Leveraging Foundation Model Adapters to Enable Robust and Semantic Underwater Exploration

Changkyu Choi[1,*], Arangan Subramaniam[2], Nils Olav Handegard[3], Ali Ramezani-Kebrya[2] and Robert Jenssen[1,4,5]

[1]Machine Learning Group, Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway
[2]University of Oslo, Oslo, Norway
[3]Institute of Marine Research, Bergen, Norway
[4]Pioneer Centre for AI, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
[5]Norwegian Computing Center, Oslo, Norway

Abstract

This position paper presents a framework for intelligent underwater exploration by marrying foundation models (FMs) with multi-frequency echosounder data. Echosounder data capture backscattered acoustic signals across a range of frequencies, providing rich insights into underwater environments by exploiting the frequency-dependent scattering properties of underwater targets. However, their heterogeneity and complex structure complicate analysis. To address these challenges, the paper introduces four key innovations aimed at improving echosounder data analysis under dynamic ocean conditions: (1) aligning multi-frequency echosounder data with FMs via lightweight FM adapters, (2) enabling continual adaptation to temporal distribution shifts in dynamic marine environments, (3) designing semantic tokenizers that preserve spatial structures, and (4) effectively leveraging sparse annotations to minimize dependence on costly labeled data. For each research direction, we map recent artificial intelligence (AI) methodologies to marine acoustic challenges and outline concrete pathways for technology transfer. Preliminary experiments demonstrate that a Vision Transformer (ViT), pretrained on natural images in a self-supervised manner, can segment sandeel schools from multi-frequency echosounder data without task-specific retraining. These results substantiate the proposed framework and illustrate the potential of cross-disciplinary AI methods for ecologically informative underwater exploration.

Keywords

Marine intelligence, foundation models, distribution shifts, semantic tokenizers, learning with limited labels

CEUR-WS.org/Vol-3975/paper8.pdf

1. Introduction

As the geopolitical and environmental importance of the ocean continues to rise, the demand for intelligent, scalable, and autonomous marine monitoring systems is becoming increasingly urgent. Despite their pivotal role in Earth's environmental [1, 2] and economic stability [3], oceans remain among the least explored environments. Recognizing their strategic value, maritime nations are accelerating efforts to develop advanced monitoring systems. In this context, AI has emerged as a transformative enabler of marine science, offering new capabilities for processing, interpreting, and integrating complicated and heterogeneous marine data.

Multi-frequency echosounders [4, 5] represent one of the most effective technologies for underwater observation, and the use of AI to analyze such data is an emerging and rapidly evolving field. By emitting and receiving acoustic signals across a broad spectrum of frequencies, these instruments capture highly detailed information about underwater targets. They are now widely deployed across diverse monitoring platforms, providing valuable information across diverse underwater applications [6, 7, 8]. Despite their growing importance, echosounder data present substantial analytical challenges that require specific adaptations for AI-driven workflows. Interpretation remains heavily reliant on specialized

NAIS 2025: Symposium of the Norwegian AI Society, June 17–18, 2025, Tromsø, Norway
*Corresponding author.
changkyu.choi@uit.no (C. Choi); arangan.subramaniam@fys.uio.no (A. Subramaniam); nilsolav@hi.no (N. O. Handegard); ali@ifi.uio.no (A. Ramezani-Kebrya); robert.jenssen@uit.no (R. Jenssen)
0000-0002-7087-4518 (C. Choi); 0009-0003-8147-7678 (A. Subramaniam); 0000-0002-9708-9042 (N. O. Handegard); 0000-0002-8767-5603 (A. Ramezani-Kebrya); 0000-0002-7496-8474 (R. Jenssen)
© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

domain expertise and time-intensive manual processes [9], highlighting the need for intelligent systems that can effectively support expert decision-making. However, the data are typically heterogeneous, high-dimensional, and sparsely annotated, presenting serious limitations for the direct application of conventional AI methodologies. Addressing these challenges calls for a holistic approach that builds upon established AI methodologies, while extending them to capture complex structures and derive context-aware representations tailored to the unique characteristics of echosounder data.

In addition, while training models from scratch on echosounder data is a viable approach, it is often neither efficient nor generalizable in practice. The inherent variability of underwater environments makes it challenging for such models to generalize and perform robustly across diverse contexts [10, 11]. Compounding this issue, energy consumption presents a growing concern in AI research, with Transformer-based models [12] serving as a prominent example. Their well-known data-hungry nature requires large datasets [13], leading to substantial computational overhead and, consequently, elevated energy demands. This raises critical questions about the environmental footprint of AI systems [14], which is especially important in marine science, where the imperative for sustainable practices extends beyond the ocean itself and into the computational methods we use to study it.

In light of these challenges, recent breakthroughs in FMs [15, 16] offer a more sustainable and scalable alternative for analyzing echosounder data. Pretrained via self-supervised learning (SSL) on vast and diverse datasets [17, 18], FMs have demonstrated a remarkable ability to extract general-purpose representations that transfer well across domains [19, 20]. Leveraging such models off-the-shelf can dramatically reduce the need for energy-intensive pretraining, aligning with global efforts to reduce the environmental footprint of AI systems [21].

To fully unlock the potential of FMs for marine acoustic analysis, an essential step is the development of a lightweight adapter [22, 23, 24] that aligns multi-frequency echosounder data with the input space of these pretrained FMs. Such an adapter would serve as an efficient intermediary, translating domain-specific acoustic signals into a representation compatible with FMs originally trained on natural image data. Crucially, by being significantly smaller and more adaptable than the backbone FM itself, this adapter could enable continual learning [25, 26, 27], allowing the system to incrementally adjust to varying environmental conditions without the need for full retraining of FMs. This paradigm holds particular promise for real-world underwater monitoring, where environmental variability and computational constraints demand adaptable, energy-efficient solutions. However, adapting multi-frequency echosounder data to the input space of FMs remains a non-trivial challenge. The fundamental differences between visual and acoustic modalities necessitate new strategies for aligning these data types while preserving their semantic content. As such, developing effective adaptation methods that bridge this modality gap constitutes a vital research direction toward scalable, intelligent, and environmentally responsible underwater environment monitoring.

This position paper presents a unified framework grounded in the FM paradigm, aimed at addressing key challenges in echosounder data analysis. The framework introduces innovations in

(1) Aligning multi-frequency echosounder data with FMs using lightweight adapters,
(2) Enabling learning under limited label scenarios [4],
(3) Capturing temporal variability in dynamic marine environments [28], and
(4) Designing semantic tokenizers that preserve the spatial characteristics of echosounder data [29].

Together, these components lay the groundwork for scalable and environmentally conscious AI systems for advanced underwater environment monitoring. In the following sections, we provide a conceptual overview of each component, with preliminary results for the (1) FM adapters included in Sec.2.

2. Foundation Model Adapters

ViT-based FMs have shown the ability to extract general-purpose representations that enable high performance on downstream tasks with minimal supervision [15, 17, 18]. However, these models are primarily trained on natural images with fixed spatial and channel structures, typically assuming

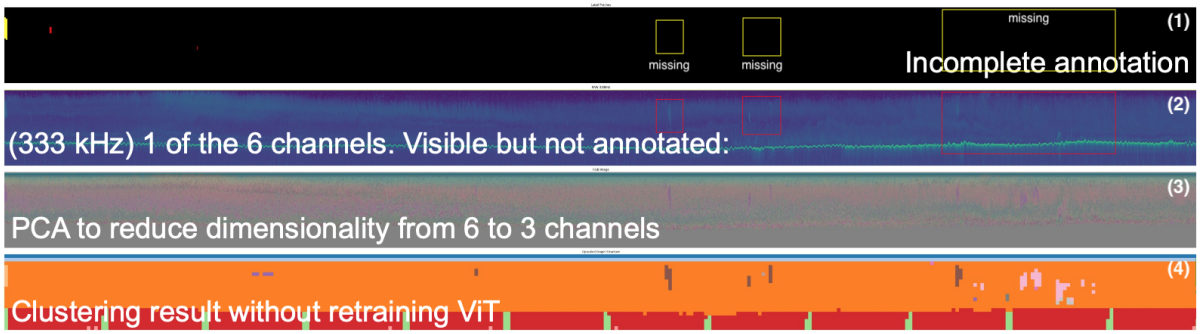

Figure 1: Graph clustering structure of multi-frequency echosounder data, where each node in the graph represents an embedding of an 8×8 patch: (1) The ground truth label mask, which is sparse and incomplete, highlighting the limitation of relying solely on manual human labeling. (2) A single frequency channel (333 kHz) out of the six available in the raw echosounder data, showing visible acoustic signals that are unlabeled in the ground truth. (3) A three-channel representation obtained by applying PCA to the original six-channel data, serving as input to the pretrained ViT-DINO model. (4) The subgraph structure result based on patch embeddings and Louvain community detection. The model successfully segments regions of interest, including those missed by human labeling.

three-channel RGB inputs with consistent visual semantics. Applying FMs directly to multi-frequency echosounder data presents several fundamental challenges due to key differences in both modality and representation compared to natural images.

First, the input is inherently multi-channel and frequency-dependent. A single echosounder observation may consist of multiple channels, each corresponding to a different acoustic frequency band that captures distinct physical properties of underwater targets. For example, the Simrad EK80 wideband echosounder system [30], commonly used in fisheries research and ecosystem monitoring, operates across a frequency range of 10 to 500 kHz and is often configured to emit continuous wave (CW) pulses at six distinct frequencies, 18, 38, 70, 120, 200, and 333 kHz. These frequency-specific responses cannot be trivially collapsed without losing important semantic variation across channels. Second, the data representation and intensity scaling differ significantly from RGB images. Echosounder intensities are typically log-transformed into decibel (dB) scale and clipped to suppress outliers, resulting in pixel values that commonly range from −100 to 0 dB. However, the actual intensity distribution varies with environmental factors and sensor configurations, leading to inconsistent dynamic ranges across datasets. To ensure compatibility with ViT-based FMs, which are sensitive to input distribution, the acoustic data must be carefully standardized through normalization and intensity scaling. These discrepancies motivate the need for a dedicated preprocessing strategy before applying FMs to echosounder data.

To address these challenges, we explore the use of lightweight FM adapter modules designed to align multi-frequency echosounder inputs with the input space of ViT. Our ultimate goal is to enable semantic segmentation of echosounder data by combining a lightweight adapter module with FMs, without relying on manually labeled datasets. In particular, we aim to adapt specific FMs for semantic segmentation such as *Segment Anything Model (SAM)* [15], which typically consists of a ViT encoder that extracts dense visual embeddings and a decoder that produces pixel-level segmentations conditioned on flexible prompts. As a first step, we adapt unchanged ViT-DINO encoder [17], pretrained on natural images, to multi-frequency echosounder data. Specifically, we keep the ViT-DINO encoder frozen and train only a lightweight FM adapter module that aligns multi-frequency inputs with the model's three-channel input requirement. To adapt the six-channel echosounder data for input into the ViT-DINO encoder, we explore both non-parametric and parametric dimensionality reduction methods. For the non-parametric approach, we apply principal component analysis (PCA). The parametric approach involves training convolutional autoencoders without pooling layers to preserve spatial resolution across frequency bands, with the three-channel latent space serving as the input to the ViT-DINO encoder. This adapter is trained independently, and the resulting three-channel latent representation is processed directly to the frozen ViT-DINO encoder.

For evaluation, we leverage the patch embeddings from the ViT-DINO encoder to construct a weighted graph in the form of an affinity matrix, where pairwise relationships (edges) between embeddings (nodes) are computed using a radial basis function (RBF) kernel [31]. We then apply Louvain's community detection algorithm [32] to partition the graph into semantically coherent subgraphs, without the need to specify the number of clusters in advance. This graph-based approach is especially well suited to the extreme class imbalance in echosounder data, where over 99% of observations may correspond to empty water [10, 11, 4, 33], because it enables adaptive region discovery without specifying the number of clusters, unlike k-means. Figure 1 illustrates the graph clustering structures.

3. Toward Continual, Semantic, and Label-efficient Learning with Foundation Model Adapters

FM adapters represent a flexible and powerful mechanism to bridge the gap between FMs and the unique characteristics of echosounder data. As independent modules, FM adapters can also be specialized to address domain-specific challenges, offering the potential for broad applicability across diverse challenges. Drawing inspiration from relevant literature, we outline three promising research directions to enhance the utility of FM adapters: (1) enabling continual adaptation to temporal distribution shifts [28, 26, 34], (2) developing semantically cohesive tokenization strategies [29, 35] , and (3) leveraging limited annotated data effectively for prediction tasks [4, 11].

3.1. Continual Adaptation to Temporal Variability

Marine environments are inherently dynamic, with echosounder data distributions evolving over time due to seasonal, climate, and ecological shifts. Conventional AI methods, which often assume stationary data distributions, struggle to maintain performance in such non-stationary settings. To address this, we propose FM adapters equipped with continual learning capabilities to adapt to temporal distribution shifts [28, 34] in echosounder data. Continual learning [26] refers to the ability of a model to learn from a stream of incoming data while preserving knowledge from previously seen distributions. A major challenge in this setting is catastrophic forgetting [36], a phenomenon in which adapting to new data significantly impairs the model's ability to perform on previously learned tasks.

Incorporating continual learning at the FM adapters offers a modular and lightweight solution, allowing FM adapters to evolve alongside changing data distributions without altering the pretrained FM backbone. Importantly, continual learning in this context can be framed within the importance-weighted empirical risk minimization (IW-ERM) framework [37], where the importance weights serve as a quantitative measure of temporal distribution shift. These importance weights allow the model to adjust its learning process dynamically, providing greater flexibility to adapt to temporal variability in the data. This is especially relevant to echosounder data analysis, where longitudinal survey practices, conducted at regular intervals, naturally induce temporal distribution shifts. In such scenarios, the underlying data distribution can vary between survey campaigns due to changing environmental conditions, often causing static models to underperform.

In estimating importance weights within the IW-ERM framework, density ratio estimation (DSE) [38] has emerged as a promising approach. DSE offers a principled way to quantify the discrepancy between past and current data distributions by estimating the ratio $r = \frac{p_{\text{new}}(x,y)}{p_{\text{old}}(x,y)}$. Recent DSE works [34, 28] provide effective strategies for estimating these ratios under distributional shifts, particularly in distributed learning settings. Applied to FM adapter training, DSE improves the model's ability to handle data exhibiting temporal shifts, supporting resilient and adaptive learning in evolving underwater environments.

3.2. Semantic Tokenizer Design for Spatial Coherence

Tokenization is a critical component of ViT, defining how input data is divided into discrete units for downstream processing. Standard ViTs use fixed, grid-based patch tokenization [12], which assumes uniform spatial structure and is agnostic to content. While effective for structured natural images, this approach is ill-suited for echosounder data, where meaningful acoustic patterns, such as fish schools, tend to be sparse, morphologically irregular, and weakly localized [11]. As a result, square patches may split or dilute semantically important regions, limiting the interpretability and effectiveness of learned representations [29].

To address this challenge, we propose replacing the fixed patch tokenizer in ViT-based FMs with a semantic tokenizer integrated into the FM adapter. This strategy preserves compatibility with the frozen FMs, allowing us to retain their representation power while adapting tokenization to the structure of echosounder data. Recent works [29, 35] offer promising approaches for learning meaningful token boundaries directly from the input. Aasan et al. [29] introduce a superpixel-based tokenizer that treats tokenization as a modular, pluggable component, decoupled from the transformer backbone. Their superpixel token merger enables efficient, online, content-aware tokenization with strong attribution faithfulness and pixel-level granularity. Chen et al. [35] propose a subobject-level tokenizer using boundary detection and watershed segmentation [39, 40] to generate compact, arbitrarily shaped tokens aligned with part-level structures. These semantic tokens are then processed by the ViT encoder to produce patch embeddings, enabling tasks such as zero-shot segmentation without retraining the ViT. Ongoing research aims to develop a pipeline in which the adapter simultaneously performs dimensionality reduction, semantic tokenization, and learns positional embeddings, facilitating seamless integration with FM encoders.

3.3. Efficient Learning from Limited Annotations

Although semantic tokenizers can produce embeddings with pixel-level precision, downstream interpretation still requires aligning these representations with target semantic classes. This is challenging in echosounder data, where annotations are limited, costly, and often ambiguous [10]. Semi-supervised learning [11, 4] offers a promising direction by leveraging both sparse labels and abundant unlabeled data to learn task-relevant decision boundaries. Our goal is to guide the clustering of patch embeddings using minimal annotation, so that the resulting structure aligns with meaningful semantic categories. One published method for semi-supervised learning with echosounder data [11, 4] alternates between unsupervised clustering and refinement using limited labeled samples. However, it relies on offline pseudo-label assignment. All embeddings are first clustered to generate pseudo-labels, and the model is then trained to fit these fixed assignments. This limits adaptability and could be improved through online learning strategies [41].

4. Conclusion

This position paper outlines a unified framework for leveraging FMs in echosounder data analysis through lightweight and modular FM adapters. We address core challenges in modality adaptation, temporal variability, semantic tokenization, and annotation efficiency, each of which is critical for enabling scalable and adaptive underwater monitoring. Our preliminary experiments demonstrate that ViT-based FMs, when combined with minimal adaptation, can perform meaningful analysis of acoustic signals without task-specific supervision. The proposed research directions highlight the versatility of FM adapters and their potential to support continual learning, spatially coherent representations, and label-efficient learning in marine environments.

Acknowledgments

This work is funded by the Research Council of Norway (RCN) through two grants: Visual Intelligence (309439) and CRIMAC–Marine Acoustic Abundance Estimation and Backscatter Classification (309512), both of which are Norwegian Centres for Research-based Innovation (SFI).

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT in order to: Grammar and spelling check, Paraphrase and reword. After using this service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] C. Kuhlisch, A. Shemi, N. Barak-Gavish, D. Schatz, A. Vardi, Algal blooms in the ocean: hot spots for chemically mediated microbial interactions, Nature Reviews Microbiology 22 (2024) 138–154.

[2] N. E. Bosch, F. Espino, F. Tuya, R. Haroun, L. Bramanti, F. Otero-Ferrer, Black coral forests enhance taxonomic and functional distinctiveness of mesophotic fishes in an oceanic island: implications for biodiversity conservation, Scientific Reports 13 (2023) 4963.

[3] S. Gaines, R. Cabral, C. M. Free, Y. Golbuu, R. Arnason, W. Battista, D. Bradley, W. Cheung, K. Fabricius, O. Hoegh-Guldberg, et al., The expected impacts of climate change on the ocean economy, in: The Blue Compendium: From Knowledge to Action for a Sustainable Ocean Economy, Springer, 2023, pp. 15–50.

[4] C. Choi, M. Kampffmeyer, N. O. Handegard, A.-B. Salberg, R. Jenssen, Deep semisupervised semantic segmentation in multifrequency echosounder data, IEEE Journal of Oceanic Engineering 48 (2023) 384–400.

[5] A. Pala, A. Oleynik, K. Malde, N. O. Handegard, Self-supervised feature learning for acoustic data analysis, Ecological Informatics 84 (2024) 102878.

[6] J. K. Horne, J. A. Swan, T. J. Tracy, G. W. Holtgrieve, Automated acoustic monitoring of fish for near-real-time resource management, ICES Journal of Marine Science 81 (2024) 1412–1423.

[7] L. N. Andersen, D. Chu, N. O. Handegard, H. Heimvoll, R. Korneliussen, G. J. Macaulay, E. Ona, R. Patel, G. Pedersen, Quantitative processing of broadband data as implemented in a scientific split-beam echosounder, Methods in Ecology and Evolution 15 (2024) 317–328.

[8] V. Ntouskos, P. Mertikas, A. Mallios, K. Karantzalos, Seabed classification from multispectral multibeam data, IEEE Journal of Oceanic Engineering 48 (2023) 874–887.

[9] G. Pedersen, E. Johnsen, B. Khodabandeloo, N. O. Handegard, Broadband backscattering by Atlantic herring (clupea harengus l.) differs when measured from a research vessel vs. a silent uncrewed surface vehicle, ICES Journal of Marine Science 81 (2024) 1362–1370.

[10] O. Brautaset, A. U. Waldeland, E. Johnsen, K. Malde, L. Eikvil, A.-B. Salberg, N. O. Handegard, Acoustic classification in multifrequency echosounder data using deep convolutional neural networks, ICES Journal of Marine Science 77 (2020) 1391–1400.

[11] C. Choi, M. Kampffmeyer, N. O. Handegard, A.-B. Salberg, O. Brautaset, L. Eikvil, R. Jenssen, Semi-supervised target classification in multi-frequency echosounder data, ICES Journal of Marine Science 78 (2021) 2615–2627.

[12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, International Conference on Learning Representations (2021).

[13] L. Pandey, S. Wood, J. Wood, Are vision transformers more data hungry than newborn visual systems?, Advances in Neural Information Processing Systems 36 (2023) 73104–73121.

[14] M. Xu, D. Cai, W. Yin, S. Wang, X. Jin, X. Liu, Resource-efficient algorithms and systems of foundation models: A survey, ACM Computing Surveys 57 (2025) 1–39.

[15] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, et al., Segment anything, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 4015–4026.
[16] M. Awais, M. Naseer, S. Khan, R. M. Anwer, H. Cholakkal, M. Shah, M.-H. Yang, F. S. Khan, Foundation models defining a new era in vision: a survey and outlook, IEEE Transactions on Pattern Analysis and Machine Intelligence (2025).
[17] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, A. Joulin, Emerging properties in self-supervised vision transformers, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 9650–9660.
[18] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, et al., DINOv2: Learning robust visual features without supervision, Transactions on Machine Learning Research (2023).
[19] R. J. Chen, T. Ding, M. Y. Lu, D. F. Williamson, G. Jaume, A. H. Song, B. Chen, A. Zhang, D. Shao, M. Shaban, et al., Towards a general-purpose foundation model for computational pathology, Nature Medicine 30 (2024) 850–862.
[20] S. Qiu, B. Han, D. C. Maddix, S. Zhang, Y. Wang, A. G. Wilson, Transferring knowledge from large foundation models to small downstream models, International Conference on Machine Learning (2024).
[21] Q. Wang, Y. Li, R. Li, Ecological footprints, carbon emissions, and energy transitions: the impact of artificial intelligence (ai), Humanities and Social Sciences Communications 11 (2024) 1–18.
[22] S. Chen, G. Long, J. Jiang, C. Zhang, Personalized adapter for large meteorology model on devices: Towards weather foundation models, Advances in Neural Information Processing Systems 37 (2024) 84897–84943.
[23] S. Ahmadi, A. Cheraghian, M. Saberi, M. T. Abir, H. Dastmalchi, F. Hussain, S. Rahman, Foundation model-powered 3d few-shot class incremental learning via training-free adaptor, in: Proceedings of the Asian Conference on Computer Vision, 2024, pp. 2282–2299.
[24] F. Chen, M. V. Giuffrida, S. A. Tsaftaris, Adapting vision foundation models for plant phenotyping, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 604–613.
[25] Q. Wang, R. Wang, Y. Wu, X. Jia, D. Meng, CBA: Improving online continual learning via continual bias adaptor, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 19082–19092.
[26] L. Wang, X. Zhang, H. Su, J. Zhu, A comprehensive survey of continual learning: Theory, method and application, IEEE Transactions on Pattern Analysis and Machine Intelligence (2024).
[27] S. A. Bidaki, A. Mohammadkhah, K. Rezaee, F. Hassani, S. Eskandari, M. Salahi, M. M. Ghassemi, Online continual learning: A systematic literature review of approaches, challenges, and benchmarks, arXiv preprint arXiv:2501.04897 (2025).
[28] Z. Wu, C. Choi, X. Cao, V. Cevher, A. Ramezani-Kebrya, Addressing label shift in distributed learning via entropy regularization, International Conference on Learning Representations (2025).
[29] M. Aasan, O. Kolbjørnsen, A. S. Solberg, A. R. Rivera, A spitting image: Modular superpixel tokenization in vision transformers, European Conference on Computer Vision MELEX Workshop (2024).
[30] D. A. Demer, L. N. Andersen, C. Bassett, L. Berger, D. Chu, J. Condiotty, B. Hutton, R. Korneliussen, N. L. Bouffant, G. Macaulay, et al., 2016 USA–Norway EK80 Workshop Report: Evaluation of a wideband echosounder for fisheries and marine ecosystem science, ICES Cooperative Research Reports (CRR), 2017.
[31] B. Schölkopf, A. J. Smola, Learning with kernels: support vector machines, regularization, optimization, and beyond, MIT press, 2018.
[32] P. De Meo, E. Ferrara, G. Fiumara, A. Provetti, Generalized Louvain method for community detection in large networks, in: 2011 11th International Conference on Intelligent Systems Design and Applications, IEEE, 2011, pp. 88–93.
[33] C. Choi, S. Yu, M. Kampffmeyer, A.-B. Salberg, N. O. Handegard, R. Jenssen, DIB-X: Formulating explainability principles for a self-explainable model through information theoretic learning, in:

ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2024, pp. 7170–7174.
[34] A. Ramezani-Kebrya, F. Liu, T. Pethick, G. Chrysos, V. Cevher, Federated learning under covariate shifts with generalization guarantees, Transactions on Machine Learning Research (2023).
[35] D. Chen, S. Cahyawijaya, J. Liu, B. Wang, P. Fung, Subobject-level image tokenization, arXiv preprint arXiv:2402.14327 (2024).
[36] J. L. McClelland, B. L. McNaughton, R. C. O'Reilly, Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory., Psychological review 102 (1995) 419.
[37] A. Gretton, A. Smola, J. Huang, M. Schmittfull, K. Borgwardt, B. Schölkopf, Covariate shift by kernel mean matching, Dataset Shift in Machine Learning 3 (2009) 5.
[38] M. Sugiyama, T. Suzuki, T. Kanamori, Density ratio estimation in machine learning, Cambridge University Press, 2012.
[39] S. Beucher, Use of watersheds in contour detection, in: Proc. Int. Workshop on Image Processing, Sept. 1979, 1979, pp. 17–21.
[40] L. Vincent, P. Soille, Watersheds in digital spaces: an efficient algorithm based on immersion simulations, IEEE Transactions on Pattern Analysis & Machine Intelligence 13 (1991) 583–598.
[41] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, A. Joulin, Unsupervised learning of visual features by contrasting cluster assignments, Advances in Neural Information Processing Systems 33 (2020) 9912–9924.