Beyond Supervised Continual Learning: a Review

Benedikt Bagus[1], Alexander Gepperth[1] and Timothée Lesort[2]

[1] University of Applied Sciences Fulda - Department of Computer Science
Leipzigerstraße 123, 36037 Fulda - Germany

[2] Montréal University (UdeM), Mila
6666 St-Urbain Street, Montréal, QC - Canada

Abstract. *Continual Learning* (CL, sometimes also termed *incremental learning*) is a flavor of machine learning where the usual assumption of stationary data distribution is relaxed or omitted. When naively applying, e.g., DNNs in CL problems, changes in the data distribution can cause the so-called *catastrophic forgetting* (CF) effect: an abrupt loss of previous knowledge. Although many significant contributions to enabling CL have been made in recent years, most works address supervised (classification) problems. This article reviews literature that study CL in other settings, such as learning with reduced supervision, fully unsupervised learning, and reinforcement learning. Besides proposing a simple schema for classifying CL approaches w.r.t. their level of autonomy and supervision, we discuss the specific challenges associated with each setting and the potential contributions to the field of CL in general.

1 Introduction

Continual learning is a field of machine learning where the data distribution is not static. It is a natural framework for many practical problems where the data arrives progressively, and the model learns continuously. For example, in robotics, robots need to adapt to their environments to interact and realize actions constantly, or recommendation systems also need to adapt constantly to the new content available and the new needs of users. However, in recent years, the field of continual learning has focused mainly on one type of classification scenario: class-incremental. This scenario evaluates how models can learn a class once and remember it when new class data arrives. While it is important to solve this problem, using only one type of scenario can lead to over-specialized solutions that can not generalize to different settings. In this paper, we propose to review the literature dealing with other settings than the default one (class-incremental) and more generally, fully supervised scenarios. The goal is to shed light on efforts made to diversify the evaluation of continual learning.

We introduce the continual learning framework and the goals of continual learning (Sec 2). Then, we describe the default scenario and its characteristics (Sec 3). In addition, we introduce a scenario that goes beyond the default scenario in supervised learning (Sec 4), unsupervised learning (Sec 5) and reinforcement learning (Sec 6).

2 Framework and Goals of Continual Learning

Continual learning (CL) is a machine learning sub-field that studies learning under time-varying data distributions. This relaxes one of the fundamental assumptions of statistical learning theory ([1]), which states that the data follows a stationary distribution. One advantage of this assumption is its simplicity, whereas CL scenarios are very diverse, depending on the nature of the non-stationarity. In CL, it is therefore very important to clearly define the scenario, the goals of learning, the evaluation measures, and the loss functions. The following section describes typical non-stationarities of the data distribution that have been considered in the literature (see also [2, 3]).

2.1 Data distribution Drifts

CL under data distribution shifts needs memorization mechanisms adapted for the type of non-stationarity, which requires assumptions by the used algorithms in the face of an infinite number of possible ways a distribution can be non-stationary. For simplicity, we will list several typical definitions used for supervised learning (e.g., classification) but that can be generalized to other settings.

A simple way to categorize non-stationarities is based on class information. We may distinguish two types of shifts [2] in this case: *concept shift*, where the annotation of existing data changes [4] and *virtual shift*, where we get new data, but the annotation does not change. Usually, the term *shift* is employed for sudden changes in data distributions, whereas the term *drift* is used for gradual changes over time. In supervised CL, virtual shifts are the most common non-stationarities that have been studied. We can distinguish the special cases of *virtual concept shift*, implying new data and new labels, and *domain shift*, where new data of known labels are observed. Those two settings are also known as respectively *class-incremental* and *instance-incremental* [5, 6].

The objectives to be optimized may change over time as well [3], as in continual reinforcement learning [7, 8, 9].
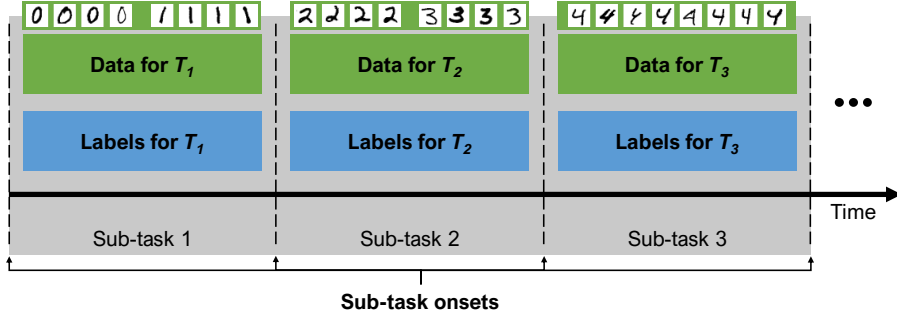
Fig. 1: The default scenario for CL. The data stream is assumed to be partitioned into *sub-tasks* defined by data and labels (targets). Data statistics within a sub-task are assumed to be stationary. In addition, sub-task data and labels are assumed to be disjoint, i.e., from different classes. The sub-task onsets are generally assumed to be known as well.

2.2 Common CL constraints

If CL were not subject to constraints, there would be a simple solution to any scenario. It involves storing all incoming samples and re-training every time a decision is expected, [10]. This entails a time and memory complexity that is at least linear in the number of processed samples. However, most CL proposals assume that memory is limited in some way, preventing this (obvious) solution. Many approaches similarly assume that the storage of samples is restricted. Other resources constraints subject of study are: *computational cost*, *memory*, *data privacy*, *fast adaptation*, *inference speed*, *transfer*. Other constraints that are more related to the reliability of approaches are *stability* and *explainability*. A discussion of CL constraints can be found in, e.g., [11], whereas evaluation measures that take these constraints into account are given in [5].

3 The default scenario for CL

A particular supervised setting, which we will refer to as the *default scenario*, is currently dominating CL literature. It is based on a classification problem divided into a small number of *sub-tasks*. Virtual concept shift occurs abruptly at sub-task boundaries by the apparition of new samples belonging to previously unseen classes, see Fig 1. Usually, the sub-task onsets/boundaries are known. A consequence of this disjointness of sub-tasks is that no pure concept shift is involved: the annotation for a given data point will never change or be subject to conflict as learning progresses.

In the default scenario, the goal of CL is usually to learn the complete statistics of all sub-tasks as if they had been processed all at once, rather than one after the other.

Sometimes, samples are processed one by one, or all samples in a given sub-

task are simultaneously available. In some works, the sub-task index is known at test time, which is used for selecting the correct head of a multi-headed DNN for inference.

The assumptions made in the default scenario are justified in many use cases. However, it is obvious that other scenarios, e.g., in robotics, may be found where they impose too severe restrictions. Moreover, many characteristics of the default scenario, such as "drift are abrupt" or, "tasks are not revisited" lend themselves to *benchmark overfitting*. As an example, consider sub-task T_2 in Fig 1: here, a DNN could punish incorrect decisions for class "1" more strongly than incorrect decisions for classes from the current sub-task since the default scenario assumes that sub-tasks are disjoint. Even if researchers do not consciously exploit these assumptions, the employed CL algorithms may still rely on them indirectly. It is thus fundamental to perform experiments in scenarios where these assumptions do not hold.

Thus, creating diverse benchmarks, as well as approaches that do not critically rely on the assumptions from the default scenario, should be an ongoing effort. This effort should be pushed notably by existing continual learning libraries such as Continuum [12], Avalanche [13] or Sequoia [7].

3.1 CL approaches for the default scenario

This section is not meant to be exhaustive, since the default scenario is not the focus of this article. Please refer to recent reviews [14, 15] for more details on CL methods for the default scenario.

Broad strategies for performing CL in the default scenario are regularization [16, 17], replay [18, 19, 20] and dynamic architectures [21, 22, 23, 24, 25]. Regularization penalizes changes to model parameters that are deemed important for past sub-tasks. This is usually achieved by adding penalty terms to the loss function, and it is implicitly assumed that new sub-tasks add only new data and classes. Dynamic architecture methods extend models over time in order to separate previously learned parameters from currently optimized ones, thus reducing cross-talk and catastrophic forgetting (but equally assuming that new sub-tasks contain only new data). Replay methods store received data for subsequent use in re-training (rehearsal). Instead of relying on stored data, re-training can also be performed using samples produced by generative models (generative replay).

Replay is known to be an effective method for preventing catastrophic forgetting (CF), especially in class-incremental settings [26, 14], but also for continual reinforcement learning [27, 28, 19] or unsupervised learning [29, 30, 31].

3.2 Metrics and evaluation procedures

In the default scenario, various measures related to the classification error are common, which have been discussed in, e.g., [32, 33, 34]. A common baseline is termed *cumulative performance*, obtained by evaluating models on the merged test sets from all sub-tasks, which corresponds to learning with stationary statistics. This baseline is often considered an upper bound for CL performance.

In addition, [17] proposed the notions of forward and backward transfer: forward transfer (FT) measures how training on sub-task i impacts performance on a future sub-task $j > i$. For backward transfer (BT), the impact on previous sub-tasks $j < i$ is considered. The common case in CL is negative BT indicating forgetting, but positive BT is theoretically possible as well.

Many authors, e.g., [16], assume that (although sub-tasks are presented sequentially) all sub-task data are available for model selection and hyper-parameter tuning. For example, tuning EWC's regularization strengths λ_i for each sub-task is often done in hindsight.

Some authors [35, 36, 37], especially in works using multi-head DNNs, assume that the sub-task ID is known during testing, although this does not seem to be the current consensus. In the limit where each sub-task contains only a single class, providing the sub-task ID at test time means providing the class label. Even if sub-tasks are more diverse, the sub-task ID contains significant information and may thus confer unfair advantages. The question of evaluation protocols in CL is discussed in [11, 38].

3.3 Benchmarks

Benchmarks for the default CL scenario are mostly derived from datasets such as MNIST, CIFAR10/100, Imagenet, SVHN etc. to create *class-incremental* or *domain-incremental* scenarios. The *permuted MNIST* benchmark, where successive sub-tasks are created by permuting all pixels according to a sub-task specific permutation scheme, was initially popular [16, 39, 40] but is less so now because it can, to good accuracy, be solved even without dedicated CL schemes [11]. Some authors used Atari games [41], Mujoco [42] or Meta-World [43] as benchmarks. CL specific variants of standard benchmarks such as, e.g., *colored MNIST* [44] are widely used as well since they can be used to investigate specific aspects of CL, see [45, 46].

4 Generalizations of the default scenario

The default scenario is convenient for evaluation and represents a rather controlled setting for CL. In less controlled settings, fully annotated data may not be available, or supplementary constraints may be imposed. We find it convenient to introduce a new taxonomy of CL approaches based on their level of autonomy.

4.1 Classifying the autonomy level of CL algorithms

The various applications of continual learning can be classified in autonomy levels, as for autonomous vehicles [47]. Obviously, CL should get harder as less and less human supervision is supplied. We identify two dimensions of autonomy, both of which will be discussed in-depth to characterize generalized supervised CL approaches better.

Objective Autonomy denotes the autonomy regarding the objective to achieve (labels, targets, rewards), which we group into 4 levels:

- **Level 0:** Full data annotation: supervised.

- **Level 1:** Sparse labelization: RL, active learning, sparse training.

- **Level 2:** No annotation for training, query for fast adaptation.

- **Level 3:** No annotation for training, zero-shot adaptation.

For objective supervision levels 2 and 3, continual training can be seen as pre-training for the unknown future objective task. We can note that if the scenario objective is unsupervised, then we can assume that it is similar to full data annotation.

Continual Learning Autonomy is concerned with autonomy regarding the distribution shifts (task label, task boundaries):

- **Level 0:** Full task annotation at train and test.

- **Level 1:** Full task annotation at train, no test task labels.

- **Level 2:** Sparse task annotation at train, no test task labels: example task boundaries only without train task label.

- **Level 3:** No task labels at all: task-agnostic.

This classification still holds for smooth transitions (concept drift). We can note that for class-incremental problems, the task-agnostic setting does not make sense since the task information is in the class labels. We can now characterize CL contributions by a pair "(i j)", where i represents the objective autonomy level, and j stands for the CL autonomy level. Hence, the pair (0 1) describes a class-incremental scenario, i.e., a classification setting with fully annotated data and task labels for training but not for test data. It is important to note that those ratings classify the complexity of a given scenario and not approaches. For example, the default scenario (class-incremental) assesses a complexity level of (0 1), domain incremental without task labels would be (0 2), task agnostic continual reinforcement learning [48] (1 3). To validate approaches' autonomy, they should be evaluated on adequate scenarios.

Investigating the cost of lowering or increasing a task's complexity level is fundamental for applications of CL. We want to make our algorithms scale up to arbitrary complexity levels, but in practice, we would always choose the lowest possible complexity. If permitted by a given application, solving a scenario with a complexity level (0 1) is obviously more efficient than solving the same scenario with a level of (3 3).

4.2 Towards a Generalization of the Default Setting

Some variants of supervised CL exist that alleviate the need for annotated data. The reduction of annotations can apply for restricted access to the task labels as in task-agnostic CL [49, 50], or by reducing the labels' availability as in continual active learning [51, 52], or in semi-supervised continual learning, [53]. As described in Sec 4.1, reducing access to task supervision lead to evaluating a CL autonomy level of 2, and removing all access to it lead to evaluating a CL autonomy level of 3. On the other hand, reducing data annotation lead to an objective autonomy of 1 instead of 0 when the full annotation is available.

Among potential supplementary constraints, data can be streamed without the possibility of multi-epoch training as in online training [54, 55], or the data can be imbalanced [56, 57], or mixed with spurious features [58]. Scenarios where the annotations change over time (real concept drift) have been investigated in [3, 59, 4]. Some contributions [60] relax the condition of disjoint sub-tasks and assess the impact of several fundamental CL strategies such as regularization and replay. Yet others [61, 62] demonstrate that detecting sub-task boundaries autonomously is generally feasible by using density estimation methods.

To the generalized settings cited above, supplementary constraints (as discussed in Sec 2.2) may be added, making them even harder. If this brings CL closer to real-life applications, solutions are required that do not overfit a particular CL setting. Currently, the field of CL is fundamentally meta: implicitly, the goal is not to train the best possible model w.r.t. non-continual baselines, but rather to create algorithms that show maximal generalization to other CL settings. Therefore, experimenting with generalized supervised scenarios can assess algorithms' robustness and improve generalized CL.

5 Unsupervised Continual Learning

Whether a machine learning task is considered supervised or not depends on the formulation of the loss function. In fact, no assumptions whatsoever are made concerning the loss in the definition of CL given in Sec 1. Therefore, CL is naturally transferred to unsupervised methods of machine learning, typical examples of which are density modeling, clustering, generative learning, and unsupervised representation learning.

5.1 Density modeling

Density modeling aims at approximating the probability density of a given set of data samples directly by minimizing a log-likelihood loss. Typically, this is achieved using *mixture models*, which model the data density as a weighted sum of N parameterized component densities, e.g., multivariate Gaussian densities or Dirichlet distributions. Density modeling allows performing, among other functions, Bayesian inference and sampling. These functionalities spawned increased interest in mixture modeling a few years back, particularly in robotics [63, 64, 65, 66]. The main issue for CL in mixture modeling is that concept

drift or shift may require an adaptation of N, which motivates heuristics for adding and removing components. Current approaches using generative replay are proposed in, e.g., [30]. Mixture models usually adapt only a small subset of components for each update step due to their intrinsic reliance on distances instead of scalar products. This is why they are less prone to catastrophic forgetting than DNNs, an effect that has been demonstrated for self-organizing maps in [67] which are an approximation to Gaussian Mixture Models (GMMs), see [68]. Modeling the data density allows partitioning data space into Voronoi cells, in each of which a separate linear classifier model can be trained. This is the essence of the popular Locally Weighted Projection Regression (LWPR) algorithm [69] which was explicitly constructed for continual classification in robotics.

5.2 Clustering

Clustering is, in a certain sense, an approximation to density modeling, although the inference is limited to determining the precise component a given data sample was generated from. Clustering methods are normally trained using a k-means type of algorithm, which approximates gradient descent on a loss function that again approximates a GMM log-likelihood. CL for clustering algorithms faces the same basic issue as in density modeling: a potentially variable number N of cluster centers during concept drift or shift. This has been demonstrated in, e.g., [70, 71, 72].

5.3 Generative learning

Generative learning aims to generate realistic samples (typically images) that are similar to a set of training data. Typical models are generative adversarial networks (GANs), variational auto-encoders (VAEs), PixelCNN, FLoW or GLoW, but many other variants have been proposed, see, e.g., [73] for a review. Training such generative models can be performed, e.g., in the CL default scenario introduced in Sec 3 (apart from the supervision information), which leads to catastrophic forgetting (CF) without additional measures. Several of the approaches used in supervised learning have been successfully applied to training generative models: knowledge distillation [74], EWC [75] and replay [76, 77, 78, 79, 29]. To our knowledge, no generic approaches that are specific to generative learning have been proposed, apart perhaps [80] where it is proposed to learn specific transformations. This, however, is very specific to a particular kind of (image) data and would have to be adapted if other kinds of data were targeted.

5.4 Continual Representation Learning

Unsupervised training for learning representations for downstream applications is a common use case for unsupervised learning. It was one of the motivations to develop various types of auto-encoders and generative models in the early days

of deep learning. In CL, using an unsupervised criterion to learn representations might be useful to avoid representations that overfit a specific task and, at the same time, improve performance on downstream tasks [81, 30, 31]. Unsupervised pre-training can also be useful for learning a general feature extractor that can be frozen for future tasks [82, 83, 84].

5.5 Challenges of unsupervised CL

Unsupervised learning offers general learning criteria that can avoid the over-specialization of supervised training and reduce forgetting. Nevertheless, unsupervised CL faces the same challenges as supervised CL, and the default scenario for supervised CL can be transferred. Moreover, in practice, unsupervised training tends to be more complex than supervised training, especially for generation and density modeling, since it is harder to model a distribution than to determine a separating hypersurface between classes in data space. With the added complexity of CL, unsupervised learning can be a formidable problem, especially w.r.t. model and hyper-parameter selection.

6 Continual Reinforcement Learning

In reinforcement learning (RL), an agent learns to interact with its environment by choosing a specific action for each state based on a reward signal. The (unknown) underlying process is formalized as a Markov decision process (MDP), where an optimal policy maximizes an expected reward. This scenario is inherently a CL setting, since the distribution of the observed data depends on the specific policy. The evolution of the policy throughout the learning process will mechanically lead to the non-stationarity of the data distribution. Hence, RL requires the ability to cope with non-stationary data. However, supplementary non-stationary, for example, in the environment or in the objective to fulfill, can increase the training difficulty and lead to a continual setting. We will use the term *Continual Reinforcement Learning* (CRL) for denoting RL in settings that go beyond the usual assumptions of non-stationarity made in conventional RL.

6.1 Existing Approaches

The works presented in [85, 86] introduce the importance of CL at an early stage and especially investigated them in the context of reinforcement learning. More recent works, e.g., [87, 19] revisit this area and consider additional aspects such as catastrophic forgetting. Some frameworks to guide future research have also been published [88, 8]. Both provide a comprehensive overview of the synergies between continual and reinforcement learning.

RL Approaches Experience replay [89, 19, 90] is the most common approach to counter non-stationarities in RL. Several variants are introduced, e.g., [91, 92, 93, 94, 95].

Continuous control, multi-task, and multi-goal are also research topics intersecting with continual reinforcement learning, but their scenarios are not

always defined in a consistent fashion in the literature. In general, the goal is to enable transfer learning between policies, which, however, omits the capacity for forgetting or re-adaptation. Some works assume a static objective [96, 97, 98, 99, 100], others a static agent and/or environment [101, 102, 103] or none of both [104, 105, 106]. In contrast, multi-agent reinforcement learning is mostly related to some kind of joint training and is hence not related to CL.

For CRL, the agent needs to acquire new skills to handle time-varying conditions, such as changes in environment [107], observations or actions, and additionally must retain the old knowledge. A variety of approaches has been published, among which knowledge-based distillations [27, 28] and context-based decompositions [108, 109] are popular. Other works are concerned with the employed model [110, 111, 112, 113], off-policy algorithms [114], policy gradient [115] or a task-agnostic perspective [48]. Evaluations of known CL methods (e.g., GEM, A-GEM, and replay) are also applied in the RL domain [116, 9].

Benchmarks An overview of CRL environments can be found in [117]. Dedicated benchmarks which allow a systematical assessment are: *Meta-World* [43], *Continual World* [118] and *L2Explorer* [119].

Libraries Some libraries aim at unifying CRL development to improve comparability and accelerate progress: *Sequoia* [7], *Avalanche rl* [120], *SaLinA* [121], *Reverb* [122] and *CORA* [123].

6.2 Assumptions in CRL

Three assumptions are commonly made in CRL: Foremost, a decomposition into sub-tasks is assumed, even if their onset is unknown since most dedicated CL methods (see Sec 3.1) assume the existence of distinct sub-tasks. Another assumption concerns samples, which are assumed to be non-contradictory within sub-tasks, meaning the assessment of rewards changes only between sub-tasks, if it changes at all. Finally, it is a common assumption that knowledge of sub-task boundaries is provided. Most existing works are using information about sub-task boundaries as if they were provided by an oracle, without the possibility to recognize or determine them autonomously.

6.3 Challenges

In CRL, various types of drifts/shifts can appear:

Environment-related The agent successively observes its environment. Therefore, on a short timescale, observations will always be non-stationary, even if the environment is. In addition, the environment itself can change over time, or rapid modifications can be encountered (*environment shift*). This would result in novel states or transitions between these, resulting in an enlargement of the actual involved MDP.

Goal-related By maximizing the reward signal, the agent attains a defined objective. If the reward function changes, the agent experiences divergent information, leading to an inconsistent policy. In this setting, the definition of

states, actions and transitions does not change, so the underlying MDP remains structurally intact. However, other rewards are assigned to previously learned mappings, enforcing changes of transition probabilities.

Agent-related The decreasing influence of exploration, regardless of whether off-policy methods such as Q-Learning or on-policy methods such as policy gradient are used, temporarily creates a source of non-stationarity, resulting in a time-varying sampling of the state-action space even with a static policy and a static environment. Additionally, it is easily possible that sensors or actuators degrade or undergo deliberate manipulations. Affecting possible actions, immediate effects on the MDP, while a changed perception of states also impacts transitions.

Sub-tasks and data acquisition For scenarios where the environment changes in a discrete fashion, we can introduce the notion of sub-tasks as in the default scenario for supervised CL, see Sec 3. A general challenge stems from the fact that samples are acquired as an online time series and have no balancing guarantees at all. Moreover, it is possible that similar states and actions appear in various sub-tasks, but with different assigned rewards, so sub-tasks are usually not disjoint and may even be contradictory, requiring un- or re-learning, a concept absent from the default supervised CL scenario. Depending on the type of non-stationarity, sub-task onset can be unknown, and the detection of boundaries may be difficult if the drifts are gradual rather than abrupt. In addition, the number of sub-tasks can be significantly higher than in supervised scenarios, up to a point where the entire concept of sub-tasks becomes questionable. Lastly, actions must be explicitly performed to transition to the appropriate subsequent state. Therefore, a generative or offline sampling is of limited usefulness, at least for exploration.

7 Discussion

The field of CL has expanded rapidly in recent years, which is why many aspects of CL are still fluid and not subject to a common consensus among researchers. This is evidenced by a wide variety of assumptions, evaluation metrics, see Sec 3.2 and constraints, see Sec 2.2. The so-called default scenario, see Sec 3, is the nearest thing to a commonly agreed scenario, yet many details fluctuate strongly between contributions. This leads to several interesting consequences and opportunities for further research:

CL comparability A direct consequence is the difficulty to directly compare results of different articles. This underscores the need, in CL more than in other domains of machine learning, to precisely describe evaluation procedures and, where possible, make use of existing libraries (see Sec 3 and 6.1) and evaluation procedures. Furthermore, as stated in Sec 2.2, CL is a multi-objective problem where achieving the cumulative baseline is important, but where other measures (see Sec 3.2) matter as well.

CL autonomy As explained in Sec 4.1, CL approaches should also be evaluated

based on the complexity and autonomy of the scenario they can generalize to, to prevent them from overfitting to a specific CL scenario or assumptions.

CL scalability An aspect that is often omitted in current works in favor of quantitative performance measures is scalability. Depending on a potential application context, CL, even in the default scenario, may be faced with a huge number of sub-tasks, each again containing enormous amounts of samples. If this were not the case, the cumulative baseline, or equivalently some variant of GDumb (see Sec 3.2), would be a much less costly and superior (w.r.t. performance) alternative to using dedicated CL methods. So time and memory complexity for the case where the number of sub-tasks is large should be included in all new works on CL algorithms to ensure comparability, at least in this respect.

CL generalization As was shown in Sec 4, 5 and 6, the default CL scenario of Sec 3 can be generalized in many ways. Moreover, these chapters show that many open issues remain, both technical and conceptual, when attempting to generalize CL.

8 Conclusion

This review article attempts to give an overview of the current state of CL beyond the purely supervised default scenario, see Sec 3. We describe the various complexification of the default scenario and the different learning paradigms, and propose a classification based on the autonomy characteristics of algorithms. We believe that attempts to generalize CL pose important questions about the fundamental assumptions behind CL. We thus encourage CL researchers to carefully reflect upon the implicit, hidden assumptions in each CL approach they are dealing with and whether they can (and should) be relaxed. In a still-fluid field such as CL, a continuous re-examination of assumptions may lead to new solutions that strongly contribute to the advancement of the field.

References

[1] Vladimir Vapnik. *The nature of statistical learning theory.* Springer science & business media, 1999.

[2] Alexander Gepperth and Barbara Hammer. Incremental learning algorithms and applications. In *European Symposium on Artificial Neural Networks (ESANN)*, Bruges, Belgium, 2016.

[3] Timothée Lesort, Massimo Caccia, and Irina Rish. Understanding continual learning settings with data distribution drift analysis. *arXiv preprint arXiv:2104.01678*, 2021.

[4] Massimo Caccia, Pau Rodriguez, Oleksiy Ostapenko, Fabrice Normandin, Min Lin, Lucas Caccia, Issam Laradji, Irina Rish, Alexandre Lacoste, David Vazquez, and Laurent Charlin. Online fast adaptation and knowledge accumulation: a new approach to continual learning. *NeurIPS*, 2020.

[5] Timothée Lesort, Vincenzo Lomonaco, Andrei Stoian, Davide Maltoni, David Filliat, and Natalia Díaz-Rodríguez. Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges. *Information Fusion*, 58:52 – 68, 2020.

[6] Gido M van de Ven and Andreas S Tolias. Three scenarios for continual learning. *arXiv preprint arXiv:1904.07734*, 2019.

[7] Fabrice Normandin, Florian Golemo, Oleksiy Ostapenko, Pau Rodriguez, Matthew D Riemer, Julio Hurtado, Khimya Khetarpal, Dominic Zhao, Ryan Lindeborg, Timothée Lesort, et al. Sequoia: A software framework to unify continual learning research. *arXiv preprint arXiv:2108.01005*, 2021.

[8] Khimya Khetarpal, Matthew Riemer, Irina Rish, and Doina Precup. Towards continual reinforcement learning: A review and perspectives. *arXiv preprint arXiv:2012.13490*, 2020.

[9] Benedikt Bagus and Alexander Gepperth. A study of continual learning methods for Q-learning. In *International Joint Conference on Neural Networks (IJCNN)*, 2022.

[10] Ameya Prabhu, Philip HS Torr, and Puneet K Dokania. Gdumb: A simple approach that questions our progress in continual learning. In *European Conference on Computer Vision*, pages 524–540. Springer, 2020.

[11] B Pfülb and A Gepperth. A comprehensive, application-oriented study of catastrophic forgetting in dnns. In *International Conference on Learning Representations (ICLR)*, 2019.

[12] Arthur Douillard and Timothée Lesort. Continuum: Simple management of complex continual learning scenarios. *arXiv preprint: 2102.06253*, 2021.

[13] Vincenzo Lomonaco, Lorenzo Pellegrini, Andrea Cossu, Gabriele Graffieti, and Antonio Carta. Avalanche: an end-to-end library for continual learning. *Github repository*, 2021.

[14] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Ales Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. Continual learning: A comparative study on how to defy forgetting in classification tasks. *arXiv preprint: 1909.08383*, 2019.

[15] Eden Belouadah, Adrian Popescu, and Ioannis Kanellos. A comprehensive study of class incremental learning algorithms for visual tasks. *Neural Networks*, 135:38–54, 2021.

[16] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proc. of the national academy of sciences*, 2017.

[17] David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30, 2017.

[18] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. iCARL: Incremental classifier and representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017.

[19] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. *Advances in Neural Information Processing Systems*, 32, 2019.

[20] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. In *Advances in Neural Information Processing Systems*, pages 2990–2999, 2017.

[21] A. A. Rusu, N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, and R. Hadsell. Progressive neural networks. *ArXiv e-prints*, jun 2016.

[22] Chrisantha Fernando, Dylan Banarse, Charles Blundell, Yori Zwols, David Ha, Andrei A. Rusu, Alexander Pritzel, and Daan Wierstra. Pathnet: Evolution channels gradient descent in super neural networks. *CoRR*, abs/1701.08734, 2017.

[23] Tom Veniat, Ludovic Denoyer, and MarcAurelio Ranzato. Efficient continual learning with modular networks and task-driven priors. In *International Conference on Learning Representations*, 2021.

[24] Oleksiy Ostapenko, Pau Rodriguez, Massimo Caccia, and Laurent Charlin. Continual learning via local module composition. In *Advances in Neural Information Processing Systems*, 2021.

[25] Jorge A Mendez, Harm van Seijen, and Eric Eaton. Modular lifelong reinforcement learning via neural composition. In *International Conference on Learning Representations(ICLR)*, 2022.

[26] Timothée Lesort, Thomas George, and Irina Rish. Continual learning in deep networks: an analysis of the last layer. *arXiv preprint arXiv:2106.01834*, 2021.

[27] René Traoré Kalifou, Hugo Caselles-Dupré, Timothée Lesort, Te Sun, Natalia Diaz-Rodriguez, and David Filliat. Continual reinforcement learning deployed in real-life using policy distillation and sim2real transfer. In *ICML Workshop on Multi-Task and Lifelong Learning*, volume 4, 2019.

[28] René Traoré, Hugo Caselles-Dupré, Timothée Lesort, Te Sun, Guanghang Cai, Natalia Díaz-Rodríguez, and David Filliat. DisCoRL: Continual reinforcement learning via policy distillation. *arXiv preprint arXiv:1907.05855*, 2019.

[29] Timothée Lesort, Hugo Caselles-Dupré, Michael Garcia-Ortiz, Jean-François Goudou, and David Filliat. Generative models from the perspective of continual learning. In *International Joint Conference on Neural Networks(IJCNN)*, 2019.

[30] Dushyant Rao, Francesco Visin, Andrei A. Rusu, Yee Whye Teh, Razvan Pascanu, and Raia Hadsell. Continual unsupervised representation learning. *arXiv preprint: 1910.14481*, 2019.

[31] Divyam Madaan, Jaehong Yoon, Yuanchun Li, Yunxin Liu, and Sung Ju Hwang. Representational continuity for unsupervised continual learning. In *International Conference on Learning Representations(ICLR)*, 2021.

[32] Ronald Kemker, Marc McClure, Angelina Abitino, Tyler Hayes, and Christopher Kanan. Measuring catastrophic forgetting in neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[33] Natalia Díaz-Rodríguez, Vincenzo Lomonaco, David Filliat, and Davide Maltoni. Don't forget, there is more than forgetting: new metrics for continual learning. *arXiv preprint: 1810.13166*, 2018.

[34] Davide Maltoni and Vincenzo Lomonaco. Continuous learning in single-incremental-task scenarios. *Neural Networks*, 116:56–73, 2019.

[35] Joan Serra, Didac Suris, Marius Miron, and Alexandros Karatzoglou. Overcoming catastrophic forgetting with hard attention to the task. *ICML*, 80:4548–4557, 10–15 Jul 2018.

[36] Thang Doan, Mehdi Abbana Bennani, Bogdan Mazoure, Guillaume Rabusseau, and Pierre Alquier. A theoretical analysis of catastrophic forgetting through the NTK overlap matrix. In *International Conference on Artificial Intelligence and Statistics*, pages 1072–1080, 2021.

[37] Mehrdad Farajtabar, Navid Azizan, Alex Mott, and Ang Li. Orthogonal gradient descent for continual learning. In *International Conference on Artificial Intelligence and Statistics*, pages 3762–3773. PMLR, 2020.

[38] Benedikt Pfülb. *Continual Learning with Deep Learning Methods in an Application-Oriented Context*. PhD thesis, University of Applied Sciences Fulda, Germany, 2022.

[39] David Lopez-Paz and Marc-Aurelio Ranzato. Gradient episodic memory for continual learning. In *Advances in Neural Information Processing Systems*. 2017.

[40] Mitchell Wortsman, Vivek Ramanujan, Rosanne Liu, Aniruddha Kembhavi, Mohammad Rastegari, Jason Yosinski, and Ali Farhadi. Supermasks in superposition. In *Advances in Neural Information Processing Systems*, 2020.

[41] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing Atari with deep reinforcement learning. *arXiv preprint: 1312.5602*, 2013. cite arxiv:1312.5602Comment: NIPS Deep Learning Workshop 2013.

[42] Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *IROS*, pages 5026–5033. IEEE, 2012.

[43] Tianhe Yu, Deirdre Quillen, Zhanpeng He, Ryan Julian, Karol Hausman, Chelsea Finn, and Sergey Levine. Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning. In *Conference on Robot Learning*, pages 1094–1100. PMLR, 2020.

[44] Byungju Kim, Hyunwoo Kim, Kyungsu Kim, Sungjin Kim, and Junmo Kim. Learning not to learn: Training deep neural networks with biased data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9012–9020, 2019.

[45] Itai Gat, Idan Schwartz, Alexander Schwing, and Tamir Hazan. Removing bias in multi-modal classifiers: Regularization by maximizing functional entropies. *Advances in Neural Information Processing Systems*, 33:3197–3208, 2020.

[46] Yi Li and Nuno Vasconcelos. Repair: Removing representation bias by dataset resampling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9572–9581, 2019.

[47] B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A Al Sallab, Senthil Yogamani, and Patrick Pérez. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[48] Massimo Caccia, Jonas Mueller, Taesup Kim, Laurent Charlin, and Rasool Fakoor. Task-agnostic continual reinforcement learning: In praise of a simple baseline. *arXiv preprint arXiv:2205.14495*, 2022.

[49] Chen Zeno, Itay Golan, Elad Hoffer, and Daniel Soudry. Task agnostic continual learning using online variational bayes. *arXiv preprint: 1803.10123*, 2018.

[50] Xu He, Jakub Sygnowski, Alexandre Galashov, Andrei A. Rusu, Yee Whye Teh, and Razvan Pascanu. Task agnostic continual learning via meta learning. *ArXiv*, abs/1906.05201, 2019.

[51] Martin Mundt, Yong Won Hong, Iuliia Pliushch, and Visvanathan Ramesh. A wholistic view of continual learning with deep neural networks: Forgotten lessons and the bridge to active and open world learning. *arXiv preprint arXiv:2009.01797*, 2020.

[52] Matthias Perkonigg, Johannes Hofmanninger, and Georg Langs. Continual active learning for efficient adaptation of machine learning models to changing image acquisition. In *International Conference on Information Processing in Medical Imaging*, pages 649–660. Springer, 2021.

[53] James Smith, Jonathan Balloch, Yen-Chang Hsu, and Zsolt Kira. Memory-efficient semi-supervised continual learning: The world is its own replay buffer. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

[54] Arslan Chaudhry, Marc'Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with A-GEM. In *International Conference on Learning Representations(ICLR)*, 2019.

[55] Jonathan Schwarz, Jelena Luketina, Wojciech M Czarnecki, Agnieszka Grabska-Barwinska, Yee Whye Teh, Razvan Pascanu, and Raia Hadsell. Progress & compress: A scalable framework for continual learning. In *ICML*, 2018.

[56] Chris Dongjoo Kim, Jinseo Jeong, and Gunhee Kim. Imbalanced continual learning with partitioning reservoir sampling. In *Proceedings of the IEEE European Conference on Computer Vision (ECCV)*, 2020.

[57] Aristotelis Chrysakis and Marie-Francine Moens. Online continual learning from imbalanced data. In *International Conference on Machine Learning*, pages 1952–1961. PMLR, 2020.

[58] Timothée Lesort. Continual feature selection: Spurious features in continual learning. *arXiv preprint arXiv:2203.01012*, 2022.

[59] Mohamed Abdelsalam, Mojtaba Faramarzi, Shagun Sodhani, and Sarath Chandar. Iirc: Incremental implicitly-refined classification. *CVPR*, pages 11038–11047, 2021.

[60] N Dzemidovich and A Gepperth. An empirical comparison of generators in replay-based continual learning. In *European Symposium on Artificial Neural Networks(ESANN)*, 2022.

[61] B Pfülb, B Bagus, and A Gepperth. Continual learning with fully probabilistic models. In *CVPR Workshop CLVISION Findings paper*, 2021.

[62] B Pfülb and A Gepperth. Overcoming catastrophic forgetting with gaussian mixture replay. In *International Joint Conference on Neural Networks(IJCNN)*, 2021.

[63] Rafael Coimbra Pinto and Paulo Martins Engel. A fast incremental gaussian mixture model. *PloS one*, 10(10):e0139931, 2015.

[64] Konstantin Shmelkov, Cordelia Schmid, and Karteek Alahari. Incremental learning of object detectors without catastrophic forgetting. *CoRR*, abs/1708.06977, 2017.

[65] Dragoljub Pokrajac, Aleksandar Lazarevic, and Longin Jan Latecki. Incremental local outlier detection for data streams. In *2007 IEEE symposium on computational intelligence and data mining*, pages 504–515. IEEE, 2007.

[66] Matej Kristan, Danijel Skocaj, and Aleš Leonardis. Incremental learning with gaussian mixture models. In *Computer vision winter workshop*, pages 25–32. Slovenian Pattern Recognition Soc. Ljubljana, Slovenia, 2008.

[67] A Gepperth and B Pfülb. A rigorous link between self-organizing maps and gaussian mixture models. In *International Conference on Artificial Neural Networks (ICANN)*, 2020.

[68] A Gepperth. Incremental learning with a homeostatic self-organizing neural architecture. *Neural Computing and Applications*, 2019.

[69] Sethu Vijayakumar and Stefan Schaal. Locally weighted projection regression: An o (n) algorithm for incremental real time learning in high dimensional space. In *International Conference on Machine Learning (ICML)*, volume 1, pages 288–293. Morgan Kaufmann, 2000.

[70] Duc Truong Pham, Stefan Simeonov Dimov, and CD Nguyen. An incremental k-means algorithm. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 218(7):783–795, 2004.

[71] Bryant Aaron, Dan E Tamir, Naphtali D Rishe, and Abraham Kandel. Dynamic incremental k-means clustering. In *2014 international conference on computational science and computational intelligence*, volume 1, pages 308–313. IEEE, 2014.

[72] Adil M Bagirov, Julien Ugon, and Dean Webb. Fast modified global k-means algorithm for incremental cluster construction. *Pattern recognition*, 44(4):866–876, 2011.

[73] Ceren Güzel Turhan and Hasan Sakir Bilge. Recent trends in deep generative models: a review. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, pages 574–579. IEEE, 2018.

[74] Ari Seff, Alex Beatson, Daniel Suo, and Han Liu. Continual learning in generative adversarial nets. *arXiv preprint: 1705.08395*, 2017.

[75] Mengyao Zhai, Lei Chen, Frederick Tung, Jiawei He, Megha Nawhal, and Greg Mori. Lifelong gan: Continual learning for conditional image generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2759–2768, 2019.

[76] Felix Wiewel and Bin Yang. Continual learning for anomaly detection with variational autoencoder. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3837–3841, 2019.

[77] Damian Campo, Giulia Slavic, Mohamad Baydoun, Lucio Marcenaro, and Carlo Regazzoni. Continual learning of predictive models in video sequences via variational autoencoders. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 753–757, 2020.

[78] Timothée Lesort, Alexander Gepperth, Andrei Stoian, and David Filliat. Marginal replay vs conditional replay for continual learning. In *International Conference on Artificial Neural Networks*, pages 466–480. Springer, 2019.

[79] Jason Ramapuram, Magda Gregorova, and Alexandros Kalousis. Lifelong generative modeling. *arXiv preprint arXiv:1705.09847*, 2017.

[80] Sakshi Varshney, Vinay Kumar Verma, PK Srijith, Lawrence Carin, and Piyush Rai. Cam-GAN: Continual adaptation modules for generative adversarial networks. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

[81] Enrico Fini, Victor G Turrisi da Costa, Xavier Alameda-Pineda, Elisa Ricci, Karteek Alahari, and Julien Mairal. Self-supervised models are continual learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9621–9630, 2022.

[82] René Traoré, Hugo Caselles-Dupré, Timothée Lesort, Te Sun, Guanghang Cai, Natalia Díaz Rodríguez, and David Filliat. DisCoRL: Continual reinforcement learning via policy distillation. *CoRR*, abs/1907.05855, 2019.

[83] Oleksiy Ostapenko, Timothee Lesort, Pau Rodríguez, Md Rifat Arefin, Arthur Douillard, Irina Rish, and Laurent Charlin. Foundational models for continual learning: An empirical study of latent replay, 2022.

[84] Lucas Caccia and Joelle Pineau. Special: Self-supervised pretraining for continual learning. In *IJCAI, Workshop on Continual Semi-Supervised Learning*, 2021.

[85] Mark Bishop Ring. *Continual learning in reinforcement environments*. PhD thesis, University of Texas at Austin Austin, Texas 78712, 1994.

[86] Sebastian Thrun and Tom M Mitchell. Lifelong robot learning. *Robotics and autonomous systems*, 15(1-2):25–46, 1995.

[87] Ju Xu and Zhanxing Zhu. Reinforced continual learning. *Advances in Neural Information Processing Systems*, 31, 2018.

[88] Timothée Lesort, Vincenzo Lomonaco, Andrei Stoian, Davide Maltoni, David Filliat, and Natalia Díaz-Rodríguez. Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges. *Information fusion*, 58:52–68, 2019.

[89] Shangtong Zhang and Richard S Sutton. A deeper look at experience replay. *arXiv preprint arXiv:1712.01275*, 2017.

[90] William Fedus, Prajit Ramachandran, Rishabh Agarwal, Yoshua Bengio, Hugo Larochelle, Mark Rowland, and Will Dabney. Revisiting fundamentals of experience replay. In *International Conference on Machine Learning*, pages 3061–3071. PMLR, 2020.

[91] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. *arXiv preprint arXiv:1511.05952*, 2015.

[92] Marcin Andrychowicz, Filip Wolski, Alex Ray, Jonas Schneider, Rachel Fong, Peter Welinder, Bob McGrew, Josh Tobin, OpenAI Pieter Abbeel, and Wojciech Zaremba. Hindsight experience replay. *Advances in Neural Information Processing Systems*, 30, 2017.

[93] David Isele and Akansel Cosgun. Selective experience replay for lifelong learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[94] Guido Novati and Petros Koumoutsakos. Remember and forget for experience replay. In *International Conference on Machine Learning*, pages 4851–4860. PMLR, 2019.

[95] Hao Hu, Jianing Ye, Guangxiang Zhu, Zhizhou Ren, and Chongjie Zhang. Generalizable episodic memory for deep reinforcement learning. *arXiv preprint arXiv:2103.06469*, 2021.

[96] Haitham Bou Ammar, Eric Eaton, Paul Ruvolo, and Matthew Taylor. Online multi-task learning for policy gradient methods. In *International conference on machine learning*, pages 1206–1214. PMLR, 2014.

[97] Yee Teh, Victor Bapst, Wojciech M Czarnecki, John Quan, James Kirkpatrick, Raia Hadsell, Nicolas Heess, and Razvan Pascanu. Distral: Robust multitask reinforcement learning. *Advances in Neural Information Processing Systems*, 30, 2017.

[98] Artyom Y Sorokin and Mikhail S Burtsev. Continual and multi-task reinforcement learning with shared episodic memory. *arXiv preprint arXiv:1905.02662*, 2019.

[99] João Ribeiro, Francisco S. Melo, and João Dias. Multi-task learning and catastrophic forgetting in continual reinforcement learning. *arXiv preprint arXiv:1909.10008*, 2019.

[100] Robin Schiewer and Laurenz Wiskott. Modular networks prevent catastrophic interference in model-based multi-task reinforcement learning. In *International Conference on Machine Learning, Optimization, and Data Science*, pages 299–313. Springer, 2021.

[101] Rui Zhao, Xudong Sun, and Volker Tresp. Maximum entropy-regularized multi-goal reinforcement learning. In *International Conference on Machine Learning*, pages 7553–7562. PMLR, 2019.

[102] Ruihan Yang, Huazhe Xu, Yi Wu, and Xiaolong Wang. Multi-task reinforcement learning with soft modularization. *Advances in Neural Information Processing Systems*, 33:4767–4777, 2020.

[103] Abhishek Gupta, Justin Yu, Tony Z Zhao, Vikash Kumar, Aaron Rovinsky, Kelvin Xu, Thomas Devlin, and Sergey Levine. Reset-free reinforcement learning via multi-task learning: Learning dexterous manipulation behaviors without human intervention. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 6664–6671. IEEE, 2021.

[104] Zhiyuan Xu, Kun Wu, Zhengping Che, Jian Tang, and Jieping Ye. Knowledge transfer in multi-task deep reinforcement learning for continuous control. *Advances in Neural Information Processing Systems*, 33:15146–15155, 2020.

[105] Dmitry Kalashnikov, Jacob Varley, Yevgen Chebotar, Benjamin Swanson, Rico Jonschkowski, Chelsea Finn, Sergey Levine, and Karol Hausman. Mt-opt: Continuous multi-task robotic reinforcement learning at scale. *arXiv preprint arXiv:2104.08212*, 2021.

[106] Stephen Kelly, Tatiana Voegerl, Wolfgang Banzhaf, and Cedric Gondro. Evolving hierarchical memory-prediction machines in multi-task reinforcement learning. *Genetic Programming and Evolvable Machines*, 22(4):573–605, 2021.

[107] Sindhu Padakandla. A survey of reinforcement learning algorithms for dynamically varying environments. *ACM Computing Surveys (CSUR)*, 54(6):1–25, 2021.

[108] Jorge A Mendez, Harm van Seijen, and Eric Eaton. Modular lifelong reinforcement learning via neural composition. In *International Conference on Learning Representations*, 2021.

[109] Tiantian Zhang, Xueqian Wang, Bin Liang, and Bo Yuan. Catastrophic interference in reinforcement learning: A solution based on context division and knowledge distillation. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[110] Christos Kaplanis, Murray Shanahan, and Claudia Clopath. Continual reinforcement learning with complex synapses. In *International Conference on Machine Learning*, pages 2497–2506. PMLR, 2018.

[111] Christos Kaplanis, Murray Shanahan, and Claudia Clopath. Policy consolidation for continual reinforcement learning. *arXiv preprint arXiv:1902.00255*, 2019.

[112] Yat Long Lo and Sina Ghiassian. Overcoming catastrophic interference in online reinforcement learning with dynamic self-organizing maps. *arXiv preprint arXiv:1910.13213*, 2019.

[113] Yizhou Huang, Kevin Xie, Homanga Bharadhwaj, and Florian Shkurti. Continual model-based reinforcement learning with hypernetworks. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 799–805. IEEE, 2021.

[114] Annie Xie, James Harrison, and Chelsea Finn. Deep reinforcement learning amidst lifelong non-stationarity. *arXiv preprint arXiv:2006.10701*, 2020.

[115] Jorge Mendez, Boyu Wang, and Eric Eaton. Lifelong policy gradient learning of factored policies for faster training without forgetting. *Advances in Neural Information Processing Systems*, 33:14398–14409, 2020.

[116] Craig Atkinson, Brendan McCane, Lech Szymanski, and Anthony Robins. Pseudo-rehearsal: Achieving deep reinforcement learning without catastrophic forgetting. *Neurocomputing*, 428:291–307, 2021.

[117] Khimya Khetarpal, Shagun Sodhani, Sarath Chandar, and Doina Precup. Environments for lifelong reinforcement learning. *arXiv preprint arXiv:1811.10732*, 2018.

[118] Maciej Wołczyk, Michał Zajac, Razvan Pascanu, Lukasz Kucinski, and Piotr Miłoś. Continual world: A robotic benchmark for continual reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

[119] Erik C Johnson, Eric Q Nguyen, Blake Schreurs, Chigozie S Ewulum, Chace Ashcraft, Neil M Fendley, Megan M Baker, Alexander New, and Gautam K Vallabha. L2explorer: A lifelong reinforcement learning assessment environment. *arXiv preprint arXiv:2203.07454*, 2022.

[120] Nicoló Lucchesi, Antonio Carta, Vincenzo Lomonaco, and Davide Bacciu. Avalanche rl: A continual reinforcement learning library. In *International Conference on Image Analysis and Processing*, pages 524–535. Springer, 2022.

[121] Ludovic Denoyer, Alfredo de la Fuente, Song Duong, Jean-Baptiste Gaya, Pierre-Alexandre Kamienny, and Daniel H Thompson. Salina: Sequential learning of agents. *arXiv preprint arXiv:2110.07910*, 2021.

[122] Albin Cassirer, Gabriel Barth-Maron, Eugene Brevdo, Sabela Ramos, Toby Boyd, Thibault Sottiaux, and Manuel Kroiss. Reverb: A framework for experience replay. *arXiv preprint arXiv:2102.04736*, 2021.

[123] Sam Powers, Eliot Xing, Eric Kolve, Roozbeh Mottaghi, and Abhinav Gupta. CORA: Benchmarks, baselines, and metrics as a platform for continual reinforcement learning agents. *arXiv preprint arXiv:2110.10067*, 2021.