MBARI Monterey Bay Aquarium Research Institute

From Detection to Direction: Guiding Adaptive Sampling with Onboard Visual Inference in an LRAUV

Korneel Somers, Delft University of Technology

Mentors: Paul Roberts, Kakani Katija, Joost Daniels

Summer 2025

Keywords: adaptive control, AUV, real-time, object detection

ABSTRACT

Exploring midwater ecosystems with autonomous underwater vehicles is constrained by the need to preprogram the vehicle's trajectory and having limited control once the mission has started. Because ocean conditions are constantly changing, the vehicle's control system must be able to adapt to its surroundings in real time. This project aimed to use wide-view imaging on MBARI's LRAUV to inform adaptive targeted sampling of *Nanomia septata*. A YOLOv11-based detection model ran onboard a Jetson Orin, sending detection rates to the vehicle controller. There, the rates were used to stay within a high-density zone by detecting peaks and adapting the YoYo depth limits accordingly. Three field trials in Monterey Bay were conducted to test system performance and validate the adaptive behavior. There was a clear correlation between depth and detection rate, and the system successfully adjusted dive limits in response to detections of *Nanomia septata* on several occasions, which confirmed proof of concept. On the other hand, the sparse distributions also sometimes caused minor fluctuations to be interpreted as peaks, triggering frequent unnecessary direction changes. This highlights the need for further finetuning of the signal processing to improve robustness. Overall, this work demonstrated a first step toward autonomous, visually driven sampling strategies for studying sparse midwater organisms.

INTRODUCTION

The deep sea is a challenging operating environment that is still largely unexplored. MBARI has been developing a number of vehicles that allow us to venture into the deep. The remotely operated vehicles (ROV) enable a variety of benthic and midwater missions, but their need for a skilled pilot and a ship limit their scalability. The long-range autonomous underwater vehicles (LRAUV) offer a solution to tackle this challenge of covering a large space and time period, as they can go out for several weeks at a time. Their downside is that they lack the maneuverability of ROVs and that we can only communicate with them during the planned surface pop-ups unless they go out in combination with, for example, a wave glider. That means the missions are preprogrammed using specific waypoints, but since the ocean ecosystems are dynamic and unpredictable, you never know where and when you will find certain things. This calls for the need for real-time adaptive control to ensure the vehicle finds and stays within regions of interest.

In the past, MBARI has conducted several experiments with autonomous adaptive vehicle control. In studies of harmful algae blooms (HABs) [1], LRAUVs equipped with real-time chlorophyll sensors were used to detect and track bloom structures both vertically and horizontally. The vehicle executed a combined behavior: vertically it performed YoYo profiles to locate the chlorophyll peak, and horizontally it searched for and remained within the patch center by following chlorophyll gradients. This allowed the LRAUV to autonomously maintain position in the bloom without requiring constant operator input. Zhang et al. [2] further developed an adaptive YoYo algorithm developed for LRAUVs equipped with the "Planktivore" in situ microscopy system. Here, real-time counts of plankton detections from the microscope's low-magnification camera were used to dynamically set the yo-yo depth limits based on peak and edge detection. This approach concentrated sampling effort within layers of high particle concentration, increasing the efficiency of data collection.

The Bio-inspiration Lab has also experimented with using wide-view lenses to run detection algorithms that enable autonomous detection, classification, and tracking of specific animals using the miniROV [3]. Building on this work, we propose to integrate wide-view imaging into the LRAUV platform, combining it with the adaptive YoYo approach to detect and follow sparse, vertically migrating organisms. In this project, Nanomia will serve as a case study. These siphonophores are abundant in Monterey Bay and are key predators in the local food web. Some open questions regarding their behaviour, like the details of their diurnal migration, could

be addressed through adaptive sampling. However, the primary objective of this work is to demonstrate the feasibility of using wide-view lenses to dynamically adapt LRAUV behaviour for targeted sampling of specific animals.

MATERIALS AND METHODS

Case Study Species: *Nanomia septata*

Nanomia septata is a physonect siphonophore in the family Agalmatidae. It is a colonial hydrozoan composed of specialized zooids for feeding, reproduction, and movement. They can be up to 30 centimeter in length and occur widely in Monterey Bay between 0 and 700 meter depth. Nanomia are suspected to consume more krill than the local whale population, making them an important component of the midwater food web [4], which means sudden changes in their abundance may influence ecosystem structure. While the hypothesis is that they undertake diurnal vertical migration, direct evidence is limited. Since tagging these animals is not feasible, performing adaptive sampling with the LRAUV could help solidify this claim.

The Nanomia are well-suited for demonstrating adaptive target sampling using wide-angle lenses. They are much sparser than the plankton species previously analyzed with an in situ microscopy system for adaptive YoYo operations [2], but still sufficiently abundant. Visiting researcher Marc Allentoft-Larsen studied the *Nanomia septata* in spring 2025 by conducting box searches between 150 and 380 meters at the M1 location in Monterey Bay using an LRAUV equipped with the Optim camera payload. Marc recorded up to 280 observations within a single hour (at a cruising speed of 0.5 m/s) and up to 8 individuals in a single image. In addition, he manually labeled Nanomia instances in many of these images, thereby expanding MBARI's existing dataset of the species that can be used for training detection algorithms. Figure 1 presents two example images.

Hardware

Long-Range Autonomous Underwater Vehicle (LRAUV): The LRAUV was designed by MBARI engineers in 2012. The goal was to bridge the endurance gap between buoyancy-driven gliders and regular AUVs [5]. A summary of its specifications can be found in Table 1. The vehicle is designed to accommodate different payloads in its nose, which will be the Triton Imaging System in our case.

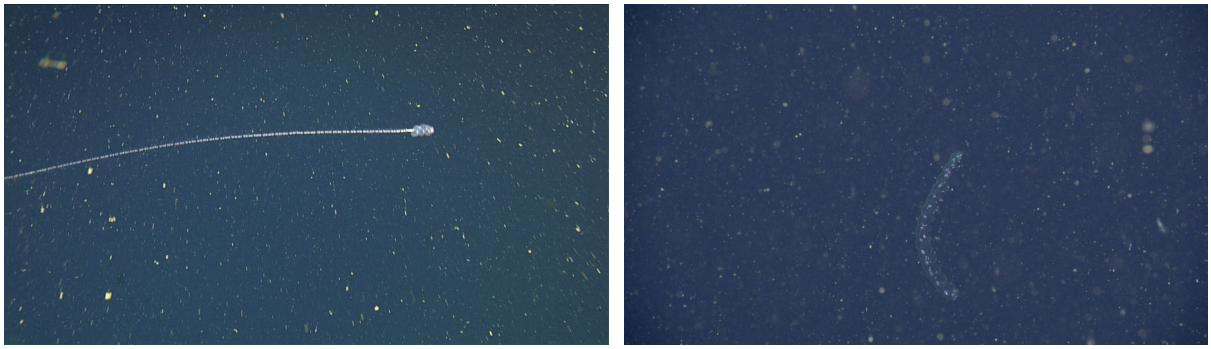

Figure 1: *Example images of Nanomia septata captured during pre-programmed box searches at M1 in Monterey Bay.*

Table 1: *LRAUV specifications.*

Parameter	Specification
Length / Diameter	3.2 m / 0.3 m
Speed	0.5–1.2 m/s
Range (primary battery)	∼1800 km over 3 weeks at 1 m/s
Range (rechargeable)	∼800 km over 12 days at 0.8 m/s
Control	State-configured layered control
Navigation	DVL-aided dead reckoning (+ GPS fixes via USBL/LBL)
Actuators	Elevators, moving internal mass, variable buoyancy

Triton Imaging Payload: The Triton payload was developed by the Bioinspiration Lab at MBARI. It has three forward looking, wide-lens cameras and includes a dedicated onboard processing unit that operates independently from the vehicle's main computer. This separation is important in case of software or hardware errors, because it makes sure that any issues with the payload do not interfere with the vehicle's core control or navigation systems. Table 2 gives a detailed overview of the payload's specifications and Figure 2 shows the entire hardware setup.

Table 2: *Triton Imaging Payload specifications.*

Parameter	Specification
Cameras	3 × Sony IMX546 sensors
Interface	CSI-2 (MIPI)
Resolution (used)	2160 × 1440 px
Shutter type	Global
Frame rate (operational)	5–10 fps
Horizontal Field of Viewc	ca. 90°
Processing unit	NVIDIA Jetson AGX Orin

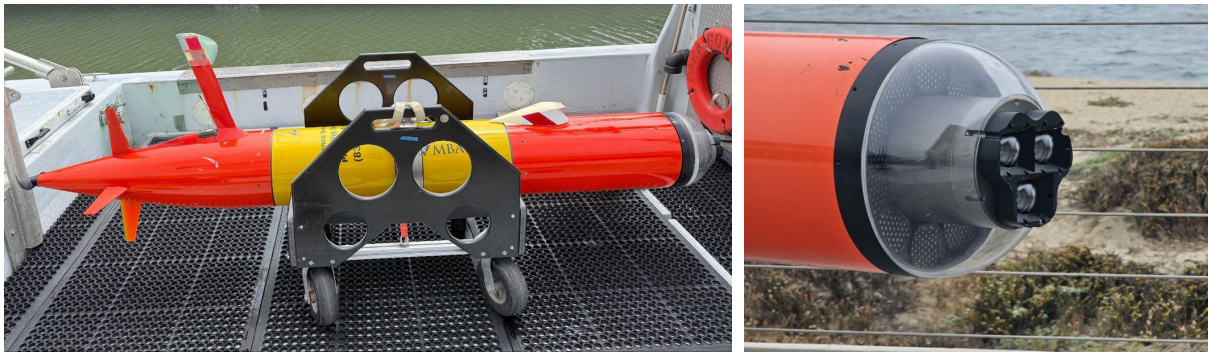

Figure 2: *Left: LRAUV equipped with the Triton Imaging Payload. Right: Close-up view of the Triton payload (by Paul Roberts).*

Software Architecture

The software architecture consists of two main flows. The vision pipeline runs on the payload's processing unit and is responsible for performing inference on incoming images. It uses Lightweight Communications and Marshalling (LCM) channels to communicate the detection rates of the targeted organism to the vehicle computer. The vehicle computer then smooths this spiky signal and uses it for adaptive control of the vehicle.

Vision Pipeline:

1. **Image acquisition:** Three synchronized camera streams operating at 10 fps.
2. **Preprocessing:** Gain adjustment to normalize image brightness.
3. **Inference:** Object detection with YOLOv11 model trained on siphonophore imagery.
4. **Output:** Per-class detection counts, inference speed, and timestamps are published to an LCM channel, which the LRAUV backseat controller subscribes to.

Vehicle Control Logic: The incoming detection signal is smoothed before being used for control. Several smoothing methods are evaluated. Zhang et al. [2] previously applied a 5-point median filter (for outlier rejection) followed by an 8-s moving-average window (for low-pass filtering). Here, alternative approaches such as the Exponential Moving Average (EMA) and Kernel Density Estimation (KDE) are also tested. The EMA provides a weighted average that emphasizes recent detections, while KDE estimates the underlying probability distribution of detections. After smoothing, the signal can be used for vehicle control. Three control logics are proposed and visualized in Figure 3:

- **Fixed YoYo:** The baseline method, with preprogrammed upper and lower depth bounds. The vehicle follows a predetermined trajectory without adaptation to sensor inputs.
- **Adaptive YoYo:** The algorithm by Yanwu [2] is implemented here to work with the irregular and sparser input signal. The depth bounds are continuously adjusted based on detection peaks.
- **Informed Adaptive YoYo:** A proposed hybrid method combining exploratory and adaptive behaviors. An initial exploration phase builds a broad environmental profile, and the vehicle then switches to adaptive sampling in target-rich layers. As confidence in the environmental model decays over time, the system re-enters exploratory mode to refresh its understanding of the environment. This can, for example, help identify multiple layers of high organism presence and provide more ecological context for later analysis.

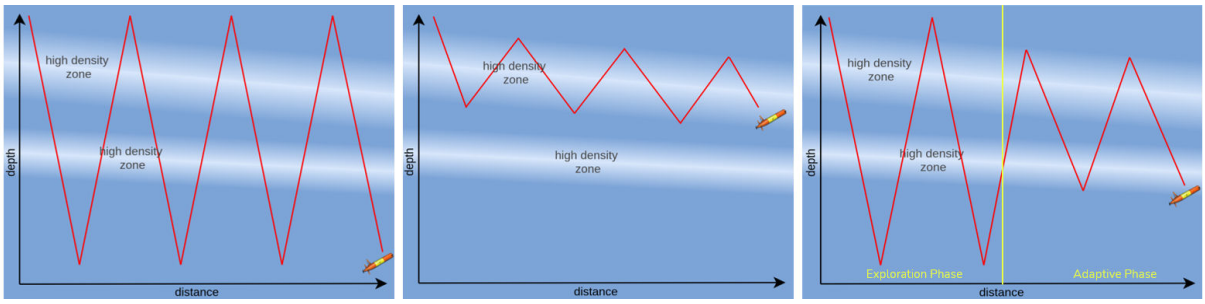

Figure 3: *Three control logics, with the white bands representing the zones of interest and the red line the LRAUV path (Left: Fixed YoYo - Middle: Adaptive YoYo - Right: Informed Adaptive YoYo)*

Simulation

Before deployment, a set of simulations was created to evaluate the three control logics. The simulations were run in a simplified 2D world where each spawned dot represents an organism. To loosely simulate the dynamics of the ocean, all organisms were assigned varying speeds and directions based on a Perlin force field (a technique commonly used in gaming to simulate natural flows) capped at a maximum speed. In this world, an LRAUV could be spawned with a cone-shaped field of view that followed one of the control logics, documenting how many organisms appeared within the field and reacting accordingly. To test performance under different conditions, three distinct Nanomia distributions were used: homogeneous, single Gaussian layer, and double Gaussian layer (see Figure 4).

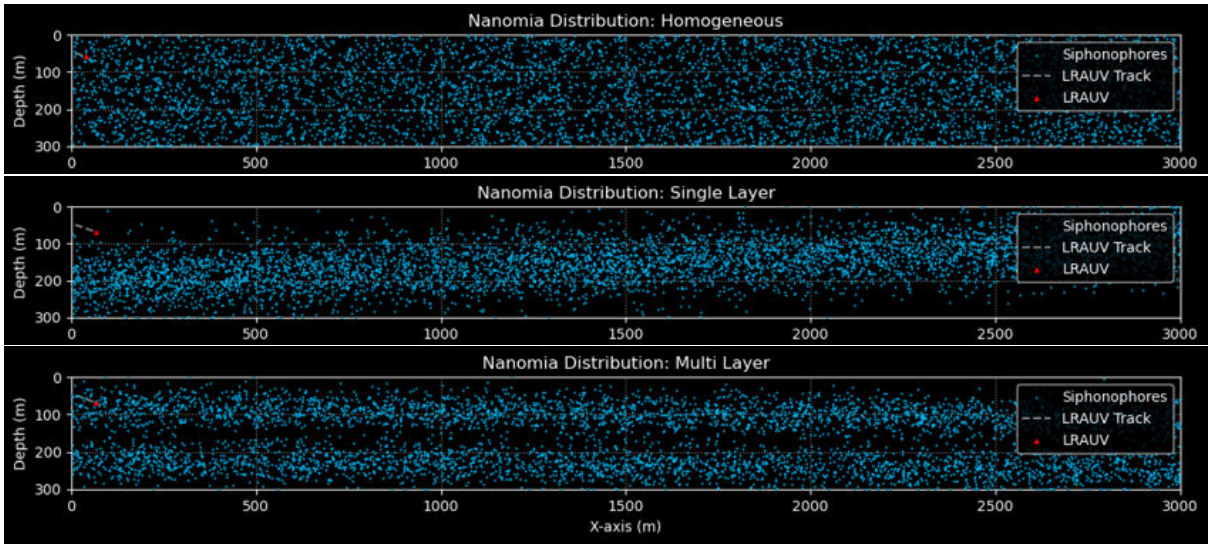

Figure 4: *Three simulation scenarios with different Nanomia distributions*

The initial comparison metric was the number of detections per unit of time, but this turned out to be too much of a simplification and can also be statistically estimated without simulation, as discussed in the results section. Instead, the goal here was to identify relevant parameters that need fine-tuning for the adaptive algorithm to function effectively at sea.

Deployments

Three at-sea deployments that were conducted in Monterey Bay between June and July 2025 are discussed in this report. The goal was to test and refine the adaptive sampling algorithm progressively. Each deployment was built upon the outcomes of the previous one, starting with basic system validation and leading to real-time adaptive control of the vehicle. Imagery and vehicle data were archived in MBARI's Video and Annotation Reference System (VARS) and Thalassa servers. The data from these dives is later used to analyze the effect of signal smoothing methods and the effectiveness of the adaptive sampling.

For all three deployments, the vehicle was deployed from the R/V Paragon in the Monterey Bay and operated near the C1 field site along the upper Monterey Canyon. Figure 5 shows the transects of the vehicle for each deployment. Table 3 summarizes their specifications.

RESULTS & DISCUSSION

The results in this section were generated with the triton-detection repository that can be found in the MBARI GitHub.

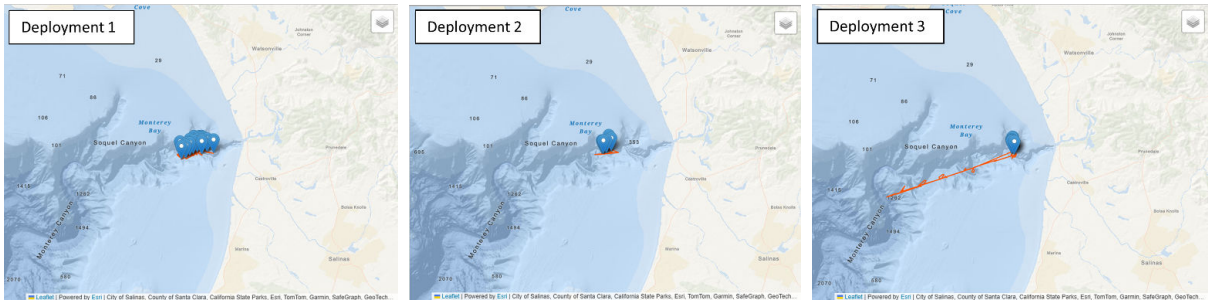

Figure 5: *Map of LRAUV Triton deployment sites in Monterey Bay during summer 2025. All three missions were conducted near the C1 field site along the upper Monterey Canyon and launched from the R/V Paragon.*

Table 3: *Summary of LRAUV Triton deployments in Monterey Bay during summer 2025.*

#	Date	Duration (h)	Depth Range (m)	Objective
1	Jun 16–19, 2025	48	0–70	Validate Triton system setup
2	Jul 7–10, 2025	69	0–159	Validate setup; perform regular YoYo dives
3	Jul 16–18, 2025	64	0–250	Perform adaptive YoYo sampling

System Preparation

Before deployment, the detection model must be fine-tuned on the target class, which is *Nanomia septata* in our case. Marc Allentoft-Larsen and Kevin Bernard fine-tuned the YOLOv11 detection model from Ultralytics with labeled Nanomia imagery from various ROV and LRAUV missions. This was done for all model sizes, allowing them to be compared in terms of performance and inference speed, which is the trade-off that must be made when deploying detection models in real-time.

For performance evaluation, we used the F1 score of the best-trained weights. The F1 score balances precision and recall, providing a meaningful indication of overall detection reliability. The results are summarized in Table 4. As can be seen, larger models perform slightly better, though the differences are small. It is important to note that the training dataset was collected over several missions and with different cameras, but it does not yet include data from the Triton cameras. For future work, a labelled test set from Triton imagery would be valuable to assess the detection performance of this payload more accurately.

While performance should be maximized, inference speed has a strict lower bound, as we want each image to be processed in real time. Because the vehicle moves relatively fast, individual *Nanomia* colonies may only appear in one image. That is especially true for those that are observed at edge regions of the frame. The camera operates at 10 fps, so the inference

Table 4: *Performance comparison of different YOLO models trained for Nanomia septata detection. The nano model is trained a second time with a high number of epochs.*

Model	F1 Score	Confidence Threshold
sipho_v4_yolo11n	0.70	0.318
sipho_v4_yolo11n_ep213	0.72	0.294
sipho_v4_yolo11s	0.73	0.364
sipho_v4_yolo11m	0.74	0.355
sipho_v4_yolo11l	0.74	0.364

rate must meet or exceed this to avoid frame loss.

Inference benchmarking was performed on a Jetson AGX Orin that was set up to be identical to the onboard system in the Triton imagery payload (see the triton-detection repository for setup instructions). Both the regular YOLO weights and the TensorRT-optimized engine versions of the models were tested. The resulting trade-off between inference speed and model accuracy is shown in Figure 6, with a dotted red line to show the required minimum inference speed. This shows that only the nano and small models are candidates for deployment.

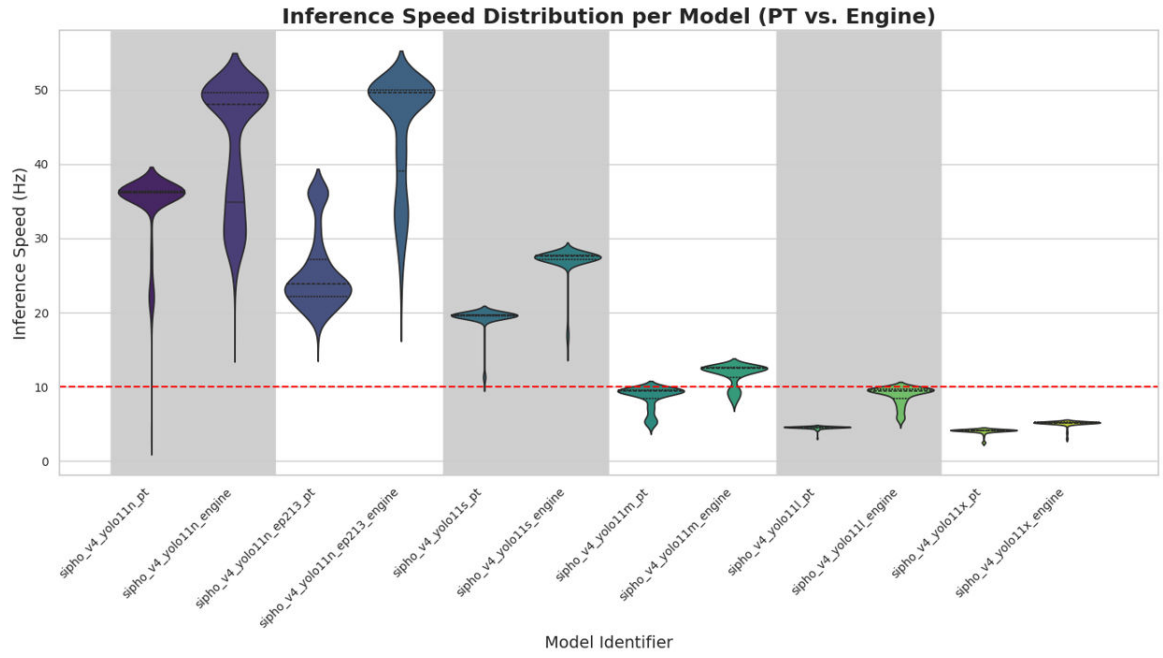

Figure 6: *Inference speed vs. model performance on the Jetson AGX Orin shown as a violin plot. Dotted red line shows the 10 Hz threshold which is the speed at which the detection model should operate.*

Simulation

The simulations assumed Gaussian depth distributions and a functioning adaptive algorithm. While the number of detections per unit of time can indicate whether the vehicle spends more time in hotspots, this metric alone adds little beyond confirming expected behavior. In the homogeneous case, all control logics resulted in similar detection rates, as any path statistically encounters the same number of targets. With single or double layers, the adaptive and informed-adaptive YoYos get more because they focus on the denser regions. Of course, since some time is spent exploring, the informed-adaptive has fewer detections, but that trade-off between quantity and adding a broader context shows that detections per unit time alone are insufficient for evaluation.

The simulations did help to explore the main parameters to tune, which are those that determine when the vehicle thinks it has passed a detection peak and should reverse direction. This is controlled by a percentage drop threshold, which is basically how much the detection rate must fall before the system considers it has left a dense region. Because detections can fluctuate a lot and often drop to zero when organisms are sparse, the signal must be smoothed appropriately. Depending on the chosen smoothing technique different parameters have to be finetuned, for example for KDE the bandwith is most critical. In the end, the gap between simulation and reality was too large to carry the exact numbers for these thresholds reliably into deployment, so finetuning the parameters at sea will be an important part of future work.

Deployments

Deployment 1

Following test-tank validation, the Triton payload was deployed at sea for the first time. This initial run was primarily done to validate the payload in the field and collect test footage for inspection. The captured images were generally underexposed, needing gain adjustments after the mission. This improved brightness, but it also compressed the dynamic range and resulted in highlight clipping as shown in Figure 7. Finetuning was done by Paul Roberts, which included fixing focusing issues and fully opening the sensor aperture to increase baseline exposure and brightness. These adjustments resulted in higher-quality imagery in the following deployment.

Figure 7: *Frames taken by the Triton payload during Deployment 1 with post-processed increased gain*

Deployment 2

The second mission aimed to perform standard (non-adaptive) YoYo profiles. Figure 9 shows snapshots of the deployment from the *R/V Paragon*. The vehicle collected clear imagery of Nanomia and several other species with an improved image quality compared to the first deployment. Some examples are shown in Figure 9. One remaining issue is the lack of illumination around the edges of the frame.

Figure 8: *Deployment 2 from R/V Paragon*

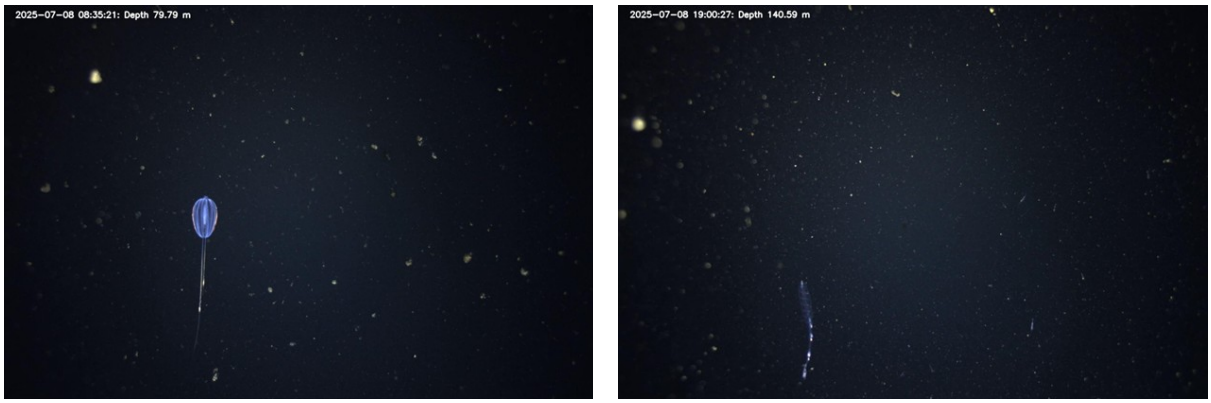

Figure 9: *Frames taken by the Triton payload during Deployment 2*

Even though the vehicle control was not yet determined by the detection rate, we ran the detection algorithm to evaluate the output of the vision pipeline. Figure 10 shows both the depth (in red) and the Nanomia detection rate (in blue) over time. In the top graph, the detection rate is shown as a binned signal, where each data point represents the number of detections per second calculated over the last ten seconds. Even here, it becomes apparent that the detection rate increases with depth. This trend is even clearer in the bottom graph, where the raw, spiky signal, that often drops to zero when no detections occur, is smoothed using kernel density estimation (KDE). As an alternative, the exponential moving average (EMA) was also tested, but that introduced a slight phase shift by delaying the peaks, which is less favorable when aiming to react as instantaneously as possible to the incoming signal.

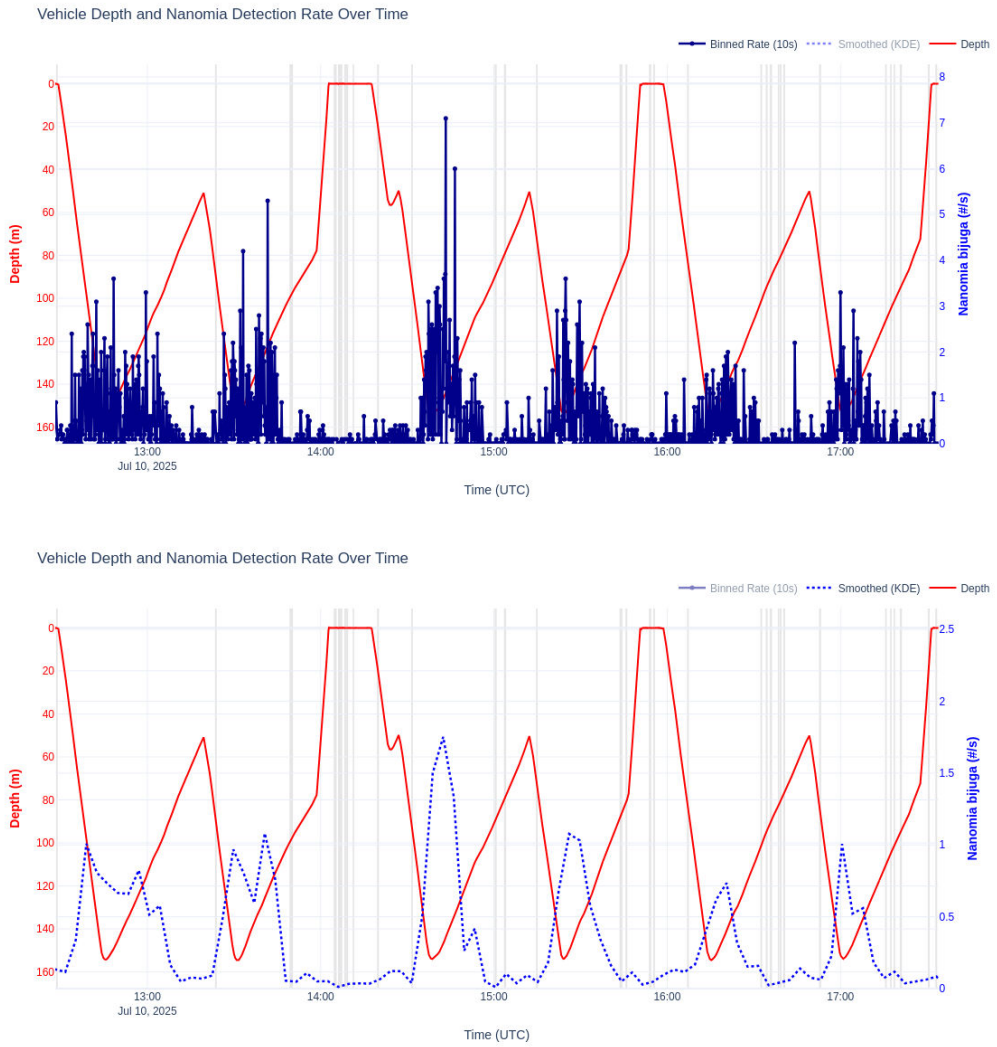

Figure 10: *Top: Raw binned detection rate over time and depth. Bottom: Smoothed detection rate using KDE over time and depth*

The detection–depth relationship indicates the zones where most Nanomia are present at a given time, which can help with figuring out the details of their diurnal migration. However, since the YoYo profiles were preprogrammed, the vehicle could not adjust its trajectory in real time. This highlights the need for adaptive YoYo's in future deployments, allowing the vehicle to autonomously search for and remain in regions that maximize detections, rather than being constrained by fixed mission boundaries or requiring human intervention.

Deployment 3

The third deployment introduced adaptive targeted sampling, where the YoYo depth limits were dynamically adjusted based on the observed detection density. Again, the vehicle was deployed from the R/V Paragon (see Figure 11), but this time the vision pipeline output was connected to the control logic of the vehicle. Figure 12 shows some example detections. For operational safety, the maximum depth was set to 250 m. That is needed for the case where no detection peak is found and ensures that the vehicle goes beyond 150 m, but remains well within its maximum operational depth.

Figure 11: *Deployment 3 from R/V Paragon*

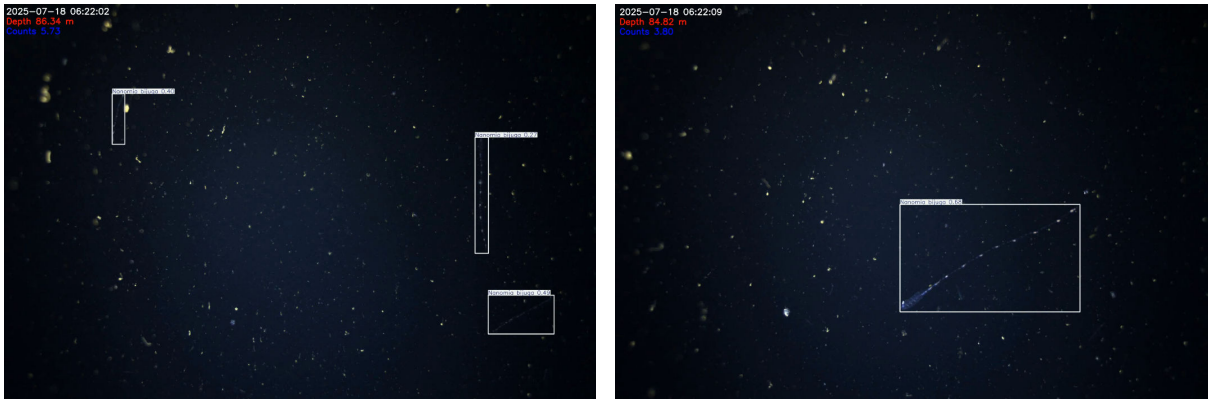

Figure 12: *Example frames from Deployment 3 showing in-situ detections of Nanomia*

Several examples demonstrate that the system successfully adjusted its YoYo depth boundaries according to the real-time detection rate, demonstrating proof of concept for adaptive control. Figure 13 shows a graph where the YoYo direction reverses when there is a clear drop in detection rate. However, on various occasions the system did not behave as expected, stressing the need for further parameter tuning.

For example, in the upper layer (approximately the first 50 meters) the vehicle appears to get stuck, likely because of the large amount of particles that interfere with the detection signal, causing the control logic to keep changing direction. Overall, the system seemed too sensitive. It often interpreted minor fluctuations as peaks, which triggered frequent direction changes. To mitigate these issues, it is important to set the upper boundary for the adaptive YoYo at about 50 meters and to reduce the aggressiveness of the peak detection by incorporating more detection history during smoothing.

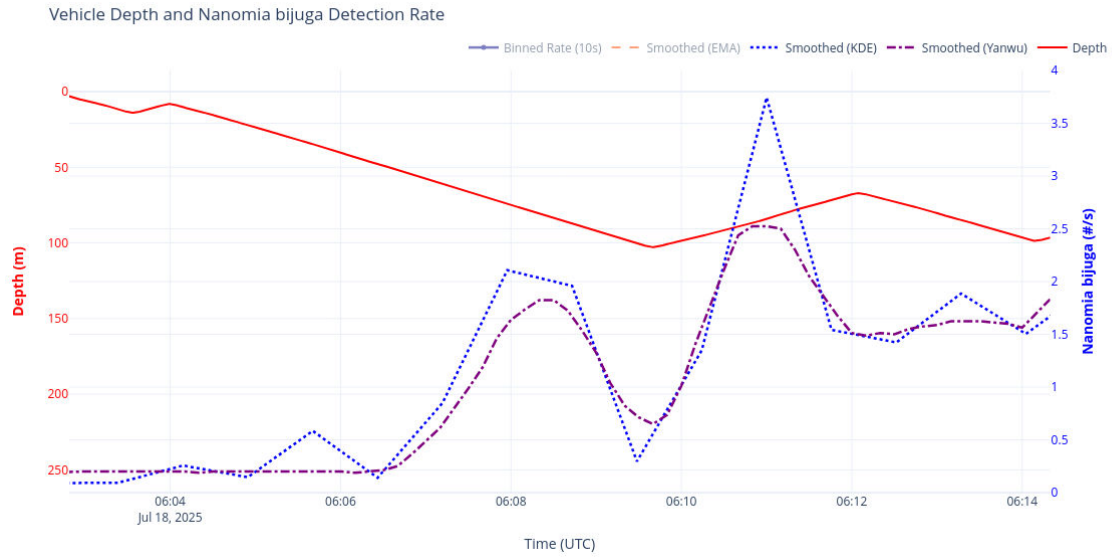

Figure 13: *Example of adaptive YoYo depth adjustments during Deployment 3*

Limitations

The field deployments demonstrated the feasibility of onboard adaptive control driven by in-situ visual detections. However, several limitations and challenges open the way for further research to make this approach truly robust. These can be grouped into detection and control limitations.

Detection Limitations: Marine snow and other particles in the water affected the reliability of the detection signal in both directions. In some cases, dense particle layers caused false

positives and produced an untrustworthy signal. In other cases, the particles mask the tail of Nanomia, which is mostly transparent but has colored dots. When visually inspecting the detection results, you see that as the tail starts to blend into the background, detection confidence drops noticeably. Solutions could be to add more occluded and lower-contrast examples to the training data. Overall, more analysis is needed to evaluate the detection performance and the effect of double-counting of individuals as they move through the images. Labeling some of the Triton imagery would allow us to use it both during training to improve the model and during testing to get accurate performance results.

Control Limitations: Even though some of the adaptive sampling was successful, it will require post-mission signal analysis and some trial and error to identify the edge cases where the system does not behave as expected, and make it more robust. The informed-adaptive control still needs to be tested at sea, with the goal not only to determine how many layers of animals are present but also to integrate multi-modal priors such as chlorophyll or temperature, depending on the mission. The informed logic could help address another limitation, which is the need to manually fine-tune the adaptive algorithm. Since it is difficult to model the distribution accurately beforehand, you have to go out and re-adjust the parameters for every new species. This is demanding if you want to scale this adaptive control to many use cases. It would therefore be beneficial to automate this process by using the context and sparsity detected during the exploration phase of the informed-adaptive logic to set the parameters for adaptive control.

CONCLUSIONS & RECOMMENDATIONS

We demonstrated the integration of the Triton wide-view imaging payload with adaptive control on the LRAUV platform, building on the previous work from multiple MBARI efforts. The at-sea deployments validated the Triton payload, and its three cameras generated over 300 hours of midwater transect video that is added to MBARI's existing database. My contribution to this project was proposing the informed-adaptive control logic, implementing the real-time vision pipeline, which connected the incoming video stream from the cameras to the adaptive YoYo control logic on the main vehicle, and post-deployment analysis of the data. The missions showed a clear correlation between *Nanomia septata* detection rates and depth, supporting the potential of using adaptive YoYo's for studying sparse midwater organisms. Some succesful adaptive YoYo's showed the feasibility of visually driven adaptive sampling, but at the same

time, the field trials showed several challenges that should be addressed in future work. Detection performance was only evaluated in a limited way and was sometimes affected by marine snow and low-contrast conditions. Generating a labeled dataset from the Triton imagery would allow for quantitative evaluation and retraining of detection models for this particular payload. Incorporating object tracking and using all three Triton cameras for cross-validation could also reduce false positives. The potential of the three cameras should in general be further explored, as they could be used for depth estimation and other things. Future deployments can test the informed-adaptive control logic at sea, switching between exploration (standard YoYo dives) and adaptive (adaptive YoYo dives) phases. Interesting priors would be temperature, chlorophyll or co-occurring species to inform about likely target regions. This approach could, for example, be used to locate a rare species by first identifying the conditions it tends to inhabit and then remaining in that zone to maximize chances of finding it. Additionally, getting an understanding of the context first could inform the parameter fine-tuning process for the adaptive control, which is currently done manually.

ACKNOWLEDGEMENTS

The MBARI Summer Internship Program is generously supported through a gift from the Dean and Helen Witter Family Fund and the Rentschler Family Fund in memory of former MBARI board member Frank Roberts (1920-2019) and by the David and Lucile Packard Foundation. Additional funding is provided by the Maxwell/Hanrahan Foundation. This project was funded by a NOAA grant. I was wonderfully guided through the process by my mentors Paul Roberts, Joost Daniels, Kakani Katija, and the entire Bioinspiration Lab. It also builds on the great work of Paul Roberts, who built the payload; Yanwu Zhang, who developed the adaptive targeted sampling algorithm; Kevin Barnard and Marc Allentoft-Larsen, who trained the detection model; and Akshay Hinduja, who developed the camera drivers. More broadly, the entire learning experience was amazingly supported by George Matsumoto, Megan Bassett, Jessica Chapman, and the fellow interns.

References

[1] Y. Zhang, B. Kieft, B. W. Hobson, B. Raanan, W. Ussler, C. M. Preston, R. M. Errera, P. A. Den Uyl, A. V. Woude, G. J. Doucette, S. A. Ruberg, K. D. Goodwin, J. M. Birch,

and C. A. Scholin, “Using a long-range autonomous underwater vehicle to find and sample harmful algal blooms in lake erie,” *Limnology and Oceanography: Methods*, vol. 22, no. 7, p. 473–483, Apr. 2024. [Online]. Available: http://dx.doi.org/10.1002/lom3.10621

[2] Y. Zhang, B. Kieft, P. Roberts, B. W. Hobson, Q. Shemet, and T. Maughan, “Auv adaptive yo-yo based on layer peak and edge detection and its application to imaging plankton layers using an onboard microscopy system,” July 2025, manuscript submitted for publication.

[3] K. Barnard, J. Daniels, P. L. D. Roberts, E. C. Orenstein, I. Masmitja, J. Takahashi, B. Woodward, and K. Katija, “Deepstaria: enabling autonomous, targeted observations of ocean life in the deep sea,” *Frontiers in Marine Science*, vol. 11, Apr. 2024. [Online]. Available: http://dx.doi.org/10.3389/fmars.2024.1357879

[4] Nov 2023. [Online]. Available: https://www.mbari.org/animal/common-siphonophore/

[5] B. W. Hobson, J. G. Bellingham, B. Kieft, R. McEwen, M. Godin, and Y. Zhang, “Tethys-class long range auvs - extending the endurance of propeller-driven cruising auvs from days to weeks,” in *2012 IEEE/OES Autonomous Underwater Vehicles (AUV)*, 2012, pp. 1–8.