Universiteit Leiden
The Netherlands

Open-world continual learning via knowledge transfer

Li, Y.

Chapter 2

Literature Review

First, we examine continual learning (CL) methods, categorizing them into three principal families according to their mechanisms for preserving and utilizing task-specific knowledge across sequential learning episodes: (1) replay-based methods, which retain and revisit past samples; (2) regularization-based methods, which constrain parameter updates; and (3) structure-based methods, which dynamically expand or isolate model components. For each category, we provide a critical analysis of its respective advantages and fundamental limitations.

Second, we investigate open-set learning (OSL) and out-of-distribution (OOD) detection, beginning with formal definitions to then moving to an evaluation of their core methodologies. This reveals significant gaps in current approaches, particularly regarding their ability to handle incremental knowledge integration and practical deployment scenarios.

Finally, we integrate these perspectives into the emerging paradigm of open-world continual learning (OWCL), which combines the sequential adaptation of continual learning with the novelty detection capabilities of open-set learning. While recent open-world continual learning works demonstrate promising results, our analysis identifies critical shortcomings in the knowledge transfer mechanism and provides insights into the importance and uniqueness of knowledge transfer in open-world continual learning.

2.1 Continual Learning

Driven by the resurgence of neural network research, continual learning and catastrophic forgetting have received considerable attention. Common approaches to addressing these challenges include reducing representation overlap through parameter regularization [21], memory-based approaches that replay either real or synthetic historical samples [41, 42], and architectural solutions employing dual-network configurations [43].

Early continual learning research was often constrained by limited computational resources, restricting experiments to shallow architectures and modestly-sized datasets [44]. Recent empirical analyses have examined how dropout [45] and various activation functions influence forgetting in sequential tasks [46]. Theoretical perspectives have also been explored, especially in task-incremental settings [19]. Recently, there has been a shift towards practical scaling to longer task sequences and larger datasets, utilizing large-scale pre-trained foundational models as backbones with fine-tuning strategies. This evolution reflects the field's progression from constrained laboratory settings to more realistic, large-scale applications.

Next, we categorize recent continual learning methods according to their mechanisms for preserving and utilizing both task-specific and transferable knowledge during incremental learning processes. This taxonomy organizes approaches into three principal categories based on their core operational principles.

Replay-based methods. Replay-based methods address forgetting by retaining and revisiting representative examples from past tasks. Classical approaches, such as iCaRL [42], retrain models on a limited set of past samples alongside new task data. Other works incorporate constrained optimization into the replay process [47, 48]. When actual past samples are unavailable, pseudo-rehearsal via generative models serves as an alternative, though it introduces complexity, including potential mode collapse and difficulties in balancing old and new samples. Replay methods also inherently face data privacy and security challenges, significantly limiting their practical deployment when previous data is inaccessible or restricted by privacy regulations.

Regularization-based methods. Regularization-based methods apply constraints on model updates to retain knowledge across tasks. For instance, Elastic Weight Consolidation (EWC) [21] mitigates forgetting by penalizing significant changes in crucial model weights. Another classical approach, Learning without Forgetting (LwF) [49], utilizes knowledge distillation techniques to preserve previous task knowledge. Recent work further explores knowledge distillation strategies [50], though these methods remain highly dependent on task similarity. Consequently, a growing body of research has begun addressing the robustness of continual learning models against varying task distributions [51].

Structure-based methods. Structure-based methods address continual learning by expanding the model's architecture to accommodate new tasks, preserving previously learned knowledge [52, 53]. These approaches typically involve either dynamically growing the model structure for new tasks or employing task-specific masks to isolate parameters of previous tasks [54]. However, without careful management, these methods risk saturating model capacity, hindering future task learning. Additionally, given recent advancements in large-scale pre-trained models, strategies

relying heavily on model parameter reuse and storage face significant limitations in scalability, particularly when applied to extensive downstream tasks.

2.2 Open-Set Learning

The conventional closed-world assumption is inadequate for real-world applications, where models must handle unknown samples during deployment while facing restricted access to historical training data. This challenge motivates open-set learning, which addresses two core objectives: (1) accurate classification of known classes and (2) robust rejection of unknown samples during inference [15, 55]. Open-set learning explicitly models the uncertainty inherent in dynamic environments where novel objects frequently appear [8].

Formally, given a training set $D_{tr} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where each y_i belongs to the set C_B of *known classes*, a set u denoting the *unified unknown category*, and an evaluation set $D_{ev} = \{(x'_1, y'_1), (x'_2, y'_2), \ldots, (x'_m, y'_m)\}$, with each $y'_i \in (C_B \cup \{u\})$, open-set learning requires:

- Precise discrimination between known C_B and unknown $\{u\}$ samples;
- No further granularity within the unknown category.

This formulation captures the essential challenge of open-set recognition while maintaining practical applicability.

Recent advances in open-set learning have produced several innovative approaches to handle unknown samples [56, 57, 58]. For example, ORE [55] combines contrastive clustering with energy-based classification for novel object detection. Additionally, [59] develops an open-set classifier using an extreme value theory-based classifier. PointCLIP [60] integrates CLIP and GPT-3 multimodal foundation models for zero-shot open-set tasks, and CEC [61] introduces an open detection framework featuring a dual-component framework with proposal advising and class-specific expulsion. These methods demonstrate the field's progression from basic thresholding to sophisticated architectural solutions.

2.3 Out-of-Distribution Detection

Parallel developments in out-of-distribution detection address the related challenge of identifying samples from non-training distributions. Current algorithms predominantly rely on the independent and identically distributed (IID) assumption, despite prevalent distribution shifts in practical settings. Formally, given training data $D_{tr} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, drawn from a distribution $P_{tr}(X, Y)$ and test samples $x'_i \sim P_{OOD}(X')$, out-of-distribution detection evaluates model confidence to

flag distributional outliers during testing phase. Metrics commonly used for out-of-distribution detection include the model's prediction confidence or certainty, typically measured via softmax probability scores, assessing the correct identification of out-of-distribution samples.

The field has evolved from early post-hoc techniques like ODIN [62], employing temperature scaling and input perturbations as post-hoc adjustments to enhance differentiation between ID and out-of-distribution data, to modern approaches using energy-based scoring [63, 64, 65] with theoretical guarantees [66], enhanced frameworks like Generalized ODIN [67] for covariate shift robustness, extension of the ODIN's framework [67] by adopting specialized training objectives and optimizing hyperparameters, like perturbation magnitude, specifically on ID data, and hierarchical semantic organization of known classes [68] with specialized classifiers such as top-down classification [69] and group softmax training [70] that have been proven very effective within out-of-distribution.

Notably, recent work [19] emphasizes the importance of novelty detection and integrates out-of-distribution detection with continual learning, as exemplified by SOLA [8], a framework that presents a strategy to enhance novelty detection, facilitate on-the-fly task adaptation, and support incremental learning.

Despite the significant progress, existing methods still fall short in adequately addressing the representation and integration of knowledge for effective transfer to future tasks. The presence of open samples further exacerbates these challenges, posing substantial difficulties in knowledge representation and accumulation. Crucially, both open-set learning and out-of-distribution methods classify unknown samples into an *unknown class*, thus preventing models from fully utilizing the informative content within these samples, leading to incomplete exploitation of available knowledge.

2.4 Open-world Continual Learning

Continual learning [21, 49] typically operates under a closed-world assumption, meaning the model presumes that all samples encountered during testing or deployment belong to predefined classes previously seen during training [15, 16, 71]. However, such an assumption is unrealistic for dynamic, real-world environments, where systems frequently encounter novel or previously unseen data [8, 30].

While continual learning and open-set learning address complementary challenges, preserving knowledge across tasks and detecting unknowns, respectively, their integration introduces fundamental tensions. Continual learning methods prioritize stability to mitigate forgetting but assume a closed-world setting, while open-set learning emphasizes plasticity to recognize novelty at the risk of disrupting learned representations. This creates a three-way trade-off: (1) retaining past knowledge

(stability), (2) acquiring new knowledge (plasticity), and (3) dynamically detecting open-set samples without prior exposure. For instance, replay-based continual learning methods may erroneously classify unknowns as known if trained only on historical data [42], while open-set learning detectors struggle when task distributions shift incrementally [20].

Practically, continual learners need to detect, adapt to, and incrementally learn from these new, unknown classes or samples while retaining previously learned information [72, 73]. These conflicting objectives explain why open-world continual learning cannot be trivially reduced to combining existing continual learning and open-set learning techniques, but rather requires novel frameworks to unify stability, plasticity, and open-set robustness. Therefore, effectively detecting and incrementally learning novelties while preserving prior knowledge becomes essential for realistic deployment.

Open-world continual learning extends continual learning to accommodate open-world scenarios. Recently, it has gained significant attention, also because it presents substantial challenges [8, 14]. To enable existing continual learning frameworks to recognize unknown or open samples effectively, initial research in open-world continual learning has integrated methods from open-set recognition and out-of-distribution detection into standard continual learning models. [59] proposed an OSR framework utilizing extreme value theory to manage incremental tasks within dynamic environments. Similarly, [55] introduced a model employing contrastive clustering coupled with energy-based identification techniques [74], enabling the continual learner to identify and integrate novel data.

Building upon these approaches, recent open-world continual learning studies have increasingly emphasized incorporating out-of-distribution detection into continual learning frameworks. For instance, [19, 25, 30] underscored novelty detection as a critical component of open-set learning, suggesting that existing out-of-distribution detection strategies could be effectively integrated within continual learning settings. Furthermore, recent frameworks such as SOLA [8, 75] combine out-of-distribution detection methods with incremental task adaptation, enabling improved novelty detection and task-specific learning in open-world environments.

Nevertheless, despite these advances, current open-world continual learning research remains constrained by three fundamental limitations. First, the absence of standardized problem formulations and evaluation protocols [72, 76] hinders fair comparison across methods. Second, most approaches rely on simplistic integrations of continual learning with out-of-distribution detection techniques [25, 30, 31, 77], failing to address the critical challenge of asymmetric knowledge transfer between known and unknown samples. While existing methods effectively transfer knowledge for known classes (via replay or regularization), they largely ignore the informational value of unknowns, creating both experimental and theoretical gaps: Experimentally, evaluations disproportionately focus on known-class accuracy while lacking benchmarks

to assess whether detected unknowns improve future task learning [73]. Unknown samples are typically discarded rather than analyzed for transferable features (e.g., hierarchical relationships to known classes [5]). Theoretically, no framework exists to quantify how unknowns should contribute to knowledge transfer. Unlike known classes, where distillation or replay provides clear mechanisms, unknowns present unique challenges: (i) preserving latent representations without destabilizing known classes, (ii) refining decision boundaries for future tasks, and (iii) balancing plasticity for new unknowns with stability for knowns. This ambiguity forces ad-hoc solutions like threshold-based rejection [30] or heuristic clustering [25] without theoretical guarantees. Ultimately, the lack of robust theoretical foundations reduces open-world continual learning to a superficial combination of continual learning and open-set techniques, rather than a principled framework for unified knowledge transfer and updating.

2.5 Concluding Remarks

Table 2.1 summarizes the core distinctions between continual learning, open-set learning, and open-world continual learning across four dimensions: objectives, strengths, limitations, and representative works. While continual learning excels at sequential knowledge retention and open-set learning specializes in novelty detection, open-world continual learning must reconcile its competing demands—preserving stability for known classes while maintaining plasticity for unknowns, all within dynamically expanding task distributions. This reveals open-world continual learning's critical gap: current methods [8, 30] lack mechanisms to transfer knowledge from unknown samples while balancing the stability-plasticity trade-off.

This chapter shows critical gaps in current open-world continual learning research that our work addresses: (1) asymmetric knowledge transfer because of the prevailing neglect of unknown samples' informational value (solved through *HoliTrans*' unified knowledge transfer in Chapter 6), (2) dynamic evaluation because of lack of benchmarks for unknown reuse (tackled via few-shot protocols in Chapter 5), and (3) theoretical guarantees because of the absence of formal bounds for open-set continual learning (resolved by *Pro-KT*'s stability-plasticity analysis in Chapter 4). These advancements are designed with practical deployment constraints in mind, as demonstrated through our real-world fraud detection applications in Chapter 7.

Table 2.1: Comparison of continual learning, open-set learning, and open-world continual learning.

Aspect	**Continual Learning**	**Open-Set Learning**	**Open-world Continual Learning**
Primary Goal	Avoid forgetting known tasks	Detect unknowns in static settings	Learn incrementally and adapt to unknowns
Key Strength	Task-specific stability [21]	Strong open-set detection [15]	Unified openness and continuity [8]
Limitation	Assumes a closed world	No incremental learning	Lack of theoretical framework for unknown utilization
Representative Work	EWC [21]	ORE [55]	SOLA [8]