Rapid Over-Horizon Awareness of AUV Image Datasets over Low Communication Bandwidths

Blair Thornton*, Miquel Massot-Campos, Adrian Bodenmann, Samuel Simmons, Chihiro Hirai
Centre for In Situ and Remote Intelligent Sensing, University of Southampton, UK. Email: B.Thornton@soton.ac.uk

***Abstract*—We present a method for flexible over-horizon awareness of autonomous underwater vehicle (AUV) gathered seafloor imagery. The method uses self-supervised learning to create compact dataset summaries. These consist of georeferenced latent representations of images and a subset of human-viewable dataset representative images. The summaries are small enough to send over low-bandwidth satellite networks such as Iridium. This allows seafloor spatial patterns to be visualised and flexibly interpreted within tens of minutes of the AUV surfacing, without the need for physical recovery or proximity to the robot. The method is demonstrated using the University of Southampton's Smarty200 AUV in a UK Marine Protected Area. Over 4000 images (140 GBytes) were summarised into 51 Iridium messages (102 kBytes), achieving a reduction to 1.4 millionth of the original size. Transmission of the summaries takes ∼20 mins once the AUV has surfaced, at which point the data can be remotely interpreted by geographically dispersed experts and operators.**

I. Introduction

Recent ship-free, multi-week autonomous underwater vehicle (AUV) imaging surveys [1] highlight an opportunity for flexible, remote interpretation of seafloor image datasets. This would allow camera-equipped AUVs to be re-tasked in a similar way to the gliders and Argo floats [2], [3] that transmit physical oceanographic datasets and receive mission updates during planned surfacing intervals. However, global communication networks like Iridium (2.4 kbit/s) cannot transmit multi-gigabyte datasets from camera surveys due to bandwidth limits. Faster satellite networks, e.g., Starlink and VSAT achieving 10 Mbit/s to 100 Mbit/s, need large antennae (0.5 to 2 m) that cannot be easily adapted for submersible pressure tolerance. Alternative channels like acoustic, optical, mobile and wifi networks offer speeds from 10 kbit/s to 20 Gbit/s but require specialised infrastructure to be installed within hundreds of meters to kilometers. As such, wireless transmission of large image datasets is likely to remain impractical in most of the ocean for the foreseeable future.

One way to enable remote awareness over low-bandwidths is to analyse images on the fly using pre-programmed classifiers [4]–[7] and transmit classified outputs via Iridium [8], [9]. A limitation is that outputs are constrained to fixed classification schemes, which may be challenging to define prior to multi-week AUV deployments in diverse environments. Furthermore, varying environmental and operational conditions (e.g., water clarity, lighting, altitude) limit the robustness of classifiers [10], making it difficult to trust outputs without access to viewable images for validation. More general approaches include curiosity [11], scene-complexity [12], anomaly [13] and topic [14] based interpretation that have been used for adaptive path planning of AUVs. Kaeli et al., [15] demonstrated online summaries (representative images and cluster grouping) for acoustic transmission of compact semantic summaries during camera surveys. Their approach used local binary patterns (LBP) to extract image features, with online clustering and cluster representative image identification that allowed clusters to be merged into classes.

This research builds on recent advances in self- and semi-supervised learning [16], [17] to generate flexible remote awareness of seafloor images over low communication bandwidths. The contributions of our method are:

- Generation and transmission of georeferenced latent representation spaces (i.e., characteristic features) of large numbers of newly acquired images
- Identification and transmission of dataset (i.e., latent space) representative images to enable flexible remote classification of received latent representations

Advantages over methods that directly transmit classification results are that labelling schemes can be decided after reviewing representative images, which is useful for exploration where classification schemes and criteria may be hard to decide upfront. The approach also allows multiple classification schemes to be applied to the same latent representations, which allows for generation of various semantic maps tailored to different survey goals. The method also improves the robustness to environmental and operational variables, since it correlates classes to characteristics of the acquired images as opposed to pre-defined thresholds. Unlike cluster-based methods where boundaries are constrained by cluster definitions (which might not suit the desired classification scheme), this approach allows for more precise class delineation based on the actual data. We demonstrate our approach with results of remote-awareness achieved during field trials using the University of Southampton's Smarty200 AUV. A 140 GByte image dataset was summarised in ∼100 kByte of image-derived information, consisting of 1500 georeferenced latent representations and 16 compressed representative images. This allows interpretation tasks (e.g., classification) to be completed in tens of minutes of an AUV surfacing, allowing for timely adjustments of operational parameters and data-informed re-tasking between AUV dive cycles without the need for physical recovery or any additional support infrastructure.

This research was funded by TechOceanS (EU H2020 : 101000858). *Blair Thornton is adjunct at IIS, The University of Tokyo, Japan.

II. Method

Our method adapts the semi-supervised workflow presented in [16], which achieves state-of-the-art performance for off-line classification of geo-spatial imagery. In offline semi-supervised workflows, the process begins with self-supervised feature learning, where a neural network is trained to embed images into a compact latent representation space with compression ratios between 10^{-3} to 10^{-6}. This space captures the diverse characteristics of the dataset, with the key advantage being that networks can learn intrinsic patterns from unlabelled images. Next, a subset of labelled images is used to correlate labels with regions of the latent representation space. This allows the entire dataset to be classified. Equivalent performance to supervised learning can be achieved using orders of magnitude fewer human-labelled image examples. Optimal performance is usually achieved when the dataset being classified is also used for feature learning, after pre-processing to reduce the effects of colour attenuation, scale variation, and geometric distortions on image appearance [18]. Classification performance can be further enhanced by selecting images for human labelling that represent diverse regions of the latent representation space. This approach maximises the value of each labelled image [17].

Fig. 1 (top) illustrates the traditional off-line workflow, differentiating between purely automated tasks (machine) of image pre-processing, feature learning and representative image identification, from the manual labelling of identified images (human). Fig. 1 (bottom) illustrates the proposed remote on-line workflow. The main on-line workflow constraints are:

- Self-supervised feature learners cannot be trained on the target dataset as training typically takes several hours to days on a workstation. Instead, existing datasets pre-train the feature encoder to generate the georeferenced latent representations for low-bandwidth transmission.
- Labelling to correlate classes with regions of the latent representation space requires identification of a dataset representative image subset that is sufficiently small for low-bandwidth transmission.

For the first point, previous studies [16] have shown that generic training datasets have limited effectiveness for interpretation of seafloor imagery. Performance improves when images in the dataset used to pre-train feature learners have similar appearance to the interpretation targets [19]. For repeat surveys, this may be a simple as using previous site survey data, but even in this case it is important that camera, lighting, imaging altitudes and water turbidity conditions are similar [10]. In [19], the authors showed improved classification performance by pre-processing images (through colour and geometric correction) collected with different hardware and under varying acquisition conditions before using them for feature learning. In scenarios such as exploration where image appearance is difficult to predict, it may be possible to prepare an ensemble of feature encoders and use dataset similarity metrics [20] to determine the most appropriate encoder to

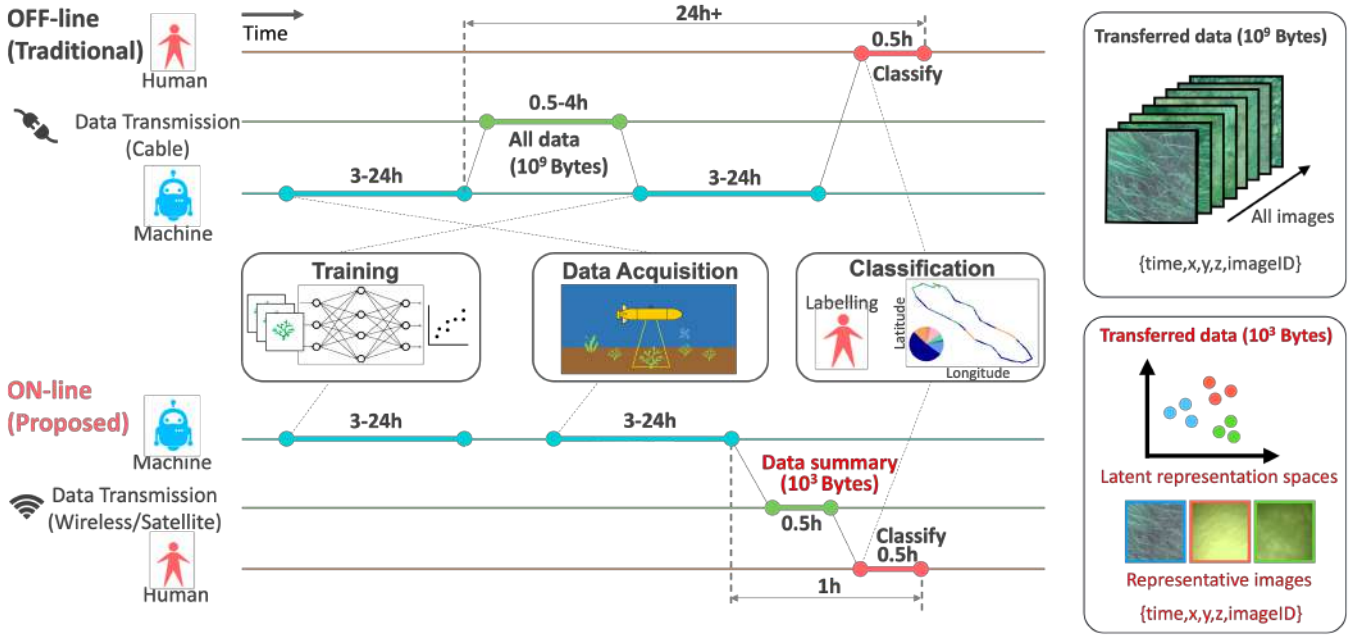

Fig. 1: Off-line and proposed remote on-line classification workflows. The top illustrates state-of-the-art off-line workflows, where large image datasets retrieved via a cable after physical robot recovery are used to train self-supervised machine learning. Typical timescales for sub-processes illustrate the delay to classification result availability. The proposed on-line framework uses pre-trained encoders to create compact feature representations of newly acquired images and identifies a dataset representative image subset. Transmitting these image derived summaries via AUV compatible satellite services allows for flexible remote interpretation within tens of minutes of the AUV surfacing, without the need for physical recovery or additional infrastructure.

embed the latent representation space. Other considerations for remote awareness are that encoding images to the latent representation space must be fast enough for realtime processing with onboard processors, and sufficiently compact to send sufficient numbers of representations to capture spatial distributions in the remotely generated semantic maps.

For the second point, kernel based methods such as Support Vector Machines (SVMs) and Gaussian Processes can efficiently model the correlation between class labels and different regions of the latent space. However, this can only be done if humans can review images of sufficient quality to identify unique class characteristics. Our approach identifies images that sample different regions of the latent representation space using k-means clustering, identifying each cluster's central image for compression and transmission over Iridium short burst data (SBD) messages. The BPG image compression format [21] is used to fit each image into the SBD message size of 1.96 kBytes. Since compressed image sizes cannot be accurately predicted, we apply a range of compression ratios to each identified image, and transmit each representative image's largest BPG compression that remains below 1.96 kBytes. The k-means clustering results, used to identify the representative images, can be visualised prior to any manual labelling to understand spatial patterns. Once labels are assigned to the representative images, a SVM with a radial basis function (RBF) kernel is used to classify all the transmitted latent representations and generate a corresponding semantic map. Key advantages of this approach are that class boundaries are not limited to the cluster boundaries, and since the location of representative images in the latent space is known, processing can take place using only the transmitted information. This allows multiple different classification schemes to be applied without further interaction with the AUV. Additional considerations include whether the compressed images retain enough information for humans to accurately determine class labels, and whether the available images are sufficiently diverse and numerous to capture the desired level of semantic detail. The computational time required to identify and compress the representative images using onboard vehicle processors also needs to be taken into account.

III. Experiments with Smarty200

We demonstrated the method using the University of Southampton's Smarty200 AUV during sea trials at the Studland bay Marine Protected Area (MPA) in September 2023. The MPA is a shallow seagrass meadow located off the coast of Dorset, UK. The AUV (see table I) is equipped with a 12M pixel strobe illuminated camera and gathered images along a 1 km long transect around 10 eco-moorings that have been installed to protect the seagrass from anchor damage. The dive duration was 1 h 7 min, and a total of 4007 images (140 GB) were gathered from a low target altitude of 1.0 m, set due to poor visibility conditions. The remote awareness framework was initiated when the AUV surfaced after the dive.

Images were processed using the AUV's onboard CPU and generated summaries were sent using Iridium SBD messages.

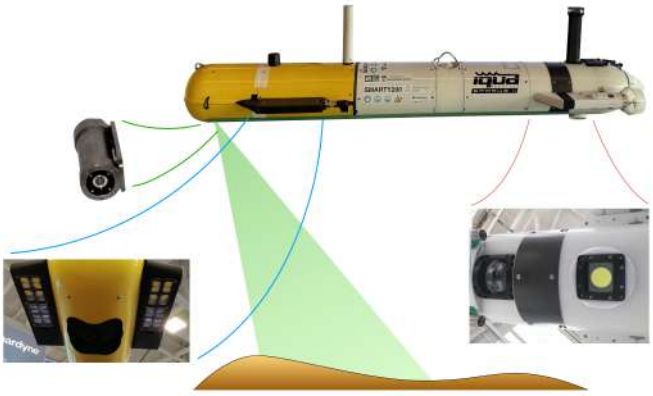

Fig. 2: Smarty200 is a 200 m depth rated seafloor imaging AUV equipped with a strobe illuminated camera and line laser scanning system. It operates at altitudes of 1 to 3 m to gather mm-resolution seafloor images and bathymetry.

TABLE I: Smarty200 system specifications

Length, mass	2.0 m, 70 kg (in air)
Endurance, range, depth	12 h, 12 km, 200 m (max.)
Main CPU	Intel i7-4700-EQ (2.4GHz)
Mapping speed, altitude range	0.3 m/s, 1 to 3 m
Swath, resolution	1.5 m to 4.5 m, <1 mm
3D imaging	Recon LS 12M pixel camera 500,000 Lumen LED strobe Line laser for bathymetry
Navigation	Sprint Nav Mini DVL-INS USBL (Avtrak Nano)
Obstacle avoidance	Micron scanning sonar
Communication	Acoustic (Avtrak Nano in water) Wifi, Iridium (at surface)

Feature encoding was performed using the Location-Guided Autoencoder (LGA) [16], [18], [22]. The LGA architecture is based on the AlexNet convolutional neural network (CNN), which takes 227×227 pixel images as input and has a forward path of 0.71 GFLOP to encode each image into a 16-dimensional latent representation space. The LGA encoder was pre-trained using images gathered during a previous survey of the same region that used a different camera setup [23]. These were pre-processed to correct for light attenuation, geometric distortion and were rescaled to a fixed pixel resolution of 5 mm/pixel based on vehicle altitude and the camera's field of view. Newly acquired images were pre-processed in the same way as the training data, which required a further 45 MFLOP of onboard computation following the method in [19].

Table II summarises the key parameters of the trials. A total of 1500 images were randomly sampled on mission completion, pre-processed and encoded. The total number of images was chosen to balance the number of data points in the semantic maps, where a larger number is desirable to better understand seafloor spatial distributions, and the time taken to generate the summaries onboard Smarty200's CPU (10 min 33 s, averaging 0.42 s/image) and transmit over satellite. 16 representative images were identified by applying k-means clustering to the latent representation space, where k=16.

TABLE II: Summary of remote awareness field trials

Imaging Survey	
Images acquired	4,007
Distance travelled	1.04 km
Duration	1 h 7 min
Raw image dataset size	140 GBytes
On-line Summary Generation	
Latent representations	1500
Pre-processing & encoding	10 min 33 s (0.42 s/image)
Representative images	16
Rep. image ID & BPG compression	48 s (3 s/image)
Total time to generate summary	11 min 20 s
Summary Transmission	
Iridium SBDs (Latents: Rep. images)	51 (35:16)
Time to transmit summaries	∼17 min
Transmitted data size	102 kBytes

The images closest to each cluster centroid were chosen and compressed using the BPG format. This number of images was chosen considering the total time taken for processing and satellite transmission. Clustering to identify 16 images and compression to ∼2 kBytes BPG format took 48 s onboard Smarty200's CPU.

The summarised information (i.e., 1500 latent image representations and 16 representative images) totalled 102 kBytes, representing a size reduction of 1400000:1 compared to the raw images. This was packed into 51 Iridium SBDs using the DCCL v4 protocol, where 35 SBDs contained latent image representations (i.e., 43 image latent representations per SBD), and 16 SBDs contained representative images (i.e., 1 BPG image per SBD). The time required to generate (11 min 20 s) and transmit (∼17 min) the summaries totalled 28 min 20 s, which is broadly similar to the time spent by Argo-floats and gliders at the surface. Once received, representative images were labelled according to the classification scheme described in [23]. The human effort required to assign labels to the 16 images is small, and the computation to train and infer class boundaries requires just a few minutes on a standard laptop for datasets of this size.

IV. Results and Discussion

Figs. 3 to 5 show the results of the remote awareness, where the same colour scheme has been used across the figures. Fig. 3 (left) shows a semantic map generated using the 1500 latent image representations and the location of the 16 representative images. The representative images are shown in Fig. 4, where even at high compression sufficient detail is preserved to identify unique class characteristics. The corresponding 2-dimensional t-SNE projection of the 16-dimensional latent representation space is shown in Fig. 5 (top). Fig. 3 (right) shows the semantic map generated after manual labelling of the 16 representative images. Fig. 4 shows the manual labels, where each representative image appears under the class label (coloured text box) assigned during our experiments. The results of the SVM-RBF classifier are shown in the latent representation space in Fig. 5 (bottom).

The pie chart in Fig. 3 (right) illustrates the site's unbalanced class distribution, with "sand" comprising just 1.6 % of the dataset while "seagrass 80-100 %" accounts for 36.5 %. Despite this, the method successfully identify a subset of representative images that evenly populates each class. This would not be the case if images were naively sampled (e.g., random or spatially stratified) for transmission. The reason for this can be seen in Fig. 5, where classes such as "algae & rock" (11.7 %) and "water column" (0.3 %), which were taken during the AUV's initial descent, form clusters in different regions of the latent representation space from the rest of the dataset. On the other hand, the classes from "sand" to "seagrass 80-100 %" form a continuum of increasing seagrass cover over a sandy substrate. The continuous nature can be seen with these classes forming a sequence of increasing seagrass cover in the t-SNE representation. Unlike methods that send clustering results for merging into relevant classes, the SVM-RBF classifier is not limited to delineate along the cluster boundaries as it considers distances between image representations in the latent space and the provided labelled examples. This is noticeable in Fig. 5, in particular representative image N lies on the edge of the class boundary between sand and "seagrass 0-20 %". Closer inspection of Fig. 4 shows that while the image was manually classified as sand, a few seagrass shoots are present in the image. Advantages over pre-programming classifiers on-board AUV processors is that multiple different classification schemes can be applied, and the ability to view images and review labels increases trust in the semantic outputs.

A limitation of this approach is that the images used to train the latent space encoders are not guaranteed to be similar in appearance to the real-time acquired data. However, since the pre-trained models do not delineate class boundaries directly, but instead compress images into low-dimensional latent vectors, as long as features in the acquired data can be distinguished in the latent representation space, appropriate labels can still be assigned by humans reviewing the dataset representative images. This maintains flexibility and robustness over having pre-trained classifiers that assume appropriately matched and calibrated labelling schemes. Another limitation is that while the BPG compression preserves much of the detail in images (smaller than 2 kBytes) degradation of finer textures in the images potentially limits the ability to identify unique class characteristics, and the limited number of representative images means that only relatively simple classification schemes can be used.

Increasing the size of the summaries (e.g., larger numbers of latent image representations, more representative images, less compressed representative images) is possible. Since AUVs typically takes images at lower than 1 Hz to allow their strobes to recharge, pre-processing and encoding latent feature spaces can be achieved in realtime to generate latent representations of all acquired images. However, it takes ∼20 s to transmit a single SBD message and transmission can only be achieved once the AUV has surfaced. Increasing the number, or quality of representative images also increases the time spent at the surface to transmit the compressed images. Furthermore, the current approach to identify dataset representative images requires dataset acquisition to have been completed. While

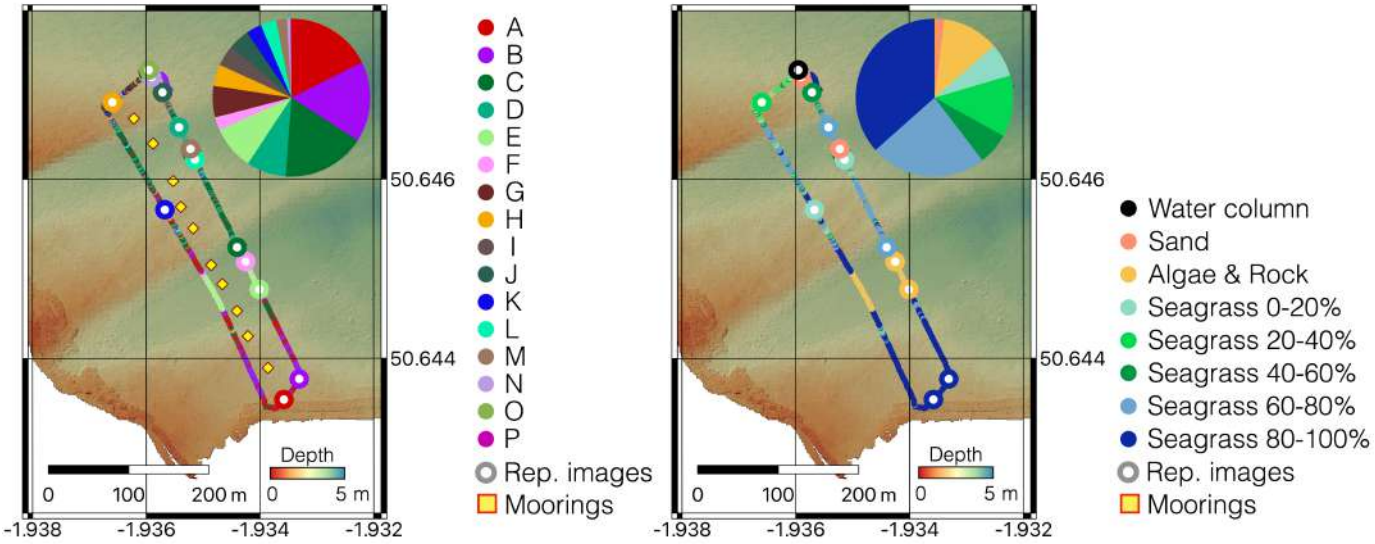

Fig. 3: Semantic maps generated from satellite transmitted summaries of a box survey around eco-moorings at the Studland bay MPA. The left shows locations of the 1500 latent representations and 16 representative images, where colour has been assigned according to intrinsic grouping of regions in the latent space. The pie chart shows the relative occurrence of the types of scene imaged by the robot. The right shows results after classification based on manually labelling of representative images.

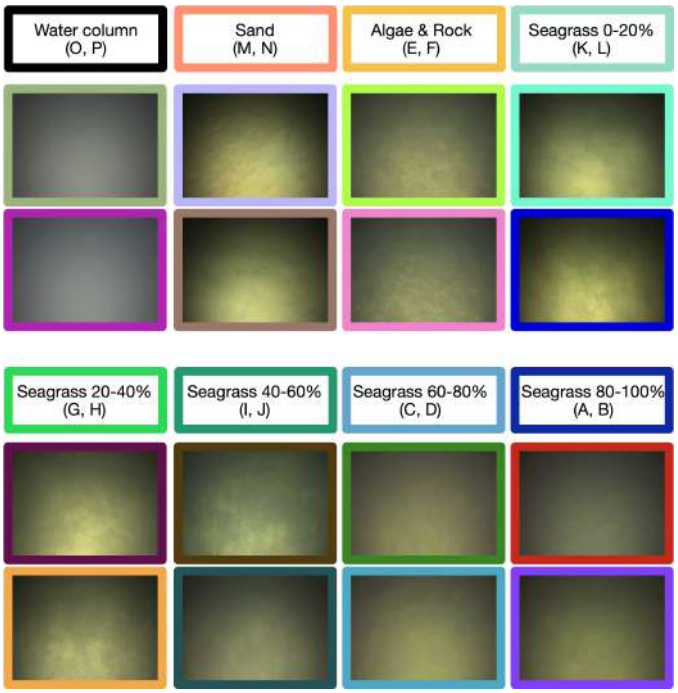

Fig. 4: 16 satellite transmitted representative images. Each image is 350×350 pixels with a resolution of 0.5 mm/pixel. The BPG format is used to compress each image to smaller than 2 kBytes.

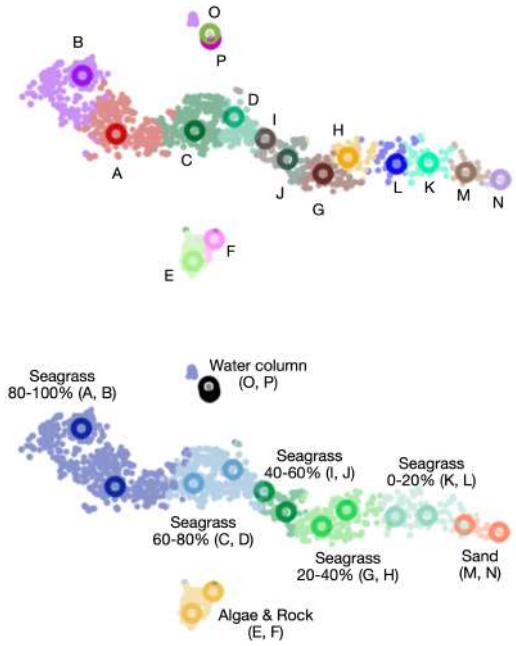

Fig. 5: 2-dimensional t-SNE projections of the 16-dimensional latent representation space. The top shows k-means groupings used to determine dataset representative images (circles). The bottom shows results of the SVM-RBF classifier trained on the manual labels assigned to the 16 representative images.

improvements in efficiency are possible by scheduling part of the processing to happen during acquisition or during the AUV's ascent from depth, communication bottlenecks mean that increased summary sizes inherently increase transmission time, and so the operational risks of the surface environment must also be considered.

V. Conclusion

Flexible remote awareness of seafloor imagery can be achieved over low communication bandwidths by combining self-supervised feature learning encoders with remote semi-supervised classifiers. Our field trials demonstrate this by processing 140 GBytes of seafloor imagery to generate a 102 kByte dataset summary that can be transmitted over the Iridium satellite network in $\sim$17 mins. This represents a 10^6 reduction in data size. Our result show that despite the large reduction in data volume, semantic maps and representative images can allow operators and experts to understand characteristic patterns of spatial distribution on the seafloor within tens of minutes of an AUV surfacing, without the need for physical recovery or proximity to the AUV. This approach can be used to better manage long-endurance AUV imaging missions, and has potential use in multivehicle deployment scenarios, where vehicles often end up queued and immediate recovery is not possible. Although we have demonstrated dataset summarising, the same framework could be adapted to return images that fit a particular target description (e.g., search missions), or images that are anomalous (e.g., in exploration).

References

[1] A. Bodenmann *et al.*, "High-resolution visual seafloor mapping and classification using long range capable auv for ship-free benthic surveys," in *2023 IEEE Underwater Technology (UT)*, pp. 1–6, IEEE, 2023.

[2] R. B. Wynn *et al.*, "MASSMO 4 project ocean glider and autonomous surface vehicle data," 2019.

[3] A. Wong *et al.*, "Argo data 1999–2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats," *Frontiers in Marine Science*, vol. 7, 2020.

[4] A. Balasuriya and T. Ura, "Vision-based underwater cable detection and following using auvs," pp. 1582 – 1587 vol.3, 11 2002.

[5] A. Ortiz, J. Antich, and G. Oliver, "A particle filter-based approach for tracking undersea narrow telecommunication cables," *Machine Vision and Applications*, vol. 22, pp. 283–302, Mar. 2011.

[6] L. Zacchini, A. Ridolfi, A. Topini, N. Secciani, A. Bucci, E. Topini, and B. Allotta, "Deep learning for on-board auv automatic target recognition for optical and acoustic imagery," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 14589–14594, 2020. 21st IFAC World Congress.

[7] N. Palomeras, T. Peñalver, M. Massot-Campos, G. Vallicrosa, P. Negre, J. Fernandez, P. Ridao, P. Sanz, G. Oliver, and A. Palomer, "I-auv docking and intervention in a subsea panel," 09 2014.

[8] M. Picheral, L. Guidi, L. Stemmann, D. Karl, G. R. Id Daoud, and G. Gorsky, "The underwater vision profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton," *Limnology and oceanography, methods*, vol. 8, pp. 462–473, 09 2010.

[9] M. Picheral, C. Catalano, D. Brousseau, H. Claustre, L. Coppola, E. Leymarie, J. Coindat, F. Dias, S. Fevre, L. Guidi, J. O. Irisson, L. Legendre, F. Lombard, L. Mortier, C. Penkerch, A. Rogge, C. Schmechtig, S. Thibault, T. Tixier, A. Waite, and L. Stemmann, "The underwater vision profiler 6: an imaging sensor of particle size spectra and plankton, for autonomous and cabled platforms," *Limnology and Oceanography: Methods*, vol. 20, no. 2, pp. 115–129, 2022.

[10] D. Langenkämper, R. van Kevelaer, A. Purser, and T. W. Nattkemper, "Gear-induced concept drift in marine images and its effect on deep learning classification," *Frontiers in Marine Science*, vol. 7, 2020.

[11] Y. Girdhar and G. Dudek, "Modeling curiosity in a mobile robot for long-term autonomous exploration and monitoring," *Autonomous Robots*, vol. 40, pp. 1267–1278, Oct. 2016.

[12] Y. Otsuki, B. Thornton, T. Maki, Y. Nishida, A. Bodenmann, and K. Nagano, "Real-time autonomous multi resolution visual surveys based on seafloor scene complexity," in *2016 IEEE/OES Autonomous Underwater Vehicles (AUV)*, pp. 330–335, 2016.

[13] Y. Girdhar, W. Cho, M. Campbell, J. Pineda, E. Clarke, and H. Singh, "Anomaly detection in unstructured environments using bayesian nonparametric scene modeling," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2651–2656, 2016.

[14] Y. Girdhar, P. Giguère, and G. Dudek, "Autonomous adaptive exploration using realtime online spatiotemporal topic modeling," *The International Journal of Robotics Research*, vol. 33, no. 4, pp. 645–657, 2014.

[15] J. W. Kaeli and H. Singh, "Online data summaries for semantic mapping and anomaly detection with autonomous underwater vehicles," in *OCEANS 2015 - Genova*, pp. 1–7, 2015.

[16] T. Yamada, M. Massot-Campos, A. Prügel-Bennett, O. Pizarro, S. B. Williams, and B. Thornton, "Guiding labelling effort for efficient learning with georeferenced images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 593–607, 2022.

[17] T. Yamada, A. Prügel-Bennett, S. B. Williams, O. Pizarro, and B. Thornton, "Geoclr: Georeference contrastive learning for efficient seafloor image interpretation," *Field Robotics*, vol. 2, pp. 1134–1155, 2022.

[18] T. Yamada, A. Prügel-Bennett, and B. Thornton, "Learning features from georeferenced seafloor imagery with location guided autoencoders," *Journal of Field Robotics*, vol. 38, no. 1, pp. 52–67, 2021.

[19] M. Massot-Campos, T. Yamada, and B. Thornton, "Towards sensor agnostic artificial intelligence for underwater imagery," in *2023 IEEE Underwater Technology (UT)*, pp. 1–6, 2023.

[20] M. Stolte, A. Bommert, and J. Rahnenführer, "A review and taxonomy of methods for quantifying dataset similarity," 2023.

[21] "Better portable graphics (BPG) image format." https://bellard.org/bpg/. Accessed: 2023-11-26.

[22] T. Yamada, M. Massot-Campos, A. Prügel-Bennett, S. B. Williams, O. Pizarro, and B. Thornton, "Leveraging metadata in representation learning with georeferenced seafloor imagery," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7815–7822, 2021.

[23] M. Massot-Campos, T. Yamada, B. Walker-Rouse, K. Collins, J. Leyland, H. Kassem, and B. Thornton, "Shallow water seagrass survey at studland bay with the auv smarty200," in *2023 IEEE Underwater Technology (UT)*, pp. 1–5, 2023.