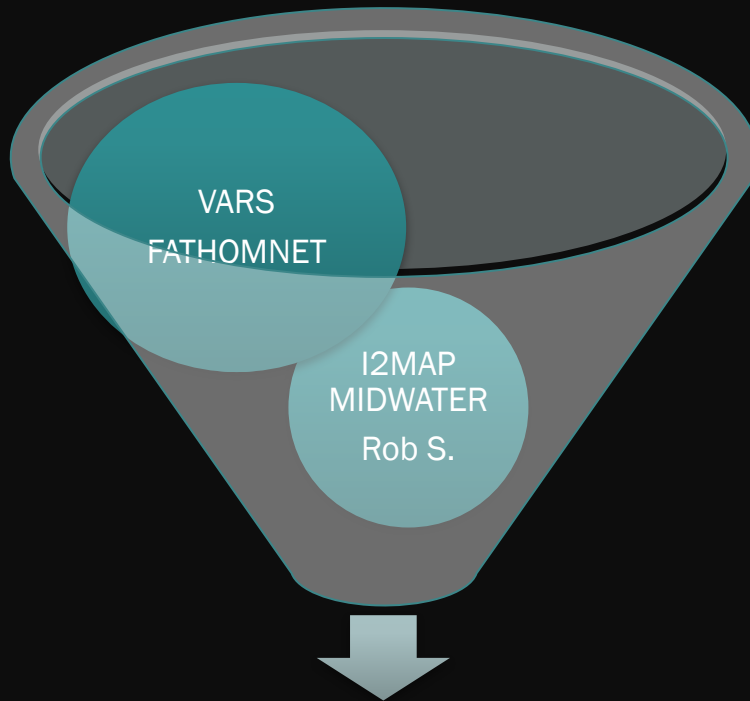
Data

**A detour into what data is
needed for ML**

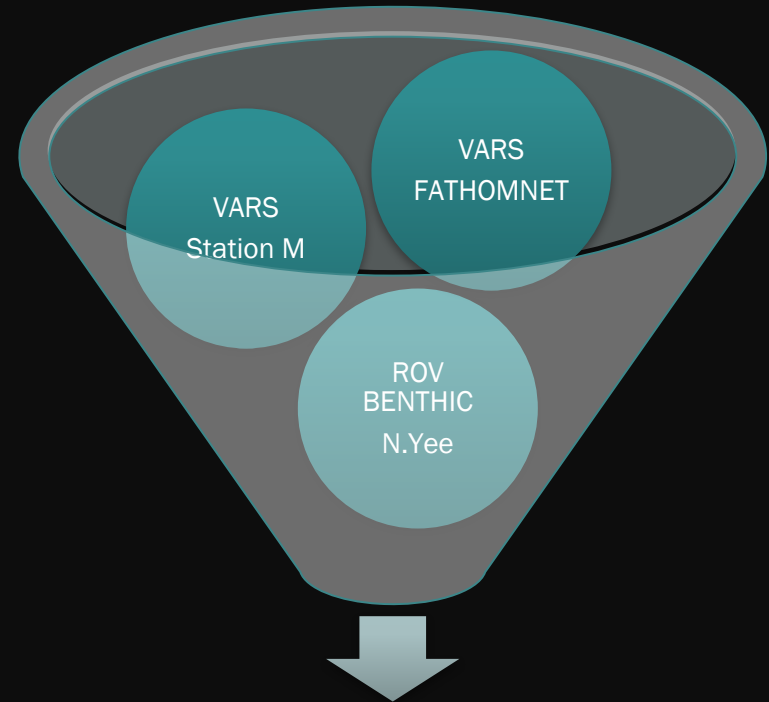
What is a good training,
testing, and development
set for the CNN methods we
are using?

Data

Data pool



Midwater



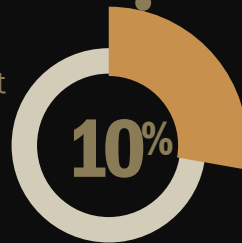
Benthic

Data

Data breakdown

Testing Data

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

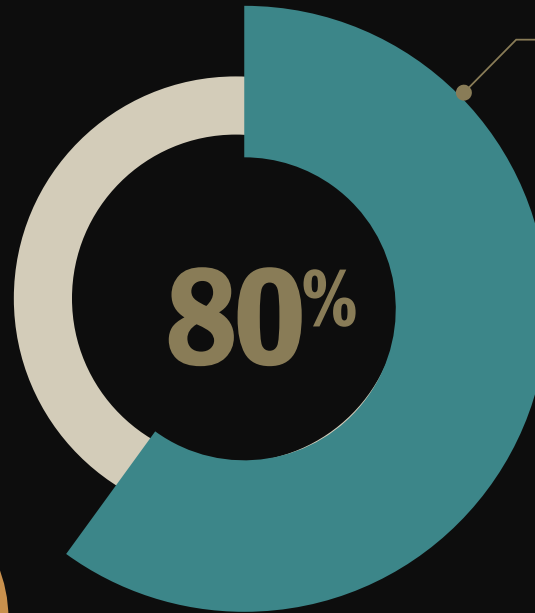


Distribution Matters

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

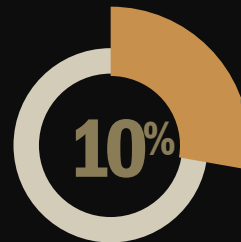
Training Data

Images and labels the models are trained on



Development/ cross-validation Data

Images and labels used to make decisions about the model. This should be what you want to perform well on.



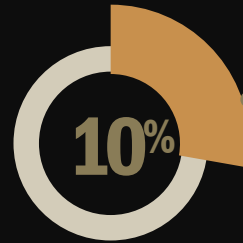
Data

Development data set needed

Select dives currently working on

Bootstrap label with 3-stage pipeline

Label all parts that are annotated, e.g. ascent/descent/transect



Development/cross-validation Data

Images and labels used to make decisions about the model. This should be what you want to perform well on.

Metrics

A discussion about the metrics for evaluation and criteria for success

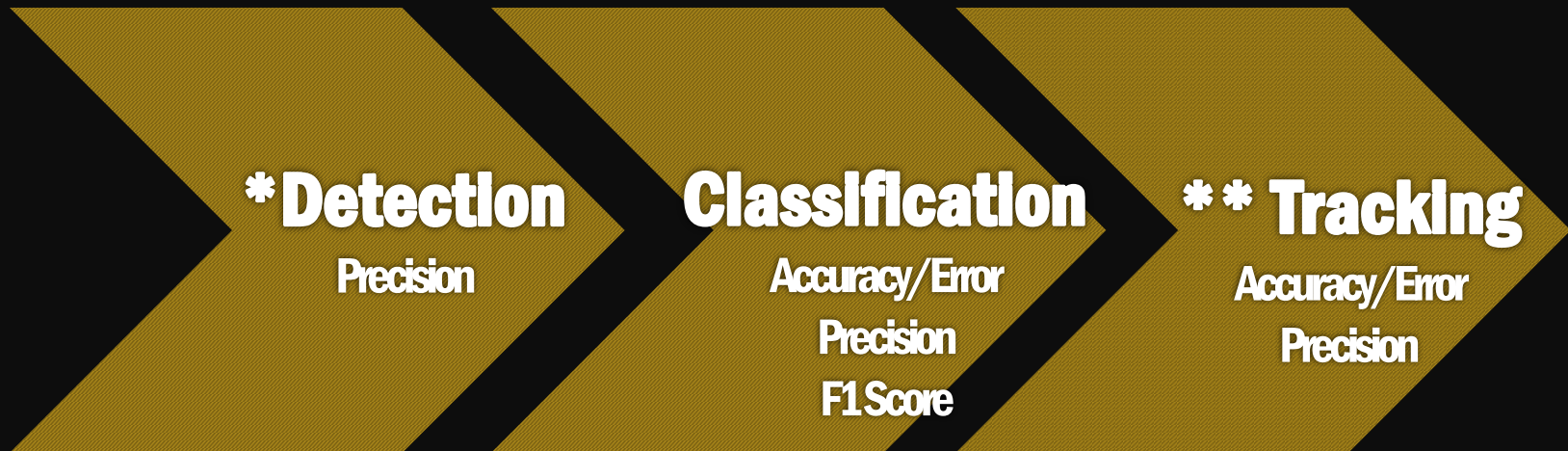
How do we know if have enough training data, or if further focus is needed on models? What is the desired performance?

Metrics

“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

Metrics are task dependent



* mean Average Precision (mAP) at 50% Intersection Over Union (IOU)

** Multiple Object Tracking Accuracy and Precision

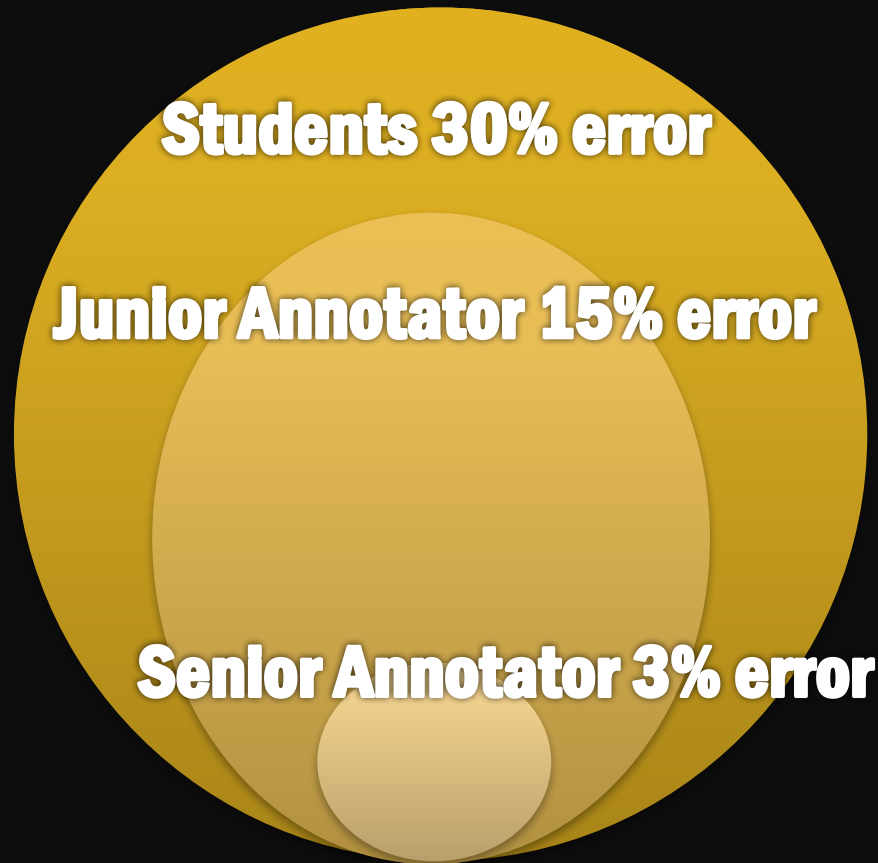
Metrics

We will look at all of these metrics, but tracking accuracy/error is perhaps the most important for video



Metrics

What is the desired performance?

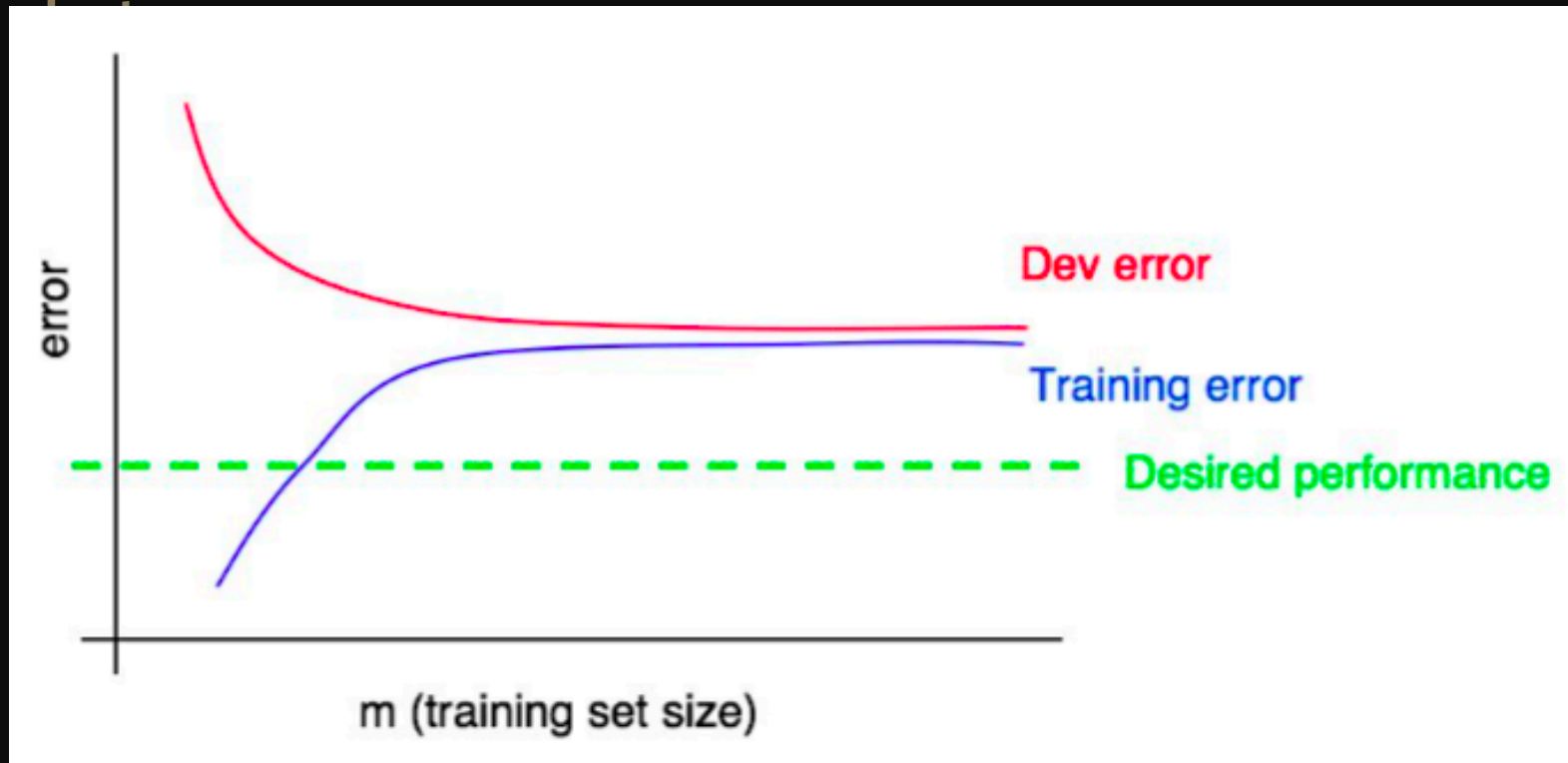


This only matters if the pipeline has greater than 30 % error

Metrics

The value of learning curves

Add more training won't help. New model?



Metrics

The value of learning curves

Add training data will help close the gap



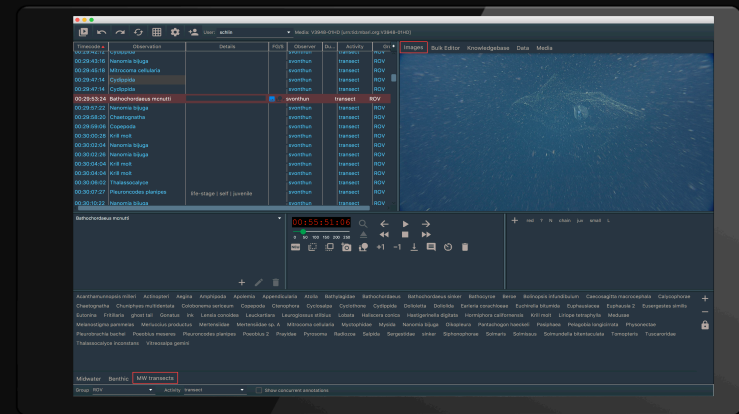
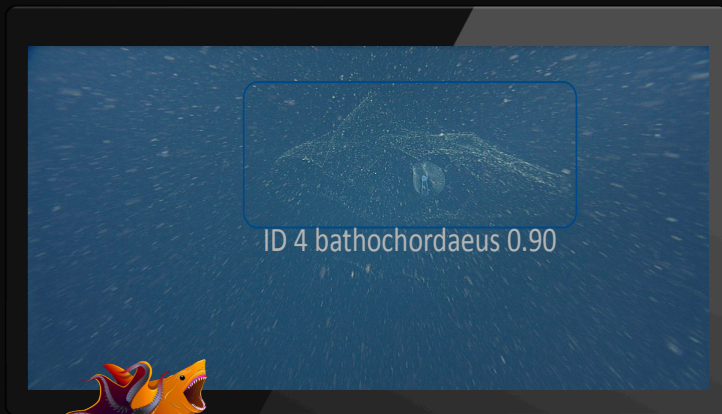
Annotation Assistance

**Steps to develop annotation
assistance in VARS**

Milestones

Annotation Assistance

Users perspective



Annotation Assistance

CAUTION



Expect few bumps in the road

Improving ML performance is a highly iterative process and the tools for production level ML are in the nascent stages.

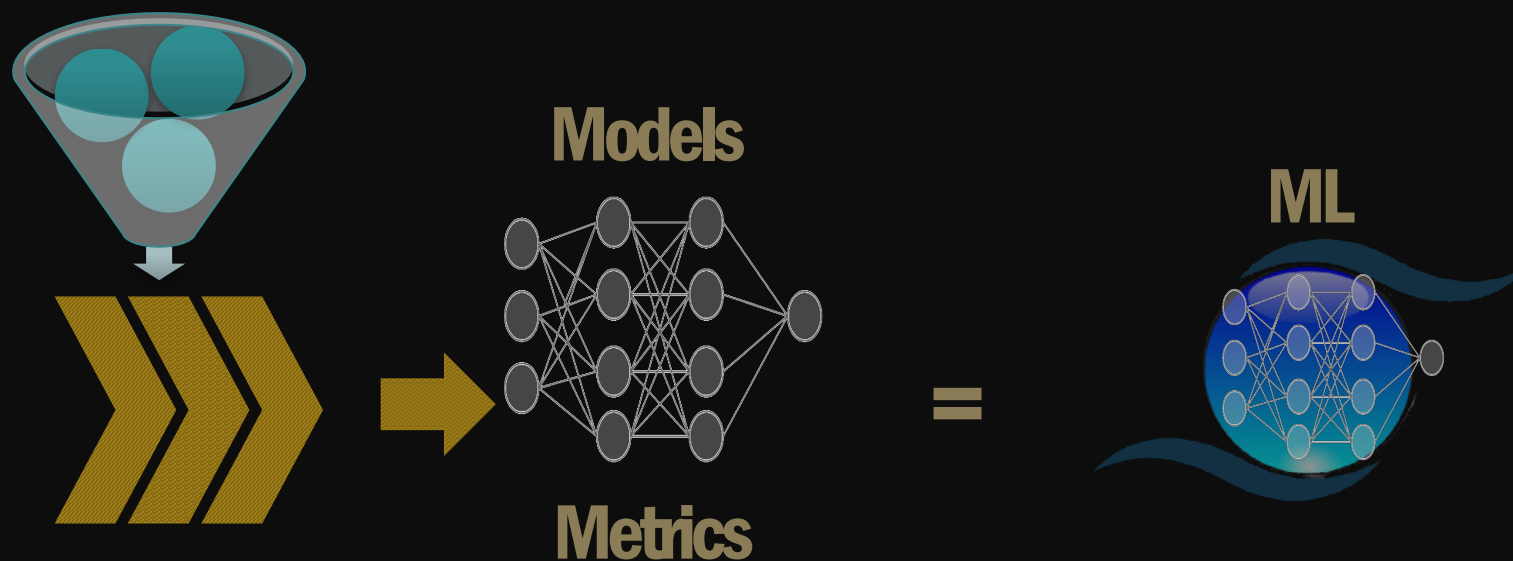
Annotation Assistance - Milestone 1

VARs import and export of labeled data



Annotation Assistance - Milestone 2

Reproducible metrics, models and workflow with DVC



DVC

Open-source Version Control System for Machine Learning Projects

Data Version Control core concepts



Metadata stored in Git for **reproducibility**

Supports workflows for **rapid iteration**

Pipeline – directed acyclic graph (DAG) command to reproduce a state

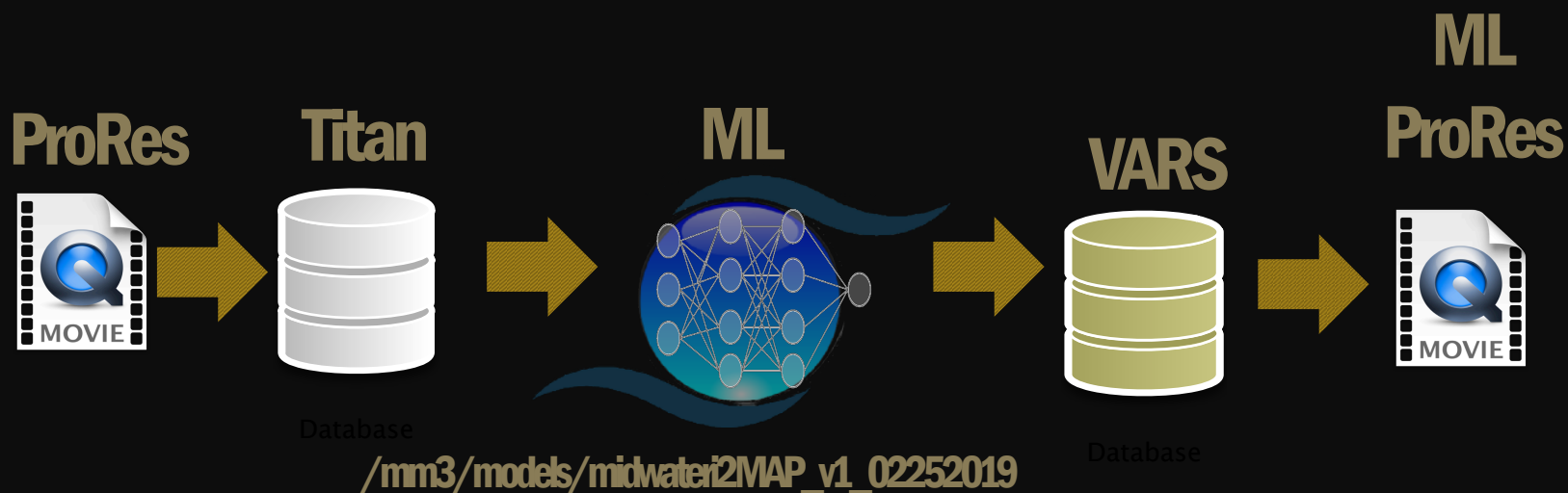
No external service required

Cached data files

Cloud storage supported to share or deploy models remotely

Annotation Assistance - Milestone 3

3-stage pipeline, VARS ingest and playout

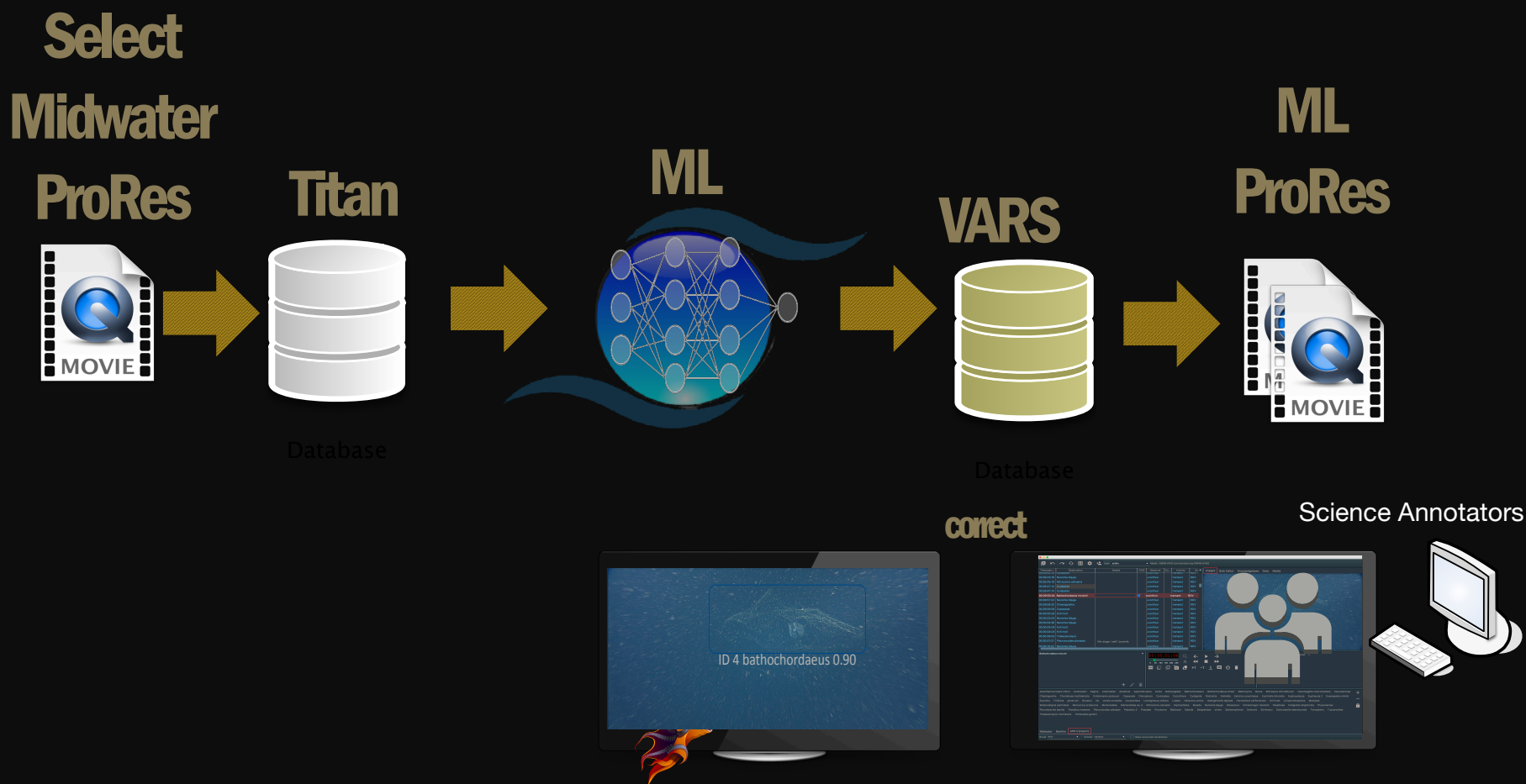


Sharktopoda playout



Annotation Assistance - Milestone 4

Bootstrap development data labeling



Annotation Assistance - Milestone 4

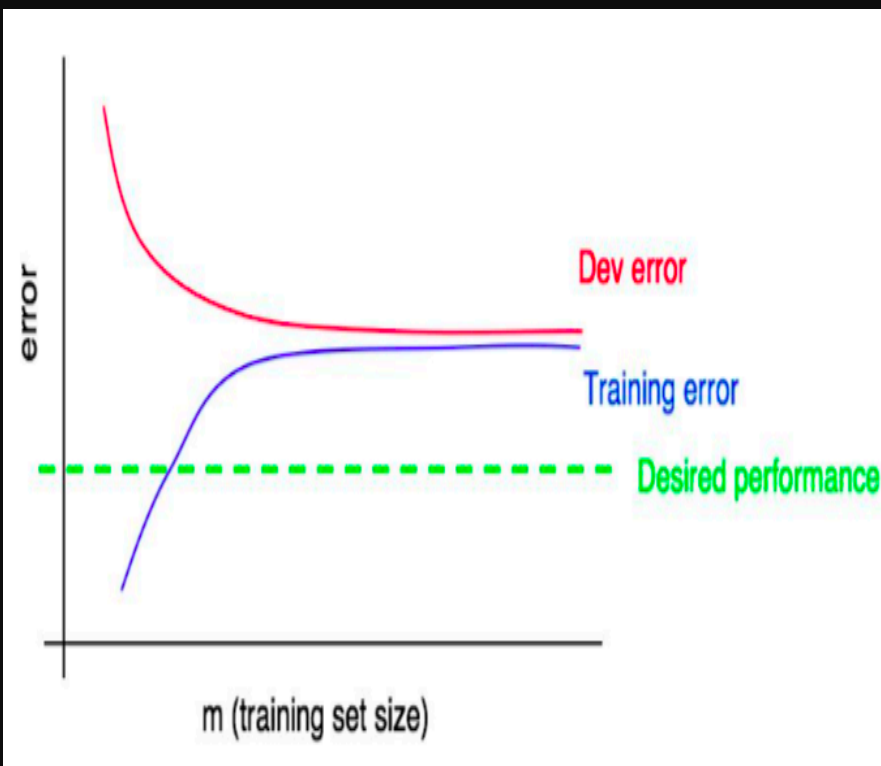
Error analysis

Address common errors (bias and variance)

Get insight from the development data

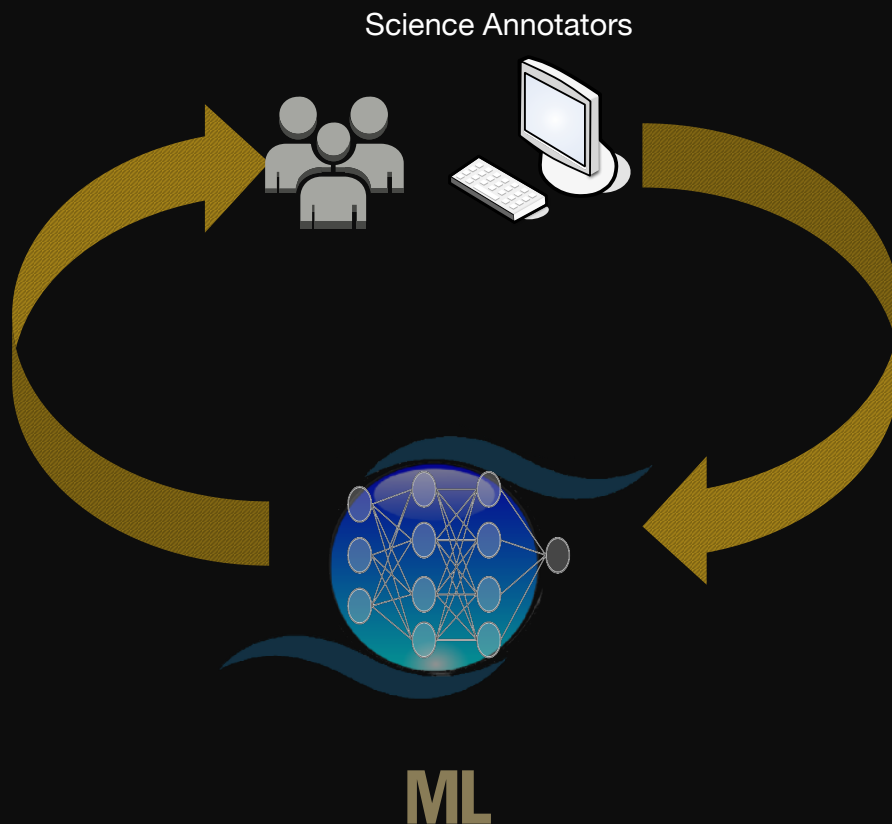
Change single number metric

Correct development data



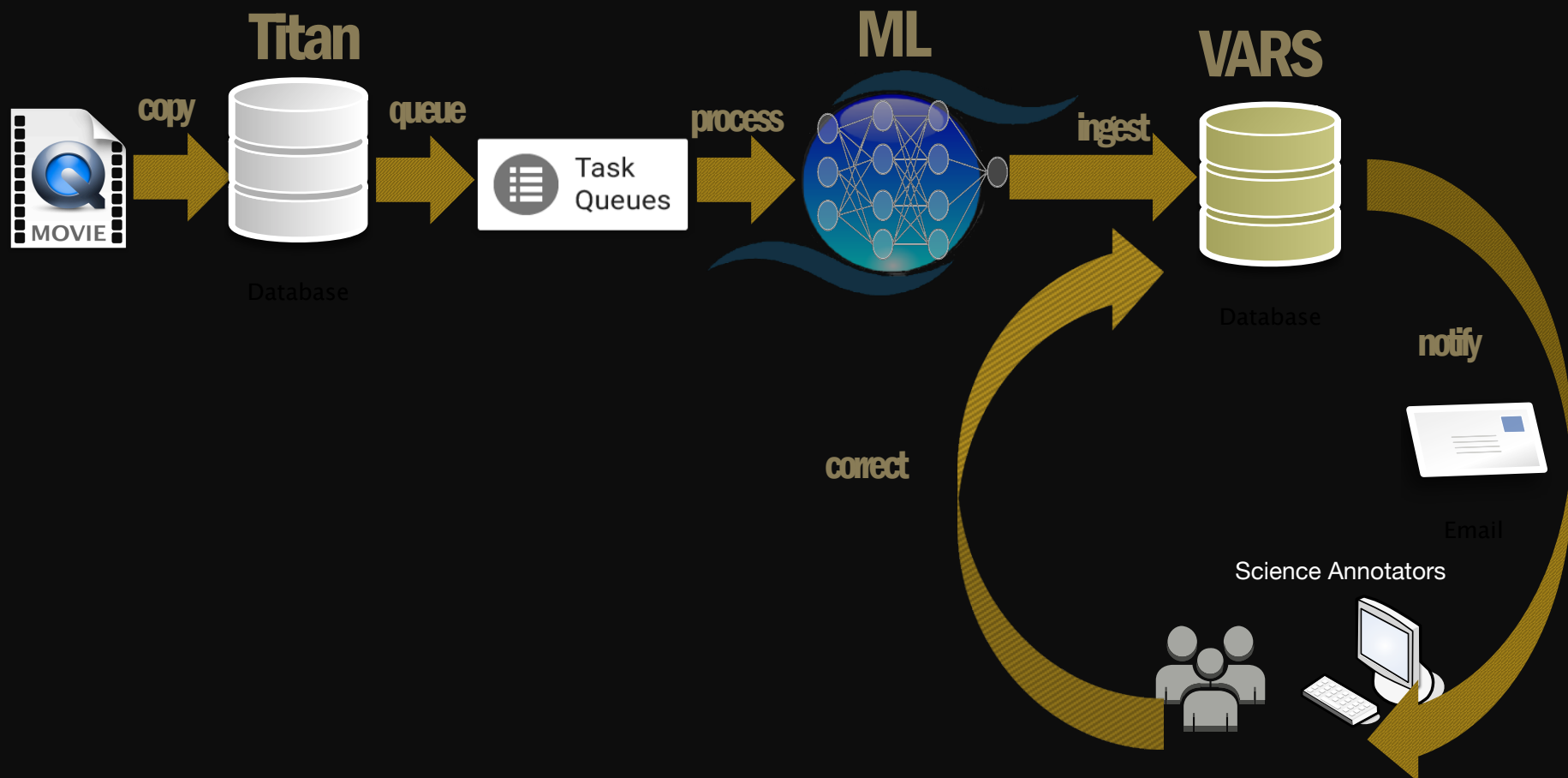
Annotation Assistance - Milestone 5

Rapid iteration for improvement



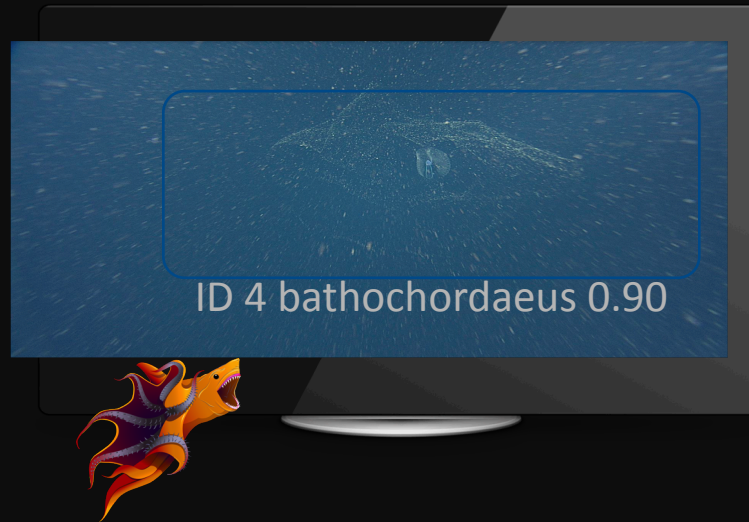
Annotation Assistance - Final

Final workflow



Prototype - Stretch Goal

Interactive Sharktopoda



Methods and Discussion

A more in-depth look at the
methods considered for the
pipelines

Methods and Discussion

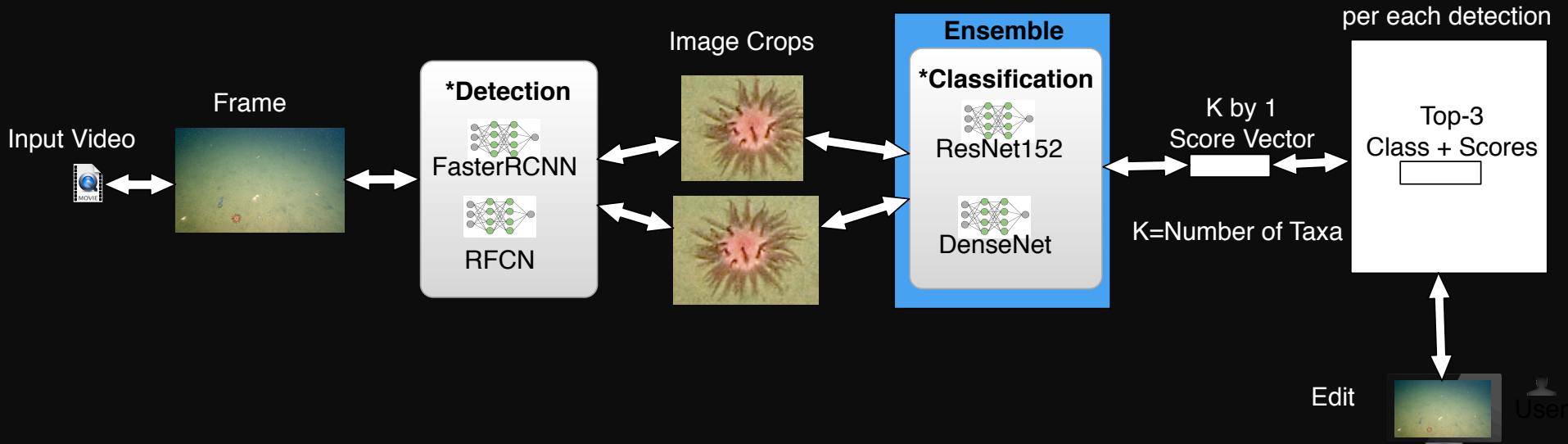
ML decision matrix – theoretical. We don't know what this is just yet

Factors	Data Dimension	Models per frame	Compute time	Algorithm complexity	Score (lower is easier)
Video, outline	3D	1	high	high	1
Video quantitative/transcript	3D	2	high	high	3
Still/quantitative	2D	3	low	medium	2

Methods and Discussion

2-stage detection and classification

Ensemble Detection and Classification



Combining models will improve predictive performance at the cost of increased time per frame

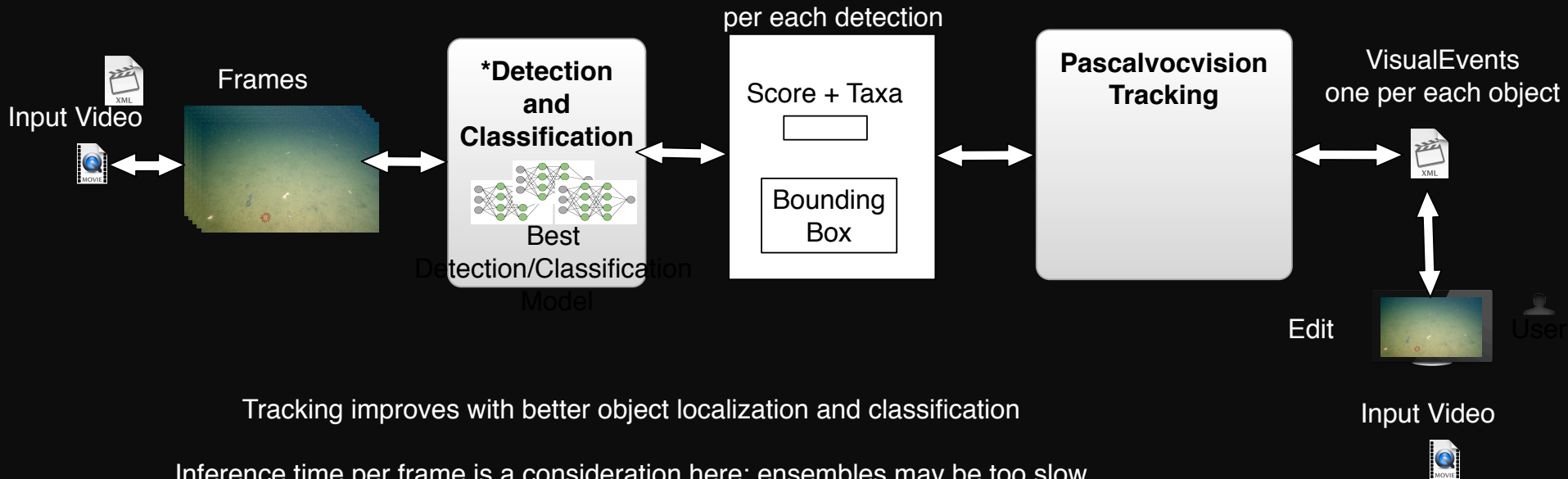
May be more suitable for still-frame analysis tasks

*AutoML to speed-up model tuning

Methods and Discussion

3-stage detection and classification

Detection, Classification and Tracking



Tracking improves with better object localization and classification

Inference time per frame is a consideration here; ensembles may be too slow

*AutoML to speed-up model tuning

thank you

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