frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 05 April 2024
DOI 10.3389/fmars.2024.1357879

Check for updates

OPEN ACCESS

EDITED BY
Daniel Wagner,
Ocean Exploration Trust, United States

REVIEWED BY
Shaowei Zhang,
Chinese Academy of Sciences (CAS), China
Christoph Waldmann,
University of Bremen, Germany

*CORRESPONDENCE
Kakani Katija
kakani@mbari.org

RECEIVED 18 December 2023
ACCEPTED 19 March 2024
PUBLISHED 05 April 2024

CITATION
Barnard K, Daniels J, Roberts PLD, Orenstein EC, Masmitja I, Takahashi J, Woodward B and Katija K (2024) *DeepSTARia*: enabling autonomous, targeted observations of ocean life in the deep sea. *Front. Mar. Sci.* 11:1357879. doi: 10.3389/fmars.2024.1357879

DeepSTARia: enabling autonomous, targeted observations of ocean life in the deep sea

Kevin Barnard[1], Joost Daniels[1], Paul L. D. Roberts[1], Eric C. Orenstein[1], Ivan Masmitja[1,2], Jonathan Takahashi[3], Benjamin Woodward[3] and Kakani Katija[1]*

[1]Research and Development, Monterey Bay Aquarium Research Institute, Moss Landing, CA, United States, [2]Institut de Ciències del Mar, Consejo Superior de Investigaciones Científicas (CSIC), Barcelona, Spain, [3]Research & Development, CVision AI, Medford, MA, United States

The ocean remains one of the least explored places on our planet, containing myriad life that are either unknown to science or poorly understood. Given the technological challenges and limited resources available for exploring this vast space, more targeted approaches are required to scale spatiotemporal observations and monitoring of ocean life. The promise of autonomous underwater vehicles to fulfill these needs has largely been hindered by their inability to adapt their behavior in real-time based on what they are observing. To overcome this challenge, we developed Deep Search and Tracking Autonomously with Robotics (*DeepSTARia*), a class of tracking-by-detection algorithms that integrate machine learning models with imaging and vehicle controllers to enable autonomous underwater vehicles to make targeted visual observations of ocean life. We show that these algorithms enable new, scalable sampling strategies that build on traditional operational modes, permitting more detailed (e.g., sharper imagery, temporal resolution) autonomous observations of underwater concepts without supervision and robust long-duration object tracking to observe animal behavior. This integration is critical to scale undersea exploration and represents a significant advance toward more intelligent approaches to understanding the ocean and its inhabitants.

KEYWORDS
ocean, autonomy, machine learning, computer vision, robotics, tracking

1 Introduction

The world's ocean, particularly the deep ocean, is one of the least accessible places on the planet, and represents nearly 98% of the habitable living space by volume (Haddock et al., 2017). Due to its importance in regulating climate (Smith et al., 2018), support of ecosystems that sustain sources of food (Pikitch et al., 2014; Vigo et al., 2021), and other ecological services (Thurber et al., 2014), understanding the ocean and how it changes with time is vitally important. However, conducting observations at spatiotemporal scales that meaningfully characterize a changing ocean is no small feat (Capotondi et al., 2019). The chemical and physical oceanography communities are beginning to meet this challenge by successfully implementing programs that rely on large-scale autonomy, robotics, and data sharing to achieve their goals (McKinna, 2015; Claustre et al., 2020). For a number of reasons, biological observations have fallen behind, where long-term observations cover only 7% of the ocean's surface waters, and are focused largely in coastal regions (Hughes et al., 2021; Satterthwaite et al., 2021). This lack of observational capacity creates large knowledge gaps in our accounting for and understanding of marine biodiversity, creating challenges for regulation and monitoring of human activities in the ocean (Hughes et al., 2021). Ocean scientists and stakeholders must improve our ability to observe the ocean as the Blue Economy (Bennett et al., 2019) — ocean-related industries and resources from renewable energy generation to food harvesting and culturing — grows and the marine environment continues to shift as the climate changes (Danovaro et al., 2020).

Relying on fully manual, labor-intensive approaches to exploration and monitoring in the ocean are too costly to execute at the necessary scale; established ship-based protocols require hours of highly trained human effort on specialized vessels. Expanding our biological observational capacity requires new autonomous sampling strategies that respond to the environment by adapting behavior or opportunistically targeting organisms (Costello et al., 2018; Ford et al., 2020). Here we present *DeepSTARia* (Searching and Tracking Autonomously with Robotics), a class of algorithms that enables autonomous underwater vehicles to execute targeted sampling tasks based on real-time visual signals, a strategy previously only available to human operators. *DeepSTARia* represents a significant advance in deep sea autonomy, illustrating the potential for autonomous underwater vehicles to effectively scale up our ability to study marine organisms by reducing the need for costly ship time and limiting reliance on manual operation.

Non-extractive biological observations can be conducted in many ways using various modalities, including imaging, environmental DNA (or eDNA), and acoustics (Benoit-Bird and Lawson, 2016; Masmitja et al., 2020; Chavez et al., 2021). Of these modalities, imaging is the most direct approach, and its use has grown with various platforms, imaging systems, and sampling missions (Durden et al., 2016; Lombard et al., 2019). Benthic landers, cabled observatories, and drop cameras for example can provide temporal data of animal distributions at a fixed location (Danovaro et al., 2017; Giddens et al., 2020). Other approaches using remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUVs) have the benefit of mobility to provide varying views in time and space of biological communities in the ocean (Robison et al., 2017). While AUVs have the benefit of autonomy (Schoening et al., 2015; Ohki et al., 2019), most of these platforms do not have adaptive and targeted sampling capabilities when compared with manually controlled ROVs (Durden et al., 2021).

Biological observations using ROVs and AUVs traditionally involve quantitative transects (Howell et al., 2010). *Transects* are missions where imaging parameters and vehicle behavior are kept constant (e.g., position relative to the seafloor for benthic missions, observation depth for missions in the water column, vehicle speed, vehicle heading, sampling duration, imaging field of view, camera exposure, illumination power) while sampling a particular location in the ocean. *Transects* can be conducted at different locations or time intervals to address a number of ecological questions (Robison et al., 2017). At the conclusion of transect missions, researchers download and review the collected visual data to identify animals, quantify species occurrence and counts, and denote the physical environment to characterize the biological community (Howell et al., 2010; Aguzzi et al., 2021). Such missions are often conducted for marine biodiversity monitoring but are not sufficient to properly account for all organisms, especially those that are small in body size, relatively rare, or patchily distributed (Brandt et al., 2014). Adaptive sampling strategies are necessary to properly account for marine biodiversity, especially in the difficult-to-access deep sea (Costello et al., 2018).

Oftentimes, research goals dictate a more opportunistic approach, seeking out and capitalizing on rare encounters. This necessitates a very different sampling strategy, usually requiring a closer look to identify animals or observe their behavior (Ford et al., 2020). These *Discovery* missions involve pausing a *Transect* to collect close-up or extended recordings of an animal to facilitate identification (Figure 1). These missions are usually directed by scientists, viewing the *in situ* video feed on a topside monitor, and adjusting vehicle behavior when they see an animal or phenomenon of interest and need more time or additional perspective views for study and evaluation. More recently, researchers have been interested in understanding not only presence and absence of animal systems, but also their fine-scale behavior to understand their ecomechanics (Katija et al., 2020). These studies require *Follow* missions to keep the target in view for longer periods of time. Both *Discovery* and *Follow* operations are typically run on an ROV flown by a skilled human pilot, which we define here as an individual with many hours of experience and who operates ROVs in a professional capacity.

Thanks to recent improvements in AUV capabilities and performance (e.g., power, control, and on-board computational resources), the research community has begun developing targeted and adaptive biological observation capabilities for these autonomous robotic platforms (Zhang et al., 2021). By switching from ROVs – which require significant physical infrastructure and personnel that cost on the order of tens of thousands of dollars per day to operate – to vehicles like AUVs, we could enable large-scale, global surveys of ocean life capable of meeting the endurance, depth range, and maneuverability requirements for such missions (Reisenbichler et al., 2016). Making these sampling strategies entirely autonomous involves leveraging vehicle sensor data (imaging, acoustics, or both) to locate animals of interest and

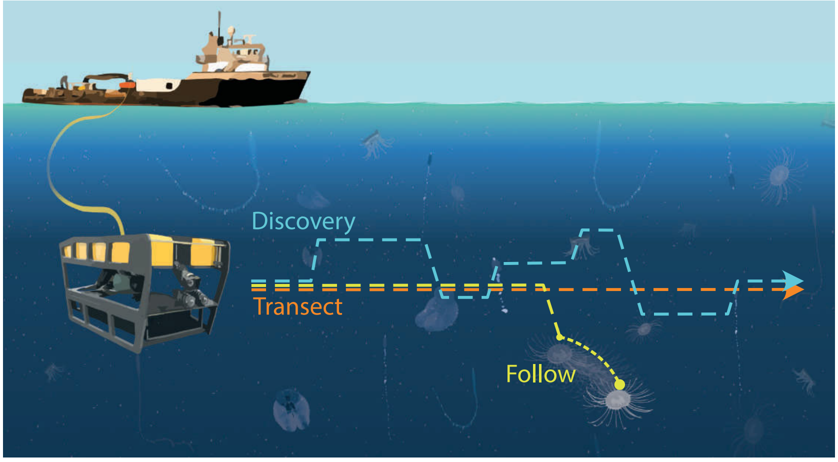

FIGURE 1
Integrating machine learning (ML) algorithms into vehicle controllers (or *DeepSTARia*) enables a suite of underwater observational missions. By varying the duration of various modes (*search*, *acquire*, *track*) of *DeepSTARia*, an autonomous underwater vehicle can conduct a variety of underwater observational missions: (Orange) *Transect*, where the vehicle moves at a constant speed and depth at specified time intervals; (Blue) *Discovery*, where the vehicle moves at a prescribed depth and changes vehicle behavior (e.g., range, bearing, depth) to slow down and observe a detected object for a specified duration before continuing on its sampling mission; and (Yellow) *Follow*, where the vehicle again moves at a prescribed depth, and slows down and continues to shadow a detected object for as long as needed for the specified mission. The ML algorithms enable selection of detected objects, enabling targeted sampling during *Discovery* and *Follow* missions.

maintain the position of the vehicle relative to the target for as long as possible (Yoerger et al., 2021). Using these signals for visual tracking and control (or visual servoing) has a long history (Wu et al., 2022), and more modern algorithms (Girdhar and Dudek, 2016; Katija et al., 2021) show promise in enabling the entire range of vehicle missions described here. However, *Discovery* and *Follow* behaviors remain challenging to implement in the open ocean and require significant algorithmic improvements before they can be conducted without humans-in-the-loop.

To address this challenge, we developed *DeepSTARia* to expand the opportunistic and adaptive sampling capabilities of remote and autonomous vehicles in the deep ocean. *DeepSTARia* consists of four modules: an object detection and classification model, a 3D stereo tracker, a vehicle controller, and a Supervisor module. By integrating real-time machine learning models operating on visual data into vehicle controllers, *DeepSTARia* has achieved a range of biological observation missions (e.g., *Transect*, *Discovery*, and *Follow*) completely autonomously for the first time. We demonstrate that vehicles using *DeepSTARia* can conduct traditional and adaptive biological observation missions without human intervention. Field tests were conducted in Monterey Bay, California, USA with a flyaway ROV as a proxy for any AUV carrying a stereo camera system. An object detection model, trained on 15 taxonomic groups, enabled near-real-time iterative improvements to the *DeepSTARia* algorithm and timely human intervention if required. Minimal user input to the algorithm enabled a suite of autonomous observations that either match or improve our biological observation capabilities during fully remote missions. Our results demonstrate the potential for *DeepSTARia* and similar tracking-by-detection algorithms to enable future autonomous missions to ply the ocean for known and unknown life. These approaches are an important step toward scaling biological observations in the ocean by reducing the human, fiscal, and environmental costs of fully manual operations. The valuable resulting data could inform intelligent, sustainable management of our shared ocean resources and inspire the future of large-scale ocean exploration.

2 Materials and methods

In order to evaluate the effectiveness of *DeepSTARia*, we conducted field trials using a deep sea robotic platform in Monterey Bay. After field trials, data were reviewed to compare various water column exploration missions using the metrics described below.

2.1 Robotic platform used to demonstrate *DeepSTARia*

Field trials of *DeepSTARia* were conducted in the Monterey Bay National Marine Sanctuary at Midwater Station 0.5 (latitude: 36.781 N, longitude: 122.012 W) with bottom depths exceeding 500 m. We used a tethered remotely operated vehicle (ROV) for our field trials as a proxy for an autonomous vehicle so as to enable real-time iterative improvements to the algorithm during trials and utilize human intervention if the need arose. Five dives were made with the 1500 m-rated ROV *MiniROV* (Figure 2) as part of these trials; results reported here were all obtained within a 6-hour window during a single dive on May 24th, 2021 to a maximum depth of 293 m. In these trials, the science/pilot camera (Insite Pacific Inc. Mini Zeus II) and white lights were complemented by a fixed stereo imaging system (based on Yoerger et al., 2021) to provide repeatable position measurements and red lights to reduce interference with animal behavior for these trials (Allied Vision

G-319B monochrome cameras and Marine Imaging Technologies underwater housings with glass dome ports, and Deep Sea Power and Light MultiRay LED Sealite 2025 at 650–670 nm). The stereo imaging system (baseline approximately 190 mm) was mounted such that the port (left) side camera was aligned with the vertical plane of the science camera, and the center of the vehicle. The center of this camera view was chosen as the origin of the vehicle's orthogonal reference frame for the purposes of *DeepSTARia* (Figure 2A). The machine learning models and vehicle control algorithms (Figure 2B) were operated on a shipboard (or topside) Tensorbook laptop (Lambda Labs, Inc.) outfitted with an Nvidia RTX 2070 GPU to allow for rapid switching between pilot control and autonomous operation.

2.2 Overview of *DeepSTARia*

Deep Search and Tracking Autonomously with Robotics (*DeepSTARia*) enables robust autonomous *Transect*, *Discovery*, and *Follow* missions in the ocean based on visual signals by combining machine learning models with vehicle control algorithms. *DeepSTARia* integrates a multi-class RetinaNet object detection model (Lin et al., 2017), a 3D Stereo Tracker, and a Supervisor module that makes vehicle control decisions to be actuated by the vehicle controller (Figure 2). The object detector is run on each of the stereo cameras, and bounding boxes of target classes are then matched within the Tracker module to estimate their position in 3D space. The object class, location, and track are passed to the Supervisor module (Figure 3), which can adjust behavior of the vehicle based on current and past 3D Stereo Tracker information. Lightweight Communications and Marshaling [LCM; Huang et al. (2010)] is used to share data between modules and save all information for later analysis.

The implementation of DeepSTARia is under active research and development and as such is not intended as a plug-and-play solution. Researchers interested in utilizing the subsequent methods of DeepSTARia should be aware that significant adaptation from the current research implementation may be required to support its deployment. Further details about the initial development of DeepSTARia (known as ML-Tracking), including the challenges encountered, are described in Katija et al. (2021).

2.2.1 Multi-class object detector and 3D stereo tracker modules

Still images from past ROV deployments were used to train the multi-class detector; (Katija et al., 2021) this included both typical color images from the science cameras of several ROVs (drawn from the underwater image database *FathomNet* (Katija et al., 2022), and monochrome images obtained with the stereo camera setup described here. Images of animals commonly observed in the Monterey Bay area were used to form 17 classes (15 taxonomic and 2 semantic categories) using visually distinct taxonomic groups of varying taxonomic levels (e.g., *Aegina*, *Atolla*, *Bathochordaeus*, *Bathocyroe*, *Beroe*, *Calycophorae*, *Cydippida*, *Lobata*, *Mitrocoma*, *Physonectae*, *Poeobius*, *Prayidae*, *Solmissus*, *Thalassocalyce*, *Tomopteridae*; see Figure 4). In addition, parts or associated elements were defined in some cases to enable more precise tracking objectives [e.g., *Bathochordaeus* inner filter, *Bathochordaeus* outer filter, *Calycophorae* (nectosome), *Physonectae* (nectosome), and *Prayidae* (nectosome)]. Labeled images were annotated and localized by experts using a variety of tools (VARS Annotation (Schlining and Stout, 2006), VARS Localize (Barnard, 2020), GridView (Roberts, 2020), RectLabel (Kawamura, 2017), and Tator (CVision AI, Inc, 2019)).

We obtained between 205 and 6,927 images per class in the labeled set, for a total of 28,485 images. This annotated image set was used to fine-tune a RetinaNet model with a ResNet50 (He et al., 2016) backbone pre-trained on ImageNet (Deng et al., 2009). Labeled training data and the MBARI Midwater Object Detector can be accessed via FathomNet at www.fathomnet.org (Katija et al., 2022) and www.github.com/fathomnet/models (Woodward et al., 2022).

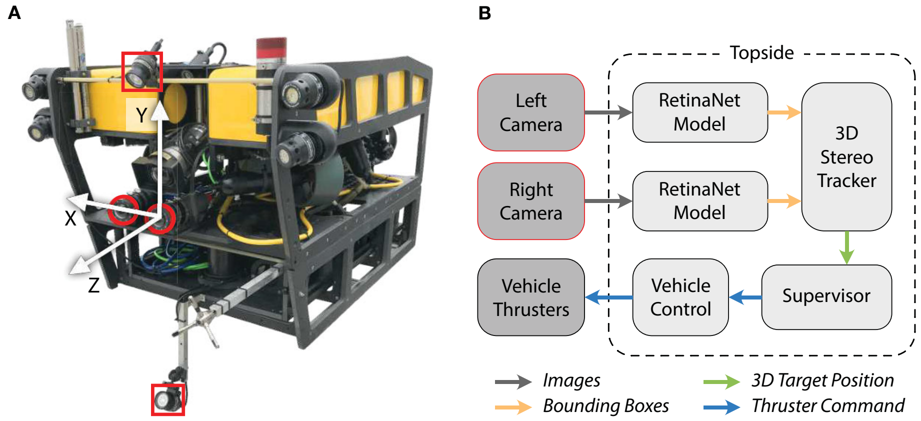

FIGURE 2
Demonstration vehicle and high-level diagram describing the *DeepSTARia* algorithm. **(A)** The ROV *MiniROV* was used as a proxy for an autonomous vehicle, with lighting (red squares) and imaging (red circles) integrated for the field trials. The vehicle reference frame is indicated by the white arrows. **(B)** Images from the left and right camera were transferred up the ROV tether for processing topside. The *DeepSTARia* algorithm involves processing images with a RetinaNet detection model Woodward et al. (2022), where detected object positions in the vehicle reference frame were computed in the 3D Stereo Tracker module. The Supervisor module then uses these inputs and prescribed logic to issue commands to the vehicle controller, which is visualized in real-time during vehicle operations. Modifications to the Supervisor module can then elicit a spectrum of vehicle missions as described here.

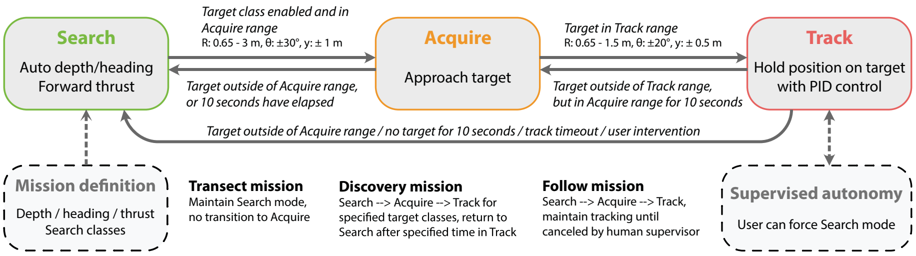
Search
Auto depth/heading
Forward thrust
Target class enabled and in
Acquire range
R: 0.65 - 3 m, θ: ±30°, y: ± 1 m
Target outside of Acquire range,
or 10 seconds have elapsed
Acquire
Approach target
Target in Track range
R: 0.65 - 1.5 m, θ: ±20°, y: ± 0.5 m
Target outside of Track range,
but in Acquire range for 10 seconds
Track
Hold position on target
with PID control
Target outside of Acquire range / no target for 10 seconds / track timeout / user intervention
Mission definition
Depth / heading / thrust
Search classes
Transect mission
Maintain Search mode,
no transition to Acquire
Discovery mission
Search --> Acquire --> Track for
specified target classes, return to
Search after specified time in Track
Follow mission
Search --> Acquire --> Track,
maintain tracking until
canceled by human supervisor
Supervised autonomy
User can force Search mode

FIGURE 3
Supervisor module within *DeepSTARia*. The *DeepSTARia* Supervisor handles transition between three modes, each associated with different vehicle behavior: *search*, *acquire*, and *track*. Thresholds for transition between modes include values for range (*R*), bearing (*θ*), and vertical position (*y*, defined to be consistent with the *y*-axis in Figure 2). Once initial thresholds and set points were set, only minimal input was required to perform any of three mission types (*Transect*, *Discovery*, *Follow*).

The object detector provides information from each camera to the 3D Stereo Tracker, designed to distinguish individual objects (e.g., organisms), by including their positional history (or trajectory) relative to the vehicle. The 3D Stereo Tracker module is a multi-target tracking algorithm based on the unscented Kalman filter [UKF; Wan and Van Der Merwe (2000); Katija et al. (2021)]. It provides the best estimate of the target location relative to the left-camera on the vehicle based on stereo video, target state estimation, and vehicle inertial measurements. It uses stereo intersection over union (IOU) to solve for correspondence between pairs of bounding boxes from the object detector. If a pair has a valid stereo IOU, the Tracker searches for an existing trajectory to update with a measurement using Mahalanobis Distance (Mahalanobis, 2018) as the matching criterion. If no matching trajectory is found, the Tracker starts a new trajectory. At each iteration, the Tracker updates each trajectory with a score based on whether or not a new measurement was assigned to the trajectory. The trajectory with the highest score is used to estimate the target location and output to the Supervisor module (Figure 3). When the tracked object leaves the field of view or is no longer detected by the object detector, the Tracker will "coast" the trajectory for a given number of iterations before deleting the trajectory. In these cases, the Supervisor module will defer to the raw object detector output or Kernelized Correlation Filter (Henriques et al., 2015) Tracker initialized on the latest object detector bounding box.

2.2.2 Supervisor module

Given the positions and classifications of detected objects in the vehicle stereo cameras, a series of commands can be issued to the vehicle controller to progress through various mission modes, a process handled by the *DeepSTARia* Supervisor module (Figure 3). The Supervisor consists of three modes – *search*, *acquire*, and *track* – that loop continuously until transitions are initiated by input from the object detectors, 3D Stereo Tracker, mode timeouts, and external communications (user intervention or Supervised Autonomy). The Supervisor interfaces with the vehicle controller to adjust the vehicle behavior. The vehicle controller has been adapted from (Rife and Rock, 2006; Yoerger et al., 2021) for our system.

In the *search* mode, the vehicle closes the loop on heading (as measured by a compass) and depth (as measured by a pressure sensor) using proportional, integral, and derivative (PID) controllers on all vehicle axes. The user can then specify a desired forward speed in the form of percent thruster effort, with a default of 20% used in our field trials (Figure 3). This mode is analogous to how an ROV pilot would typically fly a vehicle during a midwater or benthic transect mission. For *Transect* missions with *DeepSTARia*, the Supervisor never leaves the *search* mode. For *Discovery* and *Follow* missions, the Supervisor remains in the *search* mode until a particular object or list of objects is detected within the predefined acquisition range of 0.65 m to 3.0 m (Figure 3), and thereby triggering a transition to *acquire* mode. Note that this range can be adjusted depending on your mission requirements.

Upon entering *acquire* mode, the vehicle's behavior is changed, slowing down and centering the detected object in the cameras' field of view. The Supervisor achieves this by slewing the heading and depth setpoints towards the estimated target bearing and vertical offset from the vehicle origin. The same PID control and gains are used in this mode as in *search*. The vehicle forward effort is set proportionally to the range of the object such that as the vehicle approaches the object the forward effort decreases until it becomes zero when the object is within the tracking range (defined below). The Supervisor will remain in the *acquire* mode until the target enters the tracking range (and transitions to *track*) or the target remains outside of the acquisition range for more than 10 seconds (and transitions to *search*).

In *track* mode, the vehicle will attempt to hold its position relative to the target object constant. This is done by enabling the target tracking controller, which closes the loop on range, bearing, and vertical offset of the target with a defined range setpoint (typically set between 0.65 m and 1.5 m; Figure 3) and bearing and vertical offset of 0 (i.e., centered on the left stereo camera). A different set of gains is used in this mode (compared to *search* and *acquire*) to enable more precise tracking of the target with the faster response time to target movement. The Supervisor will remain in the *track* mode until one of four conditions is met: (1) The target drifts outside of the tracking range but remains in the acquisition range for more than 10 seconds (and returns to *acquire*); (2) the Supervisor receives an external command to end the tracking (and returns to *search*, 'supervised autonomy'); (3) The target remains outside the acquisition range for more than 10 seconds (and returns to *search*; 'target lost'); or (4) the track duration exceeds a predefined time limit (and returns to *search*). In *Discovery* missions, this time limit was set to 15 seconds in our field trials, forcing the vehicle to move on in search of new animals that matched the selected classes. In order to prevent reacquiring the previous target in this case, the *search* mode is locked for 1 second after leaving *track* mode. In *Follow* missions, the system can be set to remain in *track* mode indefinitely (until interruption by human intervention); for the purposes of our field demonstrations, this duration was limited to 15 minutes. Note that we distinguish human intervention as an emergency precaution during our field trials to ensure the safety of the vehicle and its operators, whereas human supervision is done during normal operations of the vehicle in an autonomous mode only when prompted by the vehicle.

The Supervisor module implements a list of target classes to track out of the total set of classes the object detector was trained on (Figure 4). During the supervisor loop, detected targets are compared to the list of selected classes and mode transitions occur only when the detected target is in the list of classes of interest, or when the Supervisor is set to ignore class label during target acquisition. This final mode enables the Supervisor to acquire any detected target, but only track targets that belong to a subset of all possible targets.

2.3 Metrics for evaluating *DeepSTARia* field trials

The raw trajectories produced by the 3D stereo Tracker module were subject to several errors common to tracking-by-detection algorithms; due to erroneous detections (false positives and false negatives), misclassifications, and false associations of new detections with existing object tracks, these raw trajectories needed correction. Two post-processing steps were performed for the sake of more meaningful quantitative analysis. The first step aimed to resolve the issue of falsely-joined trajectories comprised of several distinct objects. As these trajectories corresponded to significant time gaps between detections of the distinct objects, all trajectories with gaps of more than 2 seconds between successive detections were split accordingly. Once split, all resulting trajectories with at least 4 frames were maintained. The values reported in Table 1 are representative of the post-processed trajectories. Each trajectory was included in a mission if the timestamp of its first detection fell within the mission time bounds. The duration represents the time between the first and last detections of a trajectory. We report the number of trajectories meeting or exceeding a duration of 15 seconds as a point of comparison with the 15-second tracking timeout for *Discovery* missions.

A trajectory T can be represented as a sequence of n detections, where each detection d_i consists of a timestamp in seconds $t_i \in \mathbb{R}$ and 3D position in the vehicle frame $\mathbf{p}_i \in \mathbb{R}^3$:

$$T = (d_1, d_2, \ldots, d_n)$$

$$d_i = (t_i, p_i)$$

$$\mathbf{p}_i = \left[x_i \; y_i \; z_i\right]^\top$$

The number of detections per second is computed as the frequency of detection events within a 1-second window around each point in the trajectory. As detections occur at a maximum of 10 Hz, these values may range from 1 to 11. The average vehicle-relative target speed is estimated as the sum of point-to-point Euclidean distances (in the vehicle coordinate frame), likewise within a 1-second window around each point.

At each point p_k, the time window is defined by detections d_l and d_r, with d_l minimizing t_l where $t_l \geq t_k{-}1$ and d_r maximizing t_r where $t_r \leq t_k{+}1$. Within these bounds, we arrive at the windowed subsequence $W = (d_l,\ldots,d_k,\ldots,d_r)$.

The detection frequency f is simply the size of the subsequence divided by the true window duration:

$$f = \frac{r - l}{t_r - t_l},$$

and the average vehicle-relative speed $\bar{v}$ is

$$\bar{v} = \frac{\sum_{i=l+1}^{r} \parallel \mathbf{p}_i - \mathbf{p}_{i-1} \parallel_2}{t_r - t_l}.$$

3 Results

Field trials of *DeepSTARia* were performed on ROV *MiniROV* (Figure 1) in the Monterey Bay National Marine Sanctuary over five days in May 2021. The first three days focused on iterative improvements of settings and operational interfaces, while the final two days prioritized testing and performing consecutive *Transect* (Video S1) and *Discovery* (Video S2) missions with a wide array of midwater animal targets (Figure 4). Here, we present only data from our fourth experimental day, to ensure consistency across our tests in terms of vehicle configuration, algorithm settings, and staffing. While the object detection model and 3D stereo Tracker (Figure 2) operated continuously through the entire ROV deployment, we present the results of distinct missions where the ROV pilot relinquished control of the vehicle, and no human supervisor input was provided (Figure 3). Via the Supervisor module, *DeepSTARia* cycled between three modes – *search*, *acquire*, and *track* – that dictate vehicle behavior based on input from the object detector, 3D Stereo Tracker, user-defined settings (e.g. mission type, mode timeouts), and user intervention. Table 1 summarizes the 4 *Discovery* missions performed at different depths, lasting at least 17 minutes each, and the 6 *Follow* missions exceeding 5 minutes that we conducted. Additionally, 3 *Transect* missions are also reported for comparison. We note that a human supervisor did tune the target vertical offset of the 3D Stereo Tracker in small increments over the course of 30 seconds in mission H (Table 1), but no changes to the model parameters or vehicle controller were made.

All but one of the *Follow* missions listed were purposely terminated by human intervention; mission M concluded due to a tracking failure (Table 1). In that case, the tracked object (*Physonectae* nectosome) was particularly low in contrast due to the high level of transparency in this species (*Resomia ornicephala*),

TABLE 1 Data summary of missions conducted during *DeepSTARia* field trials.

Mission		Duration [min:s]	Genus	Mean depth [m]	# of trajectories	Mean trajectory duration [s]	# of trajectories ≥ 15 s
ID	Type						
A	Discovery	22:27		252	111	7.1	21
B	Transect	12:32		252	104	3.2	1
C	Discovery	19:40		201	110	3.7	5
D	Transect	11:28		201	52	2.5	0
E	Discovery	17:22		151	27	5.8	4
F	Transect	10:25		151	16	1.2	0
G*	Discovery	21:11		101	62	5.6	9
H	Follow	11:16	*Solmissus*	247	77	27.8	14
I	Follow	05:00	*Solmissus*	251	27	16.4	2
J	Follow	35:50	*Bolinopsis*	267	152	19.5	11
K†	Follow	08:20	*Bathochordaeus*	250	32	19.4	3
L†	Follow	08:13	*Bathochordaeus*	246	44	19.0	6
M*	Follow	09:10	*Resomia*	111	31	19.0	1

The mission duration is defined as the time between enabling the search behavior and the next human intervention (canceling the mission), with the exception of the Follow missions: here the duration in track mode (without any human input) is reported. While each Follow mission tracked one individual animal (identified to the genus level by expert annotators), other objects entered the field of view, and the associated trajectories are included here. The number of recorded trajectories and their mean duration takes into account all observations with a minimum number of 4 stereo detections. The number of trajectories greater than 15 seconds indicates the number of times track mode was successful. Missions visualized in Figures 5 and 6 are highlighted in grey.
*Exposure settings of the stereo cameras were different from the other missions, creating a brighter image and affecting object detection rates.
†Follow missions K and L tracked the same individual.

even after tuning the camera parameters specifically for this individual. We subsequently focus our analysis on *Follow* mission H (Table 1), where a *Solmissus* jellyfish was tracked for more than 11 minutes. We chose this mission because it showcases several challenges for the algorithm, including: (i) another object of the same class passing by; (ii) total occlusion; and (iii) physical interference by another object (Video S3).

Differences in animal community composition and abundance caused large variations in object detections and trajectories (sequences of 3D positions of a single object derived from detections multiple video frames) between missions at different depths. On average, between A-F (Table 1), the mean trajectory duration (e.g., number of recorded image frames per individual) increased by 187% in *Discovery* missions, and yielded 5% fewer trajectories per unit time than *Transect* missions at the same target depth. As a result, the distance covered per minute was on average 24% less in *Discovery* missions. One of the model classes, (*Physonectae* nectosome), was excluded from triggering the *acquire* and *track* modes due to the high abundance of this class, so that vehicle behavior did not change when this class was detected. However, trajectories were still being recorded, and accounted for ~ 18% of trajectories in *Discovery* missions.

Transect missions are characterized by the Supervisor remaining in *search* mode throughout the mission (i.e., not stopping to track). By comparison, *Follow* missions represent a continuous span of time spent in *track* mode following a single target organism. *Discovery* missions represent a balance between these extremes, where the vehicle repeatedly stops to acquire and track targets of interest for fixed durations before returning to *search* mode. The Supervisor modes over time for the three representative missions A, B, and H is shown in Figure 5, along with the proportion of time spent in each mode. In the *Discovery* mission A, transitions from *search* to *acquire* represent attempts to stop, and transitions to *track* represent successful acquisitions. Whenever the timeout of 15 seconds during each stop was reached, the Supervisor transitioned back to *search* mode. Unsuccessful acquisitions can be seen around 350 and 850 seconds into the mission, where the Supervisor returned back to *search* after a short time in *acquire*.

The *Discovery* and *Follow* missions were conceived to increase the amount of time and number of views per observation of an organism, allowing for a more detailed look for identification by moving the vehicle such that the animal enters the most well-resolved and illuminated area in front of the vehicle with minimal motion blur. This also provides the opportunity to observe the animal's behavior by keeping it centered in the field of view, which is rare during the relatively fast fly-by speeds associated with transects (Figure 5). During *Transect* missions, the vehicle does not respond to object detections, which therefore move radially past and out of view as the vehicle moves forward. *Discovery* and *Follow* mission instead actively align objects with respect to the image center, increasing the number of recorded views. In *Discovery*, the vehicle aligns briefly for a pre-specified duration (15 seconds) with each animal, showing a much larger fraction of bounding box observations near the image center and offering more image frames of each individual.

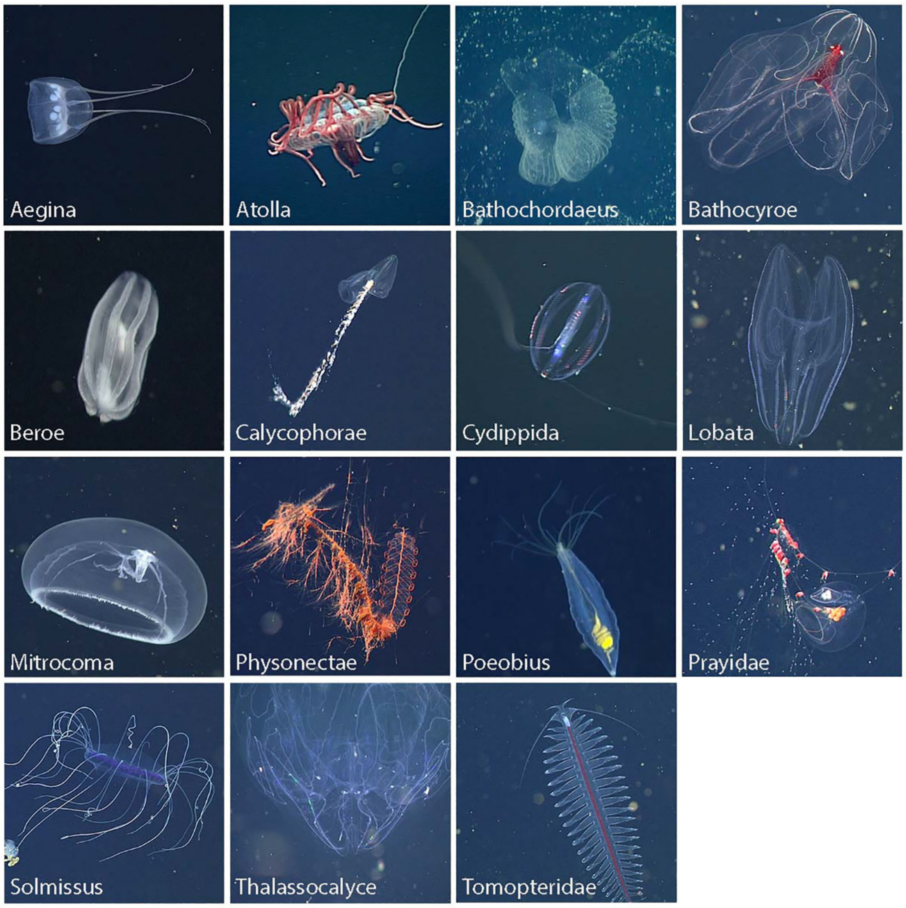

FIGURE 4
Highlight images of midwater animals that served as target objects during *DeepSTARia* field trials. Each image represents one of the 15 taxonomic groups that formed 17 separate classes in the RetinaNet model used in this work. Three classes were defined for *Bathochordaeus*: the animal, house, and outer filter, to address the different size scales of the outer structures and the small animal of interest inside, allowing initial detection of the larger structure and subsequent tracking of the animal inside.

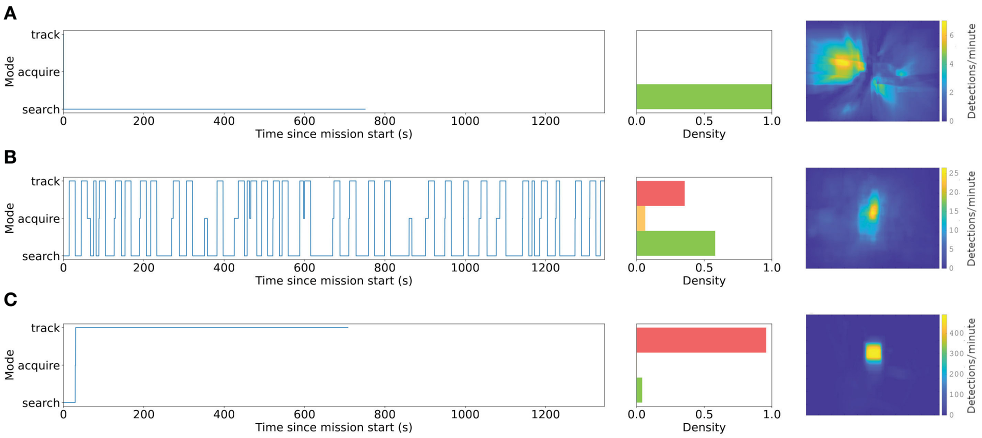

FIGURE 5
Changing modes (e.g., search, acquire, track) and cumulative distributions of detected bounding boxes from **(A)** *Transect*, **(B)** *Discovery*, and **(C)** *Follow* underwater vehicle missions during *DeepSTARia* field trials. Left column shows the mode switching over time for a representative vehicle mission and the middle illustrates the cumulative time within each mode (red = *track*, orange = *acquire*, green = *search*) for the corresponding mission. Heatmaps are based on bounding box locations in the left camera image during each mission type, where *Transect* missions B, D, and F, and *Discovery* missions A, C, and E have been combined, respectively. Note that the range of the color scale increases panels.

Finally, during *Follow*, a single object is kept in place with respect to the vehicle, resulting in higher rates of detections centered in the imaging field of view.

The 3D position of detected objects relative to the vehicle frame can be seen in more detail in Figure 6. The *Transect* mission (Figures 6A, B) sees objects passing by in nearly straight lines at constant velocity, with their detection rate increasing as the vehicle approaches (i.e., the Z position decreases; Figure 6A), until they are lost from the field of view. In *Discovery* missions, the trajectories converge as the vehicle positions itself to center the object of interest within the field of view at a fixed distance, and is associated with a reduction in relative object speeds. The *Follow* mission takes this one step further, maintaining a high rate of detection and very low relative speed throughout once the vehicle is centered on the animal. The proportion of time spent in each mode across *Discovery* missions A, C, E, and G can be seen in Figure 7.

4 Discussion

DeepSTARia enables a range of underwater vehicle missions for biological observations that are otherwise considered impossible to execute autonomously. In addition to traditional transects (where vehicle depth and heading are kept constant) (Howell et al., 2010; Robison et al., 2017), *DeepSTARia* allows for fully autonomous *Discovery* and *Follow* missions that typically require scientists and researchers to monitor and direct underwater vehicle operations (Figure 5). *Discovery* missions enable collection of more images and views – nearly 2.5 times as many on average – of individual targets (Table 1) and higher rates of detections (Figures 6A, C, E) than *Transects*. These improvements enhance the quality and composition of the imagery obtained (Figure 5), enabling extended duration animal behavior observations and more precise and accurate animal identification. *Follow* missions expand our ability to capture long duration observations of an animal in its environment as prescribed by the Supervisor *track* mode timeout setting (Table 1; missions H-M), which was defined to be 15 minutes for our field trials.

Our approach distinguishes itself from other real-time object tracking and visual servoing approaches by integrating a multi-class object detector that includes visually complex classes and the Supervisor module functionality. For *Discovery* and *Follow* missions, the multi-class approach is very effective at reducing undesired changes in vehicle behavior when compared to traditional shape-based approaches [e.g., blob detection (Yoerger et al., 2021)]. A human operator can adapt the observational focus by actively selecting or ignoring certain classes, either for research interests or to account for target abundance. For example, *Discovery* missions that continuously slow on very common species, such as the physonect siphonophore [*Nanomia bijuga*; a member within the same family (Physonectae) is shown in Figure 3] in Monterey Bay, would take a significant amount of survey time. Selective targeted vehicle behaviors like this are generally difficult to specify and control with other unsupervised methods (Girdhar and Dudek, 2016). Here, we manually defined these rejected and target classes prior to the start of a mission. Future work could involve augmenting the Supervisor module to enable the vehicle to dynamically adjust its focus, either disregarding or prioritizing classes surpassing a specified abundance threshold. Furthermore, the object detector used in *DeepSTARia* included three nested classes for the giant larvacean *Bathochordaeus* [animal, house, and outer filter (Katija et al., 2020)], which allows for initial

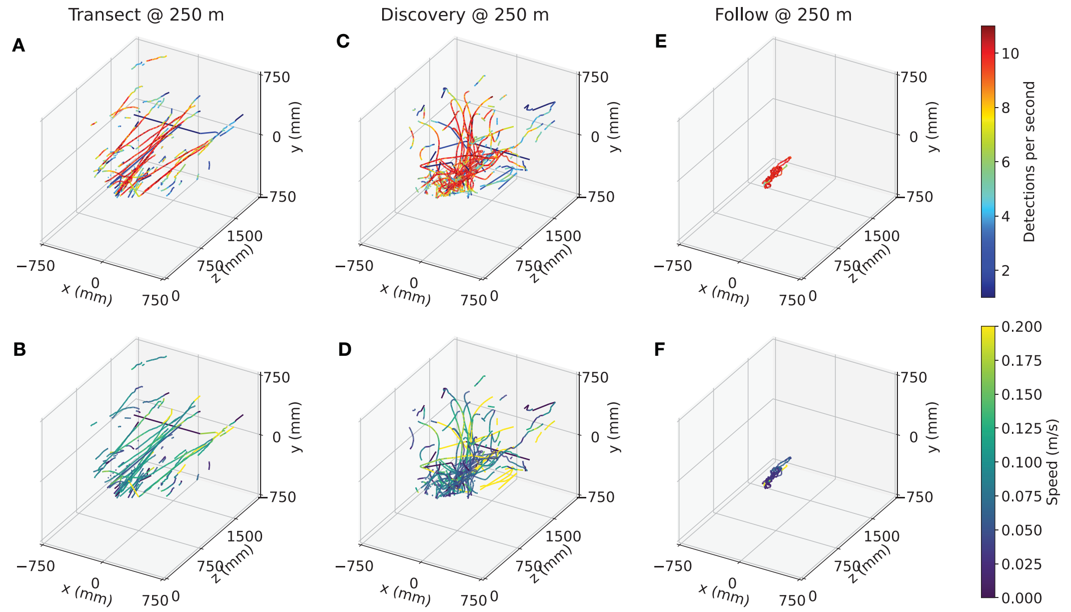

FIGURE 6
Trajectories of observed detections during representative **(A, B)** *Transect*, **(C, D)** *Discovery*, and **(E, F)** *Follow* missions. Top row shows the trajectories colored by the detection rate within a 1-second window, and bottom row shows the average vehicle-relative speed within a 1-second window, for *Transect* mission B, *Discovery* mission A, and *Follow* mission H, respectively.

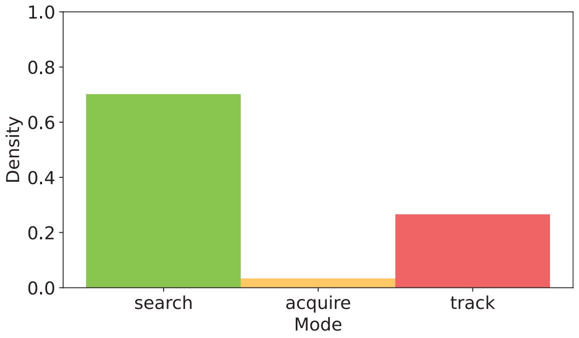

FIGURE 7
Cumulative duration within the three vehicle modes. *search* = green, *acquire* = orange, *track* = red) over four *Discovery* missions (total duration of 80:40 for missions A, C, E, G; Table 1) performed during *DeepSTARia* field trials.

detection of the large outer filter at an extended range, and tracking of the animal itself once the vehicle has approached and slowed. Not only do nested classes like these help to increase the likelihood of successful detection and subsequent tracking of smaller objects that associate with larger ones, changing vehicle behavior farther afield helps to minimize vehicle disturbance of the fragile outer mucus structures as demonstrated in (Katija et al., 2020, Katija et al., 2021) and potential changes in animal behavior.

Both the *Discovery* and *Follow* missions effectively enhance our ability to densely sample organisms of interest either with images, video, or auxiliary sensors. These sorts of long duration observations are invaluable for assessing interactions between an organism, other individuals, and the environment (Norouzzadeh et al., 2018). The missions yield data suitable for novel studies of trait-based (Orenstein et al., 2022) and movement ecology (Abrahms et al., 2021) that fundamentally rely on studying how an organism moves through space. Without bursts of images or videos, ecologists are limited to studying count data in particular spatiotemporal regions (Kennedy et al., 2019). Studying these facets of animal behavior are particularly challenging in the deep sea, where tracking individuals has historically been a labor intensive task requiring the careful attention and precise movements of a skilled ROV pilot. With consistent access to such data, scientists will be able to better assess individual biological fitness, study cryptic predator-prey interactions, and better understand migratory behavior to name a few. These missions can also generate valuable machine learning training data on new objects and animals (Katija et al., 2022), by providing a variety of perspective views on a single organism during *Track* modes that cannot be similarly achieved at the same temporal resolutions during *Transect* missions (Table 1).

Besides enabling unique ecological studies, the *Follow* mission could be used to update the behavior of an individual fully autonomous robot, inform vehicle behavior in multi-vehicle missions (Zhang et al., 2021), or coordinate robot swarms observing collections of targets (Connor et al., 2020). One potential scenario might entail a system of two vehicles: an AUV carrying an imaging system communicating acoustically with an Autonomous Surface Vehicle (ASV) tracking the subsea asset (Masmitjà Rusiñol et al., 2019). Based on the onboard *DeepSTARia* state, imagery can periodically pass between the AUV to the ASV, and be transmitted onwards via cellular or satellite networks to an onshore AUV operator. The AUV operator would monitor the AUV's behavioral changes during deployment using the Supervised Autonomy mode (Figure 2), which can be used to override the *track* mode and have the vehicle resume *search*. The workflow would function akin to our ROV-based work on a single AUV, but could be extended to trigger behavioral changes on additional vehicles carrying other sampling equipment like genomic or acoustics payloads (Zhang et al., 2021). In practice, one could expect that the energy expense of on-board computation for DeepSTARia would limit the deployment time of an AUV; however, our estimates suggest that the power budget would be more heavily impacted by the demands for illuminating the scene rather than the recording or processing of visual data. The Supervised Autonomy framework enables autonomous vehicle behavior adjustments while retaining low-latency guardrails by keeping a human in the loop.

We advocate for the selective automation of ship-borne activities, emphasizing that tasks amenable to automation, such as tedious and repetitive activities like biological monitoring via midwater transects, should be targeted for autonomous execution. Ultimately, the long-term goal is for AUVs to sample biological targets fully autonomously. However, classical supervised ML algorithms trained off-line for real-time detection and identification are unlikely to work in all scenarios in dynamic environments like the ocean: models often struggle when deployed in real world settings due to changing relative proportions of the target classes, the introduction of previously unseen concepts, or discrepancies in the pixel-level image statistics (Recht et al., 2019; Koh et al., 2021). This typically manifests in ecological applications as distribution shifts – where the statistics of the target data differ from that of the training – as a function of time or space (Koh et al., 2021). These challenges are inherent in ocean sampling and limit the ability of fully autonomous systems to adjust their behavior based on visual signals. There are several bleedingedge, pure ML solutions that are well-worth experimentation: Open World Object Detection frameworks to identify novel classes in a new domain (Joseph et al., 2021); contrastive learning to identify out-of-distribution samples and study areas (Yamada et al., 2021); and uncertainty quantification to compute robust confidence thresholds around ML outputs for hypothesis testing (Angelopoulos et al., 2022). Additionally, the promise of reinforcement learning holds potential for addressing the control problem associated with handling more complex animal behavior (e.g., swimming): an area where the current implementation of simple PID thruster-effort-based control struggles. While these approaches are promising, they are experimental, and implementation in the field will benefit from the use of Supervised Autonomy to ensure the routines are effectively acquiring the desired data and evoking the appropriate vehicle behavior.

5 Conclusion

DeepSTARia is a significant stride toward expanding the capabilities of underwater robots by actively adjusting behavior in response to real-time visual observations. Our experiments demonstrated the approach's efficacy on an ROV, allowing a human operator to completely step away from the controls. Deploying *DeepSTARia* on AUVs would fundamentally alter our approach to studying organisms in the deep sea, speeding the discovery of ocean life and processes unknown to the research community. Such a step change in observational capacity is desperately needed: estimates suggest that between 30 and 60% of marine life have yet to be described (Appeltans et al., 2012) and current methods for marine species description can take more than 21 years on average per species (Fontaine et al., 2012). The future of species discovery must someday leverage algorithms like *DeepSTARia* to autonomously run *Discovery* and *Follow* missions to continuously monitor an ocean region or explore a new one (Aguzzi et al., 2020). As algorithms and embedded hardware continue to improve on autonomous vehicles, data collected during these missions may someday lead to onboard learning of features of animals and objects without loss of performance on existing classes, identification of unknown classes (Joseph et al., 2021), and verification by human observers via Supervised Autonomy. These advances, enabled by algorithms like *DeepSTARia*, are critical to scale our ability to discover, study, and monitor the diverse animals that inhabit our ocean.

Data availability statement

The datasets presented in this study can be found in online repositories. Labeled data can be accessed through FathomNet at www.fathomnet.org, and the midwater object detector can be accessed at www.github.com/fathomnet/models. All software related to the DeepSTARia project is open-source. Code used for the May 2021 deployment of DeepSTARia is available at https://bitbucket.org/mbari/ml-boxview.

Ethics statement

The manuscript presents research on animals that do not require ethical approval for their study.

Author contributions

KB: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JD: Data curation, Investigation, Validation, Visualization, Writing – original draft, Writing – review & editing. PR: Conceptualization, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. EO: Investigation, Writing – original draft, Writing – review & editing. IM: Investigation, Writing – original draft, Writing – review & editing. JT: Methodology, Software, Writing – review & editing. BW: Conceptualization, Methodology, Software, Writing – review & editing. KK: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This effort was supported by NSF-GEO-OTIC grant #1812535 (to KK), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant #893089 (to IM), and the David and Lucile Packard Foundation (to KK), the Spanish National Program Ramón y Cajal grant #RYC2022-038056-I (to IM).

Acknowledgments

The authors would like to thank Brett Hobson, Dale Graves, Alana Sherman, Frank Flores, and the crew of RV *Rachel Carson* (MBARI) for their engineering and operational contributions to this effort, and Steve Rock (Stanford Univ.) for his invaluable input and guidance. Additional thanks goes to Alexandra Lapides (MLML) for her assistance generating labeled data using our stereo imaging system.

Conflict of interest

Authors JT and BW were employed by company CVision AI.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2024.1357879/full#supplementary-material

References

Abrahms, B., Aikens, E. O., Armstrong, J. B., Deacy, W. W., Kauffman, M. J., and Merkle, J. A. (2021). Emerging perspectives on resource tracking and animal movement ecology. *Trends Ecol. Evol.* 36, 308–320. doi: 10.1016/j.tree.2020.10.018

Aguzzi, J., Bahamon, N., Doyle, J., Lordan, C., Tuck, I. D., Chiarini, M., et al. (2021). Burrow emergence rhythms of nephrops norvegicus by uwtv and surveying biases. *Sci. Rep.* 11, 5797. doi: 10.1038/s41598-021-85240-3

Aguzzi, J., Flexas, M., Flögel, S., Lo Iacono, C., Tangherlini, M., Costa, C., et al. (2020). Exo-ocean exploration with deep-sea sensor and platform technologies. *Astrobiology* 20, 897–915. doi: 10.1089/ast.2019.2129

Angelopoulos, A. N., Kohli, A. P., Bates, S., Jordan, M., Malik, J., Alshaabi, T., et al. (2022). "Imageto-image regression with distribution-free uncertainty quantification and applications in imaging," in *International Conference on Machine Learning (PMLR)*. 717–730. Available at: https://proceedings.mlr.press/v162/angelopoulos22a.html.

Appeltans, W., Ahyong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., et al. (2012). The magnitude of global marine species diversity. *Curr. Biol.* 22, 2189–2202. doi: 10.1016/j.cub.2012.09.036

Barnard, K. (2020). VARS-localize. Available at: https://github.com/mbari-org/vars-localize.

Bennett, N. J., Cisneros-Montemayor, A. M., Blythe, J., Silver, J. J., Singh, G., Andrews, N., et al. (2019). Towards a sustainable and equitable blue economy. *Nat. Sustainability* 2, 991–993. doi: 10.1038/s41893-019-0404-1

Benoit-Bird, K. J., and Lawson, G. L. (2016). Ecological insights from pelagic habitats acquired using active acoustic techniques. *Annu. Rev. Mar. Sci.* 8, 463–490. doi: 10.1146/annurev-marine-122414-034001

Brandt, A., Griffiths, H., Gutt, J., Linse, K., Schiaparelli, S., Ballerini, T., et al. (2014). Challenges of deep-sea biodiversity assessments in the southern ocean. *Adv. Polar Sci.* 25, 204–212. doi: 10.13679/j.advps.2014.3.00204

Capotondi, A., Jacox, M., Bowler, C., Kavanaugh, M., Lehodey, P., Barrie, D., et al. (2019). Observational needs supporting marine ecosystems modeling and forecasting: from the global ocean to regional and coastal systems. *Front. Mar. Sci.* 623. doi: 10.3389/fmars.2019.00623

Chavez, F. P., Min, M., Pitz, K., Truelove, N., Baker, J., LaScala-Grunewald, D., et al. (2021). Observing life in the sea using environmental dna. *Oceanography* 34, 102–119. doi: 10.5670/oceanog

Claustre, H., Johnson, K. S., and Takeshita, Y. (2020). Observing the global ocean with biogeochemicalargo. *Annu. Rev. Mar. Sci.* 12, 23–48. doi: 10.1146/annurev-marine-010419-010956

Connor, J., Champion, B., and Joordens, M. A. (2020). Current algorithms, communication methods and designs for underwater swarm robotics: A review. *IEEE Sensors J.* 21, 153–169. doi: 10.1109/JSEN.7361

Costello, M. J., Basher, Z., Sayre, R., Breyer, S., and Wright, D. J. (2018). Stratifying ocean sampling globally and with depth to account for environmental variability. *Sci. Rep.* 8, 1–9. doi: 10.1038/s41598-018-29419-1

CVision AI, Inc. (2019). Tator. Available at: https://github.com/cvisionai/tator.

Danovaro, R., Aguzzi, J., Fanelli, E., Billett, D., Gjerde, K., Jamieson, A., et al. (2017). An ecosystembased deep-ocean strategy. *Science* 355, 452–454. doi: 10.1126/science.aah7178

Danovaro, R., Fanelli, E., Aguzzi, J., Billett, D., Carugati, L., Corinaldesi, C., et al. (2020). Ecological variables for developing a global deep-ocean monitoring and conservation strategy. *Nat. Ecol. Evol.* 4, 181–192. doi: 10.1038/s41559-019-1091-z

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255. doi: 10.1109/CVPR.2009.5206848

Durden, J. M., Putts, M., Bingo, S., Leitner, A. B., Drazen, J. C., Gooday, A. J., et al. (2021). Megafaunal ecology of the western clarion clipperton zone. *Front. Mar. Sci.* 722. doi: 10.3389/fmars.2021.671062

Durden, J. M., Schoening, T., Althaus, F., Friedman, A., Garcia, R., Glover, A. G., et al. (2016). Perspectives in visual imaging for marine biology and ecology: from acquisition to understanding. *Oceanography Mar. Biology: Annu. Rev.* 54, 1–72. Available at: https://www.semanticscholar.org/paper/PERSPECTIVES-IN-VISUAL-IMAGING-FOR-MARINE-BIOLOGY-Smith-Dale/ffbb3137f421ef416bd2f7f746304fa61cc66b23.

Fontaine, B., Perrard, A., and Bouchet, P. (2012). 21 years of shelf life between discovery and description of new species. *Curr. Biol.* 22, R943–R944. doi: 10.1016/j.cub.2012.10.029

Ford, M., Bezio, N., and Collins, A. (2020). Duobrachium sparksae (incertae sedis Ctenophora Tentaculata Cydippida): A new genus and species of benthopelagic ctenophore seen at 3,910 m depth off the coast of Puerto Rico. *Plankton Benthos Res.* 15, 296–305. doi: 10.3800/pbr.15.296

Giddens, J., Turchik, A., Goodell, W., Rodriguez, M., and Delaney, D. (2020). The national geographic society deep-sea camera system: A low-cost remote video survey instrument to advance biodiversity observation in the deep ocean. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.601411

Girdhar, Y., and Dudek, G. (2016). Modeling curiosity in a mobile robot for long-term autonomous exploration and monitoring. *Autonomous Robots* 40, 1267–1278. doi: 10.1007/s10514-015-9500-x

Haddock, S. H. D., Christianson, L. M., Francis, W. R., Martini, S., Dunn, C. W., Pugh, P. R., et al. (2017). Insights into the biodiversity, behavior, and bioluminescence of deep-sea organisms using molecular and maritime technology. *Oceanography* 30, 38–47. doi: 10.5670/oceanog

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778.

Henriques, J. F., Caseiro, R., Martins, P., and Batista, J. (2015). High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 583–596. doi: 10.1109/TPAMI.2014.2345390

Howell, K. L., Davies, J. S., and Narayanaswamy, B. E. (2010). Identifying deep-sea megafaunal epibenthic assemblages for use in habitat mapping and marine protected area network design. *J. Mar. Biol. Assoc. United Kingdom* 90, 33–68. doi: 10.1017/S0025315409991299

Huang, A. S., Olson, E., and Moore, D. C. (2010). "LCM: lightweight communications and marshalling," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IEEE)*. 4057–4062. doi: 10.1109/IROS.2010.5649358

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., et al. (2021). Sampling biases shape our view of the natural world. *Ecography* 44, 1259–1269. doi: 10.1111/ecog.05926

Joseph, K., Khan, S., Khan, F. S., and Balasubramanian, V. N. (2021). "Towards open world object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5830–5840.

Katija, K., Orenstein, E., Schlining, B., Lundsten, L., Barnard, K., Sainz, G., et al. (2022). FathomNet: A global image database for enabling artificial intelligence in the ocean. *Sci. Rep.* 12, 1–14. doi: 10.1038/s41598-022-19939-2

Katija, K., Roberts, P. L. D., Daniels, J., Lapides, A., Barnard, K., Risi, M., et al. (2021). "Visual tracking of deepwater animals using machine learning-controlled robotic underwater vehicles," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 859–868. doi: 10.1109/WACV48630.2021.00090

Katija, K., Troni, G., Daniels, J., Lance, K., Sherlock, R. E., Sherman, A. D., et al. (2020). Revealing enigmatic mucus structures in the deep sea using *DeepPIV*. *Nature* 583, 1–5. doi: 10.1038/s41586-020-2345-2

Kawamura, R. (2017) RectLabel. Available online at: https://rectlabel.com/.

Kennedy, B. R. C., Cantwell, K., Malik, M., Kelley, C., Potter, J., Elliott, K., et al. (2019). The unknown and the unexplored: Insights into the pacific deep-sea following NOAA CAPSTONE expeditions. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00480

Koh, P. W., Sagawa, S., Marklund, H., Xie, S. M., Zhang, M., Balsubramani, A., et al. (2021). "WILDS: A benchmark of in-the-wild distribution shifts," in *International Conference on Machine Learning (PMLR)*. 5637–5664. Available at: https://proceedings.mlr.press/v139/koh21a.html.

Lin, Y. H., Wang, S. M., Huang, L. C., and Fang, M. C. (2017). Applying the stereo-vision detection technique to the development of underwater inspection task with PSO-based dynamic routing algorithm for autonomous underwater vehicles. *Ocean Eng.* 139, 127–139. doi: 10.1016/j.oceaneng.2017.04.051

Lombard, F., Boss, E., Waite, A. M., Vogt, M., Uitz, J., Stemmann, L., et al. (2019). Globally consistent quantitative observations of planktonic ecosystems. *Front. Mar. Sci.* 6, 196. doi: 10.3389/fmars.2019.00196

Mahalanobis, P. C. (2018). Reprint of: P. C. Mahalanobis, (1936) "On the generalised distance in statistics". *Sankhya A* 80, 1–7. doi: 10.1007/s13171-019-00164-5

Masmitja, I., Navarro, J., Gomariz, S., Aguzzi, J., Kieft, B., O'Reilly, T., et al. (2020). Mobile robotic platforms for the acoustic tracking of deep-sea demersal fishery resources. *Sci. Robotics* 5, eabc3701. doi: 10.1126/scirobotics.abc3701

Masmitjà Rusiñol, I., Gomáriz Castro, S., Río Fernandez, J., Kieft, B., O'Reilly, T. C., Bouvet, P.-J., et al. (2019). Range-only single-beacon tracking of underwater targets from an autonomous vehicle: From theory to practice. *IEEE Access* 7, 86946–86963. doi: 10.1109/Access.6287639

McKinna, L. I. (2015). Three decades of ocean-color remote-sensing trichodesmium spp. in the world's oceans: a review. *Prog. Oceanography* 131, 177–199. doi: 10.1016/j.pocean.2014.12.013

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., et al. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci.* 115, E5716–E5725. doi: 10.1073/pnas.1719367115

Ohki, T., Nakatani, T., Nishida, Y., and Thornton, B. (2019). "Unmanned seafloor survey system without support vessel and its recent operations in sea trials," in *2019 IEEE Underwater Technology (UT) (IEEE)*. 1–4.

Orenstein, E. C., Ayata, S.-D., Maps, F., Becker, É.C., Benedetti, F., Biard, T., et al. (2022). Machine learning techniques to characterize functional traits of plankton from image data. *Limnology Oceanography* 67, 1647–1669. doi: 10.1002/lno.12101

Pikitch, E. K., Rountos, K. J., Essington, T. E., Santora, C., Pauly, D., Watson, R., et al. (2014). The global contribution of forage fish to marine fisheries and ecosystems. *Fish Fisheries* 15, 43–64. doi: 10.1111/faf.12004

Recht, B., Roelofs, R., Schmidt, L., and Shankar, V. (2019). "Do imagenet classifiers generalize to imagenet?," in *International Conference on Machine Learning (PMLR)*. 5389–5400. Available at: http://proceedings.mlr.press/v97/recht19a.html.

Reisenbichler, K. R., Chaffey, M. R., Cazenave, F., McEwen, R. S., Henthorn, R. G., Sherlock, R. E., et al. (2016). "Automating MBARI's midwater time-series video surveys: The transition from ROV to AUV," in *OCEANS 2016 MTS/IEEE Monterey*. 1–9. doi: 10.1109/OCEANS.2016.7761499

Rife, J. H., and Rock, S. M. (2006). Design and validation of a robotic control law for observation of deep-ocean jellyfish. *IEEE Trans. Robotics* 22, 282–291. doi: 10.1109/TRO.2005.862484

Roberts, P. L. D. (2020). GridView. Available at: https://bitbucket.org/mbari/gridview/.

Robison, B. H., Reisenbichler, K. R., and Sherlock, R. E. (2017). The coevolution of midwater research and ROV technology at MBARI. *Oceanography* 30, 26–37. doi: 10.5670/oceanog

Satterthwaite, E. V., Bax, N. J., Miloslavich, P., Ratnarajah, L., Canonico, G., Dunn, D., et al. (2021). Establishing the foundation for the global observing system for marine life. *Front. Mar. Sci.* 8, 1508. doi: 10.3389/fmars.2021.737416

Schlining, B., and Stout, N. (2006). "MBARI's video annotation and reference system," in *OCEANS 2006*. 1–5.

Schoening, T., Langenkämper, D., Steinbrink, B., Brün, D., and Nattkemper, T. W. (2015). "Rapid image processing and classification in underwater exploration using advanced high performance computing," in *OCEANS 2015 - MTS/IEEE Washington*. 1–5. doi: 10.23919/OCEANS.2015.7401952

Smith, K. L., Ruhl, H. A., Huffard, C. L., Messié, M., and Kahru, M. (2018). Episodic organic carbon fluxes from surface ocean to abyssal depths during long-term monitoring in ne pacific. *Proc. Natl. Acad. Sci.* 115, 12235–12240. doi: 10.1073/pnas.1814559115

Thurber, A. R., Sweetman, A. K., Narayanaswamy, B. E., Jones, D. O., Ingels, J., and Hansman, R. (2014). Ecosystem function and services provided by the deep sea. *Biogeosciences* 11, 3941–3963. doi: 10.5194/bg-11-3941-2014

Vigo, M., Navarro, J., Masmitja, I., Aguzzi, J., García, J. A., Rotllant, G., et al. (2021). Spatial ecology of Norway lobster nephrops norvegicus in mediterranean deep-water environments: implications for designing no-take marine reserves. *Mar. Ecol. Prog. Ser.* 674, 173–188. doi: 10.3354/meps13799

Wan, E. A., and Van Der Merwe, R. (2000). "The unscented Kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373) (IEEE)*, Vol. 31. 153–158. doi: 10.1109/ASSPCC.2000.882463

Woodward, B. G., Katija, K., Roberts, P. L., Daniels, J., Lapides, A., Barnard, K., et al. (2022). MBARI midwater object detector. doi: 10.5281/zenodo.5942597

Wu, J., Jin, Z., Liu, A., Yu, L., and Yang, F. (2022). A survey of learning-based control of robotic visual servoing systems. *J. Franklin Institute* 359, 556–577. doi: 10.1016/j.jfranklin.2021.11.009

Yamada, T., Massot-Campos, M., Prügel-Bennett, A., Williams, S. B., Pizarro, O., and Thornton, B. (2021). Leveraging metadata in representation learning with georeferenced seafloor imagery. *IEEE Robotics Automation Lett.* 6, 7815–7822. doi: 10.1109/LRA.2021.3101881

Yoerger, D. R., Govindarajan, A. F., Howland, J. C., Llopiz, J. K., Wiebe, P. H., Curran, M., et al. (2021). A hybrid underwater robot for multidisciplinary investigation of the ocean twilight zone. *Sci. Robotics* 6, eabe1901. doi: 10.1126/scirobotics.abe1901

Zhang, Y., Ryan, J. P., Hobson, B. W., Kieft, B., Romano, A., Barone, B., et al. (2021). A system of coordinated autonomous robots for lagrangian studies of microbes in the oceanic deep chlorophyll maximum. *Sci. Robotics* 6, eabb9138. doi: 10.1126/scirobotics.abb9138