Update on improving automated detection of scallops and flounder in optical surveys with stereo detection methods

Liese Siemann, Jason Clermont, Tasha O'Hara
Coonamessett Farm Foundation

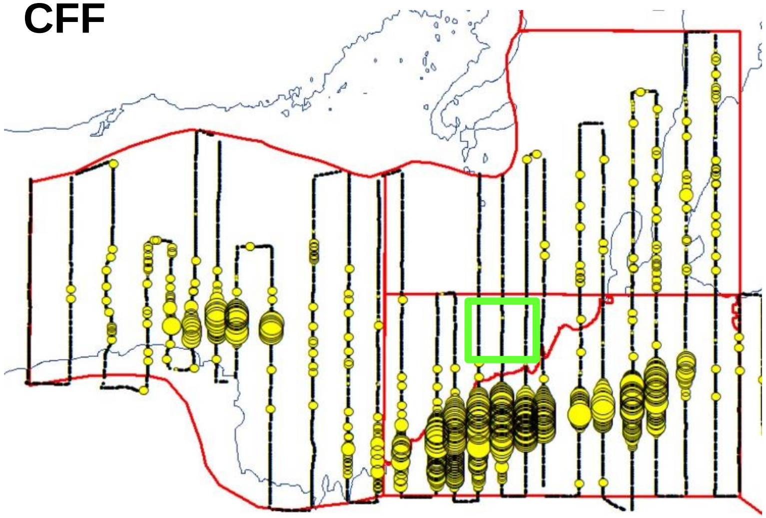
In Collaboration with
Anthony Hoogs, Matthew Dawkins, and Jon Crall
Kitware, Inc.

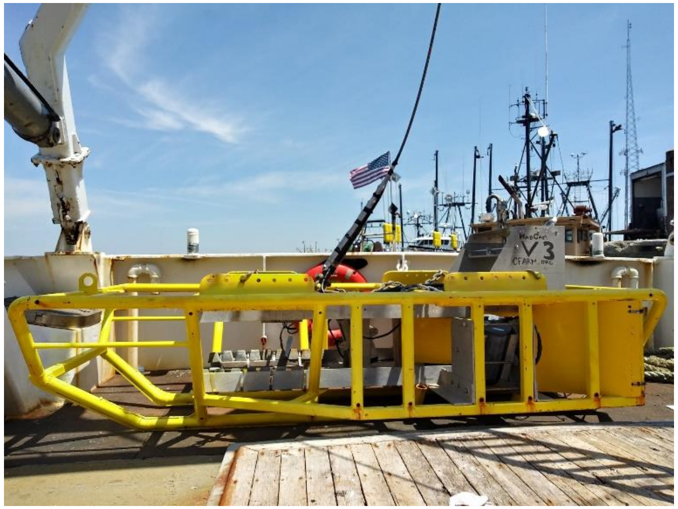
Coonamessett Farm Foundation

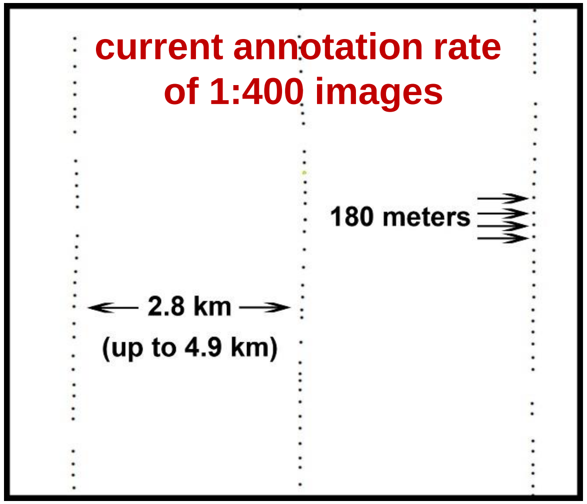
CFF HabCam surveys

Intensive optical surveys using HabCam v3 vs drop camera

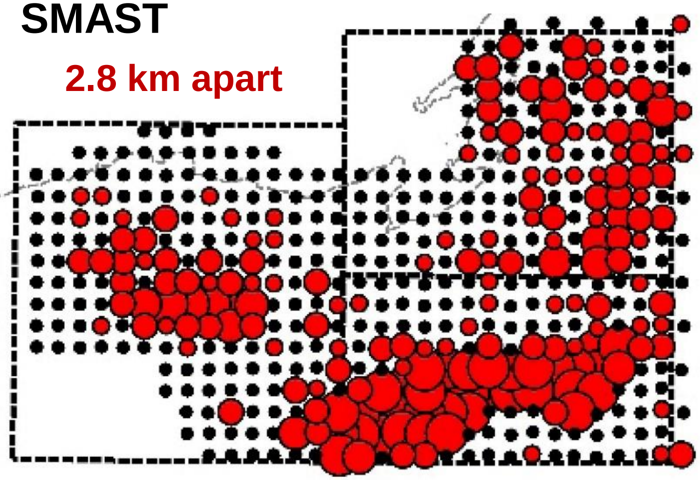
CFF







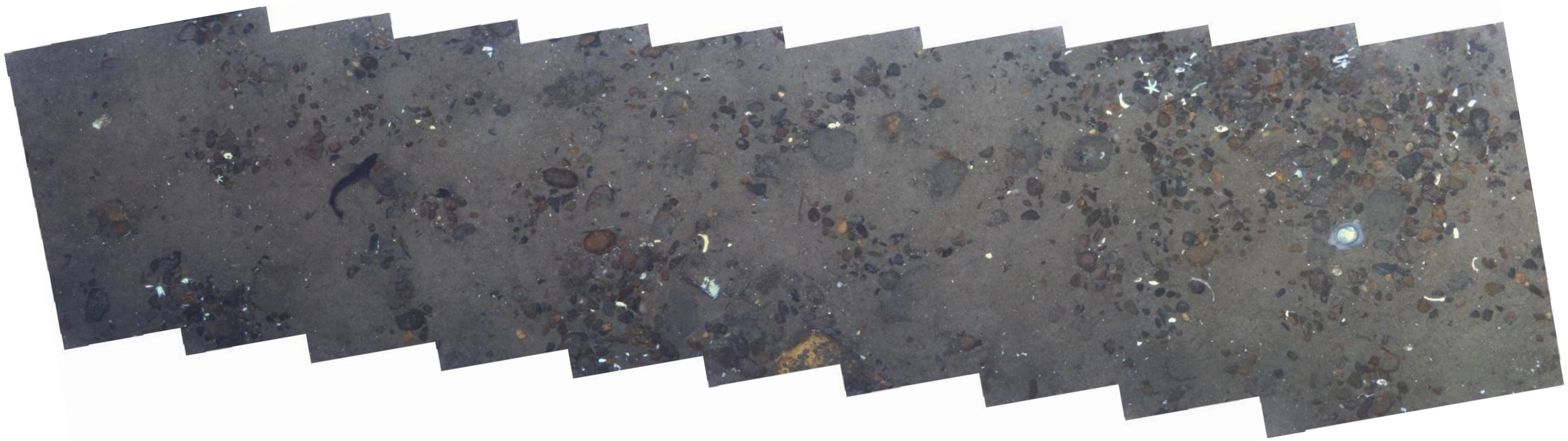
SMAST



Benefits of automated detection

- Higher annotation rates

Could be important depending on the scale of patchiness in scallop distributions and may differ between closed areas vs areas with active fishing

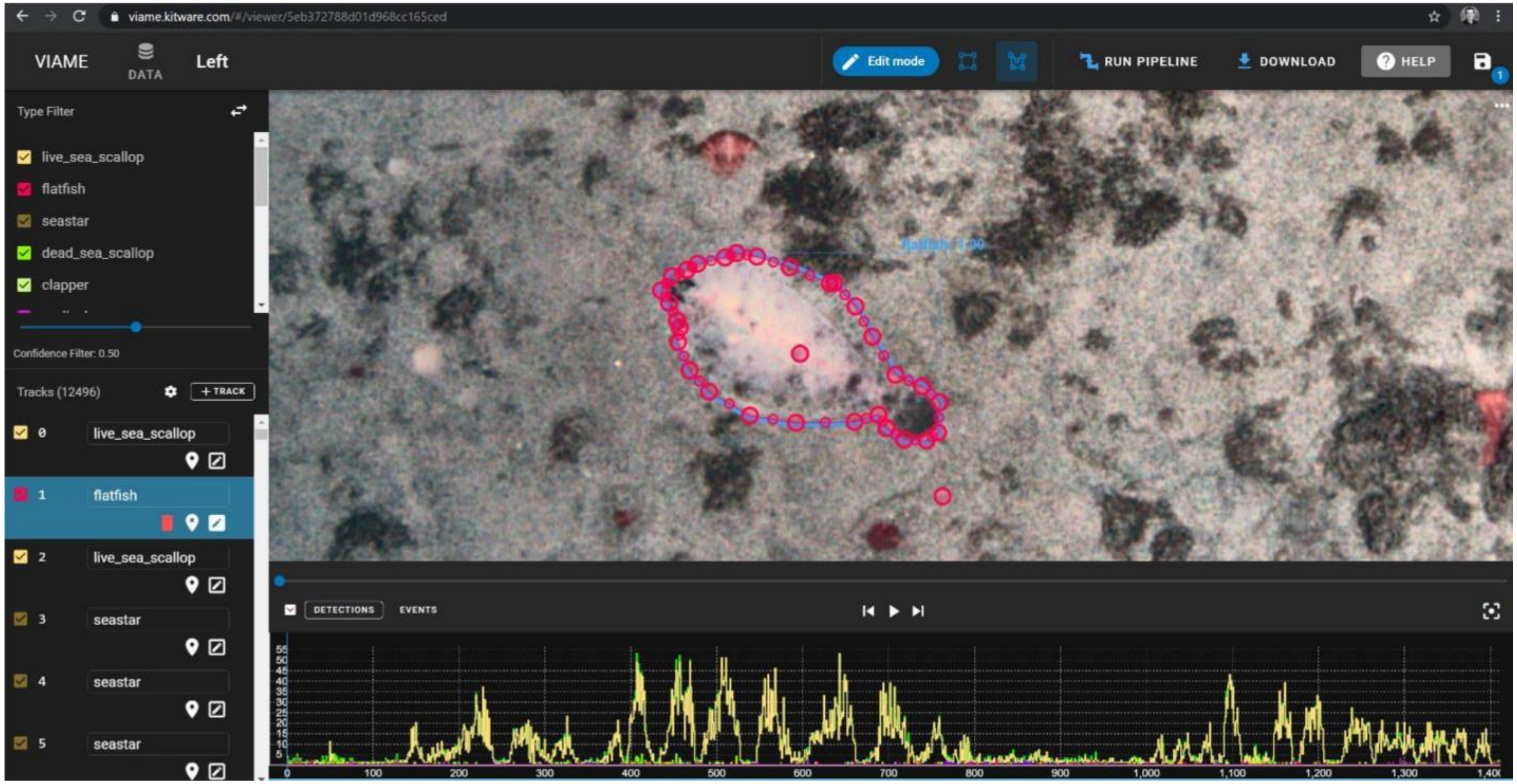


- More efficient and accurate annotation of images

Use of manual annotation is time-consuming and the skill of human annotators can vary significantly

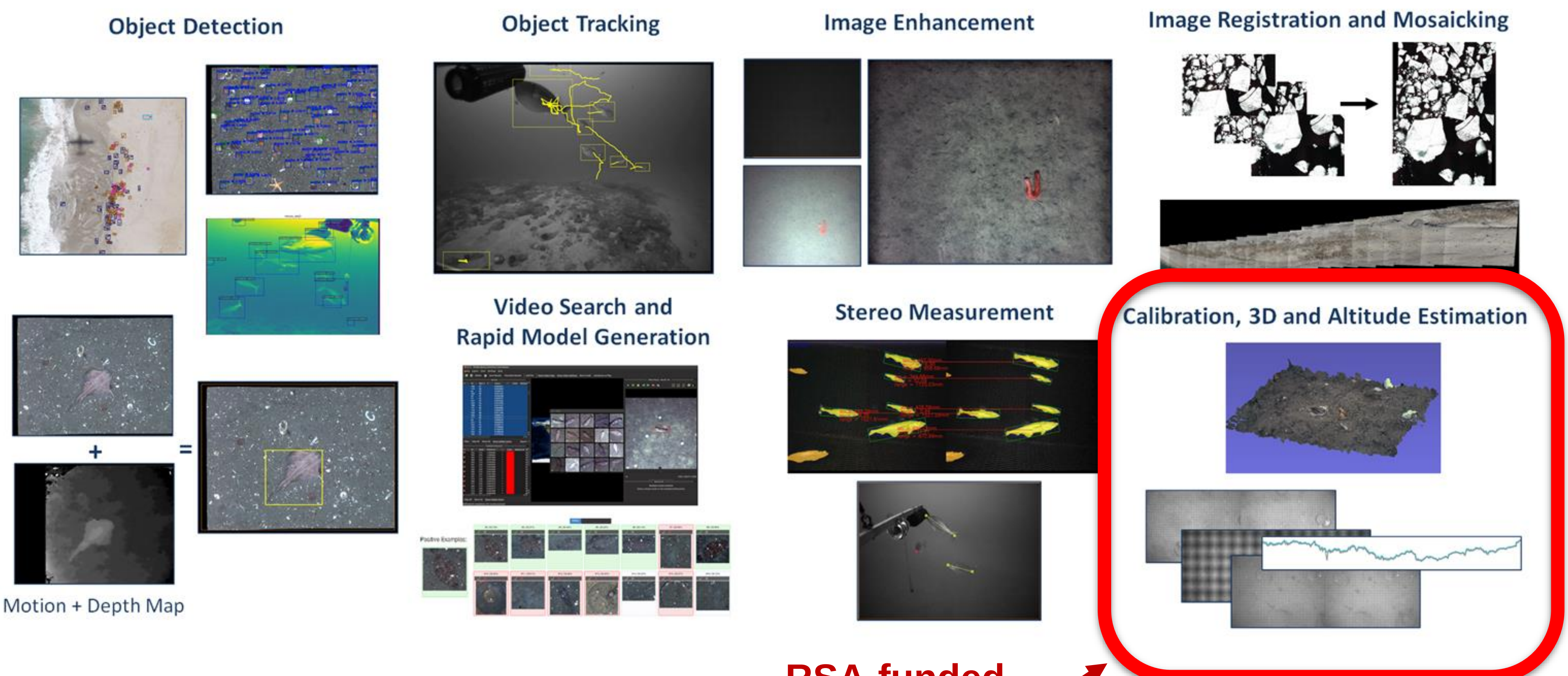
VIAME overview

Kitware began development of Video and Image Analytics for Marine Environments in 2016 with funding from a NOAA strategic initiative on automated image analysis. Its use has since expanded to multiple countries and all six fisheries science centers.



VIAME overview

It has evolved into a DIY toolkit for biologists to train up models themselves, with multiple applications in benthic image analysis (HabCam), optical fish assessments (CamTrawl), aerial image processing (right whale and sea lion surveys), and on-board catch analysis (electronic monitoring).



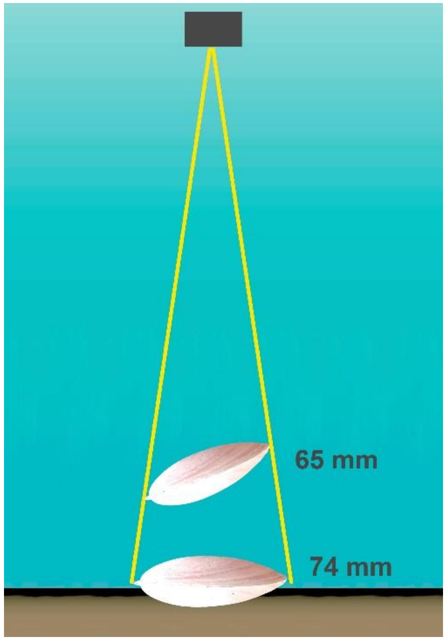
RSA-funded project

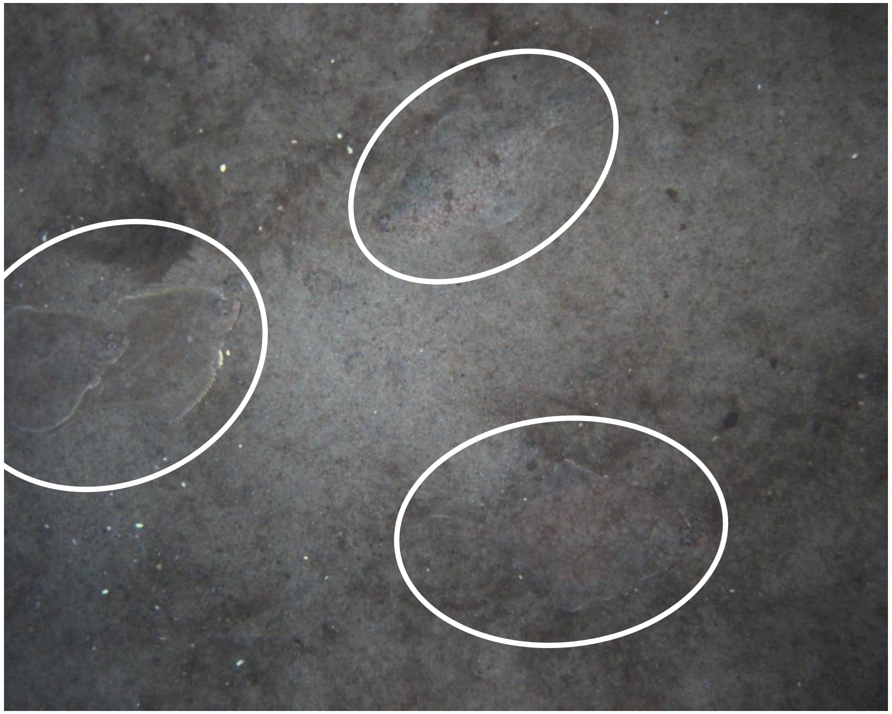
https://www.viametoolkit.org/

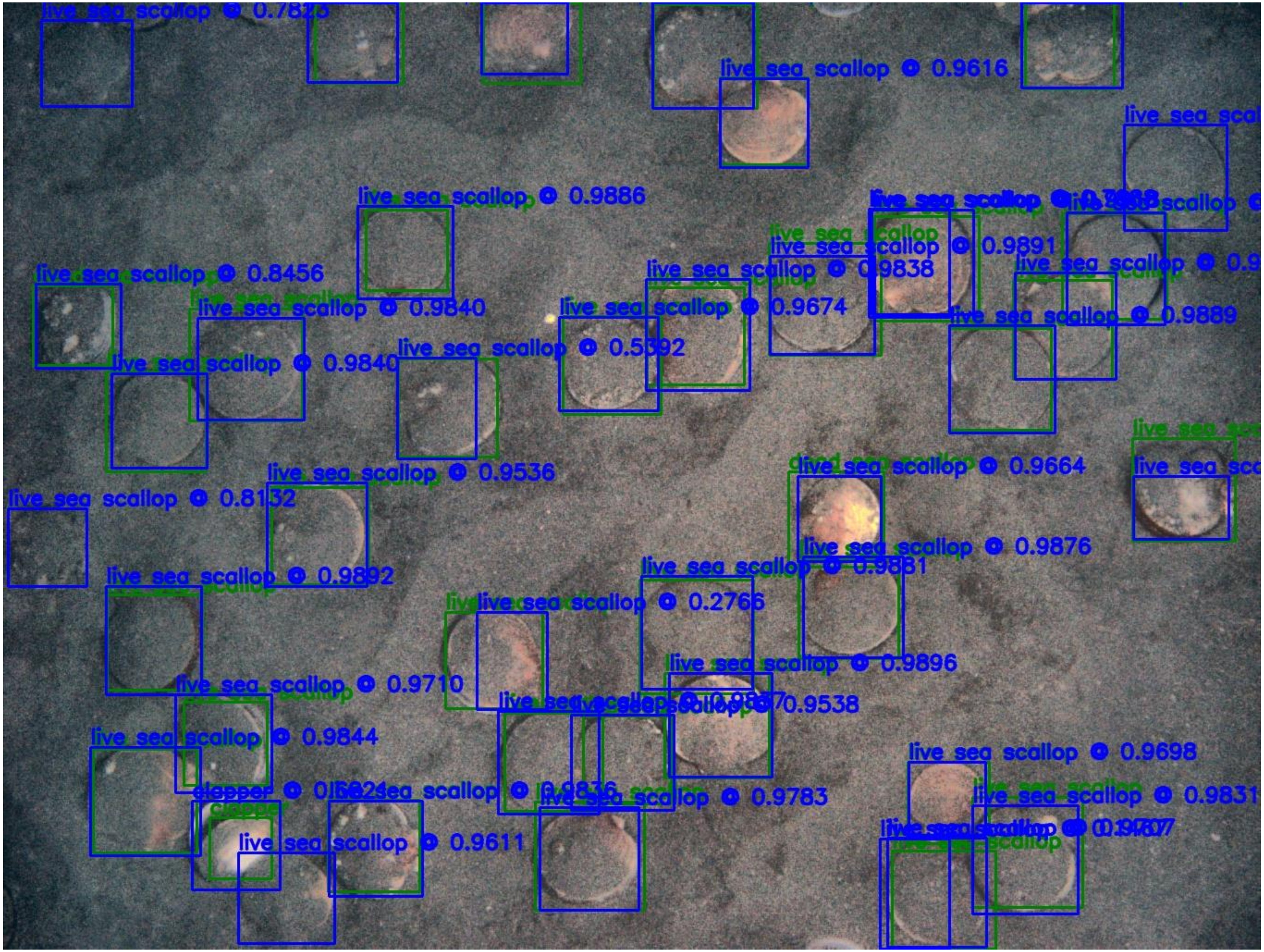
Research Objectives

Develop new algorithms and modules that will use the stereo image pairs taken during HabCam surveys to

1) Improve detection and measurements of scallops and differentiate between live, dead, and swimming scallops
2) Detect and count flounder, fish that are difficult to reliably distinguish due to their camouflage abilities, changing body patterns, and tendency to bury themselves in sandy substrates.

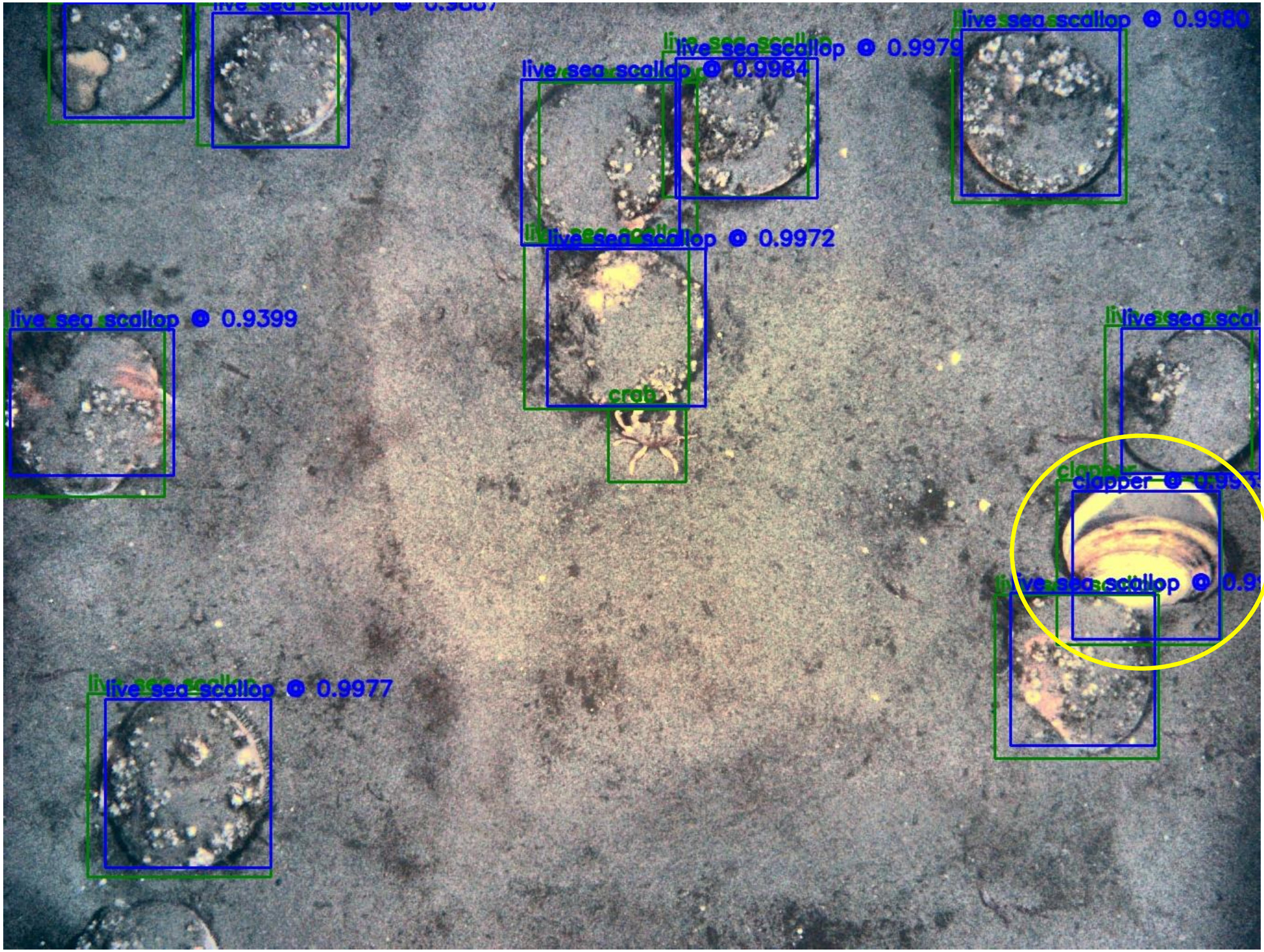






Groundtruth = human annotations

Predicted



Groundtruth = human annotations

Predicted

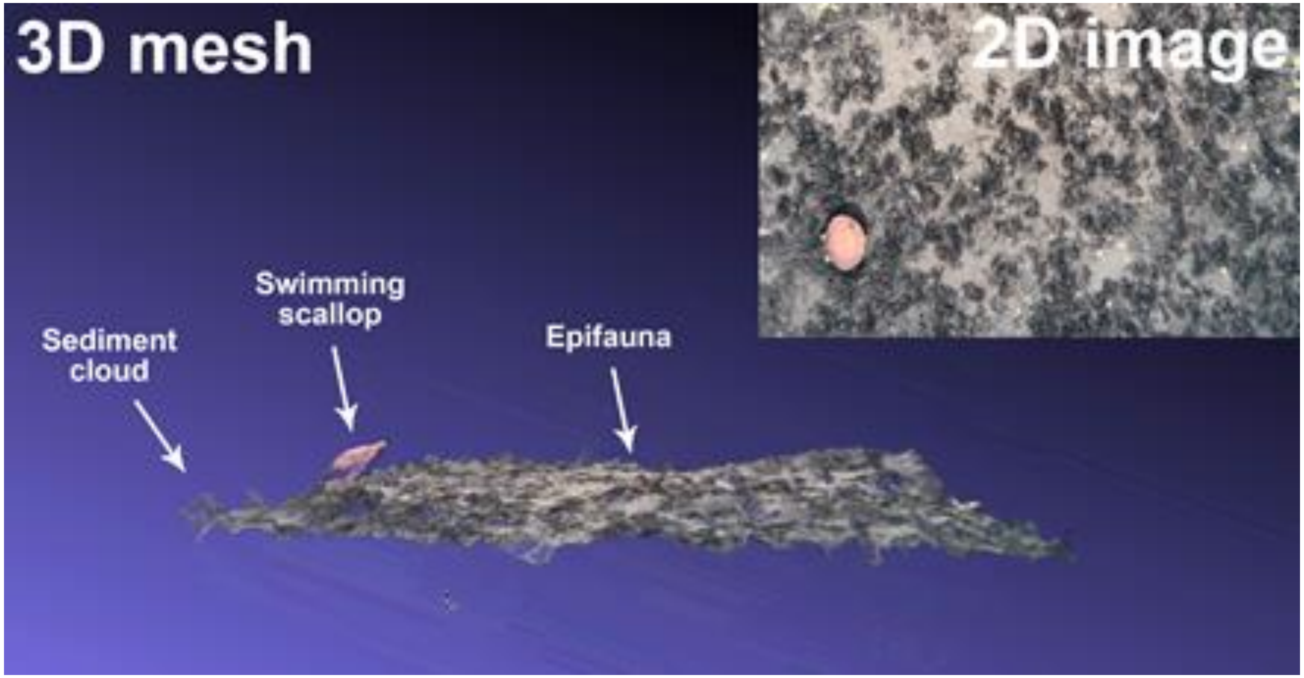
Groundtruth = human annotations

Predicted

Conclusions reported previously

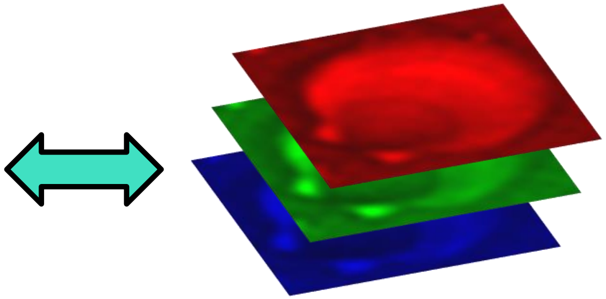
1) Scallop detection works reasonably well across all detectors, but subcategory classification (live, dead, swimming, and clapper) still needs improvement.

2) Stereo has improved detection of swimming scallops regardless of model and model version, but it has not improved detection of live scallops, clappers, or flounder.



Some computer vision basics

RGB



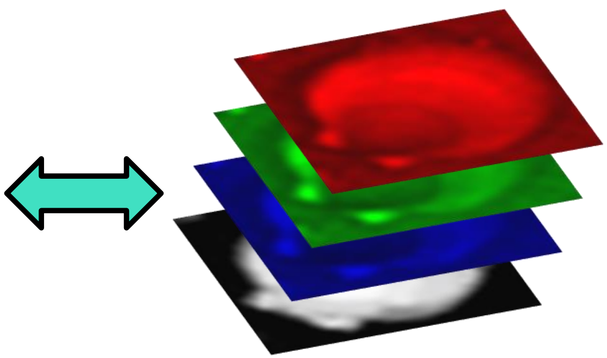
Colors as triplets

202, 162, 154 (R, G, B)

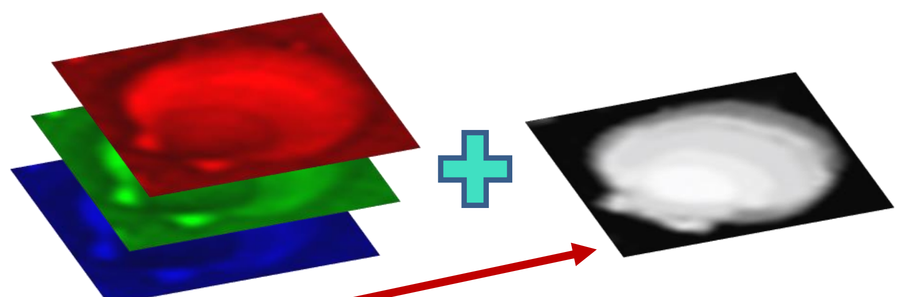
Depth maps

Depth disparities => distance from lenses

RGBD



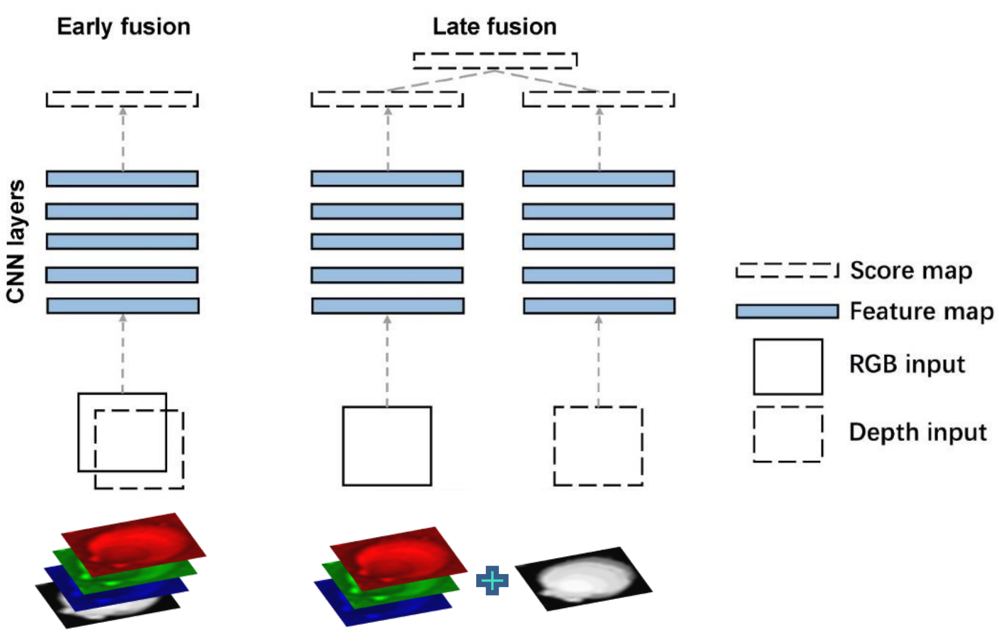
OR

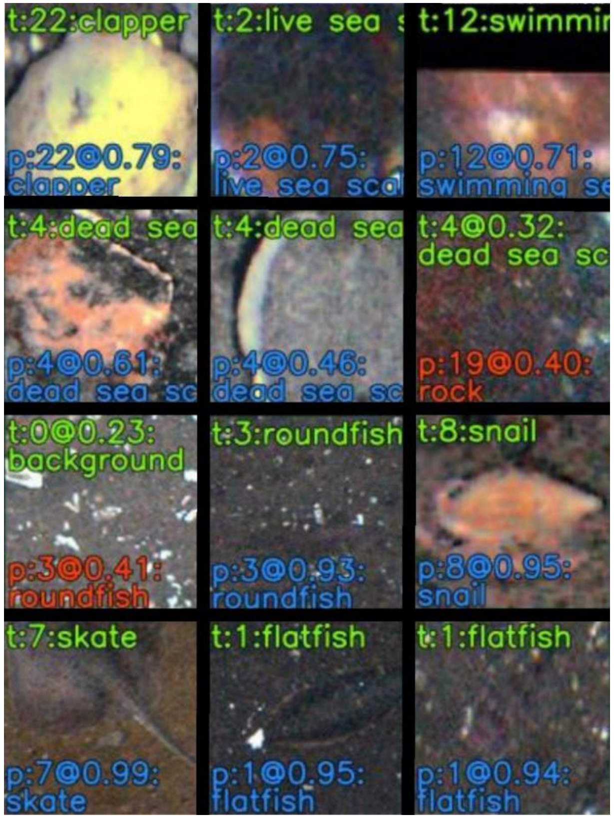


Depth map

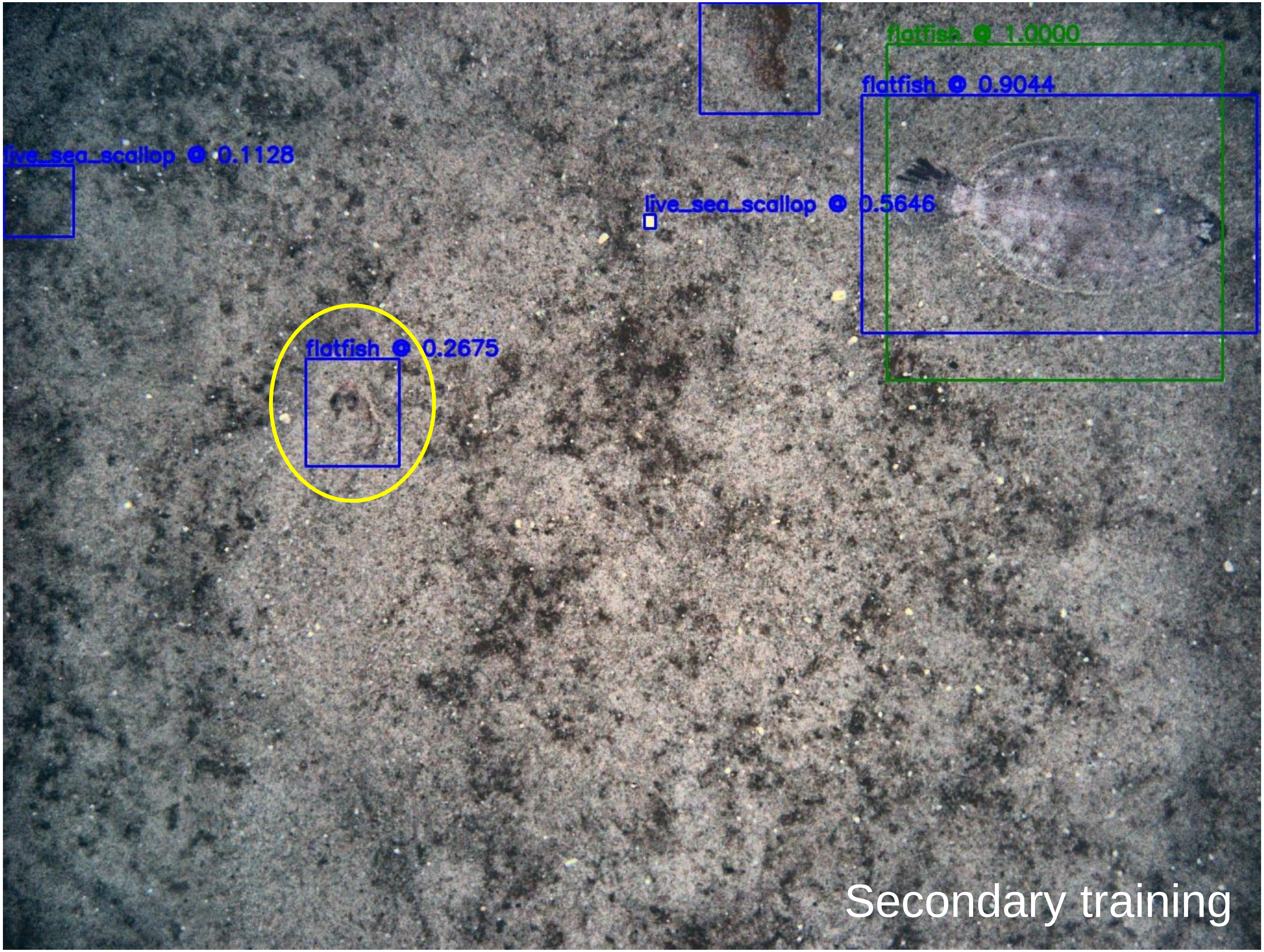
New models

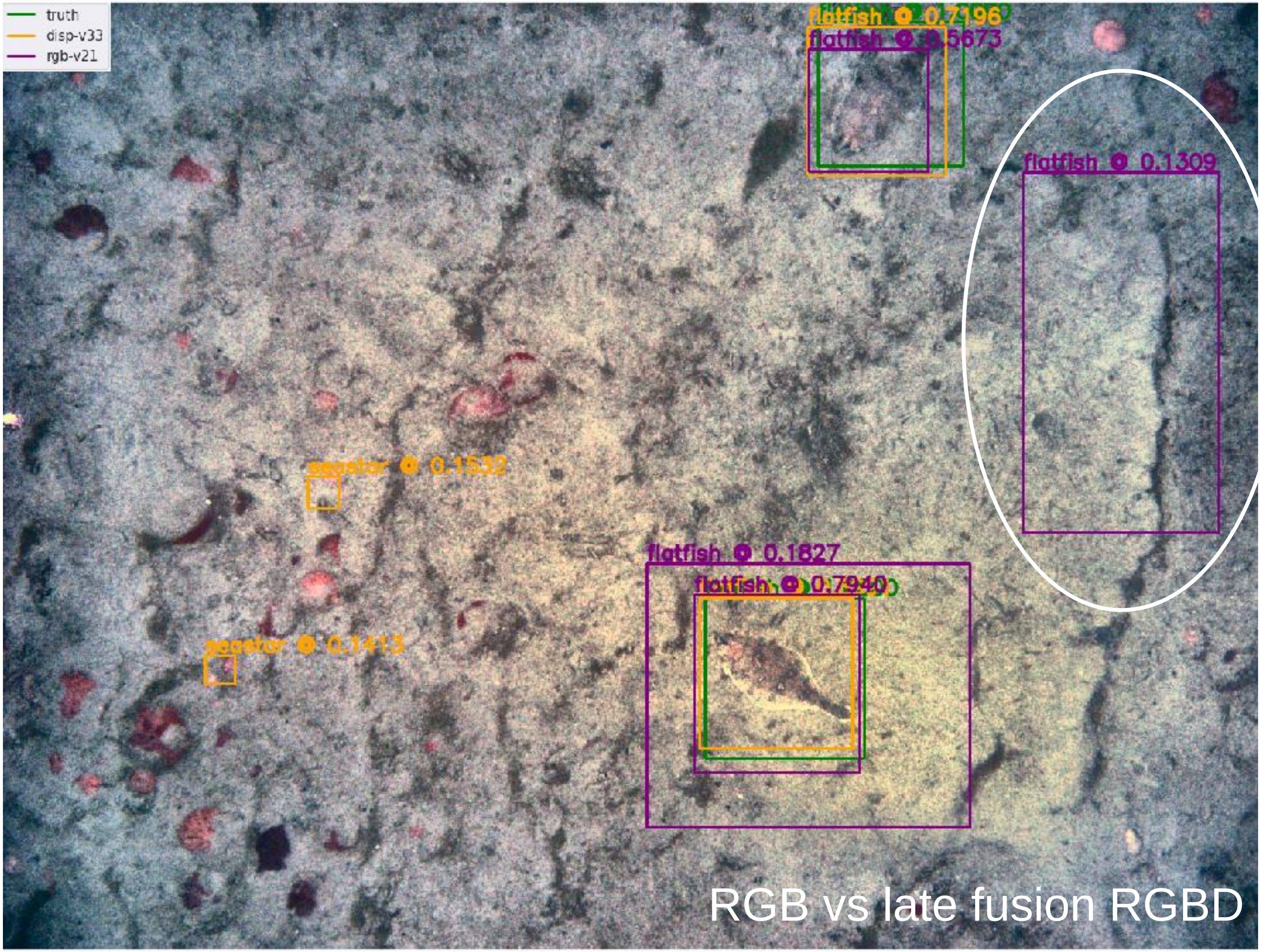
1) Late fusion of stereo-derived depth maps
2) Fine-tuning with secondary chip classifiers
3) Secondary training of multi-class models to a single classe





Adapted from Li Y, Zhang J, Cheng Y, Huang K, Tan T. 2017. Semantics-guided multi-level RGB-D feature fusion for indoor semantic segmentation. 2017 IEEE International Conference on Image Processing: 1262-1266.


flatfish @ 1.0000
flatfish @ 0.9044
live_sea_scallop @ 0.1128
live_sea_scallop @ 0.5646
flatfish @ 0.2675
Secondary training


truth
disp-v33
rgb-v21
flatfish @ 0.7196
flatfish @ 0.5673
flatfish @ 0.1309
seastar @ 0.1532
flatfish @ 0.1827
flatfish @ 0.7940
seastar @ 0.1413
RGB vs late fusion RGBD

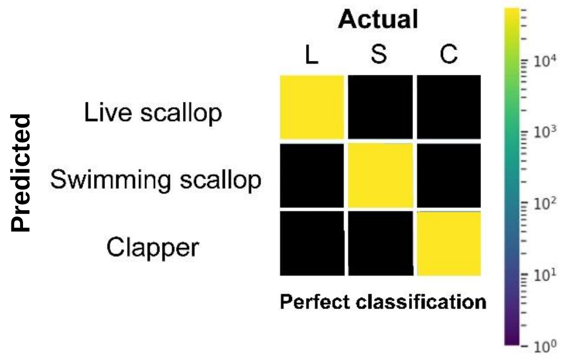
Summary of average precision results

Indicated model type vs models using just RGB images with no fine tuning

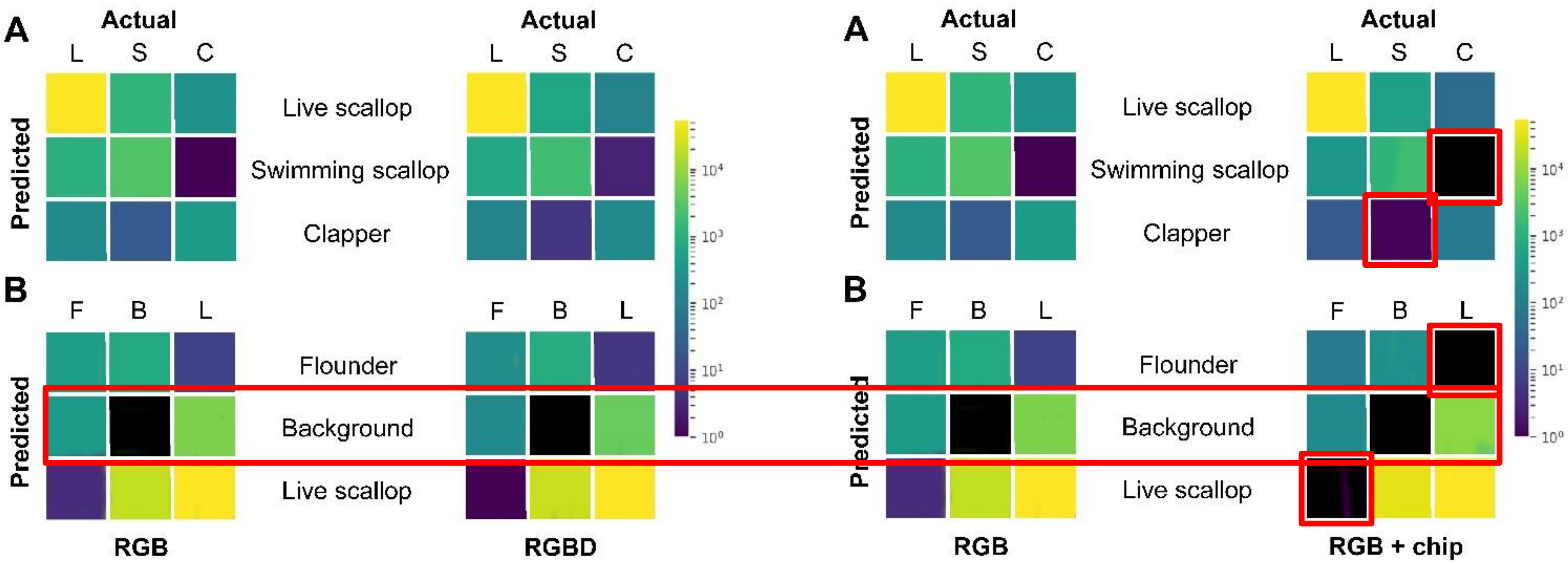
Model type	Live scallops	Swimming scallops	Clappers	Flounder
Early fusion of depth maps	-0.004	**0.04**	-0.02	-0.04
Late fusion of depth maps	**0.04**	**0.05**	NA	**0.03**
Fine tuning with secondary chips	-0.05	**0.10**	**0.20**	**0.04**
Secondary training multi to single class	-0.002	NA	NA	**0.09**

- Addition of stereo-derived depth maps did not consistently improve detection or classification of scallops classes or flounder
- Models with late fusion of depth maps performed better
- Fine tuning detectors based on RGB images alone may a better way to improve automated detection of these target organisms

Confusion matrices



Perfect classifier – predicted classifications are identical to actual classifications



Early fusion

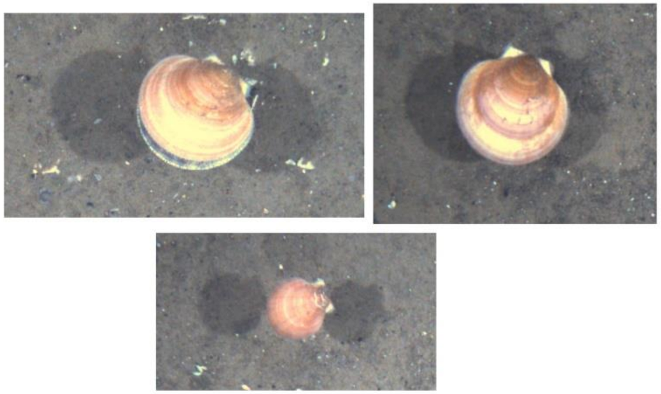
Fine-tuning with chip

Project objective conclusions

1) Scallop subcategory classification (live, dead, swimming, and clapper) still needs improvement and inclusion of stereo depth maps did not result in much more accurate classifications as we expected.

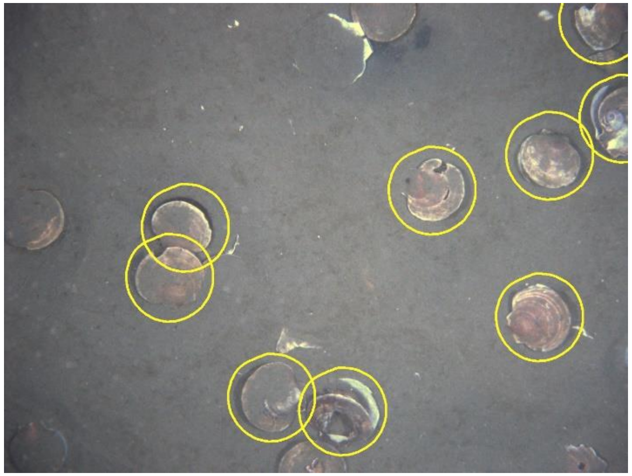
Why only a slight improvement in swimming scallop vs live on-bottom scallop detections?

Automated detectors may effectively utilize the same visual cues that human annotators use to identify swimming scallops vs live scallops on the bottom in single images e.g. shadows.



Why no improvement in clapper vs live scallop detections with depth information?

Although clappers have wider gape distances than live scallops, other characteristics like orientation in the sediment may vary and the gape may not be visible. This issue is compounded by small sample sizes for clapper and swimming scallop annotations relative to live scallops.

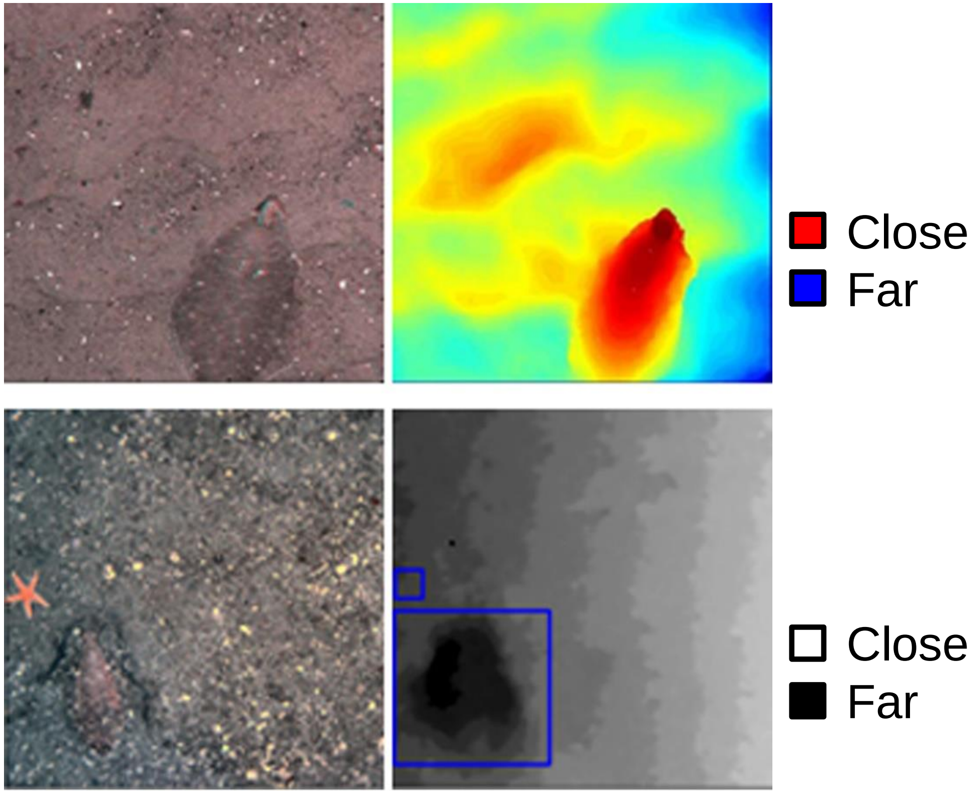


Project objective conclusions

1) Including depth maps from stereo images did not improve detection of flounder as much as expected.

Best guess for why

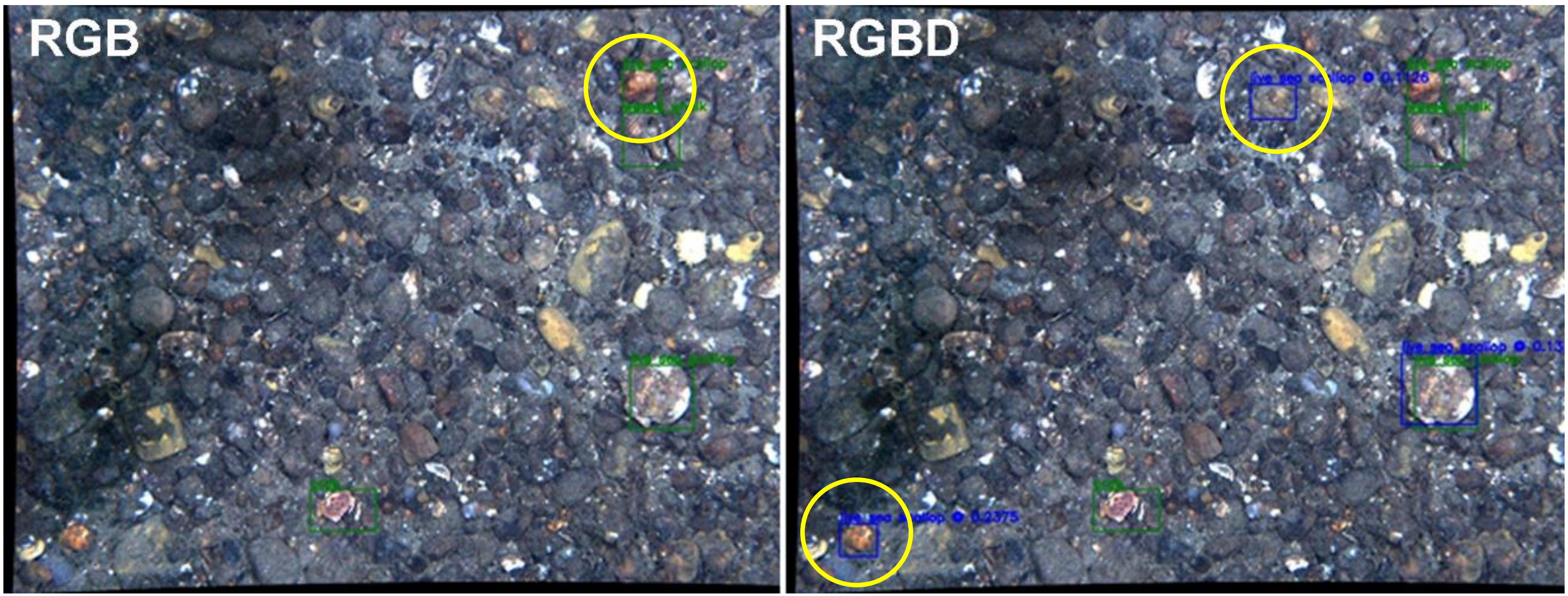
The location of flounder relative to the sediment surface is not consistent, so inclusion of depth maps does not always improve flounder detection models. This issue is compounded by small sample sizes for flounder annotations relative to live scallops.



Other project conclusions

1) Background regions were often mistaken for both scallops and flounder.

This could be partially due to human annotators missing scallops and flounder or mistakenly labelling rocks or other bivalves that aren't scored as scallops.

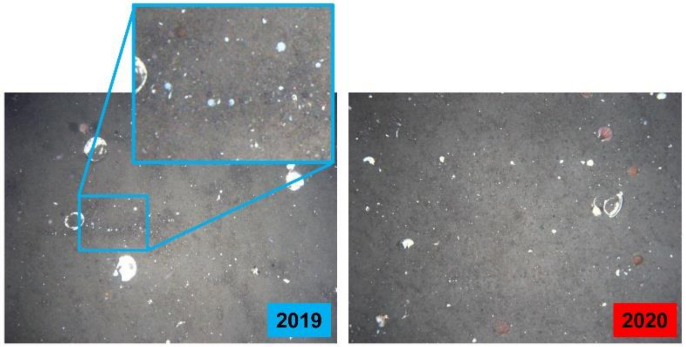


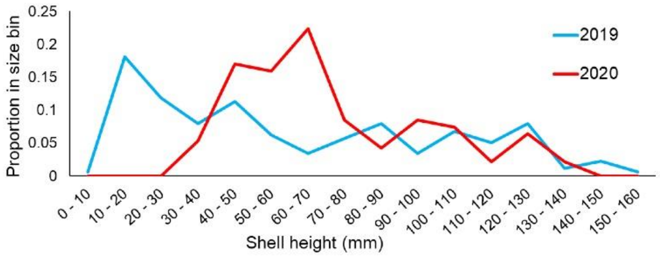
Next steps

2021-2022 Scallop RSA grant to develop VIAME new algorithms to

1) incorporate automated measurement of target organisms, including scallop shell height measurements (altitude off bottom plus key points on shell)
2) improve automated classification of benthic habitats in HabCam imagery
   - ***Reduce the number of mistaken classifications of background regions as target organisms***
   - ***Lead to better use of image databases for delving into questions about ecology and habitat changes***

Project will include examining habitat preferences of pre-recruit scallops and predictors of survivor





Scallops on the southern flank of Georges Bank near the boundary of Closed Area II

Extra slides

Image augmentation