Open-world continual learning via knowledge transfer

Li, Y.

Chapter 8

Conclusions and Future Work

This thesis is an investigation of **Open-world Continual Learning**, an emerging paradigm at the intersection of continual learning and open-set recognition, including out-of-distribution detection. By formalizing open-world continual learning with a mathematical definition and identifying four real-world scenarios, we have laid a foundational framework in this domain. Our empirical analyses reveal the unique challenges of open-world continual learning, particularly its inherent open risks and the dynamic interplay between sequential tasks, distinguishing it from conventional deep learning approaches.

From a technical perspective, we have pioneered a knowledge transfer-centric approach to open-world continual learning, addressing aspects such as knowledge representation, known-unknown boundary discovery, and forward/backward transfer mechanisms. Our proposed models *Pro-KT* and *PEAK* demonstrate the effectiveness of this perspective in both large-scale and few-shot learning scenarios. Furthermore, we introduced *HoliTrans* to provide a unified and theoretically based framework that integrates open-set detection with continual learning, offering a holistic solution to open-world continual learning challenges.

Beyond theoretical contributions, we have validated the practical relevance of open-world continual learning through a real-world application in cross-regional fraud detection. By introducing *CCL*, a continual graph learning model, we address the critical issue of distributional shifts in fraud patterns across regions, demonstrating the adaptability of open-world continual learning principles in complex, evolving environments.

Collectively, the four research questions explored in this thesis form an evolving narrative of open-world continual learning. *Pro-KT* (RQ1) initiates the study by establishing prompt-enhanced knowledge transfer for maintaining stability and adaptability. *PEAK* (RQ2) advances this direction under few-shot constraints, enhancing robustness with limited labeled data. *HoliTrans* (RQ3) unifies open-set recognition and continual learning into a single theoretical and architectural framework, addressing knowledge transfer between known and unknown samples. Finally, *CCL* (RQ4) grounds these principles in a real-world scenario, i.e., cross-regional fraud detection, showcasing the framework's practical impact. Together, these contributions chart a consistent progression from theoretical foundations to holistic methodology and applied validation.

This chapter summarizes the main contributions of this thesis and reflects on their implications for the field of open-world continual learning. We then discuss open challenges and propose future research directions to guide further progress in this evolving paradigm.

8.1 Summary of Contributions

In Chapter 1, we formulated four key research questions. We now revisit them to summarize this thesis's contributions to each.

[RQ1] *Can we develop an effective open-world continual learning model to accumulate the knowledge gained in the past and to use the knowledge to help divide the test data into knowns and unknowns/opens, especially for enlarging the open-set boundary between knowns and unknowns?*

This thesis addressed the stability-plasticity trade-off in open-world continual learning through a novel perspective on knowledge representation and transfer. We proposed *Pro-KT* in Chapter 4, a prompt-learning approach that encodes both task-specific and task-agnostic knowledge into a unified repository, enabling dynamic knowledge transfer across tasks while maintaining clear boundaries between known and unknown classes. Our experiments demonstrated that *Pro-KT* achieves superior performance in both knowledge retention (stability) and open-set detection (plasticity), while remaining practically deployable as a plug-and-play framework for pre-trained models. These results not only establish new benchmarks for open-world continual learning systems but also provide foundational insights for handling increasingly complex open-world scenarios where knowledge must evolve continuously yet robustly.

[RQ2] *How can we enhance the performance of existing open-world continual learning models under the constraint of limited labeled data by employing an improved knowledge transfer approach?*

Addressing the critical challenge of data scarcity in open-world continual learning, in Chapter 5 we introduced *PEAK*, the first dedicated framework for few-shot open-world continual learning scenarios, where data is often sparse and tasks are dynamic. Building on the foundation of Chapter 4, we identified three fundamental obstacles that emerge when learning from limited data: representation collapse under sparse data distributions, ambiguous decision boundaries for open-set detection, and degradation of knowledge transfer across tasks. *PEAK*'s architecture combines three novel components specifically designed to overcome these challenges simultaneously. Experimental results demonstrated consistent superiority over both state-of-the-art few-shot learning and open-world continual learning baselines, validating *PEAK* dual capability for robust open-set recognition and continual adaptation.

More importantly, *PEAK* establishes a methodological bridge between theoretical open-world continual learning research and practical applications where data is inherently scarce and learning requirements evolve continuously. In medical diagnosis, for instance, the ability to learn from sparse labeled examples while detecting novel pathologies would address the critical need for adaptive systems in rare disease identification. Similarly, in fraud detection, where new attack patterns emerge with minimal labeled samples, robust knowledge transfer would prevent catastrophic forgetting while flagging unseen threats. Autonomous systems operating in dynamic environments, such as disaster-response robots, would benefit from few-shot adaptation to novel objects without losing prior knowledge, a capability equally important for cybersecurity (zero-day attack detection) and biodiversity monitoring when tracking rare or even unknown species. Even in personalized AI assistants, where user preferences shift incrementally with limited feedback, it is important to balance stability and plasticity. By tackling the real-world challenges of data scarcity, open-set recognition, and non-stationary task distributions, *PEAK* opens new possibilities for deployment in resource-constrained environments.

[RQ3] *How can we leverage knowledge from both known and unknown samples to enable knowledge transfer in open-world continual learning, effectively improving open-set detection with incremental classification at both theoretical and architectural levels?*

In Chapter 6, we presented a transformative reconceptualization of open-world continual learning, improving on the limitations of existing approaches that merely combine continual learning and out-of-distribution detection techniques. Building on the foundations established in Chapter 4 and Chapter 5, we first exposed a critical shortcoming in current methods: their reliance on the flawed orthogonal decomposition assumption, which artificially separates open-set recognition from incremental classification. Through an analysis of four open-world continual learning scenarios, we demonstrated how this oversimplification leads to failures in capturing the intricate interactions and structural challenges inherent to open-world continual learning.

Our theoretical contribution in Chapter 6 started with a novel problem formulation that captures the intrinsic coupling between open-set dynamics and continual task arrival. We proposed a unified decision space for open-world continual learning and derived, through theoretical analysis, the key optimization factors that govern effective learning in open-world environments. These theoretical insights were used for the development of *HoliTrans*, an architectural solution that integrates nearest class mean principles into open-world continual learning. The elegance of *HoliTrans* lies in its dual capacity: it simultaneously achieves state-of-the-art empirical performance across all four open-world continual learning scenarios while providing provable guarantees on decision boundary stability and knowledge preservation.

By bridging theoretical insights with practical implementation, we established a new foundation for developing robust learning systems capable of operating in truly open-ended environments. We demonstrated the success of *HoliTrans* across diverse experiments, showing its potential to address adaptive yet stable learning, with applicability from evolving cybersecurity threats to dynamic medical diagnostic systems.

[RQ4] *How can we improve the model performance by formulating practical problems with an open-world continual learning framework and developing effective models for challenging real-world applications?*

In Chapter 8, we applied our theoretical and algorithmic results on open-world continual learning to the financial domain of cross-regional fraud detection. By formulating this task as an incremental learning problem, we introduced *CCL*, a framework that addresses two challenges: adapting to region-specific fraud patterns while preserving knowledge across geographically shifting distributions. The integration of prototype-based knowledge replay, parameter smoothing, and heterogeneous graph modeling enabled CCL to achieve simultaneous forward adaptability and backward stability, a balance that is missing in conventional fraud detection systems.

Experimental results demonstrated that *CCL* reduced catastrophic forgetting when compared to baseline methods while improving cross-regional knowledge transfer accuracy. These findings not only validate the practical utility of open-world continual learning principles but also reveal how knowledge-centric designs can overcome distributional shifts in real-world scenarios. More broadly, this part of our work shows that the theoretical frameworks developed earlier in the thesis translate effectively to applications where data evolves continuously.

Note. All methods proposed in this thesis were released as open-source software. We believe that open release promotes reproducibility, enables practical deployment, and provides an inspiration for further research and practical applications.

8.2 Future Work

Building upon the contributions of this thesis, we identify several promising directions to advance open-world continual learning under modern deep learning paradigms.

a) Modular OWCL with Sparse Expert LLMs

The rapid evolution of sparse Mixture-of-Experts (MoE) language models, such as Mixtral, GLaM, and DBMoE, presents a transformative opportunity for open-world continual learning. While these models have demonstrated scalability in static NLP tasks, their potential for dynamic, open-ended learning is still largely unexplored.

Current open-world continual learning methods face two critical limitations when scaled to modern LLMs: monolithic architectures hinder efficient knowledge reuse, and full fine-tuning becomes computationally prohibitive. Although MoE-based LLMs provide a modular structure that can be exploited for dynamic task specialization, it is unclear how expert activation, routing strategies, and parameter sharing behave under sequential and open-world task distributions.

To address these challenges, future research could explore the design of a paradigm where MoE-based LLMs serve as modular knowledge repositories, with task-conditioned gating mechanisms dynamically routing known inputs and emerging tasks to specialized experts. By combining prompt routing with task-adaptive gating networks, it should be possible to selectively activate subsets of experts, reducing computation time while promoting transferability. Additionally, incorporating temporal-aware gating mechanisms, such as dynamic routing (as in DRMoE) or hierarchical gating (as in HyperMoE), may enhance the system's adaptability to evolving task sequences and distribution shifts.

Another promising direction is in the integration of MoE architectures with parameter-efficient adaptation techniques. By combining dynamic expert routing with adapter tuning or expert masking strategies, we could achieve both flexible knowledge organization and computationally sustainable continual learning. This hybrid approach would enable on-demand expert reconfiguration for novel tasks, preservation of core competencies without catastrophic forgetting, and elimination of the need for expensive data replay. The development of such architectures should be accompanied by novel evaluation frameworks, including quantitative metrics for expert utilization diversity (measuring how effectively specialized knowledge is deployed) and transfer stability (assessing consistency of performance across task sequences). These metrics would provide quantitative ways to evaluate the efficiency and robustness of modular knowledge routing under open-world continual learning regimes.

This direction, bridging sparse LLM architectures with open-world continual learning, could yield systems capable of incremental specialization, where new experts are added for novel tasks while maintaining access to foundational knowledge. Such architectures would offer scalable, adaptable, and interpretable solutions to many real-world applications.

b) OWCL under Heterogeneous Multimodal Environments

Another underexplored direction involves extending open-world continual learning to handle heterogeneous task streams with multimodal data distributions. Current open-world continual learning frameworks mainly operate under the limiting assumption of homogeneous, unimodal task sequences, a condition rarely satisfied in practical deployment scenarios. While some recent works in continual learning have started to explore multimodal task streams, to the best of our knowledge, there has been no dedicated research within the open-world continual learning setting addressing heterogeneity and multimodality (e.g., visual, textual, graph-structured) with divergent statistical properties and annotation protocols.

The main challenge is in developing architectures capable of maintaining stable knowledge transfer across modality boundaries while preserving sensitivity to open-set novelty within each data domain. Potential solutions may involve integrating cross-modal prompt embeddings with multimodal transformers to facilitate shared representations across tasks involving text, vision, and graph-based data. Graph-structured memory modules or modality-aware task encoders can further assist in modeling the structural relationships between tasks and promoting generalization.

Another challenge is the mitigation of modality interference and ensuring stable transfer across heterogeneous tasks. Adaptive task selection mechanisms, curriculum-based training strategies, or graph-guided expert activation may help address this issue. Equally important is the development of evaluation methodologies for multimodal open-world continual learning systems. Current benchmarks fail to capture essential dimensions such as cross-modal interference effects, modality-specific catastrophic forgetting rates, and open-set detection performance across heterogeneous data types. New metrics and testing protocols should account for these composite challenges introduced by simultaneous modality shifts and task boundary ambiguity.

This research direction carries significant implications for domains where multimodal data evolves continuously, including clinical decision support systems integrating imaging and electronic health records, or autonomous platforms processing sensor data in non-stationary environments. The insights gained from studying heterogeneous open-world continual learning would be of interest to the broader machine learning community because of the recent interest in multimodal representation learning.

c) Scalable and Privacy-Preserving OWCL

The increasing deployment of large-scale machine learning models necessitates the development of parameter-efficient adaptation strategies that support continual learning while maintaining computational efficiency. Current approaches based on full model fine-tuning are impractical for open-world continual learning scenarios due to their substantial computational and storage requirements. Recent advances in parameter-efficient fine-tuning techniques, including low-rank adaptation, AdapterFusion, and Prefix Tuning, may be promising alternatives to investigate within the open-world continual learning paradigm.

Future research on integrating parameter-efficient fine-tuning mechanisms with prompt-based learning architectures could achieve both efficient task adaptation and stable knowledge retention. Potential solutions may include dynamic growing adapter modules, selectively routing adapters based on task identity, or compressing knowledge into modular parameter blocks for knowledge consolidation. The theoretical properties of such hybrid architectures, as well as their capacity bounds and forgetting dynamics, are important open questions.

The combination of parameter-efficient fine-tuning methods with federated open-world continual learning frameworks presents another promising research direction where models are updated across distributed clients under data privacy constraints. By implementing parameter-efficient fine-tuning's lightweight update mechanisms, it should be possible to enable continual knowledge sharing across edge nodes without transmitting sensitive data. In particular, designing parameter-efficient fine-tuning-compatible synchronization protocols and communication-efficient update strategies will be crucial for scalable and secure open-world continual learning under federated settings. Furthermore, exploring how expert modularity or adapter routing can be personalized across users or client populations offers a compelling direction for federated personalization in open-world contexts.

By uniting the principles of open-world continual learning, parameter-efficient fine-tuning, and federated learning, this research direction has particular relevance for application domains requiring both continuous adaptation and strict data privacy in a decentralized environment, such as healthcare frameworks, industrial IoT systems, and financial services.

d) Explainability and Trustworthiness in OWCL

The development of explainable and trustworthy learning systems represents an important challenge in open-world continual learning, particularly for deployment in safety-critical domains. Current open-world continual learning methodologies predominantly focus on improving task performance metrics while not considering model interpretability and decision transparency. This is increasingly problematic as open-world continual learning systems encounter novel inputs and evolving task distributions, but understanding model behavior is often essential for practical adoption.

The interaction between continual adaptation mechanisms and open-set recognition components creates complex dynamics that existing interpretation methods are not yet able to analyze. Also, the dynamic evolution of decision boundaries and feature representations in continual learning scenarios lacks any formal frameworks for temporal analysis.

Future research should develop theoretical approaches to explainability that address these OWCL-specific challenges. Promising directions include the incorporation of explainability tools such as attention entropy visualization, gating path trace analysis, or counterfactual generation to present a decision-making process. By identifying which experts or prompts are activated under different tasks, one could gain insights into the model's internal reasoning and evaluate the consistency of knowledge reuse. Causal inference techniques may also play a role in uncovering the underlying factors driving prediction shifts or classification errors. Finally, visual explanations that highlight influential features or task-specific representations can enhance interpretability and inform model debugging.

As for many deep learning systems, ultimately, explainability-enhanced open-world continual learning systems will bridge the gap between black-box model behavior and human understanding, allowing humans in the loop, interactive learning, and regulatory transparency needed, for example, in healthcare diagnostics, autonomous systems, and financial risk assessment.

e) Meta-Memory and Adaptive Sampling for Long-Horizon OWCL

In real-world dynamic and evolving environments, open-world continual learning systems must go beyond passive adaptation and develop mechanisms for strategic memory management and sample prioritization. One promising direction is to integrate meta-memory architectures and adaptive sampling strategies to guide long-term knowledge retention and continual adaptation under constrained resources.

Meta-memory architectures offer a promising solution by maintaining abstracted, task-generalizable knowledge representations that support efficient retrieval during subsequent learning phases. These architectures differ from conventional memory buffers by enabling selective access to relevant prior knowledge without requiring extensive storage of raw data or brute force replay. When combined with adaptive sampling strategies (for example, guided by task uncertainty, representation novelty, or learning progress), the model can dynamically select the most informative or high-impact samples for fine-tuning, thus improving both data efficiency and generalization.

The unique requirements of open-world continual learning introduce several important considerations for these approaches. First, unlike traditional active learning that assumes access to an external oracle, adaptive sampling in open-world continual learning can be self-supervised or uncertainty-aware, making it particularly valuable in environments where annotations are costly or delayed. Second, the interplay between meta-memory and adaptive sampling requires an architectural design of open-world continual learning as autonomous systems that learn efficiently, prioritize effectively, and operate reliably over extended time horizons. Third, the integration of curriculum learning principles with adaptive sampling could provide systematic control over task difficulty progression while mitigating catastrophic forgetting.

The development of such architectures requires addressing fundamental questions about memory compression, retrieval efficiency, and sample selection criteria in open-world scenarios. Theoretical analysis of these systems' long-term behavior and empirical validation under realistic non-stationary conditions are important open challenges in the field that could find applications in domains characterized by gradual distribution shifts and sparse supervision signals. Examples include personalized education systems where meta-memory could enable long-term student modeling while adaptive sampling would focus on knowledge gaps, and autonomous agents operating in physical environments that could benefit from prioritized experience retention and efficient sample selection for continuous adaptation.