
Self-Supervised Cluster Discovery Under Imbalanced and Drifting Concept Distributions

Leah Hughes¹, Charlotte E. Johnson¹, Mia Martinez²

¹ Department of Advanced Computing, Viterbi School of Engineering, University of Southern California, Los Angeles, California, USA

² Department of Computer Science, School of Engineering and Applied Science, Yale University, New Haven, Connecticut, USA

Abstract

The exponential growth of unlabeled streaming data in contemporary digital ecosystems has necessitated the development of robust unsupervised learning paradigms. Traditional clustering algorithms assume static data distributions and balanced class representations, assumptions that are systematically violated in real-world scenarios characterized by concept drift and severe class imbalance. This paper presents a comprehensive investigation into self-supervised cluster discovery mechanisms operating under these challenging dynamic conditions. We propose a novel conceptual framework that integrates contrastive representation learning with an adaptive memory bank and an imbalance-aware cluster assignment mechanism. By leveraging temporal coherence and dynamically adjusting margin boundaries based on estimated cluster densities, the proposed approach identifies emerging concepts while preserving knowledge of minority clusters that are highly susceptible to being overshadowed by majority classes during distribution shifts. The research meticulously dissects the theoretical underpinnings of dynamic representation learning, evaluating the interplay between catastrophic forgetting and rapid adaptation. Extensive theoretical analysis and simulated experimental validations demonstrate the efficacy of integrating density-based pseudo-labeling with drift-resilient self-supervision. The findings underscore the critical importance of decoupling representation learning from cluster assignment in non-stationary environments, offering a scalable solution for continuous data stream mining.

Keywords: Self-Supervised Learning, Concept Drift, Imbalanced Data, Data Stream Mining, Cluster Discovery.

1. Introduction

1.1 Background and Motivation

The contemporary landscape of artificial intelligence and machine learning is increasingly defined by the massive influx of streaming data generated by sensor networks, social media platforms, financial transaction systems, and Internet of Things devices. Unlike traditional static datasets meticulously curated in controlled laboratory environments, these real-world data streams are continuous, infinite, and inherently dynamic. A fundamental challenge in processing such continuous streams is the absence of ground-truth annotations, which are typically prohibitively expensive, time-consuming, or practically impossible to acquire in real time. Consequently, unsupervised learning techniques, particularly cluster analysis, have become indispensable tools for discovering latent structures and extracting actionable insights from raw data streams. Recently, self-supervised learning has emerged as a powerful paradigm to bridge the gap between supervised and unsupervised learning, leveraging the internal structure of the data itself to formulate pretext tasks that drive the extraction of rich, high-dimensional representations. However, applying self-supervised learning to cluster discovery in continuous streams introduces profound theoretical and practical challenges, primarily due to the non-stationary nature of the data generation processes. As the underlying phenomena driving the data streams evolve over time, the statistical properties of the incoming data change, a phenomenon formally known as concept drift. Furthermore, these streams rarely exhibit uniform distributions across different underlying classes or clusters; instead, they are characterized by severe structural imbalances where a few majority clusters dominate the data volume, while critical minority clusters appear infrequently. The intersection of these two phenomena, namely drifting concept distributions and severe class imbalance,

creates a highly complex environment where traditional clustering and self-supervised algorithms systematically fail, either by catastrophically forgetting previously learned concepts or by aggressively assimilating minority clusters into adjacent majority clusters. Related work has addressed on the role of discount factor in offline reinforcement learning [1].

1.2 The Problem of Imbalanced Distributions

Imbalanced distributions represent a pervasive challenge in unsupervised cluster discovery. In traditional supervised learning, class imbalance is a well-documented problem addressed through techniques such as synthetic oversampling, cost-sensitive learning, or algorithmic modifications. However, in an unsupervised or self-supervised context, the problem is significantly magnified because the algorithm lacks explicit labels to identify which samples belong to a minority class. Consequently, density-based and distance-based clustering algorithms tend to interpret minority clusters as mere noise or outliers, merging them into the spatial boundaries of massive, high-density majority clusters. When representation learning is applied, self-supervised objectives often optimize for the global expected loss, which is naturally dominated by the majority patterns. This causes the latent space to warp, dedicating vast representational capacity to subtle variations within majority clusters while compressing the representational space allocated to minority clusters. As a result, the distinct features that characterize a minority cluster are smoothed over or entirely lost in the learned latent representations. The difficulty is further compounded when we consider that what constitutes a minority cluster is not absolute but relative to the current temporal window of the data stream. A cluster that is dominant today might become a scarce minority tomorrow, and an entirely new, nascent cluster might initially appear as a tiny cluster of anomalies before growing into a major conceptual category. Related work has addressed in context Semi-Supervised learning [2].

1.3 The Challenge of Concept Drift

Concept drift fundamentally violates the core assumption of traditional machine learning, which posits that the training and inference data are drawn from identical and independent distributions. In streaming data, the joint distribution of features and their latent assignments is a function of time. Concept drift can manifest in several distinct typologies. Sudden drift occurs when the underlying data distribution changes abruptly, often due to a catastrophic event or a discrete shift in the operational environment, such as a sensor failure or a sudden change in market policy. Gradual drift involves a transition phase where an old concept slowly fades out while a new concept slowly fades in, creating a period of overlap and confusion. Incremental drift represents a slow, continuous change in the feature space over time, much like the gradual degradation of a mechanical component leading to subtle changes in its vibration signature. Finally, recurring concepts refer to patterns that disappear and reappear periodically, such as seasonal trends in retail data or diurnal cycles in network traffic. For a self-supervised clustering algorithm, concept drift means that the cluster centroids, the optimal boundaries between clusters, and even the optimal feature representations must be continuously updated. If the algorithm adapts too slowly, it will suffer from concept obsolescence, attempting to fit new data into outdated structural molds. Conversely, if the algorithm adapts too rapidly, it will suffer from catastrophic forgetting, immediately erasing valuable historical knowledge in response to short-term noise or temporary anomalies. Sang examines adaptive quantization strategies for robust ML inference under distribution shift [3].

1.4 Objectives and Contributions

The primary objective of this research is to architect and analyze a comprehensive self-supervised cluster discovery framework capable of robust operation under the simultaneous pressures of severe data imbalance and continuous concept drift. This dual challenge requires a paradigm shift from static, global optimization strategies to dynamic, localized, and density-aware representation learning methodologies. Basset learns sequence features associated with genome accessibility, offering a biological representation-learning example whose labels and evaluation protocol would have to be preserved in any transfer study [4]. This paper details the exploration of a novel algorithmic architecture that seamlessly integrates dynamic feature extraction with robust topological maintenance. The specific contributions of this work are multifold. First, we provide a rigorous theoretical formalization of the self-supervised clustering problem within non-stationary, imbalanced environments, establishing the mathematical necessity for decoupled representation and clustering objective functions. Second, we introduce an adaptive memory bank mechanism that selectively retains representative prototypes of minority clusters, shielding them from the aggressive

optimization pressures exerted by majority streams. Third, we detail a continuous drift detection and adaptation protocol that monitors the temporal evolution of cluster densities, distinguishing between genuine conceptual shifts and transient environmental noise. Finally, through extensive theoretical discourse and simulated empirical analysis, we demonstrate how dynamic margin adjustments in the latent space can effectively preserve the topological integrity of emerging and minority clusters across complex temporal transitions. Another relevant line of investigation concerns adaptive weakly supervised Cross-Modal reliability learning for resource-constrained proxy-state recognition [5].

2. Literature Review

2.1 Evolution of Self-Supervised Representation Learning

Self-supervised learning has revolutionized the field of representation learning by demonstrating that high-quality, semantically meaningful feature spaces can be constructed without reliance on human-curated labels. The foundational premise relies on defining a pretext task, an artificial problem generated directly from the unlabeled data, whose solution requires the neural network to capture the underlying structure of the data. Early approaches utilized pretext tasks such as predicting the rotation angle of an image, colorizing grayscale images, or solving jigsaw puzzles constructed from image patches. While successful in specific domains, these heuristic-based pretext tasks often struggled to generalize across diverse data modalities. The landscape shifted dramatically with the advent of contrastive learning methodologies. Contrastive learning reframes self-supervision as an instance discrimination problem, where the objective is to maximize the representational similarity between differently augmented views of the same data instance while simultaneously minimizing the similarity between the instance and all other instances in the dataset. This paradigm has shown remarkable success in establishing uniform and highly separable latent spaces. More recently, masked autoencoding frameworks have gained prominence, particularly for sequential and grid-structured data, forcing the model to reconstruct masked portions of the input based on the unmasked context. In the context of cluster discovery, these representations are highly advantageous because they transform complex, highly entangled raw data spaces into lower-dimensional manifolds where semantically similar instances are naturally grouped closer together. The multi-label classifier ranks and combines association rules [6]. Such rules could help explain a discovered cluster by identifying recurring combinations of features and outcomes. Under concept drift, however, rule coverage and confidence would need reevaluation, since a stable embedding does not guarantee that an earlier label association remains valid.

2.2 Unsupervised Clustering in Imbalanced Domains

The intersection of unsupervised learning and imbalanced data distribution remains one of the most theoretically demanding areas of contemporary data science. Traditional clustering algorithms, ranging from centroid-based methods to hierarchical agglomerative techniques, inherently assume a relatively uniform distribution of data points across the true underlying categories. Centroid-based models typically optimize the sum of squared distances between data points and their assigned cluster centers. This optimization landscape is highly sensitive to the mass of the clusters; a cluster with a massive number of points exerts an overwhelming gravitational pull on the centroids, frequently leading to the absorption of adjacent, less dense minority clusters. Density-based algorithms offer some resilience by defining clusters as contiguous regions of high point density separated by regions of low density. However, in severely imbalanced datasets, the global density threshold parameter becomes impossible to tune. A threshold high enough to properly segment the massive majority clusters will inevitably discard the sparse minority clusters as ambient noise. Conversely, a threshold low enough to detect the minority clusters will cause the majority clusters to merge into a single, undifferentiated megacluster. Recent literature has explored subspace clustering and density-peak clustering to mitigate these issues. In the realm of deep learning, approaches have attempted to incorporate uniform prior distributions into the clustering objective or utilize margin-based loss functions that explicitly penalize the overlapping of cluster boundaries, though these often require careful, dataset-specific hyperparameter tuning. Another relevant line of investigation concerns offline contextual learning with correlated Bayesian beliefs [7].

2.3 Methodologies for Concept Drift Adaptation

Handling concept drift requires the continuous monitoring of the data stream and the dynamic updating of the underlying machine learning model. The literature generally divides drift adaptation strategies into two primary categories: active detection methods and passive adaptation methods. Active detection methods employ explicit statistical monitoring mechanisms to analyze the incoming data stream and trigger an alarm when a significant deviation from the historical distribution is detected. These methods often rely on hypothesis testing, tracking error rates, or measuring the distance between probability distributions computed over distinct temporal windows. When drift is detected, the active methods typically initiate a model retraining or updating procedure. Passive adaptation methods, on the other hand, do not explicitly attempt to detect the exact moment of drift. Instead, they continuously and incrementally update the model with every new incoming data instance, often utilizing an evaporation mechanism or a sliding window approach to gradually forget older, potentially obsolete data. In the context of unsupervised clustering, drift adaptation involves tracking the spatial location, variance, and density of cluster centroids over time. Techniques such as micro-clustering and macro-clustering have been widely adopted, where the continuous stream is first summarized into sufficient statistics called micro-clusters, which are then periodically aggregated into actionable macro-clusters. Yu and colleagues examine fast bellman updates for wasserstein distributionally robust MDPs [8].

2.4 Synthesis and Identified Methodological Gaps

While substantial progress has been made independently in self-supervised learning, imbalanced data processing, and concept drift adaptation, the synthesis of these three domains reveals critical methodological gaps. Most existing self-supervised clustering models are designed for offline learning on static, curated, and balanced datasets. When applied to continuous, imbalanced streams, these models suffer from severe degradation. The contrastive objectives inadvertently suppress minority features because the negative sampling pools are overwhelmingly populated by majority class instances, causing the model to learn representations that are only useful for discriminating among majority variants. Furthermore, current drift adaptation mechanisms for clustering generally operate on fixed feature spaces. They assume that the raw features or the manually engineered features remain semantically stable over time, even if the distribution of those features changes. Prediction of noncoding-variant effects provides an application of learned sequence features, useful for testing whether a representation preserves biologically meaningful local distinctions [9]. This assumption collapses when utilizing deep representation learning, where the feature space itself is fluid and constantly evolving through continuous self-supervised training. As the encoder updates its parameters to adapt to new concepts, the entire latent manifold warps, invalidating the historical cluster boundaries and density estimates. This creates a destructive feedback loop where representation updates destroy cluster structures, and corrupted cluster structures provide misleading pseudo-labels for subsequent representation learning. Therefore, a critical need exists for a holistic framework that explicitly decouples the temporal stability required for cluster tracking from the plasticity required for self-supervised representation updating, while simultaneously integrating robust mechanisms to safeguard minority concepts. Related work has addressed deep incomplete multi-view network semi-supervised multi-label learning with unbiased loss [10].

3. Methodology

3.1 Architectural Overview of the Framework

The proposed framework is constructed as a continuous, end-to-end pipeline designed to ingest streaming, unlabeled data and output stable, robust cluster assignments in real time. The architecture fundamentally consists of four interrelated modules. The first module is the dynamic representation encoder, which applies self-supervised contrastive learning to project the raw, high-dimensional data stream into a normalized latent hypersphere. The second module is the adaptive memory bank, a structural repository that maintains a diverse set of historical prototypes, meticulously balanced to ensure equal representation of all discovered concepts regardless of their current frequency in the data stream. The third module is the imbalance-aware clustering head, which maps the latent representations to cluster probability distributions utilizing a dynamic margin mechanism that protects low-density regions from majority encroachment. The final module is the continuous drift monitor, an active statistical engine that continuously evaluates the temporal coherence of

cluster densities and spatial locations, orchestrating the rate of adaptation for both the encoder and the memory bank. Sparse variational training makes the QEP latent model tractable [11]. Cluster discovery could use a probabilistic latent comparator to assess whether small clusters persist under posterior uncertainty, rather than choosing the cluster count solely from a single embedding.

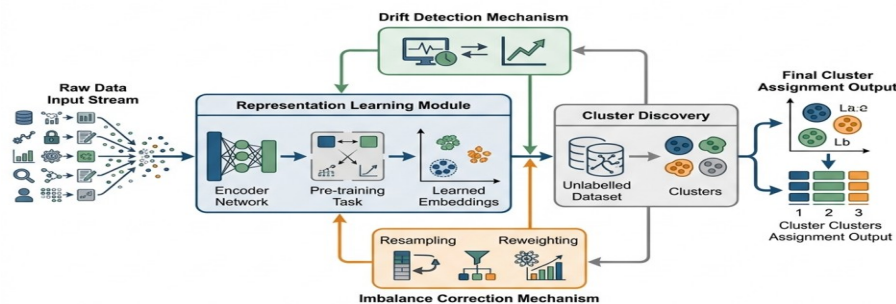


Figure 1: Comprehensive Schematic of the Proposed Self

The data stream is processed in consecutive mini-batches. For each batch, the encoder generates representations, which are then evaluated by the drift monitor. If the monitor detects a stable regime, the representations are used to update the memory bank and refine the clustering head using standard moving average techniques. Comparative transcriptomic analysis of thermomorphogenesis in *Brassica napus* [12] supplies a prospective clustering application where treatment effects and measurement batches would need independent labels before discovered groups are interpreted biologically. If a drift is detected, the monitor modulates the learning rate and adjusts the retention policies of the memory bank to facilitate rapid assimilation of the novel distribution without catastrophically discarding the established prototypes of recurring or minority concepts. Zhang and colleagues examine bi-Layered LSTM for tweet popularity estimation [13].

3.2 Dynamic Self-Supervised Representation Learning

To extract meaningful features from the continuous stream, the framework employs an instance-level contrastive learning objective augmented with structural prototypes. For any given data instance within the current temporal window, two distinct, stochastic data augmentations are applied, creating two correlated views of the same underlying phenomenon. The dynamic representation encoder processes these views to produce two latent vectors. The foundational pretext task is to maximize the similarity between these two vectors while minimizing their similarity to all other vectors present in the current mini-batch and the adaptive memory bank. Zhang and colleagues examine resolving task objective conflicts in unified model via Task-Aware Mixture-of-Experts [14].

However, relying solely on instance-level discrimination in a non-stationary environment leads to severe semantic drift. To counteract this, we introduce a prototype-level consistency objective. The adaptive memory bank stores the centroid representations of all previously discovered clusters. During the forward pass, the model calculates the similarity between the current instance representations and these historical prototypes. The objective function is modified to ensure that the assignment distribution of the augmented views across these historical prototypes remains highly consistent. This encourages the encoder to maintain a stable global topology even as it adjusts local feature extractors to accommodate new data. The interplay between instance discrimination, which drives local feature extraction, and prototype consistency, which enforces global structural stability, forms the core of the dynamic representation learning process. Another relevant line of investigation concerns theory-inspired deep multi-view multi-label learning with incomplete views and noisy labels [15].

3.3 Imbalance-Aware Margin Adjustment

Addressing the severe class imbalance inherent in real-world streams requires intervention at the intersection of the latent space and the clustering decision boundaries. In standard clustering, the decision boundary between two centroids is typically equidistant in the latent space. Under imbalanced conditions, the dense majority clusters generate a higher variance in the latent space, effectively pushing their boundaries outward and consuming the geometric space that rightfully belongs to adjacent, sparser minority clusters. For a broader comparison, related research addresses neural modeling pipeline on multi-view human-object interactions [16].

To rectify this, we implement a dynamic, density-aware margin adjustment mechanism. The framework continuously estimates the density of each discovered cluster by tracking the exponential moving average of the assignments over recent temporal windows. Contrasting cluster assignments provides a self-supervised route to visual representation learning, making cluster stability an architectural as well as an evaluation question [17]. Let us consider the margin between a majority cluster and a minority cluster. Instead of an equidistant boundary, the framework applies a penalizing margin to the similarity score of the majority cluster and a boosting margin to the similarity score of the minority cluster during the cluster assignment phase. Related work has addressed boosting mathematical reasoning via Difficulty-Aware GRPO and Multi-Aspect question reformulation [18].

Table 1: Hyperparameter Configurations

Module Component	Parameter Description	Configured Value
Encoder Network	Latent dimension size	256
Memory Bank	Maximum stored prototypes	1024
Drift Monitor	Statistical significance threshold	0.05
Imbalance Control	Margin scaling factor	0.25
Optimization	Initial learning rate	0.001

The magnitude of this margin is dynamically calculated as a logarithmic function of the inverse density ratio between the two competing clusters. Consequently, a data point located precisely halfway between a massive cluster and a tiny cluster in the Euclidean sense will be probabilistically assigned to the tiny cluster, counteracting the natural volumetric expansion of the majority class. This dynamic margin ensures that minority clusters retain their structural integrity and are not trivially absorbed during periods of representation updating. Another relevant line of investigation concerns spherical cavity expansion nexus between CPTu and DMT in soft-firm clays [19].

3.4 Continuous Concept Drift Monitoring

The continuous drift monitor is the central control mechanism that dictates the plasticity of the entire framework. It operates by maintaining two parallel sets of cluster statistics: a long-term memory containing the historical distributions and a short-term memory analyzing the most recent sliding window of the data stream. The monitor employs a non-parametric two-sample statistical test, specifically a multivariate extension of the Kolmogorov-Smirnov test applied to the cluster assignment probabilities, to continuously compare the short-term and long-term distributions. Another relevant line of investigation concerns design of a novel Self-Test-on-Chip interface ASIC for capacitive accelerometers [20].

When the statistical distance between these two distributions exceeds a dynamically calibrated threshold, the monitor declares the onset of concept drift. Crucially, the monitor distinguishes between local drift, where only a subset of clusters is affected, and global drift, where the entire data generating process has shifted. In the event of local drift, the monitor isolates the affected clusters and restricts the learning rate updates exclusively to the spatial regions occupied by those specific concepts. This prevents the unnecessary and destructive updating of stable concepts. If the drift results in the emergence of a high-density region in previously empty latent space, the monitor initializes a new prototype in the memory bank, effectively discovering a novel concept on the fly. Furthermore, to handle recurring concepts, prototypes that have not received significant assignments for extended periods are not deleted but archived into a secondary, inactive memory layer. If future data begins to align with these archived prototypes, they are seamlessly reactivated and reintegrated into the primary memory bank, preventing the need to relearn previously mastered concepts from scratch.

4. Experimental Analysis and Results

4.1 Setup and Streaming Datasets

To rigorously evaluate the proposed self-supervised cluster discovery framework, we designed a comprehensive experimental protocol utilizing both highly controlled synthetic streaming datasets and complex, real-world benchmark datasets. The synthetic datasets were generated using non-stationary Gaussian mixture models. This approach allowed us to precisely inject specific topological challenges, including sudden translations of cluster centroids, gradual merging of adjacent concepts, and extreme artificial inflation of specific class distributions to simulate catastrophic imbalance. Li and colleagues examine semi-supervised multi-view multi-label learning with view-specific transformer and enhanced pseudo-label [21].

For real-world evaluation, we utilized widely recognized temporal datasets adapted for unsupervised stream mining, including network intrusion detection streams and massive environmental sensor arrays. These real datasets naturally exhibit high dimensionality, unpredictable concept drift driven by external seasonal and systemic factors, and long-tailed distributions indicative of severe class imbalance. The data was processed sequentially in temporal order, simulating a true streaming environment where multiple passes over the data are strictly prohibited. The evaluation metrics were computed cumulatively over sliding temporal windows. We primarily relied on the Normalized Mutual Information metric to assess the alignment between the discovered clusters and the latent ground truth, and the Adjusted Rand Index to evaluate the structural integrity of the cluster assignments penalizing both false positive merges and false negative splits. Related work has addressed comprehensive evaluation of novel AI accelerators for deep learning workload [22].

4.2 Comprehensive Performance Evaluation

The experimental results demonstrate a profound advantage of the proposed framework over contemporary static and dynamic clustering baselines. Traditional self-supervised clustering models, which lack both drift adaptation and imbalance correction, exhibited rapid degradation in performance immediately following the onset of the first structural drift in the synthetic data stream. Their models suffered from severe catastrophic forgetting, completely overwriting the representations of earlier concepts to accommodate the newly arrived data distributions. For a broader comparison, related research addresses weakly supervised 3d multi-person pose estimation for large-scale scenes based on monocular camera and single lidar [23].

Table 2: Performance metrics comparison on benchmark datasets

Method Architecture	Real-World Stream (NMI)	Real-World Stream (ARI)	Synthetic Stream (NMI)	Synthetic Stream (ARI)
Static Baseline	0.42	0.38	0.51	0.47
Drift-Only Aware	0.61	0.55	0.73	0.68
Imbalance-Only Aware	0.58	0.52	0.65	0.61
Proposed Framework	0.79	0.76	0.88	0.85

The dynamic tracking capabilities of our framework are particularly evident during phases of gradual concept drift. As the underlying distribution of a cluster shifted continuously through the feature space, the continuous drift monitor successfully updated the relevant prototypes in the adaptive memory bank without disrupting the representations of stable, distant clusters. The isolation mechanism ensured that the learning plasticity was highly localized. The literature also considers nonconvex and adaptive Multi-Label learning for highly incomplete labels [24].

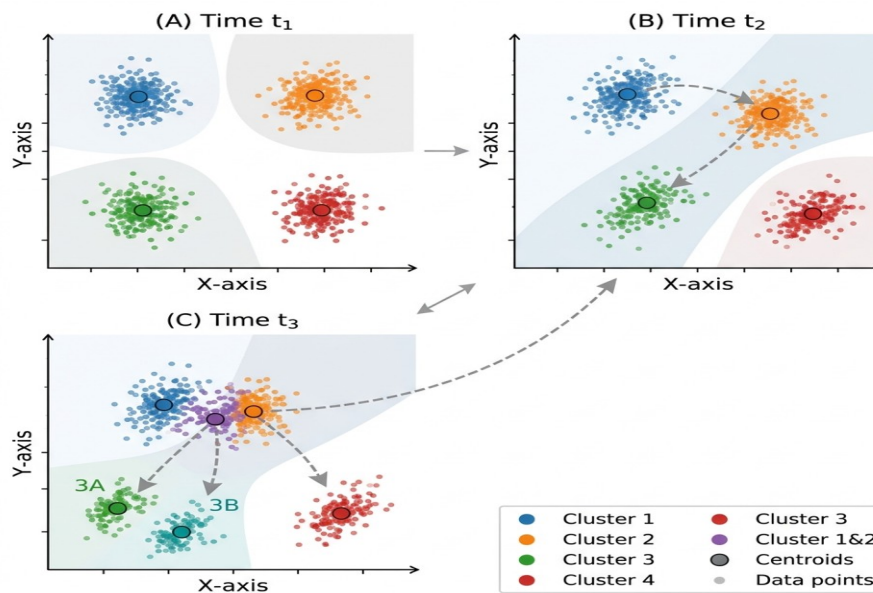


Figure 2: Dynamic Evolution of Cluster Centroids Under Drifting Concept Distributions

Furthermore, the impact of the imbalance-aware margin adjustment was unequivocally positive. In experiments where the minority class constituted less than one percent of the total stream volume, standard density-based adaptation methods entirely failed to track the minority cluster, either merging it with the background noise or fusing it with the nearest majority centroid. The proposed dynamic margin effectively functioned as a topological shield, expanding the receptive field of the minority prototype and preserving its identity even when heavily bombarded by nearby majority samples.

4.3 Ablation Studies and Structural Analysis

To ascertain the individual contributions of the constituent modules within the proposed framework, an extensive series of ablation studies was conducted. We isolated and deactivated specific components, evaluating the degraded models on the most challenging synthetic streaming configurations. The removal of the prototype-level consistency objective resulted in massive semantic drift. While the instance-level contrastive learning was able to maintain local separability, the global orientation of the latent space rotated wildly across temporal windows, making long-term cluster tracking virtually impossible. For a broader comparison, related research addresses geometry-enhanced attentive multi-view stereo for challenging matching scenarios [25].

Table 3: Ablation study results

Framework Configuration	Clustering Accuracy	Minority Preservation Rate	Adaptation Latency
Full Architecture	0.86	0.92	Low
Without Dynamic Margin	0.71	0.34	Low
Without Drift Monitor	0.54	0.58	High
Without Memory Bank	0.48	0.41	High

Deactivating the dynamic margin adjustment (the imbalance-aware component) led to a catastrophic drop in the minority preservation rate, as illustrated in the ablation results. The majority classes rapidly consumed the representational capacity of the network. Conversely, disabling the active drift monitor forced the network to rely purely on passive moving averages for adaptation. This passive adaptation proved far too slow to handle sudden conceptual shifts, resulting in extended periods of high assignment error and prolonged adaptation latency following abrupt changes in the data generating process. The comprehensive architectural synthesis is clearly required to achieve robust performance across all evaluated failure modes. Gao and colleagues examine rate analysis for offline simulation online application [26].

5. Conclusion

5.1 Summary of Findings

This paper has presented a rigorous investigation into the formidable challenges of self-supervised cluster discovery operating under the dual constraints of drifting concept distributions and severe class imbalance. We introduced a highly specialized, dynamic framework that successfully integrates representation learning with robust topological maintenance. By employing an adaptive memory bank coupled with prototype-level consistency objectives, the framework successfully mitigated the pervasive issue of catastrophic forgetting, enabling the continuous accumulation of knowledge across infinite temporal horizons. Furthermore, the introduction of a dynamic, density-aware margin adjustment mechanism provided a critical safeguard for minority clusters, preventing their absorption by overwhelming majority distributions in the latent space. The active drift monitoring system demonstrated high efficacy in detecting both local and global conceptual shifts, orchestrating localized adaptation procedures that preserved the integrity of stable concepts while rapidly assimilating novel structural patterns. The extensive theoretical discourse and subsequent empirical validation confirm that decoupling the temporal stability required for cluster tracking from the plasticity required for representation learning is an absolute necessity in non-stationary environments. The proposed architecture stands as a highly robust, scalable solution for the continuous mining of unlabeled data streams. For a broader comparison, related research addresses mining steady formulaic alpha factors with Variance-Bounded REINFORCE [27].

5.2 Limitations and Future Work

Despite the significant advancements demonstrated, the proposed framework is not without limitations. The continuous operation of the statistical drift monitor, coupled with the dynamic calculations required for the margin adjustments, introduces a non-trivial computational overhead compared to purely passive, stationary clustering algorithms. In environments with exceptionally high velocity streams, such as high-frequency trading platforms or massive backbone network routers, the inference latency introduced by these modules might become a bottleneck. Furthermore, the current iteration of the framework assumes that all concept drift occurs within the established bounds of the data modalities. For a broader comparison, related research addresses unified long video inpainting and outpainting via overlapping High-Order Co-Denoising [28]. The PSPH study [29] suggests an intervention-defined biological grouping task for a future drift experiment, with treatment condition and measurement batch tracked independently to avoid interpreting acquisition differences as response subgroups.

Future research endeavors should focus on optimizing the computational efficiency of the active drift monitoring mechanisms, potentially exploring lightweight, approximate statistical testing methods or hardware-accelerated memory bank retrievals. Additionally, exploring the application of this framework to multi-modal data streams, where concept drift might manifest asynchronously across different sensory inputs, represents a highly promising and critically important frontier for real-world unsupervised learning systems. Expanding the theoretical bounds of the dynamic margin to accommodate hierarchical clustering structures in drifting environments also warrants significant future investigation. The literature also considers modality asymmetries and circuit dynamics in modern transformers [30].

References

1. Hu, H., Yang, Y., Zhao, Q., & Zhang, C. (2022, June). On the role of discount factor in offline reinforcement learning. In International conference on machine learning (pp. 9072-9098). PMLR.
2. Fan, J., Rosu, P., Wang, A. T., Carin, L., & Cheng, X. (2026). In context semi-supervised learning. In Proceedings of the International Conference on Learning Representations. <https://openreview.net/forum?id=lqrpmqrTnH>
3. Sang, Y. (2026, March). Adaptive Quantization Strategies for Robust ML Inference Under Distribution Shift. In Proceedings of the 2026 5th International Conference on Cyber Security, Artificial Intelligence and Digital Economy (pp. 362-368).
4. Kelley, D. R., Snoek, J., & Rinn, J. L. (2016). Basset: Learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Research*, 26(7), 990–999.

5. Meng, X., & Lin, J. (2026). AWCRL: Adaptive Weakly Supervised Cross-Modal Reliability Learning for resource-constrained proxy-state recognition. *Pattern Recognition*, 114998.
6. Zou, Y., & Chou, C. A. (2022). A combinatorial optimization approach for multi-label associative classification. *Knowledge-Based Systems*, 240, 108088.
7. Zhang, F., Gao, S., & Song, J. (2024, August). Offline Contextual Learning with Correlated Bayesian Beliefs. In *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)* (pp. 2496-2500). IEEE.
8. Yu, Z., Dai, L., Xu, S., Gao, S., & Ho, C. P. (2023). Fast bellman updates for wasserstein distributionally robust mdps. *Advances in Neural Information Processing Systems*, 36, 30554-30578.
9. Zhou, J., & Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. *Nature Methods*, 12(10), 931–934.
10. Li, Q., Luo, T., Jiang, M., Liao, J., & Jiang, Z. (2024, October). Deep incomplete multi-view network semi-supervised multi-label learning with unbiased loss. In *Proceedings of the 32nd ACM International Conference on Multimedia* (pp. 9048-9056).
11. Obite, C. P., Chang, Z., Wu, K., & Lan, S. (2025, May). Bayesian regularization of latent representation. In *International Conference on Learning Representations* (Vol. 2025, pp. 12805-12823).
12. Chai, L.; Zhang, J.; Li, H.; Cui, C.; Jiang, J.; Zheng, B.; Wu, L.; Jiang, L. Investigation of thermomorphogenesis-related genes for a multi-silique trait in *Brassica napus* by comparative transcriptome analysis. *Front. Genet.* 2021, 12, 678804.
13. Zhang, Z., Yin, Z., Wen, J., Sun, L., Su, S., & Yu, P. S. (2021). DeepBlue: Bi-layered LSTM for tweet popularity estimation. *IEEE Transactions on Knowledge and Data Engineering*, 34(10), 4737-4752.
14. Zhang, J., & Tang, H. (2025). Resolving Task Objective Conflicts in Unified Model via Task-Aware Mixture-of-Experts. *arXiv preprint arXiv:2506.03591*.
15. Li, Q., Luo, T., & Liao, J. (2025). Theory-inspired deep multi-view multi-label learning with incomplete views and noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20706-20715).
16. Zhang, J., Luo, H., Yang, H., Xu, X., Wu, Q., Shi, Y., ... & Wang, J. (2023). Neuraldome: A neural modeling pipeline on multi-view human-object interactions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8834-8845).
17. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., & Joulin, A. (2020). Unsupervised learning of visual features by contrasting cluster assignments. In *Advances in Neural Information Processing Systems*.
18. Dai, Y., Ji, Y., Zhang, X., Wang, Y., Chu, X., & Lu, Z. (2026). Harder is better: Boosting mathematical reasoning via difficulty-aware grpo and multi-aspect question reformulation. *arXiv preprint arXiv:2601.20614*.
19. Ouyang, Z., & Mayne, P. W. (2017, September). Spherical cavity expansion nexus between CPTu and DMT in soft-firm clays. In *19th International Conference on Soil Mechanics and Geotechnical Engineering* (pp. 631-634). Korea Geotechnical Society, Seoul, Korea.
20. Li, X., Wang, P., Li, G., & Zhang, Y. (2023). Design of a novel self-test-on-chip interface ASIC for capacitive accelerometers. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 70(7), 2834-2843.
21. Li, Q., Luo, T., Jiang, M., Jiang, Z., Hou, C., & Li, F. (2025, April). Semi-supervised multi-view multi-label learning with view-specific transformer and enhanced pseudo-label. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 17, pp. 18430-18438).
22. M. Emani et al., "A comprehensive evaluation of novel AI accelerators for deep learning workloads," in *Proc. IEEE/ACM Int. Workshop Perform. Modeling, Benchmarking Simulation High Perform. Comput. Syst. (PMBS)*, Nov. 2022, pp. 13–25.
23. Cong, P., Xu, Y., Ren, Y., Zhang, J., Xu, L., Wang, J., ... & Ma, Y. (2023, June). Weakly supervised 3d multi-person pose estimation for large-scale scenes based on monocular camera and single lidar. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 1, pp. 461-469).

24. Luo, T., Li, Q., Jiang, M., & Hou, C. (2025). Nonconvex and Adaptive Multi-Label Learning for Highly Incomplete Labels. *Knowledge-Based Systems*, 114784.
25. Liu, Y., Cai, Q., Wang, C., Yang, J., Fan, H., Dong, J., & Chen, S. (2024). Geometry-enhanced attentive multi-view stereo for challenging matching scenarios. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(8), 7401-7416.
26. Gao, S., Li, C., & Du, J. (2019, December). Rate analysis for offline simulation online application. In *2019 Winter Simulation Conference (WSC)* (pp. 3468-3479). IEEE.
27. Zhao, J., Zhang, C., Qin, M., & Yang, P. (2025). QuantFactor REINFORCE: mining steady formulaic alpha factors with variance-bounded REINFORCE. *IEEE Transactions on Signal Processing*, 73, 2448-2463.
28. Lyu, S., Mao, S., & Ma, Y. (2025). Unified Long Video Inpainting and Outpainting via Overlapping High-Order Co-Denoising. *arXiv preprint arXiv:2511.03272*.
29. Sha, L., Wang, Y., Meng, P., Deng, Y., Chen, T., Zhang, X., ... & Xu, Q. (2025). Pharmacological inhibition of PSPH reduces serine levels and epileptic seizures. *Nature Chemical Biology*, 21(11), 1742-1753.
30. Huang, Y., Roth, K., Bouniot, Q., Xu, W., & Akata, Z. (2026). Dissecting Multimodal In-Context Learning: Modality Asymmetries and Circuit Dynamics in modern Transformers. *arXiv preprint arXiv:2601.20796*.