

# 连续学习方法与其在视觉任务中的应用

方 岩<sup>1</sup>, 魏云超<sup>1\*</sup>, 丛润民<sup>2</sup>, 左旺孟<sup>3</sup>, 赵 耀<sup>1</sup>

(1. 北京交通大学计算机科学与技术学院, 北京 100044; 2. 山东大学控制科学与工程学院, 山东济南 250100;  
3. 哈尔滨工业大学计算机学院, 黑龙江哈尔滨 150001)

**摘 要:** 连续学习广义上指智能算法与智能体学习和适应动态变化世界的能力, 这使得智能算法能够在部署周期中不断获取、更新、积累和利用知识。连续学习技术赋予了智能系统自适应发展的前景和能力。在深度学习中, 连续学习具体指的是能够从非平稳数据流中学习和适应不断变化的训练目标, 这一任务通常面临着灾难性遗忘的挑战, 即学习新任务通常会导致旧任务性能的大幅下降。近年来, 随着深度学习在语言、视觉等诸多领域的迅速发展, 涌现了诸多进展, 有效拓展了对连续学习的理解和应用。本工作对现有连续学习工作进行了较为广泛而深入的调研, 并从连续学习基础定义、代表性方法、在视觉领域的应用等多角度分析。最后, 本文也对连续学习当前的前沿发展和未来研究趋势进行探讨。基于对连续学习领域相关工作的探讨, 期待本文这一综述可以有效促进该领域和后续研究工作的进一步发展和探索。

**关键词:** 连续学习; 持续学习; 计算机视觉; 通用人工智能(AI); 深度学习; 神经网络

**基金项目:** 科技创新2030-新一代人工智能重大专项(No.2021ZD0112100)

**中图分类号:** TP18

**文献标识码:** A

**文章编号:** 0372-2112(2025)06-1713-28

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20240750

第二十七届中国科协年会学术论文

## Continual Learning Methods and Applications in Computer Vision

FANG Yan<sup>1</sup>, Wei Yun-chao<sup>1\*</sup>, Cong Run-min<sup>2</sup>, Zuo Wang-meng<sup>3</sup>, Zhao Yao<sup>1</sup>

(1. School of Computer Science and Technology, Beijing Jiaotong University, Beijing 100044, China;

2. School of Control Science and Engineering, Shandong University, Jinan, Shandong 250100, China;

3. School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China)

**Abstract:** Continual learning generally refers to the ability of intelligent algorithms and agents to learn and adapt to a dynamic and changing world, enabling these algorithms to continually acquire, update, accumulate, and utilize knowledge throughout their deployment cycle. Continual learning technologies endow intelligent systems with the prospects and capabilities of adaptive development. In the context of deep learning, continual learning specifically refers to the ability to learn from non-stationary data streams and adapt to changing training objectives. This task often faces the challenge of catastrophic forgetting, where learning new tasks typically results in a significant decline in performance on previously learned tasks. In recent years, with the rapid development of deep learning in various fields such as language and vision, numerous advancements have emerged, effectively extending the understanding and application of continual learning. This work conducts a relatively extensive and comprehensive survey of existing continual learning research, analyzing it from multiple perspectives including fundamental definition, representative methods, applications in the visual domain. Finally, this paper also discusses the current cutting-edge developments and future research trends in continual learning. Based on the discussion of relevant work in the field of continual learning, we hope this review can effectively promote the development and exploration of this field and subsequent research endeavors.

**Key words:** continual learning; incremental learning; computer vision; general artificial intelligence; deep learning; neural networks

**Foundation Item(s):** Technological Innovation 2030: New Generation Artificial Intelligence Major Project (No.2021ZD0112100)

## 1 引言

近十年来,深度学习技术快速发展,在自然语言处理、计算机视觉、强化学习等诸多领域实现了接近甚至超越人类水平的性能表现. 基于深度学习的智能算法和智能系统应运而生,受到了各界的广泛关注. 学习能力是构建强大智能系统的基础. 为了应对外部变化,进化赋予以人类为代表的智慧生物强大的适应性,使其能够不断获取、更新、积累和利用知识. 受生物智能启发,无数学者尝试以类似的学习范式设计人工智能(Artificial Intelligence, AI)系统. 其中,典型学习范式是串行学习一系列任务,并期望模型在所有任务上达到与并行训练所得的理想性能相当的水平. 这些学习任务可以是新旧技能的新样本、不同环境、不同上下文,以近似特定的现实挑战等. 由于这一范式中学习任务以串行方式逐步出现,连续学习在一些文献中也被称为增量学习(Incremental Learning)或终身学习(LifeLong Learning).

传统机器学习技术建立在建模学习“封闭集合”,即在固定的静态数据分布的前提之下,与其不同的是,连续学习旨在从动态、变化的非平稳数据流分布中学习变化的数据分布. 在这一学习目标下,连续学习面临的主要研究问题是“灾难性遗忘”,即模型适应新的数据分布的代价往往是模型拟合旧数据分布的能力大幅下降. 这一挑战通常被也表述为学习可塑性和记忆稳定性的平衡问题,简称“可塑性-稳定性平衡”:增强算法的(学习)可塑性会影响算法的(记忆)稳定性,反之亦然. 除了在可塑性和稳定性之间取得平衡外,理想的连续学习解决方案还应具备强大的泛化能力,以适应任务内和任务间的分布差异,以适应变化的、动态的真实世界环境. 针对上述连续学习目标,最直接且简单的方法是在条件允许的情况下,重新训练所有旧的训练样本. 然而,这种方法的代价是巨大的计算和存储开销,同时还会涉及潜在的数据隐私问题. 进一步地,在理想条件下,最佳的连续学习算法应当只使用新的训练样本进行学习,这对连续学习方法的设计提出了新的要求.

当前,学术界针对上述问题和挑战,不断涌现出诸多工作,这些工作主要可以归纳到四类方法的范畴中.(1)基于正则化的方法:添加正则化项;(2)基于样本重训练的方法:近似和恢复旧任务数据分布;(3)基于优化的方法:显式实现优化项;(4)基于表示学习的方法:学习稳健且泛化良好的特征表示. 基于对现有连续学习方法的大致分类,本文通过对相关工作进行简要介绍,对不同方法进行了概括,旨在为读者构建连续学习的大致版图.

同时,本文关注到连续学习在一些具体应用中也

遇到了一些特定的挑战:目前连续学习的进展主要集中在视觉分类任务上,但在如计算机视觉领域(如目标检测、语义分割与图像生成)的应用中体现出具体的独特问题,这些领域中的应用有着相应的机会和挑战. 基于这些领域内的连续学习发展与应用,本文相应地对计算机视觉领域中的连续学习问题进行简要梳理与探究. 进一步地,近期以大语言模型、多模态模型、扩散式图像生成模型等为代表的新兴领域中的连续学习也开始逐渐活跃起来,因此本文也对这些新兴研究方向中连续学习的应用进行讨论和展望.

当前,已有一些优秀中文综述对连续学习进行大致梳理,如文献[1,2]对连续学习工作按方法类型进行细粒度分类,文献[3]从脑启发的类脑连续学习角度对现有工作进行梳理,文献[4]从人脑记忆机制出发对灾难性遗忘现象进行深入分析并从理论、方法、策略等多角度进行了有效改进和总结. 也有文献[5]同样关注到视觉领域中的连续学习任务并结合基础方法进行了简单探索,但有失系统完备. 相较于这些已有文献,本文展现了更系统具体的持续学习理论与方法梳理,以及连续学习在视觉任务及前沿应用领域中的应用与探索. 此外,与现已发表的多篇前沿连续学习领域综述<sup>[6-8]</sup>相比,本文将关注的领域范围从简单地关注图像分类任务拓展到以目标检测、图像分割、条件生成等多类计算机视觉领域任务为代表的复杂挑战任务中去,增强连续学习与不同视觉任务的关联. 与最新领域前沿的综述<sup>[9]</sup>相比,本文更关注连续学习在视觉领域不同任务中的发展与应用,而不刻意强调连续学习与其他机器学习范式联系的详细讨论,使本文更聚焦于连续学习在视觉领域的应用讨论,对相应内容的展现更具体充实.

综上,本文从基本定义、主要方法、计算机视觉领域中的应用,以及前沿讨论与展望四项主要内容出发,对连续学习这一研究问题进行简要介绍和分析,期望本工作能够为读者构建起一个相对完善的连续学习认识,对后续研究工作有所启发.

## 2 任务定义

### 2.1 基础定义

连续学习不同于一般深度学习任务的重要特点是需要从动态数据流分布中学习. 在连续学习实践中,不同数据分布的训练样本按顺序到达. 在这一动态数据流分布的任务定义中,连续学习模型需要串行学习多个任务,并要求在相应测试集上有较好性能. 一般地,在连续学习任务中,模型通常只能部分访问或不可访问旧任务的训练样本. 形式上,属于任务 $t$ 的一个新到达的批量训练样本可以表示为 $D_{t,b} = \{X_{t,b}, Y_{t,b}\}$ ,其中

$X_{t,b}$  为输入数据,  $Y_{t,b}$  为对应输入数据的标签,  $t \in T = \{1, 2, \dots, k\}$  为任务标识,  $b \in B_t$  是训练批次索引 ( $T$  和  $B_t$  分别表示其对应的空间). 本文通过任务  $t$  的训练样本  $D_t$  来定义一个任务服从的数据分布  $D_t = \{X_t, Y_t\}$ . 当省略批次索引  $b$  时,  $D_t$  表示整个训练数据集, 这一表示同样地适用于  $X_t$  和  $Y_t$ . 一般情况下, 假设具体任务的训练样本和测试样本的分布没有差异. 在连续学习中, 每个任务的训练样本可以增量地以批量形式到达, 即

$$\{\{D_{t,b}, b \in B_t\}, t \in T\} \quad (1)$$

或完全同时到达, 即

$$\{D_t, t \in T\} \quad (2)$$

特殊地, 在现实应用约束下, 数据标签  $Y_t$  和任务标识  $t$  并不总是可获取的, 这也为连续学习任务衍生出了一系列具体细分.

## 2.2 连续学习类型分类

本文根据增量学习任务中增量任务批次  $b$  的划分和任务标识  $t$  的可获取性, 将现有主要的增量学习任务形式归纳细分为下述几类:

(1) 域增量学习. 不同任务具有相同的数据标签空间  $Y$ , 但输入分布  $X$  不同, 无需任务标识  $t$ . 这一增量学习任务的特点在于, 所有任务数据具有统一的类别空间  $Y$ , 但各任务的数据分布、数据形式 (如风格、环境分布) 不同.

(2) 任务增量学习. 不同任务具有非重叠的数据标签空间  $Y$ . 在训练和测试阶段均提供任务标识  $t$ . 在任务增量学习中, 每一个任务在训练和测试阶段都具有对应的任务标识  $t$  以及每个任务对应的类别标签空间  $Y_t$ , 且不同任务的类别标签空间完全不同, 即  $Y_{t_1} \cap Y_{t_2} = \emptyset, \forall t_1 \neq t_2$ . 在测试阶段, 每个测试样本均会提供对应任务标识  $t$ , 模型可根据数据的任务标识, 在数据所属任务对应的类别空间  $Y_t$  上进行预测.

(3) 类别增量学习. 不同任务具有非重叠的数据标签空间  $Y$ . 任务标识  $t$  仅在训练阶段提供. 不同于任务增量学习, 在类别增量学习中, 每一个任务仅在训练中具有对应的任务标识  $t$ , 且不同任务的类别标签空间完全不同, 即  $Y_{t_1} \cap Y_{t_2} = \emptyset, \forall t_1 \neq t_2$ . 在测试阶段, 测试数据不具有任务标识  $t$  信息, 模型需要在完整的类别标签空间  $Y$  上进行预测.

(4) 实例增量学习. 所有训练样本都属于同一任务, 以批量形式到达. 在实例增量学习中, 每个样本依次到达, 模型对每个实例依次学习. 该增量学习任务形式一般少见.

(5) 在线增量学习. 不同任务具有非重叠的数据标签空间  $Y$ . 每个任务的训练样本在数据流中单次通过. 与实例增量学习不同, 在线连续学习中的训练样本具

有明确的任务和标签空间属性.

一般假设每个增量学习任务通常都具有足够数量的标记训练样本, 即全监督连续学习. 根据每个任务训练数据  $D_t$  中提供的  $X_t$  和  $Y_t$ , 连续学习可以进一步分为零样本学习<sup>[10,11]</sup>、小样本学习<sup>[12]</sup>、半监督学习<sup>[13]</sup>、开放世界 (即识别未知类别后合并未知标签)<sup>[14,15]</sup> 和无监督/自监督学习<sup>[16,17]</sup> 等不同数据场景. 此外, 近期一些连续学习工作还考虑并融合了其他实际挑战, 例如, 多标签<sup>[18]</sup>、噪声标签<sup>[19]</sup>、域对齐<sup>[20]</sup>、域迁移<sup>[21]</sup>、即时推理<sup>[22]</sup>、新类挖掘<sup>[23]</sup>、多模态任务<sup>[24]</sup> 等. 值得注意的是, 一些学者认为通过对这些不同场景进行组合可以更好地模拟现实世界中智能系统所面临挑战的复杂性.

上文中的分类主要基于数据分布特点和任务标识的可见性, 讨论不同学术流派对连续学习任务的不同努力和尝试. 然而, 灾难性遗忘这一核心挑战在这些不同任务中广泛存在, 并具有较大影响, 这与上述任务形式无关. 即上述不同类型连续学习任务不仅要处理不同场景下的特定挑战, 还要给出克服灾难性遗忘的解决方案.

## 2.3 评价指标

一般来说, 连续学习的性能可以从三个方面进行评估: 目前为止所学习的任务总体性能、对旧任务的记忆稳定性以及对新任务的学习可塑性. 在该节中, 本文以分类任务为例, 总结了常见的评估指标. 在第  $t$  个增量任务阶段, 增量分类算法的性能可通过当前所有已知类别上的准确率  $A_t$  进行评估.

所有任务总性能一般可以由所有任务平均准确率 (Average Accuracy, AA) 以及增量任务平均准确率 (Average Incremental Accuracy, AIA) 评估. 以  $a_{k,j} \in [0, 1]$  表示在学习第  $k$  个任务后, 连续学习模型在第  $j$  ( $j \leq k$ ) 个任务测试集上的分类准确率. 则在第  $k$  个任务上的总性能可以通过下述公式计算得到:

$$AA_k = \frac{1}{k} \sum_{j=1}^k a_{k,j} \quad (3)$$

$$AIA_k = \frac{1}{k} \sum_{i=1}^k AA_i \quad (4)$$

其中, AA 为在当前时刻增量模型的总性能; AIA 进一步反映出所有增量任务上的历史平均性能.

记忆稳定性可以通过遗忘度 (Forgetting Measure, FM) 和后向迁移 (Backward Transfer, BWT) 两类指标度量. FM 用于评估增量模型在先前学习的第  $j$  个任务上的性能下降情况, 可通过计算先前该任务的最优性能与当前模型在该任务上的性能之差来得到, 即

$$f_{j,k} = \max_{i \in \{1, 2, \dots, k-1\}} (a_{i,j} - a_{k,j}), \forall j < k \quad (5)$$

而当前时刻模型在所有旧任务上的总体平均 FM 的计算式为

$$FM_k = \frac{1}{k-1} \sum_{j=1}^{k-1} f_{j,k} \quad (6)$$

而 BWT 指标则直接评估模型在学习第  $k$  个任务后对所有旧任务性能的平均影响,计算式为

$$BWT_k = \frac{1}{k-1} \sum_{j=1}^{k-1} (a_{k,j} - a_{j,j}) \quad (7)$$

近期的连续学习任务更倾向使用 BWT 指标评估模型的记忆稳定性。

学习可塑性通常用不变性指标 (Intransience Measure, IM) 和前向迁移指标 (ForWard Transfer, FWT) 衡量。IM 定义为模型学习新任务的能力,通过某项任务的联合训练性能与连续学习性能之间的差异计算得到,即

$$IM_k = a_k^* - a_{k,k} \quad (8)$$

其中,  $a_k^*$  为在所有任务  $\{D_j, j=1, 2, \dots, k\}$  上联合训练 (joint training) 所得的参考模型在第  $k$  个任务的性能。与 IM 不同, FWT 评估所有旧任务对学习当前第  $k$  个任务的影响,有

$$FWT_k = \frac{1}{k-1} \sum_{j=2}^k (a_{j,j} - \tilde{a}_j) \quad (9)$$

其中,  $\tilde{a}_j$  为在第  $j$  个任务的数据集  $D_j$  上训练得到的参考模型的分类准确率。

考虑到评价连续学习算法应当同时考虑模型的记忆稳定性和学习可塑性,本文提出一种新的连续学习算法综合评估指标 (Continual learning metric, CI), 在第  $k$  个任务时,通过下式计算这一指标:

$$CI_k = \lambda \frac{a_{k,k-1}}{a_{k,k-1}^*} + (1-\lambda) \frac{a_{k,k}}{a_{k,k}^*} \quad (10)$$

$$\lambda = \frac{|D_{1,2,\dots,k-1}|}{|D_{1,2,\dots,k-1}| + |D_k|} \quad (11)$$

其中,  $a_{k,k-1}^*$  为模型在完成最新的第  $k$  个任务学习后,其在上一个任务 (即  $k-1$ ) 数据上的性能上界,即理想条件下所能达到的最优抗遗忘性能;  $a_{k,k}^*$  为模型在完成最新的第  $k$  个任务学习后,其在当前任务数据上的性能上界,即理想条件下所能达到的最优迁移性能。这两项指标通常可以通过联合训练得到的模型测得。使用这两项分别对模型在当前阶段测得的遗忘性能和迁移性能进行数值归一化,使该指标可有效适配多种连续学习任务。通过上述计算方式,可计算得到  $k$  时刻归一化的模型遗忘性能指标与迁移性能指标,并对两者使用基于旧任务数据规模  $|D_{1,2,\dots,k-1}|$  与  $k$  时刻总数据规模  $|D_{1,2,\dots,k-1}| + |D_k|$  之比得到的系数加权求和,从而得到最终的当前时刻的模型连续学习性能评估  $CI_k$ 。当模型完成所有任务学习时,对每一任务时刻计算得到的指

标进行平均,得到模型的最终连续学习能力评估

$$CI = \sum_{k=1}^{|T|} CI_k \quad (12)$$

相较于前述对记忆稳定性或学习可塑性进行单一评估的 FM、BWT、IM、FWT 等指标,本文提出的新指标 CI 实现了对两者的同时评估与均衡考量。通过使用理论性能上界进行归一化,CI 可以验证模型性能提升的具体来源——任务难度的降低,还是模型连续学习能力的增益。这一过程减少了模型连续学习能力评估的复杂性。此外,数值归一化过程将不同任务上的指标数值统一到  $[0, 1]$  区间,使得不同任务上不同模型的连续学习能力具有可比性,因此具有广泛的统一应用价值。

无论使用上述何种连续学习评测指标,都离不开具体下游任务的评价指标  $a_{k,j}$ , 如用于目标检测任务的平均精确度 (Average Precision, AP) 指标,用于图像分割任务的交并比 (Intersection-over-Union, IoU) 指标,用于图像生成任务的 FID (Fréchet Inception Distance) 指标等。

### 3 方法分类

本节主要关注现有工作在解决连续学习核心问题与挑战 (灾难性遗忘) 时的不同解决思路,而这些思路与具体的连续学习任务类型无关。本节旨在对典型的、有代表性的连续学习方法进行整合归类,并系统性分析与介绍这些方法在解决灾难性遗忘问题时的基本思路、代表性工作和主要特点。所涉及的方法具有通用意义。在介绍不同方法分类后,本节末尾还对不同类型连续学习方法的特点、优缺点进行了总结。连续学习方法分类与代表性论文汇总见表 1,连续学习方法的经纬与分类如图 1 所示。

#### 3.1 数据重放方法

连续学习任务通常假设模型在学习新任务时无法访问旧任务的数据。然而,最近一些工作尝试在学习新任务时允许访问存储的少量旧任务样本数据,并同时使用新任务数据和存储的旧任务数据进行模型训练和参数更新。这类方法通常被称为数据重放 (data replay 或 data rehearsal)。根据数据重放方法的数据来源和重放机制的不同,这一类方法可以进一步细分为直接数据回放、生成式回放和特征回放。

直接数据重放方法 (又称经验重放) 通常在一个人设置为设置的小型、有限记忆缓冲区中存储一些旧任务的训练样本。然而,这类方法在实际应用中面临多方面的限制。一方面,有限的存储空间只能存储少量旧任务样本;另一方面,由于数据隐私和安全问题,直接存储旧任务样本在部分敏感应用场景中并不适用。在这类方

表 1 连续学习方法分类与代表性论文汇总

方法分类	代表方法
正则化方法	参数正则化 EWC <sup>[25]</sup> , SI <sup>[26]</sup> , MAS <sup>[27]</sup> , RWalk <sup>[28]</sup> , R-EWC <sup>[29]</sup> , XK-FAC <sup>[30]</sup> , ALASSO <sup>[31]</sup> , IMM <sup>[32]</sup> , ResCL <sup>[33]</sup> , P&C <sup>[34]</sup> , AFEC <sup>[35]</sup> , VCL <sup>[36,37]</sup> , NVCL <sup>[38]</sup> , CLAW <sup>[39]</sup> , GVCL <sup>[40]</sup> , KCL <sup>[41]</sup> , VAR-GPs <sup>[42]</sup> , NPC <sup>[43]</sup> , UCL <sup>[44]</sup> , AGS-CL <sup>[45]</sup>
	函数正则化 LwF <sup>[46]</sup> , LwM <sup>[47]</sup> , GD <sup>[48]</sup> , EEIL <sup>[49]</sup> , LUCIR <sup>[50]</sup> , PODNet <sup>[51]</sup> , DER <sup>[52]</sup> , FRCL <sup>[53]</sup> , FROMP <sup>[54]</sup> , S-FSVI <sup>[55]</sup>
数据重放	经验重放 iCaRL <sup>[56]</sup> , GSS <sup>[57]</sup> , CCBO <sup>[58]</sup> , OCS <sup>[59]</sup> , ASER <sup>[60]</sup> , RM <sup>[61]</sup> , AQM <sup>[62]</sup> , MRDC <sup>[63]</sup> , RAR <sup>[64]</sup>
	生成重放 DGR <sup>[65]</sup> , MeRGAN <sup>[66]</sup>
	特征重放 GFR <sup>[67]</sup> , FA <sup>[68]</sup> , DSR <sup>[69]</sup> , IL2M <sup>[70]</sup> , SNCL <sup>[71]</sup> , RER <sup>[72]</sup> , REMIND <sup>[73]</sup> , ACAE-REMIND <sup>[74]</sup> , FeTrIL <sup>[75]</sup>
优化方法	梯度投影 OWM <sup>[76]</sup> , AOP <sup>[77]</sup> , OGD <sup>[78]</sup> , GPM <sup>[79]</sup> , CGP <sup>[80]</sup> , FS-DGPM <sup>[81]</sup> , CUBER <sup>[82]</sup> , TRGP <sup>[83]</sup> , Adam-NSCL <sup>[84]</sup> , AdNS <sup>[85]</sup> , NCL <sup>[86]</sup> , RGO <sup>[87]</sup>
	损失函数 OML <sup>[88]</sup> , ANML <sup>[89]</sup> , AIM <sup>[90]</sup> , MER <sup>[91]</sup> , iTAML <sup>[92]</sup> , La-MAML <sup>[93]</sup> , OSAKA <sup>[94]</sup> , MERLIN <sup>[95]</sup> , PR <sup>[96]</sup> , MARK <sup>[97]</sup> , ARI <sup>[98]</sup>
	元学习 Stable-SGD <sup>[99]</sup> , MC-SGD <sup>[100]</sup> , Linear Connector <sup>[101]</sup>
特征表示方法	自监督学习 LUMP <sup>[102]</sup> , MinRed <sup>[103]</sup> , CaSSLe <sup>[104]</sup> , Co2L <sup>[105]</sup> , DualNet <sup>[106]</sup>
	下游预训练 Side-Tuning <sup>[107]</sup> , DLCFT <sup>[108]</sup> , TwF <sup>[109]</sup> , GAN-Memory <sup>[110]</sup> , FiLM <sup>[111]</sup> , AdaFM <sup>[112]</sup> , ADA <sup>[69]</sup> , L2P <sup>[113]</sup> , CODA-Prompt <sup>[114]</sup> , Dual-Prompt <sup>[115]</sup> , S-Prompts <sup>[116]</sup> , HiDe-Prompt <sup>[117]</sup> , Progressive Prompts <sup>[118]</sup> , CwD <sup>[119]</sup> , F2M <sup>[120]</sup> , SAM <sup>[121,122]</sup>
	持续预训练 IncCLIP <sup>[123]</sup> , ECONET <sup>[124]</sup> , IDA <sup>[125]</sup> , ORDER <sup>[126]</sup> , CTP <sup>[127]</sup> , DAP <sup>[128]</sup>

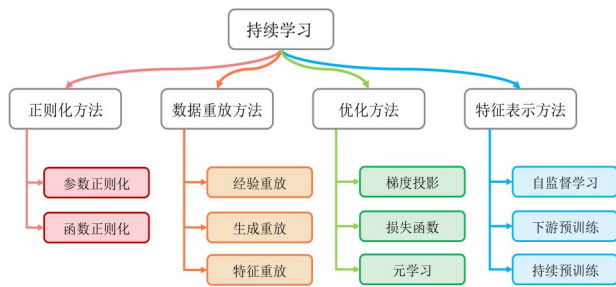


图 1 连续学习方法的经纬与分类

法的实现中, iCaRL<sup>[56]</sup>首次提出了基于直接数据重放的连续学习方法, 在每个增量任务  $D_t$  学习完成后, iCaRL 为每一类别存储固定数量的代表性样本. iCaRL 选择距离每类别平均样本向量(也称类别原型)最近的样本进行存储和重放. 然而, 随着任务数量增加, 这一类方法所需的内存缓冲大小也相应增加. 这一特性使得如何在有限存储空间条件下, 有效构建和利用内存缓冲区成为直接数据重放方法所面临的关键挑战. 同时, 为了准确有效地恢复旧任务数据分布, 数据重放方法需要有效选择、压缩、增强和更新所保留的旧训练样本. 早期的方法采用固定的样本选择原则. 例如, 采用蓄水池采样策略(reservoir sampling)<sup>[91, 129, 130]</sup>从每个训练批次中随机保留固定数量的旧训练样本; 采用环形缓冲区策略(ring buffer)<sup>[131]</sup>确保每个类别的旧训练样本数量相等; iCaRL 选择与每一类别的特征均值最接近的旧训练样本. 然而, 这些方法的性能表现在实际应用任务中并不理想, 基于梯度或可优化的策略具有更优越的性能, 例如, 通过最大化样本多样性基于梯度的样本选择(Gradient-based Sample Selection, GSS)<sup>[57]</sup>、任务性能约束通过双层优化实现的约束核心集(Constrained Coresets via Bilevel Optimization, CCBO)<sup>[58]</sup>、梯度相似性和跨批次梯度多样性在线核心集选择(Online Coreset Selec-

tion, OCS)<sup>[59]</sup>、优化潜在决策边界的能力对抗性 Shapley 值(Adversarial Shapley value, ASER)<sup>[60]</sup>和对抗扰动的鲁棒性多样性彩虹记忆(Rainbow Memory, RM)<sup>[61]</sup>等度量指标选择用于记忆的旧任务样本. 另一方面, 为了提高存储效率, 自适应量化模块(Adaptive Quantization Modules, AQM)<sup>[62]</sup>基于向量量化变分自编码器(Vector Quantized-Variational AutoEncoder, VQ-VAE)架构<sup>[132]</sup>实现在线持续压缩, 并保存压缩数据进行数据重放. 带数据压缩的记忆重放(Memory Replay with Data Compression, MRDC)<sup>[63]</sup>将数据压缩的经验重放公式化为行列式点过程(Determinantal Point Processes, DPPs)<sup>[133]</sup>, 并推导得到一种计算效率较高的在线最优压缩率确定方法. RM 沿袭先前方法, 采用基于标签混合的数据增强策略, 增强旧训练样本的多样性. 回顾性对抗重放(Retrospective Adversarial Replay, RAR)<sup>[64]</sup>在遗忘边界附近合成对抗样本, 并结合数据增强策略实现更优效果. 当前, 如何有效构建旧样本记忆缓冲池以及选择最优的旧任务样本, 仍是此类方法的重难点问题. 典型数据重放方法范式如图 2 所示, 可通过直接样本存储或生成模型实现样本重放训练.

直接数据回放方法通过直接存储和回放部分旧任

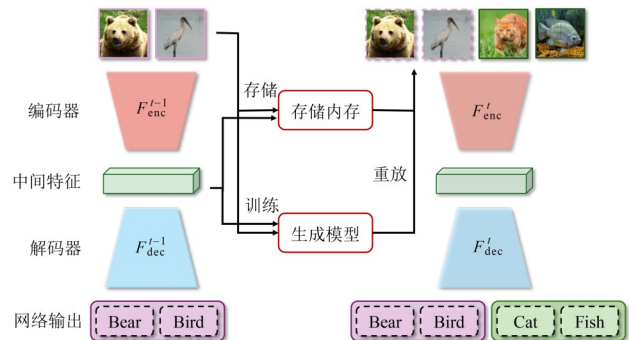


图 2 典型数据重放方法范式

务的真实数据解决灾难性遗忘的问题,但在隐私敏感的应用领域并不适用,同时旧样本存储带来的存储开销也不可忽视.为了解决这些问题,生成式数据回放方法应运而生.此类方法通过训练额外的生成模型生成重放数据,同时生成模型本身也需要进行增量更新.深度生成重放(Deep Generative Replay, DGR)<sup>[65]</sup>提出了一个简单的实现框架,即当训练每个任务的生成模型时,都会从旧生成模型中生成数据进行重放训练,以此来维持对旧任务记忆的稳定性.记忆重放生成对抗网络(Memory Replay GANs, MeRGAN)<sup>[66]</sup>进一步采用重放对齐确保旧生成模型和新生成模型生成数据的一致性,实现生成模型的稳定增量更新.不同于前两者,基于扩散的深度生成重放(Deep Diffusion-based Generative Replay, DDGR)<sup>[134]</sup>提出使用扩散生成网络 Diffusion,利用分类任务与条件生成任务在 Diffusion 框架下的相互促进关系,实现了在两项任务上的互补增强效果.为解决在增量过程中生成样本质量降低的问题,基于梯度投影的原型扩散模型(Gradient Projection-based Prototype Diffusion Model, GPPDM)<sup>[135]</sup>在 DDGR 工作基础上提出使用可学习的类别原型(class prototypes)增强 diffusion 模型的旧任务样本生成质量.近期基于引导的增量学习与扩散模型(GUIDancE-based incremental learning with diffusion models, GUIDE)<sup>[136]</sup>进一步扩展上述使用 Diffusion 的生成式重放思路,通过将 Diffusion 模型分类器引导融入扩散过程,定制化生成高质量的分类器边界决策样本,增强抗遗忘效果.

特征回放不直接存储原始旧任务训练样本,而是存储旧任务样本的中间特征,在效率和隐私方面具有部分优势.由于所使用的样本特征提取网络在增量学习过程中同样不断更新,特征表示分布发生迁移不可避免,因此该方法仍面临灾难性遗忘的挑战.基于这一挑战,生成特征重放(Generative Feature Replay, GFR)<sup>[67]</sup>、特征适应(Feature Adaptation, FA)<sup>[68]</sup>和动态自维持表示(Dynamic Self-sustaining Representation, DSR)<sup>[69]</sup>在旧模型和新模型之间进行特征蒸馏以保持特征提取网络在不同任务阶段的记忆稳定性.双记忆增量学习(Incremental Learning with dual Memory, IL2M)<sup>[70]</sup>和面向连续学习的语法感知网络(Syntax-aware Network for Continual Learning, SNCL)<sup>[71]</sup>基于经验重放恢复特征表示的均值和协方差等统计信息以保持特征提取网络的记忆稳定性.表示演化与重放(Representation Evolution and Replay, RER)<sup>[72]</sup>通过显式估计特征表示迁移更新旧特征.提醒(REMIND)<sup>[73]</sup>和辅助分类器自编码器用于 REMIND (Auxiliary Classifier Auto-Encoder for REMIND, ACAE-REMIND)<sup>[74]</sup>固定特征提取器的输入层,并重建中间特征以更新近输出层.特征转换用于增

量学习(Feature Translation for Incremental Learning, FeTrIL)<sup>[75]</sup>从初始任务中训练得到冻结的特征提取网络,在后续的任务中对提取的特征进行重放.

### 3.2 正则化方法

正则化方法在损失函数中添加显式的正则化项,以平衡记忆旧任务和学习新任务.这一类方法通常需要存储旧模型的冻结参数副本作为参考模型参数.根据正则化项的作用对象,这类方法可以细分为权重正则化和函数正则化.典型正则化方法范式如图3所示,主要通过参数正则化和函数正则化两类方式实现.

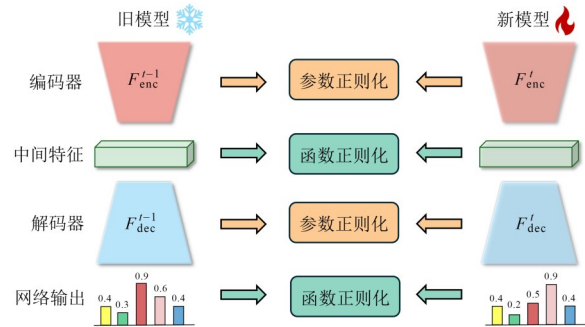


图3 典型正则化方法范式

一些研究对二次罚项的实现进行了改进.由于弹性权重巩固(Elastic Weight Consolidation, EWC)<sup>[25]</sup>中的费舍尔信息矩阵(Fisher Information Matrix, FIM)对角线近似可能会失去关于旧任务的信息,旋转弹性权重巩固(Rotated Elastic Weight Consolidation, R-EWC)<sup>[29]</sup>通过参数空间的对角化旋转对 FIM 进行分解.扩展的 Kronecker 分解近似曲率(eXtended Kronecker-Factored Approximate Curvature, XK-FAC)<sup>[30]</sup>在近似 FIM 时进一步考虑了样本间的关系,以更好地适应批量归一化.单侧过估计的非对称损失近似(Asymmetric Loss Approximation with Single-Side Overestimation, ALASSO)<sup>[31]</sup>设计了一种非对称的二次罚项,对 FIM 矩阵一侧使用较大罚项,以应对参数变化对旧任务的非对称影响.为了在旧模型的约束下学习每个新任务同时避免记忆的不稳定,“扩展-重归一化”过程被证明可以提供更好的“稳定性-可塑性”权衡<sup>[32-35, 101]</sup>.增量矩匹配(Incremental Moment Matching, IMM)<sup>[32]</sup>通过增量匹配旧任务和新任务的后验分布的矩估计,即新旧任务上参数的加权平均值来实现这一目标.残差连续学习(Residual Continual Learning, ResCL)<sup>[33]</sup>通过可学习的组合系数对这一想法进行扩展.进展与压缩(Progress & Compress, P&C)<sup>[34]</sup>使用额外的网络单独学习每个任务,然后使用 EWC 的通用正则化项将其蒸馏回旧模型.主动遗忘负迁移(Active Forgetting of negative transfer, AFEC)<sup>[35]</sup>引入了一个遗忘率以消除原始后验概率  $p(\theta|D_{1:k-1})$  带来的潜

在负向迁移,并通过二次项来惩罚网络参数与旧任务和新任务解的差异.为了可靠地均衡旧任务和新任务上的增量模型参数,一些工作通过将模型参数约束在线性低误差路径上来构建线性连接器.针对网络本身的其他形式的正则化方法,也属于这一方向.例如,后验分布的在线估计可以作为参数变化的隐式正则化,如变分连续学习(Variational Continual Learning, VCL)<sup>[36,37]</sup>、自然变分连续学习(Natural Variational Continual Learning, NVCL)<sup>[38]</sup>、自适应权重连续学习(Continual Learning with Adaptive Weights, CLAW)<sup>[39]</sup>、广义变分连续学习(Generalized Variational Continual Learning, GVCL)<sup>[40]</sup>、核连续学习(Kernel Continual Learning, KCL)<sup>[41]</sup>和变分自回归高斯过程(Variational Auto-Regressive Gaussian Processes, VAR-GPs)<sup>[42]</sup>.与聚合参数不同,神经元级可塑性控制(Neuron-level Plasticity Control, NPC)<sup>[43]</sup>估计每个神经网络单元的重要性,并有选择地降低其学习率.基于不确定性的连续学习(Uncertainty-based Continual Learning, UCL)<sup>[44]</sup>和自适应组稀疏连续学习(Adaptive Group Sparse Continual Learning, AGS-CL)<sup>[45]</sup>冻结连接重要神经元的参数,即相当于权重正则化的硬改进.

总的来说,这些方法主要关注如何通过调整网络参数或结构进行正则化来平衡记忆稳定性和学习可塑性.不同的方法在不同应用场景和任务中表现出各自优势和局限性,需要根据具体需求进行选择和改进.

函数正则化针对增量模型的中间特征或最终预测施加正则化项.这种策略通常使用先前训练的模型作为教师网络,当前训练的模型作为学生网络,通过知识蒸馏(Knowledge Distillation)<sup>[137]</sup>来缓解灾难性遗忘.理想情况下,知识蒸馏应当可使用所有旧任务样本,但受限于连续学习任务场景中对旧样本的有限可获取限制,这一方法并不可行.替代的实现方案包括使用新训练样本<sup>[46,68,138,139]</sup>、一小部分旧训练样本<sup>[49-51,56]</sup>、外部未标记数据<sup>[48]</sup>或生成数据<sup>[66,140]</sup>,这些数据在不同程度上存在分布偏移.无遗忘学习(Learning without Forgetting, LwF)<sup>[46]</sup>是该方向的开创性工作,其通过使用旧任务输出层的预测来计算蒸馏损失函数值,学习新训练样本.无记忆学习(Learning without Memorizing, LwM)<sup>[47]</sup>利用新训练样本的注意力分数进行知识蒸馏.EBLL学习特定任务的自编码器,并防止特征表示发生变化.梯度下降(Gradient Descent, GD)<sup>[48]</sup>使用可使用的未标记数据流进一步蒸馏知识.当旧训练样本被有效恢复时,函数正则化的作用可以被进一步增强.基于这一发现,函数正则化经常与重放少量旧训练样本相结合,如增量分类器与表示学习(incremental Classifier and Representation Learning, iCaRL)<sup>[56]</sup>、端到端增量学习

(End-to-End Incremental Learning, EEIL)<sup>[49]</sup>、通过再平衡增量学习统一分类器(Learning a Unified Classifier Incrementally via Rebalancing, LUCIR)<sup>[50]</sup>、池化输出蒸馏网络(Pooled Outputs Distillation Network, PODNet)<sup>[51]</sup>、暗经验重放(Dark Experience Replay, DER)<sup>[52]</sup>等.另一方面,函数空间上的序列化贝叶斯推理也可以视为一种函数正则化形式,但这一实现通常需要存储少量旧训练样本,如功能正则化连续学习(Functional Regularisation for Continual Learning, FRCL)<sup>[53]</sup>、旧能力遗忘正则化(Functional Regularisation Of Memorable Past, FROMP)<sup>[54]</sup>和序贯函数空间变分推断(Sequential Function-Space Variational Inference, S-FSVI)<sup>[55]</sup>.对于条件生成重放,先前学习的样本及旧模型和新模型输出的生成数据之间需要进行一致性正则化惩罚,如记忆重放生成对抗网络(Memory Replay Generative Adversarial Networks, MeRGANs)<sup>[66]</sup>、动态检索与想象(Dynamic Retrieval and Imagination, DRI)<sup>[141]</sup>和终身生成对抗网络(Lifelong Generative Adversarial Networks, LifelongGAN)<sup>[140]</sup>.

### 3.3 基于优化的方法

克服灾难性遗忘,实现连续学习,不仅可以通过在损失函数中添加额外的惩罚项实现,还可以通过基于优化理论的分析指导实现.常见的基于优化理论的连续学习主要可以进一步分为梯度投影法、元学习方法及损失函数平滑度方法.优化方法范式如图4所示.梯度投影类方法通过对梯度方向正交分解实最优方法优化;损失函数平滑度鼓励找到所有任务最优区域.

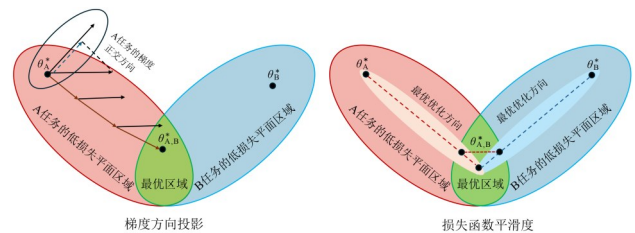


图4 优化方法范式

一种基于优化的典型方法是梯度投影法.这一类方法通过对梯度更新方向的显式约束,缓解由于模型训练过程中由于参数更新引起的旧任务性能下降.其中,基于数据重放的方法(如梯度情景记忆(Gradient Episodic Memory, GEM)<sup>[131]</sup>、平均梯度情景记忆(Averaged Gradient Episodic Memory, A-GEM)<sup>[142]</sup>、基于梯度分解的逐层优化(Layerwise Optimization by Gradient Decomposition, LOGD)<sup>[143]</sup>和情景记忆重放(Memory Episodic Replay, MER)<sup>[91]</sup>)将参数更新方向投影到与经验重放对齐的方向上,部分等效于使用部分旧任务训练样本,以实现保留旧任务的输入空间和梯度空间.与上述重放并对齐旧任务方向不同,正交权重调制(Or-

thogonal Weight Modulation, OWM)<sup>[76]</sup>和自适应正交投影(Adaptive Orthogonal Projection, AOP)<sup>[77]</sup>将参数更新方向调整为旧任务输入空间的正交方向. 正交梯度下降(Orthogonal Gradient Descent, OGD)<sup>[78]</sup>保留旧任务的梯度更新方向,并约束当前学习任务的梯度更新方向与旧任务正交,减轻新任务参数更新对旧任务性能的影响. 正交子空间方法<sup>[144]</sup>进一步通过正交的低秩向量子空间进行连续学习,并保持不同任务的梯度相互正交. 例如,梯度投影记忆(Gradient Projection Memory, GPM)<sup>[79]</sup>在更新参数时维护对旧任务有较高重要度的梯度子空间(即核心梯度空间的基向量),以实现正交投影. 类别梯度投影(Class Gradient Projection, CGP)<sup>[80]</sup>从各个学习类别计算这一核心梯度子空间,从而进一步减轻不同任务不同类别间干扰. 动态梯度投影记忆的平滑锐度(Flattening Sharpness for Dynamic Gradient Projection Memory, FS-DGPM)<sup>[81]</sup>动态释放梯度投影记忆(Gradient Projection Memory, GPM)<sup>[79]</sup>中不重要的基向量,以提高学习可塑性,同时鼓励模型收敛到平坦的损失函数超平面. 具有反向知识传递和正向知识保持的连续学习(Continual learning with Backward knowledge transfer and forward knowledge preservation, CUBER)<sup>[82]</sup>选择性地投影梯度,更新与当前任务正相关的旧任务知识. 信任域梯度投影(Trust Region Gradient Projection, TRGP)<sup>[83]</sup>通过对先前输入的子空间的梯度投影范数来定义“可信区域”,从而选择性地重用旧任务的冻结权重. 带零空间约束的Adam优化器用于连续学习(Adam with Null Space Constraints for continual Learning, Adam-NSCL)<sup>[84]</sup>将候选参数更新投影到由旧任务的非中心特征协方差近似的零向量空间中,而自适应零空间(Adaptive Null Space, AdNS)<sup>[85]</sup>考虑在前一个任务和当前任务零向量空间的共享部分进行参数更新. 此外,自然连续学习(Natural Continual Learning, NCL)<sup>[86]</sup>统一了贝叶斯权重正则化和梯度投影,鼓励在旧任务的零空间中进行参数更新,并收敛到贝叶斯近似后验的最大值. 递归梯度优化(Recursive Gradient Optimization, RGO)<sup>[87]</sup>在贝叶斯权重正则化的二次惩罚上限下,使用递归优化方法修改梯度方向以获得最优解. 正则化和样本重放最终都是通过修正当前的梯度方向来实现的,而梯度投影则是在参数更新过程中显式地直接进行的.

另一类代表性方法是元学习方法(meta-learning). 此类方法试图为各种任务场景获取数据驱动的通用归纳偏置<sup>[145]</sup>,捕捉任务之间的共性,学习一种通用学习规则(如初始参数、优化算法、更新策略),从而构建多任务共享的特征表示,以较小的训练成本实现向新任务的迁移. 这些方法中,在线元学习(Online Meta-Learning, OML)<sup>[88]</sup>提供了一种简易的元训练策略实现,

其对顺序到达的样本进行在线更新并最小化样本干扰(灾难性干扰),获取适用于连续学习任务的稀疏表示,实现快速学习新任务的同时具有鲁棒的抗遗忘性能. 进一步地,自适应神经记忆学习(Adaptive Neural Memory Learning, ANML)<sup>[89]</sup>基于元学习方法实现一个上下文相关的门控函数,实现与增量任务相关神经元的选择性激活. 进一步地,注意力独立机制(Attentive Independent Mechanisms, AIM)<sup>[90]</sup>实现混合专家模型(Mixture Of Experts, MOE),使用OML或ANML的特征表示进行预测,进一步在架构层面稀疏化特征表示优化模型连续学习能力. 此外,元学习方法同样可以与经验重放方法兼容使用,更好地利用旧任务和新任务的训练样本. 内存高效重放(Memory-Efficient Replay, MER)<sup>[91]</sup>对齐新旧任务的梯度方向,而增量式任务无关元学习(incremental Task-Agnostic Meta-Learning, iTAML)<sup>[92]</sup>应用一个元学习更新规则,使新旧任务两者保持平衡. 前瞻式元学习(Look-ahead Meta Learning, La-MAML)<sup>[93]</sup>使用自适应调节的学习率在线优化OML目标,增强所获取稀疏表征对连续学习任务的适应性. 在线快速适应与知识积累(Online fast Adaptation and Knowledge Accumulation, OSAKA)<sup>[94]</sup>提出了一个知识积累和快速适应的混合策略,通过使用元训练获得高效的预训练参数初始化,并以在线更新形式将新知识融入初始化参数中实现新知识更新. 元学习还可以用于优化专用架构. 元巩固连续学习(Meta-consolidation for continual learning, MERLIN)<sup>[95]</sup>结合了每个任务表示的模型参数的元分布,为每个任务生成特定的模型并将这些特定任务模型参数组合用于推理. 同样,后验元重放(Posterior meta-Replay, PR)<sup>[96]</sup>采用贝叶斯策略,使用共享的元模型学习特定任务的后验. 元学习可重用知识(Meta-learning for Reusable Knowledge, MARK)<sup>[97]</sup>维护一组共享模型权重并通过元学习方法逐步更新,并选择性屏蔽以应用于特定任务. 类似地,抗逆向干扰(Anti-Retroactive Interference, ARI)<sup>[98]</sup>将对抗攻击与经验回放相结合以得到用于特定任务的模型,并通过元训练技术将这些任务特定模型进行融合.

此外,还有一些工作尝试从损失函数平滑度的角度对优化过程进行指向性细化. 例如,稳定随机梯度下降(Stable Stochastic Gradient Descent, Stable-SGD)<sup>[99]</sup>通过调整训练方案中的优化因子(如Dropout、学习率衰减和批量大小),使SGD优化器找到一个平滑的局部最小值,具有更好的抗干扰能力,从而优异的抗遗忘性能. 模态连通性随机梯度下降(Mode Connectivity Stochastic Gradient Descent, MC-SGD)<sup>[100]</sup>证明了多任务学习(即所有增量任务的联合训练)和连续学习所获得的局部最

小值可以通过一个低误差的线性路径连接,且可以应用经验回放方法沿该低误差线性路径找到更好的用于连续学习的模型参数.线性连接器(Linear Connector)<sup>[101]</sup>采用Adam-NSCL和特征蒸馏来获得由低误差线性路径连接的旧任务和新任务的模型参数,并使用线性平均获取最后模型参数.此外,无监督/自监督学习(相对于传统的监督学习)<sup>[102,146,147]</sup>和大规模预训练(相对于随机初始化)<sup>[16,121,148-150]</sup>在近年来的实践中被证明能够减少灾难性遗忘.从技术思想出发,这两者都可以归因于获得更稳健的特征表示(例如正交、稀疏和均匀散布特征表示<sup>[16,102,119,149]</sup>),并收敛到更平坦的损失函数平面<sup>[16,102,121,151,152]</sup>,使模型具备较强的抗干扰能力,从而实现优异的连续学习能力.

### 3.4 基于特征表示的方法

本文关注到也有部分方法聚焦于将某些预训练模型固有的较强特征表示能力作用于连续学习任务.早期一些工作关注于通过元学习(meta-learning)<sup>[88]</sup>技术实现稀疏特征学习,不同于这一类早期的稀疏特征建模方法,近期一些工作试图结合自监督学习(Self-Supervised Learning, SSL)<sup>[102,106,153]</sup>、大规模预训练技术<sup>[119,121,150]</sup>中涌现的优秀特征表达能力,以实现更好的连续学习能力.

常见的基于特征表示的方法为连续学习任务实现自监督学习,通过自监督学习技术可以实现对灾难性遗忘较为鲁棒的特征表示.无监督记忆与投影学习(Learning with Unsupervised Memory and Projection, LUMP)<sup>[102]</sup>通过对新旧任务的训练样本实例间进行插值实现连续学习任务性能的改进.在数据重放范式中,最小冗余(Minimum Redundancy, MinRed)<sup>[103]</sup>对内存池中存储的旧样本进行自监督去相关从而有效提高了经验重放的重放样本多样性.持续自监督学习(Continual Self-Supervised Learning, CaSSLe)<sup>[104]</sup>通过将训练样本特征表示的当前状态映射到旧任务特征表示状态,将自监督损失函数转换为蒸馏损失函数.对比连续学习(Contrastive continual Learning, Co2L)<sup>[105]</sup>采用监督对比损失函数来学习每个任务,并采用自监督损失函数提取新旧模型的知识.双网络(Dual Network, DualNet)<sup>[106]</sup>分别训练一个使用监督损失函数的快速学习模型和一个使用自监督损失函数的慢速学习模型,通过慢速学习模型帮助快速学习模型达到更具泛化性的特征表示.

在下游连续学习任务中使用预训练技术.研究结果表明:下游连续学习任务明显受益于预训练技术的使用,既带来了强大的知识迁移能力,也带来了对灾难性遗忘的鲁棒性<sup>[121,149,153,154]</sup>.特别地,当使用较大的数据量<sup>[149,154]</sup>、较大的模型容量<sup>[110]</sup>和对比学习损失<sup>[148,153]</sup>

进行预训练时,下游连续学习任务的性能提升往往更加明显.然而,这一类方法面临的关键挑战是,模型需要自适应地利用预训练中的知识学习当前任务,并保持对未来新任务的通用泛化能力.针对这个问题有多种策略实现,可根据预训练主干网络是否冻结进行分类.预训练模型主干网络冻结情景下,侧调(Side-Tuning)<sup>[107]</sup>和深度线性连续微调(Deep Linear Continual Fine-Tuning, DLCFT)<sup>[108]</sup>训练与主干网络并行的轻量级网络,并线性融合两者输出.转移而不遗忘(Transfer without Forgetting, TwF)<sup>[109]</sup>进一步训练一个并行网络,以分层的方式从主干网络中提取知识.生成对抗网络记忆(GAN-Memory)<sup>[110]</sup>利用特征级线性调制(Feature-wise Linear Modulation, FiLM)<sup>[111]</sup>和自适应特征调制(Adaptive Feature Modulation, AdaFM)<sup>[112]</sup>学习预训练生成模型每一层的任务特定的参数,而自适应数据增强(Adaptive Data Augmentation, ADA)<sup>[69]</sup>使用结合知识蒸馏的适配器对预先训练的Transformer网络进行调整.

近期基于提示词(prompt)的方法使用一些提示词向量来适配使用预训练的Transformer网络.此类方法通常涉及通过探索提示词架构来适应任务共享和任务特定的知识,构建任务自适应提示并推断适当的提示词.代表性的策略包括从提示词库(prompt pool)中选择最相关的提示学习提示(Learning to Prompt, L2P)<sup>[113]</sup>,对提示词库与注意力因子(attention factor)进行加权求和(持续分解注意力提示(COntinual Decomposed Attention-based Prompting, CODA-Prompt)<sup>[114]</sup>),使用明确的任务共享和任务特定的提示双重提示(DualPrompt)<sup>[115]</sup>或仅任务特定的提示(S-Prompts)<sup>[116]</sup>,层次分解提示(Hierarchical Decomposition Prompt, HiDe-Prompt)<sup>[117]</sup>,任务特定提示词的渐进扩展(Progressive Prompts)<sup>[118]</sup>等.此外,通过保存类别原型,将最近的类均值分类器附加到主干网络著有成效<sup>[155,156]</sup>,可以通过FiLM适配器<sup>[111]</sup>等迁移学习技术进一步增强.此外,还优化可更新的主干网络,F2M<sup>[120]</sup>在预训练阶段搜索损失函数景观中平坦的局部最小值,然后在平坦区域内学习增量任务.类别间去相关性(Class-wise Decorrelation, CwD)<sup>[119]</sup>将初始阶段表示正则化为均匀分散.锐度感知最小化(Sharpness-Aware Minimization, SAM)<sup>[121,122]</sup>鼓励通过优化损失函数景观平坦度指标在下游连续学习任务中找到平坦区域.针对上述方法中存在着示词库和查询向量重构导致的参数量和和计算量增加的问题,单提示与虚拟异常值正则化(OnePrompt with Virtual Outlier Regularization, OVOR)<sup>[157]</sup>使用模拟离群样本来优化分类器决策边界,消除额外计算成本的同时实现相近的效果.卷积提示(Convolutional Prompting, ConvPrompt)<sup>[158]</sup>针对提示词方法跨任务共享知识效果较差的问题,提出使用卷积

网络构建提示词输入,使维护逐层共享的表达向量成为可能,实现特定网络层的学习和更好的任务间知识迁移.一致性提示(Consistent Prompting, CPrompt)<sup>[159]</sup>提出一致性提示词方法对此类方法中训练和测试阶段不一致性进行缓解.

慢分类器对齐方法(Slow Learner with Classifier Alignment, SLCA)<sup>[160]</sup>发现缓慢微调预训练 Transformer 模型的主干网络参数可以在连续学习中实现出色的性能,进一步结合类别分布建模与分类层对齐策略,SLCA 实现了较提示词方法更优异的性能.随机投影与预训练模型(Random projections and pre-trained models, Ranpac)<sup>[161]</sup>基于上述策略,在预训练模型的特征表示和输出端之间实现随机映射层.生成多模态模型(Generative Multi-Modal models, GMM)<sup>[162]</sup>将上述缓慢微调预训练模型方法用于生成式多模态模型,在类增量学习中用其替代判别模型,显著缓解遗忘现象.可扩展子空间集成(Expandable Subspace Ensemble, EASE)<sup>[163]</sup>在缓慢微调预训练网络的同时为每一个新任务训练轻量化适配器模块.语义偏移增量适配器调整(Semantically-Shifted Incremental Adapter-Tuning, SSIAT)<sup>[164]</sup>在预训练模型基础上额外增加并增量微调适配器模块,有效提高视觉增量分类任务效果.同时,ConvPrompt<sup>[158]</sup>、CPrompt<sup>[159]</sup>、无干扰低秩适应(Interference-free Low-Rank Adaptation, InfLoRA)<sup>[165]</sup>等方法直接或间接引入微调预训练模型的策略,实现了可观的可塑性-稳定性均衡.

同样值得关注的是持续预训练(Continual Pre-Training, CPT)或持续元训练(Continual Meta-Training, CMT).由于预训练所需的大量数据通常以增量方式收集,因此进行预训练连续学习以提高下游性能尤为重要.研究发现,对于视觉语言模型的连续学习<sup>[166]</sup>,自监督预训练比监督训练更有效,这与视觉任务中得到的结论一致<sup>[16]</sup>.由于文本通常比图像更易收集和使用,增量式 CLIP(Incremental CLIP)<sup>[123]</sup>重放以图像为条件生成的负文本,并使用多模态知识蒸馏更新对比语言-图像预训练(Contrastive Language-Image Pre-training, CLIP)<sup>[167]</sup>.对于语言模型的 CPT,有效连续预训练网络(Effective COntinual pretraining NETwork, ECONET)<sup>[124]</sup>设计了一个具有生成重放功能的自监督框架.同时,持续元训练需要解决类似的问题,即基类的预训练知识以增量形式丰富和迁移适应.间接判别对齐(Indirect Discriminant Alignment, IDA)<sup>[125]</sup>强制新旧模型的判别式相对于旧中心进行对齐,否则使用自由嵌入向量以适应新任务.ORDER<sup>[126]</sup>采用具有经验回放和特征回放的未标注的分布外数据应对较大的任务间差异.兼容拓扑保持(Compatible Topology Preservation, CTP)<sup>[127]</sup>、动态适应与剪枝(Dynamic Adaptation and Pruning,

DAP)<sup>[128]</sup>等分别关注于多模态模型和大语言模型中持续预训练技术,提出相应的持续预训练范式.在产业应用中,百度提出了 ERNIE 2 的持续预训练框架及基于 ERNIE 2.0 的文心大模型.ERNIE 2 不断引入多样的预训练任务,旨在从多个不同的任务中分别学习词汇、句法和语义信息.每当出现新任务时,ENRIE 2 的持续多任务学习方法使用先前的参数模型初始化,同时在原任务和新任务上训练,实现较好预训练效果.

最后,基于上述叙述总结的四类解决灾难性遗忘的主流方法,本文对上述几类方法进行总结与对比,结果如表 2 所示,对上述方法的核心思想及优缺点进行总结,以供读者参考审阅.

## 4 视觉任务中的连续学习

这一部分具体探讨了连续学习研究在主流的视觉任务中的存在及影响,旨在向读者介绍连续学习在计算机视觉领域的应用方式及相关方法.根据当前连续学习在视觉领域中的发展,本文主要聚焦于目标检测、语义分割和条件图像生成任务三类典型视觉任务.由于其他领域的关注量相对较少,现有工作与上述三领域的研究方法有相似之处,故本节不再进行重复讨论.

### 4.1 目标检测

针对目标检测任务(Object Detection)的连续学习应用一般称为增量目标检测(Incremental Object Detection, IOD),其中按任务增量顺序引入带有不同类别标注的训练样本,模型需要正确地定位并识别属于已学习类别的对象.与每个训练样本中只出现一个对象的视觉分类任务不同,目标检测通常有多个属于旧类和新类的对象同时出现.这一“共现”情况给 IOD 带来了额外的挑战,其中在学习新类时,旧类被标记为背景(或未标注类别),这一处理加剧了灾难性遗忘现象.同时,这使得知识蒸馏成为增量目标检测的有效方案——可以从新的训练样本中获取旧类对象知识,以约束旧模型和新模型之间的预测差异.

这一领域的早期工作目标检测器的增量学习(Incremental Learning of Object Detectors, ILOD)<sup>[168]</sup>,对旧类的预测进行蒸馏,以防止快速区域卷积神经网络(Fast Region-based Convolutional Neural Network, Fast R-CNN)<sup>[169]</sup>产生灾难性遗忘.随后的关系引导的知识迁移(Relation guided Knowledge Transfer, RKT)<sup>[170]</sup>进一步蒸馏选定候选集(proposal)中的共现知识.此后,知识蒸馏的思路被引入到其他检测框架中,如基于关键点三元组的目标检测 CenterNet<sup>[171]</sup>(选择性与相关性蒸馏(Selective and Inter-related Distilla, SID)<sup>[172]</sup>)、RetinaNet<sup>[173]</sup>(实时增量目标检测(Real-time Incremental Learning for Object Detection, RILOD)<sup>[174]</sup>)、泛化焦点损

表 2 连续学习方法分类与优缺点对比

方法分类		特点	优点	缺点
正则化方法	参数正则化	显式约束模型更新,通过正则化项减少模型对旧任务遗忘	不需要保存样本,仅需设计正则化项与优化方式,方法时间、空间复杂度低	模型偏差严重,连续学习性能差
	函数正则化			
数据重放	经验重放	显式保存旧任务信息,后续任务学习时通过重训练克服遗忘	连续学习性能优异,且方法简单易实现,简单有效	需要额外空间进行数据存储,方法时间、空间复杂度增加,且容易过拟合存储的数据,同时有一定的数据隐私性问题
	特征重放			
	生成重放	生成额外的旧任务数据进行训练以克服遗忘	连续学习性能优异,无数据隐私性问题	
优化方法	梯度投影	显式限制梯度更新方向,避免参数更新对旧任务的影响	连续学习性能优异,类别间干扰小,适用于多种网络结构	需要额外空间存储旧任务样本,时间、空间复杂度较高,对任务顺序较为敏感
	元学习	优化模型的元参数,使其能够快速适应新任务,同时保留旧任务的知识	新任务迁移能力较好,且无需存储旧任务样本,无数据隐私性问题	方法复杂度较高,对任务分布的假设依赖较强,方法扩展性不佳
	损失函数	鼓励模型收敛到平滑的共享低敏感区域,具有良好的抗扰动能力和泛化表征能力	优化过程稳定,通用性强,且无数据隐私性问题	时间复杂度较高,对任务顺序敏感,有收敛到局部最小值的风险
特征表示方法	自监督学习	通过更鲁棒的模型表征能力克服遗忘现象	连续学习性能优异,时间空间复杂度低,且无数据隐私性问题	方法对预训练网络依赖强,通用性不佳
	下游预训练			
	持续预训练	对预训练模型进行持续学习提高下游任务上的迁移能力	数据利用的效率较高,降低训练成本,新任务迁移能力强,贴合产业应用需求	当新任务数据规模较小时,有较大过拟合风险

失第一版 (Generalized Focal Loss Version 1, GFLV1)<sup>[175]</sup> (弹性响应蒸馏 (Elastic Response Distillation, ERD)<sup>[176]</sup> 和更快区域卷积神经网络 (Faster Region Convolutional Neural Network, Faster R-CNN)<sup>[177]</sup> (类别增量 Faster R-CNN (Class-Incremental Faster R-CNN, CIFRCN)<sup>[178]</sup>、更快的增量学习目标检测器 (Faster Incremental Learning Object Detector, Faster ILOD)<sup>[179]</sup>、深度模型巩固 (Deep Model Consolidation, DMC)<sup>[180]</sup>、桥接非共现 (Bridging Non Co-occurrence, BNC)<sup>[181]</sup>、基于元学习的增量目标检测 (Incremental Object Detection via Meta-Learning, IOD-ML)<sup>[182]</sup>。一些方法利用未标记数据将旧模型和新模型蒸馏到共享模型中,以桥接潜在的双非共现关系 (Bi-Non-Cooccurrence, BNC) 并实现更好的稳定性-可塑性均衡动态记忆控制 (Dynamic Memory Control, DMC)。为了进一步减轻知识蒸馏对可塑性的负面影响, IOD-ML 采用元学习将参数梯度重塑为旧类和新类之间的平衡方向。SDDGR (Stable Diffusion-based Deep Generative Replay)<sup>[134]</sup> 直接使用预训练的 Stable Diffusion 生成一定数量旧任务的类别样本,并在新任务阶段进行重放训练,有效增强 IOD 模型对旧任务知识记忆。IOD 不仅适用于 2D 图像,还适用于 3D 图像数据<sup>[183]</sup> 和视频数据<sup>[184]</sup>。此外,还有一些其他相关的任务,如增量小样本检测<sup>[185]</sup>, 预训练的检测模型仅使用少量标注数据学习

新类别;以及开放世界目标检测<sup>[14]</sup>, 模型需要识别未知类别的潜在对象,并在获取相应的标签后进行注册。

## 4.2 语义分割

持续语义分割 (Continual Semantic Segmentation, CSS) 旨在为图像产生像素级类别预测,并在旧类别的基础上学习新类别。与 IOD 类似,旧类和新类的对象可以一起出现。早期的工作在连续学习中使用了旧类和新类对象的完整标注<sup>[186,187]</sup>。然而,重新标注旧类的巨大经济和时间成本,仅使用新类的标注的方法得到了更多关注,这导致旧类被视为背景 (即“背景偏移”现象),从而加剧了灾难性遗忘。

在这一任务中,一种常见的策略是从旧模型中自适应地蒸馏知识,这可以准确地将未标注的旧类与背景区分开。例如,记忆融合批量归一化 (Memory-incorporated Batch normalization, MiB)<sup>[188]</sup> 使用旧模型的预测来校准背景像素的正规交叉熵 (Cross Entropy) 和知识蒸馏损失。自适应逻辑正则化和特征重放用于增量语义分割 (Adaptive Logit regularizer and Feature replay for incremental semantic segmentation, ALIFE)<sup>[189]</sup> 通过正则化进一步改进了正规交叉熵和知识蒸馏,并使用特征回放对分类器进行微调。增量学习中的表示补偿 (Representation Compensation for Incremental Learning, RCIL)<sup>[190]</sup> 将网络重新参数化为两个并行分支,其中

旧分支被冻结以进行中间层间的知识蒸馏. 稀疏且解耦的表示 (Sparse and Disentangled Representations, SDR)<sup>[191]</sup> 和不确定性感知对比蒸馏 (Uncertainty-aware Contrastive Distillation, UCD)<sup>[192]</sup> 将对学习引入到特征表示的知识蒸馏中, 其中同一类的像素被聚类, 不同类的像素被分离. 伪标签优化的无遗忘学习 (Pseudo-Label Optimized learning without forgetting, PLOP)<sup>[193]</sup>、RECALL<sup>[194]</sup>、未知标签的语义分割 (Semantic Segmentation with Unknown Label, SSUL)<sup>[195]</sup>、期望最大化 (Expectation-Maximization, EM)<sup>[196]</sup>、自训练 (Self-Training)<sup>[197]</sup>、UCD<sup>[192]</sup> 和带权重的语义分割增量学习优化 (Weighted Incremental Learning with Semantic segmentation Optimization, WILSON)<sup>[198]</sup> 显式使用旧模型生成旧类的伪标签. 额外数据资源如网络爬虫 (RECALL with Web Scraping, RECALL-Web<sup>[194]</sup>)、预训练的生成模型 (RECALL with Generative Adversarial Networks, RECALL-GAN<sup>[194]</sup>)、大量未标记数据 (Self-Training<sup>[197]</sup>) 和少量旧训练样本 (EM<sup>[196]</sup>、SSUL<sup>[195]</sup>、图像级标签与重放 (Image-Level Labels and Rehearsal, ILLR)<sup>[199]</sup>) 也被用于知识蒸馏方法以防止灾难性遗忘.

此外, 图像的显著性信息通常用于在 CSS 中定位未标注的对象, 以增强仅图像级标注的弱监督信息 (WILSON<sup>[198]</sup>、ILLR<sup>[199]</sup>), 此外, 在背景类中定义未知类别以有利于学习可塑性 (SSUL<sup>[195]</sup>). 微观分割 (Micro-Segmentation, MicroSeg)<sup>[200]</sup> 根据当前任务训练样本的背景中可能包含的旧任务或未来新任务中的类别, 以增强记忆稳定性和学习可塑性. CoinSeg<sup>[201]</sup> 在此基础上通过类内类间的对比学习增强了增量语义分割模型的可塑性.

### 4.3 视觉内容生成

条件生成连续学习 (Continual Learning for Conditional Generation, CLCG) 与生成式重放技术密切相关. 在 3.1 节中的生成式重放内容中, 本文讨论了一系列尝试通过使用条件生成模型生成旧任务分布样本, 以克服任务模型的灾难性遗忘. 在生成式重放任务中, 用于旧样本生成的生成模型在增量学习过程中, 由于需要通过模型参数更新以适应当前任务的数据分布, 不可避免地会伴随任务模型发生遗忘. 此外, 一些文献指出<sup>[202, 203]</sup>, 当条件生成模型以类似任务增量学习任务模型进行生成任务上的增量学习时, 生成模型中的判别器同样会发生遗忘现象. 上述两种遗忘情况本质殊途同归, 共同指向了非平稳数据流下, 条件生成模型也无法避免灾难性遗忘挑战. 关于这一领域的体系化分析研究, 文献[204]中针对基于自编码器的视觉内容生成与连续学习研究给出了更加系统和体系化的分析论述.

关注到条件生成模型的遗忘问题, 部分工作首先尝试使用如正则化方法<sup>[205]</sup>、经验重放<sup>[206, 207]</sup>、对模型参

数指数移动平均<sup>[208]</sup>及使用元学习<sup>[209]</sup>等方法以减轻遗忘问题. 受这些先驱工作启发, 一系列探索生成模型连续学习的工作涌现, 在增量训练过程中, 探索条件生成模型能够持续生成高质量样本, 而不受灾难性遗忘的解决方案. 在针对对抗生成模型 GAN 的抗遗忘探索中, 结合 3.1 节中的突触智能 (Synaptic Intelligence, SI)<sup>[26]</sup> 和 EWC<sup>[25]</sup> 权重正则化技术, 出现了解耦表示差异用于持续学习 (Disentangled Representation Discrepancy for Continual learning, DRDisCo)<sup>[13]</sup> 和文献[210]方法. 基于经验重放方法, MeRGANs<sup>[66]</sup> 与 L-VAEGAN<sup>[211]</sup> 被相继提出. 基于知识蒸馏方法, LifelongGAN<sup>[140]</sup> 与超网络终身学习生成对抗网络 (Hyper-Lifelong Generative Adversarial Network, Hyper-LifelongGAN)<sup>[212]</sup> 将先前任务的知识从旧网络蒸馏到新网络中以避免对旧任务知识的遗忘. 少样本增量学习用于标签到图像的翻译 (Few-shot Incremental Learning for Label-to-image Translation, FILIT)<sup>[213]</sup> 通过为每个新类别学习一组新参数的方式实现生成网络的类别增量学习, 类似地, GAN-Memory<sup>[110]</sup> 为新任务迁移构造预训练 GAN 网络中的全连接和卷积层. 针对变分自编码器的抗遗忘探索通常依赖于可扩展的模型结构、权重正则化 (VCL<sup>[37]</sup>) 或生成重放 (持续无监督表示学习 (Continual Unsupervised Representation Learning, CURL)<sup>[17]</sup>、共享嵌入的变分自编码器 (Variational Autoencoder with Shared Embeddings, VASE)<sup>[214]</sup>) 来克服灾难性遗忘. 其中, 变分自编码器的持续学习提升方法 (Boosting approach for continual learning of Variational AutoEncoder, BooVAE)<sup>[215]</sup> 采用固定的静态架构, 将不同任务的数据分布信息合并到一个加性聚合后验中, 并将其用作后续任务的先验, 在新任务训练时提供旧任务信息以克服遗忘. 双重一致性模型反演 (Dual-Consistency Model Inversion, DCMI)<sup>[216]</sup> 提出通过合成样本和相应类别原型之间的语义一致性, 以及新类别合成样本和真实图像之间的域一致性的双重一致性保持生成高质量的旧类合成样本, 实现 VAE 网络的抗遗忘效果. 除 GAN 和 VAE 两者外, 近期随着稳定扩散 (Stable Diffusion)<sup>[217]</sup> 生成模型的兴起, 持续低秩适应 (Continual Low-Rank Adaptation, C-LoRA)<sup>[218]</sup> 将参数正则化方法与低秩适应 (Low-Rank Adaptation, LoRA) 方法结合, 实现了 Stable Diffusion 模型的稳定能力扩增, 开始了 Stable Diffusion 模型上的连续学习探索.

## 5 连续学习前沿与展望

本文在这一节主要讨论连续学习方法最新的研究与发展方向. 关注到以在线连续学习为代表的通用连续学习近期逐渐活跃, 受到领域内学者广泛关注, 本节首先讨论当前通用连续学习技术的发展. 其次, 关注到

新近兴起的以大语言模型、视觉基础模型、多模态模型为代表的**基础模型(foundation model)**发展浪潮,本节也对连续学习在基础模型中的应用进行充分讨论,这代表着连续学习技术和深度学习领域最前沿技术的联合发展动向.最后,针对连续学习在一些卓有前景的如具身智能(Embodied AI)、AI与科学(AI for Science, AI4Sci)等新兴交叉领域中的应用,本节进行一定叙述与展望.

### 5.1 通用连续学习

通用连续学习(General Continual Learning, GCL)<sup>[52,219,220]</sup>既是指广泛应用场景下的连续学习任务,同时也是深度学习**中广义的学习范式问题**.具体地,模型以在线方式学习增量数据,没有明确的任务边界.相应地,通用连续学习会关联到两个相互关联的连续学习任务设定:无任务连续学习<sup>[219]</sup>(Task-Free Continual Learning, TFCL)与在线连续学习<sup>[57]</sup>(Online Continual Learning, OCL).无任务连续学习指任务标识在训练与测试中都无法获取,在线连续学习中训练样本在一次性数据流中进行训练.由于无任务连续学习通常一次只访问小批量训练样本,以逐步改变任务分布,而在线连续学习通常只需要数据标签而不是每个训练样本的任务标识,两者的学习设定相互兼容,代表性方法也彼此相互兼容.

具体而言,其中一些工作尝试在不断扩展的网络架构中学习特定参数.无任务连续学习中, CNPDM<sup>[221]</sup>采用狄利克雷(Dirichlet)过程混合模型来构建越来越多的专家神经单元,而同期工作<sup>[222]</sup>从概率元学习器中推导出类似混合模型.变分架构增长(Variational Architecture Growing, VariGrow)<sup>[223]</sup>采用基于能量函数的新颖度分数决定是扩展新专家网络还是更新旧专家网络.在线差异距离学习(Online Discrepancy Distance Learning, ODDL)<sup>[224]</sup>估计当前内存缓冲区和先前学习的知识之间的差异作为扩展信号.在线连续学习中,实例感知参数化(Instance-Aware Parameterization, InstA-Param)<sup>[225]</sup>为单个训练样本选择和增强合适的网络路径.另一方面,许多基于任务的持续学习(Task-based Continual Learning, TSCL)方法和大多数OCL方法都建立在经验重放之上,专注于内存缓冲区的构建、管理和开发.由于训练样本以小批量的形式到达,任务边界的信息效果较差,蓄水池采样通常是一种有效的样本选择策略.更高级的策略优先考虑信息丰富的训练样本的重放(信息论在线记忆选择(Information-theoretic online memory Selection, InfoRS)<sup>[226]</sup>、参数梯度的多样性(GSS)<sup>[57]</sup>或先前学习的知识(ODDL<sup>[224]</sup>、类标签平衡(类别平衡重放策略(Class-Balanced Replay Strategy, CBRS)<sup>[227]</sup>、GDumb<sup>[228]</sup>、持续原型演变(Continual Prototype Evolution, CoPE)<sup>[229]</sup>),并有利于潜在决策边界

(ASER<sup>[60]</sup>).同时,内存缓冲区可以动态管理,例如通过删除较次要度的训练样本(类别增量学习与模糊任务配置(Class Incremental Learning with Blurry task configuration, CLIB)<sup>[22]</sup>、编辑旧训练样本(基于梯度的记忆编辑(Gradient-based Memory Editing, GMED))<sup>[230]</sup>或其分布(分布鲁棒优化(Distributionally Robust Optimization, DRO)<sup>[231]</sup>)更有可能被遗忘,并检索容易受到干扰的旧训练样本(最大干扰检索(Maximal Interfered Retrieval, MIR)<sup>[232]</sup>、双视图一致性(Dual View Consistency, DVC)<sup>[233]</sup>).

近期,多级在线监督专长(Multi-level Online Supervised Expertise, MOSE)<sup>[234]</sup>基于对现有在线连续学习的分析,即使用旧样本缓冲区带来的过度拟合-欠拟合问题,引入多级监督和反向自监督的模型堆叠技术解决新样本的欠拟合以及旧样本多次回放引起的过度拟合问题.在线持续学习中的自适应实例相似性嵌入(Adaptive Instance Similarity Embedding for Online Continual Learning, AISEOCL)<sup>[235]</sup>采用自适应相似度向量,从由当前任务和之前任务组成的当前训练样本中获得额外有价值的相似信息,从而提高在线连续学习中经验样本重放的效果.代理的持续注意力机制(Continual Attention Mechanism for Agents, CAMA)<sup>[236]</sup>指出现有在线连续学习任务中对存储的旧任务样本依赖问题,提出以移动平均的方式基于没有任务边界的置信度分数更新存储的旧任务样本信息.协作持续学习(Collaborative Continual Learning, CCL)<sup>[237]</sup>针对在线连续学习模型的可塑性不足问题,提出基于协作学习的在线连续学习方法,旨在提高模型获取新知识的能力,有效提高了现有在线连续学习方法的**可塑性**.等角表示学习(Equi-Angular Representation Learning, EARL)<sup>[238]</sup>引导利用神经崩溃现象在模型的表示空间中形成一个单纯形等角紧框架结构,以便在表示空间中实现预备数据训练和残差矫正,使得单次训练的连续学习模型可以更好地学习新数据.动态特征空间自组织(DYnamic feature space Self-OrgAnization, DYSON)<sup>[239]</sup>为在线无任务连续学习任务提出的动态特征空间自组织方法,在出现新类时扩展类原型分布而不损失原型分布的最优性.

### 5.2 基础模型与连续学习

近两年,以生成式预训练变换器(Generative Pre-trained Transformer, GPT)<sup>[240]</sup>、DeepSeek、OpenGLM、通义千问等为代表的大语言新基础模型迅猛发展,成为了深度学习领域最大的发展趋势,这一发展趋势同样体现在视觉领域,如CLIP<sup>[167]</sup>、SAM<sup>[241]</sup>、Dinov2(Distillation with No Labels v2)<sup>[242]</sup>、Stable Diffusion<sup>[243,244]</sup>等为代表的视觉基础模型纷纷出现成为领域发展主流,在多模态领域,以视觉-语言(Intern Vision-Language, InternVL)

模型<sup>[245]</sup>、LlaVA (Large language and Vision Assistant)<sup>[246]</sup>为代表的基础模型成为领域研究热点。这些新基础模型普遍是具有大规模预训练量、超大参数量(十亿级)的庞然巨物。使用这些基础模型时,传统的全量微调方法往往会遇到过拟合问题,而新兴的以低秩适应 LoRA<sup>[247]</sup>为代表的高效参数微调技术(Parameter Efficient FineTuning, PEFT)<sup>[247-249]</sup>虽然能够通过少量参数进行微调训练即可实现任务迁移,但也不可避免遇到基础模型会遗忘丰富的预训练。在这一背景下,基础模型的连续学习技术成为近期领域发展重点。

基础模型连续学习技术的发展与微调适配器 Adapter<sup>[248]</sup>、微调提示词 Prompts<sup>[249]</sup>以及新兴的低秩适应 LoRA<sup>[247]</sup>方法等高效微调技术密不可分。其中,以 L2P<sup>[113]</sup>、DualPrompt<sup>[115]</sup>、S-Prompts<sup>[116]</sup>、Progressive-Prompts<sup>[118]</sup>、CODA-Prompt<sup>[250]</sup>、HiDE-Prompt<sup>[117]</sup>等为代表的基于提示词微调的基础模型连续学习方法发展最为迅速。L2P<sup>[113]</sup>从提示词库(prompt pool)中选择与上下文最相关的提示进行优化,CODA-Prompt<sup>[114]</sup>对提示词库与注意力因子进行加权求和,DualPrompt<sup>[115]</sup>使用明确的任务共享和任务特定的提示,S-Prompts<sup>[116]</sup>、HiDE-Prompt<sup>[117]</sup>使用任务特定的提示词,Progressive Prompts<sup>[118]</sup>使用任务特定提示词的渐进扩展等。此类方法通过对基础模型提示词进行增量微调,实现基础模型对新模型的快速迁移,同时由于未对基础模型本身进行大规模参数更新,基础模型仍能保持较好的通用知识(又称世界知识)。这一类技术广泛应用于基于 Transformer 架构的基础模型上。

同时,结合低秩适应 LoRA<sup>[247]</sup>的连续学习方法在大语言模型、多模态模型、图像生成模型等多个不同领域中也得到了广泛应用。LoRA 本质上是对模型的参数矩阵进行低秩分解,并仅训练分解得到的矩阵参数,通常应用于将大模型向具体下游任务迁移的场景,与连续学习中的可塑性理念不谋而合。在 LoRA 基础上,持续学习与低秩适应(Continual learning with Low-Rank adaptation, CoLoR)<sup>[251]</sup>利用 LoRA 来更新预训练的适配器以实现新任务上的可塑性,同时保持先前任务的性能。低秩适应与高效调整(Low-rank Adaptation with Efficient Tuning, LAE)<sup>[252]</sup>针对连续学习中参数高效微调一类方法中普遍存在的灵活性、特定参数选择等问题提出在线高效微调模块与知识积累和集成预测方法的有效统一框架,证明了多种 PEFT 方法的统一性。正交低秩适应(Orthonormal Low-Rank Adaptation, O-LoRA)<sup>[253]</sup>基于 LoRA 的矩阵参数中捕获了模型参数更新方向这一现象,即过去任务上的 LoRA 子空间可以近似梯度空间,提出在学习新任务时约束新旧任务 LoRA 子空间正交的方法,在自然语言任务上,实现了优秀的稳定性可

塑性性能。与前述方法不同之处是,文献[254]结合任务算术(task arithmetic)、LoRA 和样本记忆技术,实现了接近于完全微调的模型性能。InfLoRA<sup>[165]</sup>在旧任务的梯度空间中注入少量参数,来近似旧任务的梯度,发现微调这些注入的参数等同于在旧任务梯度子空间内微调预训练权重,可实现消除新任务对旧任务的干扰,从而在稳定性和可塑性之间实现了良好的权衡。SSIAT<sup>[164]</sup>提出在不施加参数更新约束的情况逐步调整优化共享适配器参数,增强主干网络学习能力,可消除模型扩展和旧任务样本存储需求。

当前,这一领域正在蓬勃发展,除基于 LoRA 的方法外,也有动态数据增强与选择(Dynamic Data Augmentation and Selection, DDAS)<sup>[255]</sup>针对视觉语言模型在微调迁移中的长程知识遗忘问题,提出使用混合专家(Mixture Of Experts, MOE)适配器动态扩展与训练的 CLIP 模型以适应新任务,并对 MOE 中的路由机制进行优化,将分布内和分布外的输入分别路由到 MOE 适配器和原始预训练 CLIP 模型。持续近端策略优化(Continual Proximal Policy Optimization, CPPO)<sup>[256]</sup>针对当前预训练语言模型的基于人类反馈强化学习(Reinforcement Learning from Human Feedback, RLHF)所涉及的时间、计算成本以及数据隐私问题,提出了使用加权策略判定用于增强策略学习和记忆强化的训练样本。文献[257]提出联合自适应重参数化技术,与动态任务相关知识检索结合,针对特定下游任务实现语言模型的自适应微调,有效缓解连续学习在大语言模型应用中的局限性。在可预见的未来中,连续学习在基础模型中的研究与应用有极大可能是本领域研究热点。

### 5.3 连续学习前沿展望

除在前文所述的连续学习领域或在计算机视觉领域的前沿进展外,连续学习在一些新兴、交叉研究方向中也正在发光发热,其中以具身智能和 AI 与科学两者最为耀眼。

具身智能是相对于“互联网人工智能”的一种新兴研究方向,旨在使 AI 算法和智能代理(agent)通过与物理或虚拟环境<sup>[258]</sup>的直接交互来实现学习,这一过程与人类通过感官体验和动作实践来获得知识和技能的方式更为接近。其核心在于,智能代理能够利用自我中心的感知系统来捕捉环境信息,并以此为基础,通过行为实践来优化自身的决策和行动策略。这一学习过程中要求智能算法需要不断与环境进行交互与增量学习以实现高级智能<sup>[236,259-261]</sup>。最新的研究成果表明:通过结合多模态学习<sup>[262-266]</sup>(multimodal learning)和连续学习技术<sup>[267]</sup>,具身智能代理<sup>[258]</sup>(embodied agents)能够实现更高层次的抽象和泛化能力。这意味着具身智能算法不仅能够从单一的数据源学习,还能够整合来自视觉、

听觉、触觉等多种感官的信息,从而在更广泛的任务和环境中实现有效的学习与应用.此外,对通用连续学习的研究有助于具身代理从类似于人类的自我中心感知和行为中学习<sup>[268-270]</sup>,并为研究人员提供了智能研究的新视角,即通过从广泛收集的人类活动数据中追寻终身学习的本质,实现类人的智能算法系统.当前,文献[261]基于在线连续学习与通用连续学习技术,探索连续学习技术赋能具身智能任务,开启连续学习技术迈向真实物理世界的探索尝试.

AI与科学(AI for Science)已成为近年来快速发展的科学研究范式,旨在通过AI算法来辅助生物学<sup>[271,272]</sup>、医学、材料化学<sup>[273,274]</sup>、气象学<sup>[275]</sup>等科学学科的研究.以蛋白质序列结构预测模型AlphaFold<sup>[276]</sup>为代表的成功应用,表明了AI具有重塑科学研究范式的潜力.AI不仅能够辅助科学家提出假设,还可以设计实验、收集和解释数据,从而突破现有科学理解的界限.连续学习方法作为受生物智能启发的技术,也已被有效应用于以医学图像分析为代表的生物医学课题<sup>[277-279]</sup>中,并涌现出数量可观的相关工作<sup>[280-284]</sup>.进一步地,一些典型科学应用领域中存在着难以构筑大规模的统一标注数据集、数据难以访问与获取、对域外泛化能力要求显著等问题,具有典型的动态非平稳数据流特点,基于这些特点,连续学习技术在这些典型交叉领域的探索中同样值得关注,有望在多个交叉科学领域中取得突破性进展.

## 6 结论

连续学习旨在赋予模型从动态非稳定的外部环境中不断获取、更新、积累和利用知识,最终实现具有类似人类学习能力的智能系统.本文首先介绍了连续学习的问题与评价指标,根据不同的方法实现思路,对广泛的连续学习工作及其最新进展进行了综述和分析,并对这些代表性方法如何解决连续学习中的可塑性与稳定性均衡挑战进行了广泛分析.其次,本文还介绍了计算机视觉任务中的典型连续学习工作.最后,本文以对连续学习前沿的介绍、讨论与展望结尾,为读者梳理了连续学习的最新前沿与未来发展.基于本文的系统介绍和梳理,本文期待可以有效促进连续学习领域以及跨方向跨、领域的发展探索.

**致谢** 感谢朱泓光博士、张耕维博士、王立元博士、张幸幸研究员,以及余萧同学给本文写作和成稿提供的帮助和意见.

## 参考文献

[1] 朱飞,张煦尧,刘成林.类别增量学习研究进展和性能评价[J].自动化学报,2023,49(3):635-660.

ZHU F, ZHANG X Y, LIU C L. Class incremental learning: A review and performance evaluation[J]. Acta Automatica Sinica, 2023, 49(3): 635-660. (in Chinese)

[2] 张东阳,陆子轩,刘军民,等.深度模型的持续学习综述:理论、方法和应用[J].电子与信息学报,2024,46(10):3849-3878.

ZHANG D Y, LU Z X, LIU J M, et al. A survey of continual learning with deep networks: Theory, method and application[J]. Journal of Electronics & Information Technology, 2024, 46(10): 3849-3878. (in Chinese)

[3] 杨静,李斌,李少波,等.脑启发式持续学习方法:技术、应用与发展[J].电子与信息学报,2022,44(5):1865-1878.

YANG J, LI B, LI S B, et al. Brain-inspired continuous learning: Technology, application and future[J]. Journal of Electronics & Information Technology, 2022, 44(5): 1865-1878. (in Chinese)

[4] 常一帆.人脑学习与记忆机制启发的卷积神经网络持续学习方法[D].合肥:中国科学技术大学,2023.

CHANG Y F. Research on Continual Learning of Convolutional Neural Networks Inspired[D]. Hefei: University of Science and Technology of China, 2023. (in Chinese)

[5] 张亮.针对视觉数据的连续学习方法研究[D].成都:电子科技大学,2024.

ZHANG L. Research on Continual Learning Methods for Visual Data[D]. Chengdu: University of Electronic Science and Technology of China, 2024. (in Chinese)

[6] SOUTIF-CORMERAIS A, CARTA A, COSSU A, et al. A comprehensive empirical evaluation on online continual learning[C]//2023 IEEE/CVF International Conference on Computer Vision Workshops. Piscataway: IEEE, 2023: 3510-3520.

[7] MASANA M, LIU X, TWARDOWSKI B, et al. Class-incremental learning: Survey and performance evaluation on image classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 5513-5533.

[8] ZHOU D W, WANG Q W, QI Z H, et al. Class-incremental learning: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 9851-9873.

[9] WANG Z Y, YANG E N, SHEN L, et al. A comprehensive survey of forgetting in deep learning beyond continual learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(3): 1464-1483.

[10] GHOSH S. Dynamic VAEs with generative replay for continual zero-shot learning[EB/OL]. (2021-04-26)[2024-05-30]. <https://arxiv.org/abs/2104.12468>.

- [11] SINGH P, MAZUMDER P, RAI P, et al. Rectification-based knowledge retention for continual learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 15277-15286.
- [12] TAO X Y, HONG X P, CHANG X Y, et al. Few-shot class-incremental learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 12180-12189.
- [13] WANG L Y, YANG K, LI C X, et al. ORDIsCo: Effective and efficient usage of incremental unlabeled data for semi-supervised continual learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 5379-5388.
- [14] JOSEPH K J, KHAN S, KHAN F S, et al. Towards open world object detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 5826-5836.
- [15] WANG Q F, GENG X, LIN S X, et al. Learngene: From open-world to your learning task[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(8): 8557-8565.
- [16] HU D P, YAN S P, LU Q, et al. How well does self-supervised pre-training perform with streaming data[EB/OL]. (2021-07-21)[2024-05-30]. <https://arxiv.org/abs/2104.12081>
- [17] RAO D, VISIN F, RUSU A A, et al. Continual unsupervised representation learning[EB/OL]. (2019-10-31)[2024-05-30]. <https://arXiv.org/abs/1910.14481>.
- [18] KIM C D, JEONG J, KIM G. Imbalanced continual learning with partitioning reservoir sampling[EB/OL]. (2020-09-08)[2024-05-30]. <https://arXiv.org/abs/2009.03632>.
- [19] BANG J, KOH H, PARK S, et al. Online continual learning on a contaminated data stream with blurry task boundaries[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 9265-9274.
- [20] LIN H B, ZHANG Y F, QIU Z, et al. Prototype-guided continual adaptation for class-incremental unsupervised domain adaptation[M]//Computer Vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 351-368.
- [21] SIMON C, FARAKI M, TSAI Y H, et al. On generalizing beyond domains in cross-domain continual learning[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 9255-9264.
- [22] KOH H, KIM D, HA J W, et al. Online continual learning on class incremental blurry task configuration with any-time inference[EB/OL]. (2021-10-19)[2024-05-30]. <https://arXiv.org/abs/2110.10031>.
- [23] JOSEPH K J, PAUL S, AGGARWAL G, et al. Novel class discovery without forgetting[M]//Computer Vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 570-586.
- [24] SRINIVASAN T, CHANG T Y, ALVA L L P, et al. CLiMB: A continual learning benchmark for vision-and-language tasks[EB/OL]. (2022-06-18)[2024-05-30]. <https://arXiv.org/abs/2206.09059>.
- [25] KIRKPATRICK J, PASCANU R, RABINOWITZ N, et al. Overcoming catastrophic forgetting in neural networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2017, 114(13): 3521-3526.
- [26] ZENKE F, POOLE B, GANGULI S. Continual learning through synaptic intelligence[C]//Proceedings of the 34th International Conference on Machine Learning - Volume 70. New York: ACM, 2017: 3987-3995.
- [27] ALJUNDI R, BABILONI F, ELHOSEINY M, et al. Memory aware synapses: Learning what (not) to forget[M]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 144-161.
- [28] CHAUDHRY A, DOKANIA P K, AJANTHAN T, et al. Riemannian walk for incremental learning: Understanding forgetting and intransigence[M]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 556-572.
- [29] LIU X L, MASANA M, HERRANZ L, et al. Rotate your networks: Better weight consolidation and less catastrophic forgetting[C]//2018 the 24th International Conference on Pattern Recognition. Piscataway: IEEE, 2018: 2262-2268.
- [30] LEE J, HONG H G, JOO D, et al. Continual learning with extended kronecker-factored approximate curvature[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 8998-9007.
- [31] PARK D, HONG S, HAN B, et al. Continual learning by asymmetric loss approximation with single-side overestimation[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 3334-3343.
- [32] LEE S W, KIM J H, JUN J, et al. Overcoming catastrophic forgetting by incremental moment matching[EB/OL]. (2017-03-24)[2024-05-30]. <https://arXiv.org/abs/1703.08475>.
- [33] LEE J, JOO D, HONG H G, et al. Residual continual learning[EB/OL]. (2020-02-17)[2024-05-30]. <https://arXiv.org/abs/2002.07111>.

iv.org/abs/2002.06774.

- [34] SCHWARZ J, CZARNECKI W, LUKETINA J, et al. Progress & Compress: A scalable framework for continual learning[EB/OL]. (2018-05-16)[2024-05-30]. <https://arxiv.org/abs/1805.06370>.
- [35] WANG L Y, ZHANG M T, JIA Z F, et al. AFEC: Active forgetting of negative transfer in continual learning[EB/OL]. (2021-11-04)[2024-05-30]. <https://arXiv.org/abs/2110.12187>.
- [36] SWAROOP S, NGUYEN C V, BUI T D, et al. Improving and understanding variational continual learning[EB/OL]. (2019-05-06)[2024-05-30]. <https://arXiv.org/abs/1905.02099>.
- [37] NGUYEN C V, LI Y Z, BUI T D, et al. Variational continual learning[EB/OL]. (2018-05-20)[2024-05-30]. <https://arXiv.org/abs/1710.10628>.
- [38] TSERAN H, KHAN M E, HARADA T, et al. Natural Variational Continual Learning[EB/OL]. (2018-01-01)[2024-05-30]. [https://marcpickett.com/cl2018/CL-2018\\_paper\\_51.pdf](https://marcpickett.com/cl2018/CL-2018_paper_51.pdf).
- [39] ADEL T, ZHAO H, TURNER R E. Continual learning with adaptive weights (CLAW)[EB/OL]. (2020-06-16)[2024-05-30]. <https://arXiv.org/abs/1911.09514>.
- [40] LOO N, SWAROOP S, TURNER R E. Generalized variational continual learning[EB/OL]. (2020-11-24)[2024-05-30]. <https://arXiv.org/abs/2011.12328>.
- [41] DERAKHSHANI M M, ZHEN X, SHAO L, et al. Kernel continual learning[C]//Proceedings of the 38th International Conference on Machine Learning. Cambridge: PMLR, 2021: 2621-2631.
- [42] KAPOOR S, KARALETSOS T, BUI T D. Variational auto-regressive Gaussian processes for continual learning[EB/OL]. (2021-06-12)[2024-05-30]. <https://arXiv.org/abs/2006.05468>.
- [43] PAIK I, OH S, KWAK T, et al. Overcoming catastrophic forgetting by neuron-level plasticity control[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 5339-5346.
- [44] AHN H, CHA S, LEE D, et al. Uncertainty-based continual learning with adaptive regularization[EB/OL]. (2019-11-14)[2024-05-30]. <https://arXiv.org/abs/1905.11614>.
- [45] JUNG S, AHN H, CHA S, et al. Continual learning with node-importance based adaptive group sparse regularization[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 3647-3658.
- [46] LI Z Z, HOIEM D. Learning without forgetting[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(12): 2935-2947.
- [47] DHAR P, SINGH R V, PENG K C, et al. Learning without memorizing[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 5133-5141.
- [48] LEE K, LEE K, SHIN J, et al. Overcoming catastrophic forgetting with unlabeled data in the wild[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 312-321.
- [49] CASTRO F M, MARÍN-JIMÉNEZ M J, GUIL N, et al. End-to-end incremental learning[M]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 241-257.
- [50] HOU S H, PAN X Y, LOY C C, et al. Learning a unified classifier incrementally via rebalancing[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 831-839.
- [51] DOUILLARD A, CORD M, OLLION C, et al. PODNet: Pooled outputs distillation for small-tasks incremental learning[C]//Computer Vision - ECCV 2020. Cham: Springer Nature Switzerland, 2020: 86-102.
- [52] BUZZEGA P, BOSCHINI M, PORRELLO A, et al. Dark experience for general continual learning: A strong, simple baseline[EB/OL]. (2020-10-22)[2024-05-30]. <https://arXiv.org/abs/2004.07211>.
- [53] TITSIAS M K, SCHWARZ J, DE G MATTHEWS A G, et al. Functional regularisation for continual learning with Gaussian processes[EB/OL]. (2022-02-11)[2024-05-30]. <https://arXiv.org/abs/1901.11356>.
- [54] PAN P B, SWAROOP S, IMMER A, et al. Continual deep learning by functional regularisation of memorable past[EB/OL]. (2020-01-08)[2024-05-30]. <https://arXiv.org/abs/2004.14070>.
- [55] RUDNER T G J, SMITH F B, FENG Q X, et al. Continual learning via sequential function-space variational inference[EB/OL]. (2023-12-28)[2024-05-30]. <https://arXiv.org/abs/2312.17210>.
- [56] REBUFFI S A, KOLESNIKOV A, SPERL G, et al. iCaRL: Incremental classifier and representation learning[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 5533-5542.
- [57] ALJUNDI R, LIN M, GOUJAUD B, et al. Gradient based sample selection for online continual learning[EB/OL]. (2019-10-31)[2024-05-30]. <https://arXiv.org/abs/1903.08671>.
- [58] BORSOS Z, MUTNY M, KRAUSE A. Coresets via bilevel

- optimization for continual learning and Streaming[C]// Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 14879-14890.
- [59] YOON J, MADAAN D, YANG E, et al. Online coreset selection for rehearsal-based continual learning[EB/OL]. (2022-03-18)[2024-05-30]. <https://arXiv.org/abs/2106.01085>.
- [60] SHIM D, MAI Z D, JEONG J, et al. Online class-incremental continual learning with adversarial shapley value[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(11): 9630-9638.
- [61] BANG J, KIM H, YOO Y, et al. Rainbow memory: Continual learning with a memory of diverse samples[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 8214-8223.
- [62] CACCIA L, BELILOVSKY E, CACCIA M, et al. Online learned continual compression with adaptive quantization modules[C]//Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 1240-1250.
- [63] WANG L Y, ZHANG X X, YANG K, et al. Memory replay with data compression for continual learning[EB/OL]. (2022-03-09)[2024-05-30]. <https://arXiv.org/abs/2202.06592>.
- [64] KUMARI L, WANG S, ZHOU T, et al. Retrospective adversarial replay for continual Learning[C]//Advances in Neural Information Processing Systems 35. San Diego: NeurIPS, 2022: 28530-28544.
- [65] SHIN H, LEE J K, KIM J, et al. Continual learning with deep generative replay[EB/OL]. (2017-12-12) [2024-05-30]. <https://arXiv.org/abs/1705.08690>.
- [66] WU C, HERRANZ L, LIU X, et al. Memory replay GANs: Learning to generate new categories without forgetting[C]//Advances in Neural Information Processing Systems 31. San Diego: NeurIPS, 2018: 31.
- [67] LIU X L, WU C S, MENTA M, et al. Generative feature replay for class-incremental learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2020: 915-924.
- [68] ISCEN A, ZHANG J, LAZEBNIK S, et al. Memory-efficient incremental learning through feature adaptation[M]// Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 699-715.
- [69] ZHU K, ZHAI W, CAO Y, et al. Self-sustaining representation expansion for non-exemplar class-incremental learning[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 9286-9295.
- [70] BELOUADAH E, POPESCU A. IL2M: Class incremental learning with dual memory[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 583-592.
- [71] GONG Z, ZHOU K, ZHAO X, et al. Continual pre-training of language models for math problem understanding with syntax-aware memory network[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2022: 5923-5933.
- [72] TOLDO M, OZAY M. Bring evanescent representations to life in lifelong class incremental learning[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 16711-16720.
- [73] HAYES T L, KAFLE K, SHRESTHA R, et al. REMIND your neural network to prevent catastrophic forgetting[M]// Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 466-483.
- [74] WANG K, HERRANZ L, VAN DE WEIJER J. ACAE-REMIND for online continual learning with compressed feature replay[EB/OL]. (2021-07-08) [2024-05-30]. <https://arXiv.org/abs/2105.08595>.
- [75] PETIT G, POPESCU A, SCHINDLER H, et al. FeTrIL: Feature translation for exemplar-free class-incremental learning[C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2023: 3900-3909.
- [76] ZENG G X, CHEN Y, CUI B, et al. Continual learning of context-dependent processing in neural networks[J]. Nature Machine Intelligence, 2019, 1(8): 364-372.
- [77] GUO Y D, HU W P, ZHAO D Y, et al. Adaptive orthogonal projection for batch and online continual learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(6): 6783-6791.
- [78] FARAJTABAR M, AZIZAN N, MOTT A, et al. Orthogonal gradient descent for continual learning[C]//Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics. Cambridge: PMLR, 2020: 3762-3773.
- [79] SAHA G, GARG I, ROY K. Gradient projection memory for continual learning[EB/OL]. (2021-03-17) [2024-05-30]. <https://arXiv.org/abs/2103.09762>.
- [80] CHEN C, ZHANG J, SONG J K, et al. Class gradient projection for continual learning[C]//Proceedings of the 30th ACM International Conference on Multimedia. New

York: ACM, 2022: 5575-5583.

- [81] DENG D R, CHEN G Y, HAO J Y, et al. Flattening sharpness for dynamic gradient projection memory benefits continual learning[EB/OL]. (2021-10-09) [2024-05-30]. <https://arXiv.org/abs/2110.04593>.
- [82] LIN S, YANG L, FAN D L, et al. Beyond not-forgetting: Continual learning with backward knowledge transfer[EB/OL]. (2022-11-01)[2024-05-30]. <https://arXiv.org/abs/2211.00789>.
- [83] LIN S, YANG L, FAN D L, et al. TRGP: Trust region gradient projection for continual learning[EB/OL]. (2022-02-07)[2024-05-30]. <https://arXiv.org/abs/2202.02931>.
- [84] WANG S P, LI X R, SUN J, et al. Training networks in null space of feature covariance for continual learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 184-193.
- [85] KONG Y J, LIU L, WANG Z, et al. Balancing stability and plasticity through advanced null space in continual learning[M]//Computer Vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 219-236.
- [86] KAO T C, JENSEN K T, VAN DE VEN G M, et al. Natural continual learning: Success is a journey, not (just) a destination[EB/OL]. (2021-12-15) [2024-05-30]. <https://arXiv.org/abs/2106.08085>.
- [87] LIU H, LIU H P. Continual learning with recursive gradient optimization[EB/OL]. (2022-01-29) [2024-05-30]. <https://arXiv.org/abs/2201.12522>.
- [88] JAVED K, WHITE M. Meta-learning representations for continual learning[EB/OL]. (2019-10-30) [2024-05-30]. <https://arXiv.org/abs/1905.12588>.
- [89] SHAWN B, LAPO F, THOMAS M, et al. Learning to continually learn[EB/OL]. (2020-03-04) [2024-05-30]. <https://arxiv.org/abs/2002.09571>.
- [90] LEE E, HUANG C H, LEE C Y. Few-shot and continual learning with attentive independent mechanisms[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 9435-9444.
- [91] RIEMER M, CASES I, AJEMIAN R, et al. Learning to learn without forgetting by maximizing transfer and minimizing interference[EB/OL]. (2019-03-03) [2024-05-30]. <https://arxiv.org/abs/1810.11910>.
- [92] RAJASEGARAN J, KHAN S, HAYAT M, et al. iTAML: An incremental task-agnostic meta-learning approach[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 13585-13594.
- [93] GUPTA G, YADAV K, PAULL L. Look-ahead meta learning for continual learning[C]//Advances in Neural Information Processing Systems 33. San Diego: NeurIPS, 2020: 11588-11598.
- [94] CACCIA M, RODR I P. Online fast adaptation and knowledge accumulation (OSAKA): A new approach to continual learning[C]//NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems New York: ACM, 2020: 16532 - 16545.
- [95] JOSEPH K J, BALASUBRAMANIAN V N. Meta-consolidation for continual learning[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 14374-14386.
- [96] HENNING C, CERVERA M R, D'ANGELO F, et al. Posterior meta-replay for continual learning[EB/OL]. (2021-10-21)[2024-05-30]. <https://arXiv.org/abs/2103.01133>.
- [97] HURTADO J, RAYMOND-S A A. Optimizing reusable knowledge for continual learning via metalearning[C]//Proceedings of the 35th Conference on Neural Information Processing Systems. New York: ACM, 2021: 1-13.
- [98] WANG R Q, BAO Y X, ZHANG B C, et al. Anti-retroactive interference for Lifelong learning[C]//Computer Vision - ECCV 2022. Cham: Springer, 2022: 163-178.
- [99] MIRZADEH S I, FARAJTABAR M, PASCANU R, et al. Understanding the role of training regimes in continual learning[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 7308-7320.
- [100] MIRZADEH S I, FARAJTABAR M, GORUR D, et al. Linear mode connectivity in multitask and continual learning[EB/OL]. (2020-10-09) [2024-05-30]. <https://arXiv.org/abs/2010.04495>.
- [101] LIN G L, CHU H L, LAI H J. Towards better plasticity-stability trade-off in incremental learning: A simple linear connector[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 89-98.
- [102] MADAAND, YOON J, LI Y C, et al. Representational continuity for unsupervised continual learning[EB/OL]. (2022-04-05)[2024-05-30]. <https://arXiv.org/abs/2110.06976>.
- [103] PURUSHWALKAM S, MORGADO P, GUPTA A. The challenges of Continuous self-supervised learning[C]//Computer Vision - ECCV 2022. Cham: Springer, 2022: 702-721.
- [104] FINI E, COSTA V G T DA, ALAMEDA-PINEDA X, et al. Self-supervised models are continual learners[C]//2022 IEEE/CVF Conference on Computer Vision and

- Pattern Recognition. Piscataway: IEEE, 2022: 9611-9620.
- [105] CHA H, LEE J, SHIN J. Co2L: Contrastive continual learning[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 9496-9505.
- [106] PHAM Q, LIU C, HOI S. DualNet: Continual learning, fast and slow[C]//Advances in Neural Information Processing Systems 34. San Diego: NeurIPS, 2021: 16131-16144.
- [107] ZHANG J O, SAX A, ZAMIR A, et al. Side-tuning: A baseline for network adaptation via additive side networks[M]//Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 698-714.
- [108] SHON H, LEE J, KIM S H, et al. DLCFT: Deep linear continual fine-tuning for general incremental learning[M]//Computer Vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 513-529.
- [109] BOSCHINI M, BONICELLI L, PORRELLO A, et al. Transfer without forgetting[M]//Computer Vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 692-709.
- [110] CONG Y, ZHAO M, LI J, et al. GAN memory with no forgetting[C]//Advances in Neural Information Processing Systems 33. San Diego: NeurIPS, 2020: 16481-16494.
- [111] PEREZ E, STRUB F, DE VRIES H, et al. FiLM: Visual reasoning with a general conditioning layer[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 3942-3951.
- [112] ZHAO M Y, CONG Y L, CARIN L. On leveraging pretrained GANs for generation with limited data[EB/OL]. (2020-08-08) [2024-05-31]. <https://arxiv.org/abs/2002.11810>.
- [113] WANG Z F, ZHANG Z Z, LEE C Y, et al. Learning to prompt for continual learning[EB/OL]. (2022-03-21) [2024-05-30]. <https://arxiv.org/abs/2112.08654>.
- [114] SMITH J S, KARLINSKY L, GUTTA V, et al. CODA-prompt: COntinual decomposed attention-based prompting for rehearsal-free continual learning[EB/OL]. (2023-03-30)[2024-05-30]. <https://arxiv.org/abs/2211.13218>.
- [115] WANG Z F, ZHANG Z Z, EBRAHIMI S, et al. Dual-Prompt: Complementary prompting for rehearsal-free continual learning[M]//Computer Vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 631-648.
- [116] WANG Y B, HUANG Z W, HONG X P. S-prompts learning with pre-trained transformers: An Occam's razor for domain incremental learning[EB/OL]. (2023-03-18)[2024-05-31]. <https://arxiv.org/abs/2207.12819>.
- [117] WANG L Y, XIE J Y, ZHANG X X, et al. Hierarchical decomposition of prompt-based continual learning: Rethinking obscured sub-optimality[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. New York: ACM, 2023: 69054-69076.
- [118] RAZDAIBIEDINA A, MAO Y N, HOU R, et al. Progressive prompts: Continual learning for language models[EB/OL]. (2023-01-29) [2024-05-30]. <https://arxiv.org/abs/2301.12314>.
- [119] SHI Y J, ZHOU K Q, LIANG J, et al. Mimicking the oracle: An initial phase decorrelation approach for class incremental learning[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 16701-16710.
- [120] SHI G, CHEN J, ZHANG W, et al. Overcoming catastrophic forgetting in incremental few-shot learning by finding flat minima[C]//Advances in Neural Information Processing Systems 34, San Diego: NeurIPS, 2024: 6747-6761.
- [121] MEHTA S V, PATIL D, CHANDAR S, et al. An empirical investigation of the role of pre-training in lifelong learning[J]. Journal of Machine Learning Research, 2023, 24(1): 10185-10234.
- [122] FORET P, KLEINER A, MOBAHI H, et al. Sharpness-aware minimization for efficiently improving generalization[EB/OL]. (2021-04-29) [2024-05-30]. <https://arxiv.org/abs/2010.01412>.
- [123] YAN S P, HONG L Q, XU H, et al. Generative negative text replay for continual vision-language pretraining[M]//Computer Vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 22-38.
- [124] HAN R J, REN X, PENG N Y. ECONET: Effective continual pretraining of language models for event temporal reasoning[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2021: 5367-5380.
- [125] LIU Q, MAJUMDER O, ACHILLE A, et al. Incremental few-shot meta-learning via indirect discriminant alignment[M]//Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 685-701.
- [126] WANG Z Y, SHEN L, FANG L, et al. Meta-learning with less forgetting on large-scale non-stationary task distributions[M]//Computer Vision - ECCV 2022.

- Cham: Springer Nature Switzerland, 2022: 221-238.
- [127] ZHU H G, WEI Y C, LIANG X D, et al. CTP: Towards vision-language continual pretraining via compatible momentum contrast and topology preservation[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2023: 22200-22210.
- [128] KE Z X, SHAO Y J, LIN H W, et al. Continual pre-training of language models[EB/OL]. (2023-04-12) [2024-05-30]. <https://arXiv.org/abs/2302.03241>.
- [129] CHAUDHRY A, ROHRBACH M, ELHOSEINY M, et al. On tiny episodic memories in continual learning[EB/OL]. (2019-06-04)[2024-05-30]. <https://arXiv.org/abs/1902.10486>.
- [130] RICE J R, VITTER J S. Random sampling with a reservoir[J]. *ACM Transactions on Mathematical Software*, 1985, 11(1): 37-57.
- [131] LOPEZ-PAZ D, RANZATO M. Gradient episodic memory for continual learning[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6470-6479.
- [132] VAN DEN OORD A, VINYALS O, KAVUKCUOGLU KORAY. Neural discrete representation learning[C]//Advances in Neural Information Processing Systems 30. San Diego: NeurIPS, 2017: 1-10.
- [133] KULESZA A. Determinantal point processes for machine learning[J]. *Foundations and Trends in Machine Learning*, 2012, 5(2/3): 123-286.
- [134] KIM J, CHO H, KIM J, et al. SDDGR: Stable diffusion-based deep generative replay for class incremental object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 28772-28781.
- [135] DOAN K, TRAN Q, TRAN T L, et al. Class-prototype conditional diffusion model with gradient projection for continual learning[EB/OL]. (2024-03-21) [2024-05-30]. <https://arXiv.org/abs/2312.06710>.
- [136] CYWISKI B, DEJA K, TRZCISKI T, et al. GUIDE: Guidance-based incremental learning with diffusion models[EB/OL]. (2024-03-31)[2024-05-30]. <https://arXiv.org/abs/2403.03938>.
- [137] GOU J P, YU B S, MAYBANK S J, et al. Knowledge distillation: A survey[J]. *International Journal of Computer Vision*, 2021, 129(6): 1789-1819.
- [138] DHAR P, SINGH R V, PENG K C, et al. Learning without memorizing[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE, 2019, 5138-5146.
- [139] TRIKI A R, ALJUNDI R, BLASCHKO M B, et al. Encoder based lifelong learning[EB/OL]. (2017-04-06) [2024-05-30]. <https://arXiv.org/abs/1704.01920>.
- [140] ZHAI M Y, CHEN L, TUNG F, et al. Lifelong GAN: Continual learning for conditional image generation[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 2759-2768.
- [141] WANG Z, LIU L, DUAN Y Q, et al. Continual learning through retrieval and imagination[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(8): 8594-8602.
- [142] CHAUDHRY A, RANZATO M, ROHRBACH M, et al. Efficient lifelong learning with A-GEM[EB/OL]. (2019-01-09)[2024-05-30]. <https://arXiv.org/abs/1812.00420>.
- [143] TANG S X, CHEN D P, ZHU J G, et al. Layerwise optimization by gradient decomposition for continual learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 9629-9638.
- [144] CHAUDHRY A, KHAN N, DOKANIA P K, et al. Continual learning in low-rank orthogonal subspaces[EB/OL]. (2020-12-08)[2024-05-30]. <https://arXiv.org/abs/2010.11635>.
- [145] HADSELL R, RAO D, RUSU A A, et al. Embracing change: Continual learning in deep neural networks[J]. *Trends in Cognitive Sciences*, 2020, 24(12): 1028-1040.
- [146] CHEN X L, XIE S N, HE K M. An empirical study of training self-supervised vision transformers[EB/OL]. (2021-08-16)[2024-05-30]. <https://arXiv.org/abs/2104.02057>.
- [147] ERICSSON L, GOUK H, HOSPEDALES T M. How well do self-supervised models transfer?[EB/OL]. (2021-03-29)[2024-05-30]. <https://arXiv.org/abs/2011.13377>.
- [148] DAVARI M, ASADI N, MUDUR S, et al. Probing representation forgetting in supervised and unsupervised continual learning[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 16691-16700.
- [149] RAMASESH V V, LEWKOWYCZ A, DYER E. Effect of scale on catastrophic forgetting in neural networks[C]//International Conference on Learning Representations. Appleton: ICLR. 2021: 1-33.
- [150] WU T Y, SWAMINATHAN G, LI Z Z, et al. Class-incremental learning with strong pre-trained models[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 9591-

- 9600.
- [151] HAO Y R, DONG L, WEI F R, et al. Visualizing and understanding the effectiveness of BERT[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg: ACL, 2019: 4141-4150.
- [152] NEYSHABUR B, SEDGHI H, ZHANG C Y. What is being transferred in transfer learning?[EB/OL]. (2021-01-14)[2024-05-30]. <https://arXiv.org/abs/2008.11687>.
- [153] GALLARDO J, HAYES T L, KANAN C. Self-supervised training enhances online continual learning[EB/OL]. (2021-10-22)[2024-05-30]. <https://arXiv.org/abs/2103.14010>.
- [154] OSTAPENKO O, TIMOTHY E. Continual learning with foundation models: An empirical study of latent replay[C]//Proceedings of the 1st Conference on Lifelong Learning Agents. Cambridge: PMLR, 2022: 60-91.
- [155] PELOSIN F. Simpler is better: Off-the-shelf continual learning through pretrained backbones[EB/OL]. (2022-05-03)[2024-05-30]. <https://arXiv.org/abs/2205.01586>.
- [156] JANSON P, ZHANG W X, ALJUNDI R, et al. A simple baseline that questions the use of pretrained-models in continual learning[EB/OL]. (2023-03-29) [2024-05-30]. <https://arXiv.org/abs/2210.04428>.
- [157] HUANG W C, CHEN C F, HSU H. OVOR: OnePrompt with virtual outlier regularization for rehearsal-free class-incremental learning[EB/OL]. (2024-02-26) [2024-05-30]. <https://arXiv.org/abs/2402.04129>.
- [158] ROY A, MOULICK R, VERMA V K, et al. Convolutional prompting meets language models for continual learning[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 23616-23626.
- [159] GAO Z X, CEN J, CHANG X B. Consistent prompting for rehearsal-free continual learning[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 28463-28473.
- [160] ZHANG G W, WANG L Y, KANG G L, et al. SLCA: Slow learner with classifier alignment for continual learning on a pre-trained model[EB/OL]. (2023-10-09) [2024-05-30]. <https://arXiv.org/abs/2303.05118>.
- [161] MCDONNELL M D, GONG D, PARVENEH A, et al. RanPAC: Random projections and pre-trained models for continual learning[EB/OL]. (2024-01-16) [2024-05-30]. <https://arXiv.org/abs/2307.02251>.
- [162] CAO X S, LU H R, HUANG L L, et al. Generative multimodal models are good class-incremental learners[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 28706-28717.
- [163] ZHOU D W, SUN H L, YE H J, et al. Expandable subspace ensemble for pre-trained model-based class-incremental learning[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 23554-23564.
- [164] TAN Y W, ZHOU Q H, XIANG X, et al. Semantically-shifted incremental adapter-tuning is a continual ViT-transformer[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 23252-23262.
- [165] LIANG Y S, LI W J. InfLoRA: Interference-free low-rank adaptation for continual learning[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 23638-23647.
- [166] COSSU A, TUYTELAARS T, CARTA A, et al. Continual pre-training mitigates forgetting in language and vision[EB/OL]. (2022-03-09) [2024-05-30]. <https://arXiv.org/abs/2205.09357>.
- [167] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//Proceedings of the 38th International Conference on Machine Learning. Cambridge: PMLR, 2021: 8748-8763.
- [168] SHMELKOV K, SCHMID C, ALAHARI K. Incremental learning of object detectors without catastrophic forgetting[C]//2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 3420-3429.
- [169] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 1440-1448.
- [170] RAMAKRISHNAN K, PANDA R, FAN Q F, et al. Relationship matters: Relation guided knowledge transfer for incremental learning of object detectors[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2020: 1009-1018.
- [171] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as Points[EB/OL]. (2019-04-25)[2024-05-30]. <https://arxiv.org/abs/1904.07850>.
- [172] PENG C, ZHAO K, MAKSOU D, et al. SID: Incremental learning for anchor-free object detection via se-

- lective and inter-related distillation[J]. *Computer Vision and Image Understanding*, 2021, 210: 103229.
- [173] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318-327.
- [174] LI D W, TASCI S, GHOSH S, et al. RILOD: Near real-time incremental learning for object detection at the edge[C]//*Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*. New York: ACM, 2019: 113-126.
- [175] LI X, WANG W H, WU L J, et al. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection[C]//*Proceedings of the 34th International Conference on Neural Information Processing Systems*. New York: ACM, 2020: 21002-21012.
- [176] FENG T, WANG M, YUAN H J. Overcoming catastrophic forgetting in incremental object detection via elastic response distillation[C]//*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2022: 9417-9426.
- [177] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//*IEEE Transactions on Pattern Analysis and Machine Intelligence*. Piscataway: IEEE, 2017: 1137-1149.
- [178] HAO Y, FU Y W, JIANG Y G, et al. An end-to-end architecture for class-incremental object detection with knowledge distillation[C]//*2019 IEEE International Conference on Multimedia and Expo*. Piscataway: IEEE, 2019: 1-6.
- [179] PENG C, ZHAO K, LOVELL B C. Faster ILOD: Incremental learning for object detectors based on faster RCNN[EB/OL]. (2020-10-07)[2024-05-30]. <https://arXiv.org/abs/2003.03901>.
- [180] ZHANG J T, ZHANG J, GHOSH S, et al. Class-incremental learning via deep model consolidation[C]//*2020 IEEE Winter Conference on Applications of Computer Vision*. Piscataway: IEEE, 2020: 1120-1129.
- [181] DONG N, ZHANG Y, DING M, et al. Bridging non co-occurrence with unlabeled in-the-wild data for incremental object detection[C]//*Proceedings of the 35th International Conference on Neural Information Processing Systems*. New York: ACM, 2021: 30492-30503.
- [182] JOSEPH K J, RAJASEGARAN J, KHAN S, et al. Incremental object detection via meta-learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(12): 9209-9216.
- [183] ZHAO N, LEE G H. Static-dynamic co-teaching for class-incremental 3D object detection[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(3): 3436-3445.
- [184] WANG J R, WANG X, SHANG-GUAN Y, et al. Wanderlust: Online continual object detection in the real world[C]//*2021 IEEE/CVF International Conference on Computer Vision*. Piscataway: IEEE, 2021: 10809-10818.
- [185] PEREZ-RUA J M, ZHU X T, HOSPEDALES T M, et al. Incremental few-shot object detection[C]//*2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2020: 13843-13852.
- [186] MICHIELI U, ZANUTTIGH P. Incremental learning techniques for semantic segmentation[C]//*2019 IEEE/CVF International Conference on Computer Vision Workshop*. Piscataway: IEEE, 2019: 3205-3212.
- [187] MICHIELI U, ZANUTTIGH P. Knowledge distillation for incremental learning in semantic segmentation[EB/OL]. (2021-01-20)[2024-05-30]. <https://arXiv.org/abs/1911.03462>.
- [188] CERMELLI F, MANCINI M, ROTA BULÒ S, et al. Modeling the background for incremental learning in semantic segmentation[C]//*2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2020: 9230-9239.
- [189] OH Y, BAEK D, HAM B. ALIFE: Adaptive logit regularizer and feature replay for incremental semantic segmentation[J]. *Advances in Neural Information Processing Systems* 35. San Diego: NeurIPS, 2022: 14516-14528.
- [190] ZHANG C B, XIAO J W, LIU X L, et al. Representation compensation networks for continual semantic segmentation[C]//*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2022: 7043-7054.
- [191] MICHIELI U, ZANUTTIGH P. Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations[C]//*2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2021: 1114-1124.
- [192] YANG G L, FINI E, XU D, et al. Uncertainty-aware contrastive distillation for incremental semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(2): 2567-2581.

- [193] DOUILLARD A, CHEN Y F, DAPOGNY A, et al. PLOP: Learning without forgetting for continual semantic segmentation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 4039-4049.
- [194] MARACANI A, MICIELI U, TOLDO M, et al. RECALL: Replay-based continual learning in semantic segmentation[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 7006-7015.
- [195] CHA S, KIM B, YOO Y, et al. SSUL: Semantic segmentation with unknown label for exemplar-based class-incremental learning[EB/OL]. (2021-11-19) [2024-05-30]. <https://arXiv.org/abs/2106.11562>.
- [196] YAN S P, ZHOU J L, XIE J W, et al. An EM framework for online incremental learning of semantic segmentation[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 3052-3060.
- [197] YU L, LIU X L, VAN DE WEIJER J. Self-training for class-incremental semantic segmentation[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(11): 9116-9127.
- [198] CERMELLI F, FONTANEL D, TAVERA A, et al. Incremental learning in semantic segmentation from image labels[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 4361-4371.
- [199] Fortin M P, Chaib-draa B. Continual semantic segmentation leveraging image-level labels and rehearsal[C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. California: IJCAI, 2022: 1268-1275.
- [200] ZHANG Z K, GAO G Y, FANG Z Y, et al. Mining unseen classes via regional objectness: A simple baseline for incremental segmentation[EB/OL]. (2023-01-07) [2024-05-30]. <https://arXiv.org/abs/2211.06866>.
- [201] ZHANG Z K, GAO G Y, JIAO J B, et al. CoinSeg: Contrast inter- and intra- class representations for incremental segmentation[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2023: 843-853.
- [202] THANH-TUNG H, TRAN T. Catastrophic forgetting and mode collapse in GANs[C]//2020 International Joint Conference on Neural Networks. Piscataway: IEEE, 2020: 1-10.
- [203] LIANG K J, LI C Y, WANG G Y, et al. Generative adversarial network training is a continual learning problem[EB/OL]. (2018-11-27) [2024-05-30]. <https://arXiv.org/abs/1811.11083>.
- [204] 吴美君, 杨新, 潘超凡, 等. 自编码器结合持续学习: 现状、挑战与展望[J]. *计算机学报*, 2025, 48(2): 317-357.
- WU M J, YANG X, PAN C F, et al. Autoencoders combined with continual learning: Development, challenges, and prospects[J]. *Chinese Journal of Computers*, 2025, 48(2): 317-357. (in Chinese)
- [205] CHEN H T, WANG Y H, XU C, et al. Data-free learning of student networks[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 3513-3521.
- [206] BINICI K, TRUNG PHAM N, MITRA T, et al. Preventing catastrophic forgetting and distribution mismatch in knowledge distillation via synthetic data[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2022: 3625-3633.
- [207] BINICI K, AGGARWAL S, PHAM N T, et al. Robust and resource-efficient data-free knowledge distillation by generative pseudo replay[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(6): 6089-6096.
- [208] DO K, LE H, NGUYEN D, et al. Momentum adversarial distillation: Handling large distribution shifts in data-free knowledge distillation[EB/OL]. (2022-09-21) [2024-05-30]. <https://arXiv.org/abs/2209.10359>.
- [209] PATEL G, MOPURI K R, QIU Q. Learning to retain while acquiring: Combating distribution-shift in adversarial data-free knowledge distillation[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 7786-7794.
- [210] SEFF A, BEATSON A, SUO D, et al. Continual learning in generative adversarial nets[EB/OL]. (2017-05-23) [2024-05-30]. <https://arXiv.org/abs/1705.08395>.
- [211] YE F, BORS A G. Learning latent representations across multiple data domains using lifelong VAEGAN[C]//Computer Vision - ECCV 2020. Cham: Springer, 2020: 777-795.
- [212] ZHAI M Y, CHEN L, MORI G. Hyper-LifelongGAN: Scalable lifelong learning for image conditioned generation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 2246-2255.
- [213] CHEN P, ZHANG Y K, LI Z J, et al. Few-shot incre-

- mental learning for label-to-image translation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 3687-3697.
- [214] ACHILLE A, ECCLES T, MATTHEY L, et al. Life-long disentangled representation learning with cross-domain latent homologies[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 9895-9905.
- [215] EGOROV E, KUZINA A, BURNAEV E, et al. Boosting approach for continual learning of VAE[C]//Advances in Neural Information Processing Systems 34. San Diego: NeurIPS, 2021: 17889-17901.
- [216] QIU Z H, XU Y, MENG F M, et al. Dual-consistency model inversion for non-exemplar class incremental learning[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 24025-24035.
- [217] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 10674-10685.
- [218] SMITH J S, HSU Y C, ZHANG L Y, et al. Continual diffusion: Continual customization of text-to-image diffusion with C-LoRA[EB/OL]. (2024-05-02) [2024-05-30]. <https://arXiv.org/abs/2304.06027>.
- [219] ALJUNDI R, KELCHTERMANS K, TUYTELAARS T. Task-free continual learning[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 11246-11255.
- [220] DE LANGE M, ALJUNDI R, MASANA M, et al. A continual learning survey: Defying forgetting in classification tasks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(7): 3366-3385.
- [221] LEE S, HA J, ZHANG D S, et al. A neural dirichlet process mixture model for task-free continual learning[EB/OL]. (2020-01-14)[2024-05-30]. <https://arXiv.org/abs/2001.00689>.
- [222] JERFEL G, GRANT E, GRIFFITHS T L, et al. Reconciling meta-learning and continual learning with online mixtures of tasks[EB/OL]. (2019-06-19) [2024-05-30]. <https://arXiv.org/abs/1812.06080>.
- [223] ARDYWIBOWO R, HUO Z, WANG Z, et al. Vari-Grow: Variational architecture growing for task-agnostic continual learning based on Bayesian novelty[C]//Proceedings of the 39th International Conference on Machine Learning. Cambridge: PMLR, 2022: 865-877.
- [224] YE F, BORS A G. Task-free continual learning *via* online discrepancy distance learning[EB/OL]. (2022-10-12) [2024-05-30]. <https://arXiv.org/abs/2210.06579>.
- [225] CHEN H J, CHENG A C, JUAN D C, et al. Mitigating forgetting in online continual learning via instance-aware parameterization[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 17466-17477.
- [226] SUN S Y, CALANDRIELLO D, HU H Y, et al. Information-theoretic online memory selection for continual learning[EB/OL]. (2022-04-10) [2024-05-30]. <https://arXiv.org/abs/2204.04763>.
- [227] CHRYSAKIS A, MOENS M F. Online continual learning from imbalanced data[C]//Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 1952-1961.
- [228] PRABHU A, TORR P H S, DOKANIA P K. GDumb: A simple approach that questions our progress in continual learning[M]//Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 524-540.
- [229] DE LANGE M, TUYTELAARS T. Continual prototype evolution: Learning online from non-stationary data streams[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 8230-8239.
- [230] JIN X S, SADHU A, DU J Y, et al. Gradient-based editing of memory examples for online task-free continual learning[EB/OL]. (2021-12-07)[2024-05-30]. <https://arXiv.org/abs/2006.15294>.
- [231] WANG Z Y, SHEN L, FANG L, et al. Improving task-free continual learning by distributionally robust memory evolution[C]//International Conference on Machine Learning. Cambridge: PMLR, 2022: 22985-22998.
- [232] ALJUNDI R, BELILOVSKY E, TUYTELAARS T, et al. Online continual learning with maximal interfered retrieval[C]//Advances in Neural Information Processing Systems 32. San Diego: NeurIPS, 2019: 1-12.
- [233] GU Y N, YANG X, WEI K, et al. Not just selection, but exploration: Online class-incremental continual learning via dual view consistency[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 7432-7441.
- [234] YAN H W, WANG L Y, MA K S, et al. Orchestrate latent expertise: Advancing online continual learning with multi-level supervision and reverse self-distillation[C]//2024 IEEE/CVF Conference on Computer Vision and

- Pattern Recognition. Piscataway: IEEE, 2024: 23670-23680.
- [235] HAN Y N, LIU J W. Adaptive instance similarity embedding for online continual learning[J]. Pattern Recognition, 2024, 149: 110238.
- [236] KIM B, SEO M, CHOI J. Online continual learning for interactive instruction following agents[EB/OL]. (2024-03-13)[2024-05-30]. <https://arxiv.org/abs/2403.07548>.
- [237] WANG M R, MICHEL N, XIAO L, et al. Improving plasticity in online continual learning via collaborative learning[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 23460-23469.
- [238] SEO M, KOH H, JEUNG W, et al. Learning equi-angular representations for online continual learning[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 23933-23942.
- [239] HE Y H, CHEN Y J, JIN Y H, et al. DYSON: Dynamic feature space self-organization for online task-free class incremental learning[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 23741-23751.
- [240] OPENAI, ACHIAM J, ADLER S, et al. GPT-4 technical report[EB/OL]. (2024-03-04) [2024-05-30]. <https://arxiv.org/abs/2303.08774>.
- [241] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything[EB/OL]. (2023-04-05) [2024-05-30]. <https://arXiv.org/abs/2304.02643>.
- [242] OQUAB M, DARCET T, MOUTAKANNI T, et al. DINOv2: Learning robust visual features without supervision[EB/OL]. (2024-02-02) [2024-05-30]. <https://arXiv.org/abs/2304.07193>.
- [243] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[EB/OL]. (2020-12-16) [2024-05-30]. <https://arxiv.org/abs/2006.11239>.
- [244] SONG J M, MENG C L, ERMON S. Denoising diffusion implicit models[EB/OL]. (2022-10-05) [2024-05-30]. <https://arXiv.org/abs/2010.02502>.
- [245] CHEN Z, WU J N, WANG W H, et al. Intern VL: Scaling up vision foundation models and aligning for generic visual-linguistic tasks[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 24185-24198.
- [246] LIU H, LI C, WU Q, et al. Visual instruction tuning[EB/OL]. (2023-12-11)[2024-05-30]. <https://arxiv.org/abs/2304.08485>.
- [247] HU E J, SHEN Y L, WALLIS P, et al. LoRA: Low-rank adaptation of large language models[EB/OL]. (2021-10-16)[2024-07-02]. <https://arXiv.org/abs/2106.09685>.
- [248] HOULSBY N, GIURGIU A, JASTRZEBSKI S, et al. Parameter-efficient transfer learning for NLP[C]//Proceedings of the 36th International Conference on Machine Learning. Cambridge: PMLR, 2019: 2790-2799.
- [249] LESTER B, AL-RFOU R, CONSTANT N. The power of scale for parameter-efficient prompt tuning[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2021: 3045-3059.
- [250] SMITH J S, KARLINSKY L, GUTTA V, et al. CODA-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 11909-11919.
- [251] WISTUBA M, SIVAPRASAD P T, BALLE L, et al. Continual learning with low rank adaptation[EB/OL]. (2023-11-29)[2024-07-02]. <https://arXiv.org/abs/2311.17601>.
- [252] GAO Q K, ZHAO C, SUN Y F, et al. A unified continual learning framework with general parameter-efficient tuning[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2023: 11449-11459.
- [253] BÜYÜKAKYÜZ K. OLoRA: Orthonormal low-rank adaptation of large language models[EB/OL]. (2024-11-25) [2025-05-02]. <https://arXiv.org/abs/2406.01775>.
- [254] CHITALE R, VAIDYA A, KANE A, et al. Task arithmetic with LoRA for continual learning[EB/OL]. (2023-11-04)[2024-05-30]. <https://arxiv.org/abs/2311.02428>.
- [255] YU J Z, ZHUGE Y Z, ZHANG L, et al. Boosting continual learning of vision-language models via mixture-of-experts adapters[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 23219-23230.
- [256] ZHANG H, LEI Y, GUI L, et al. CPPO: Continual learning for reinforcement learning with human feedback[EB/OL]. (2024-12-21)[2024-06-30]. <https://arxiv.org/abs/2402.14228>.
- [257] PENG B H, TIAN Z T, LIU S, et al. Scalable language model with generalized continual learning[EB/OL]. (2024-04-11)[2024-05-30]. <https://arXiv.org/abs/2404.07470>.
- [258] WANG G Z, XIE Y Q, JIANG Y F, et al. Voyager: An open-ended embodied agent with large language models [EB/OL]. (2023-10-19)[2024-05-30]. <https://arXiv.org/>

- abs/2305.16291.
- [259] TZIAFAS G, KASAEI H. Lifelong robot library learning: Bootstrapping composable and generalizable skills for embodied control with language models[C]//2024 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2024: 515-522.
- [260] MENDEZ-MENDEZ J, KAELBLING L P, LOZANO-PÉREZ T. Embodied lifelong learning for task and motion planning[EB/OL]. (2023-11-06)[2024-05-30]. <https://arXiv.org/abs/2307.06870>.
- [261] 于达. 克服灾难性遗忘的连续学习方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2024.
- YU D. Research on Continual Learning Method for Overcoming Catastrophic Forgetting[D]. Harbin: Harbin Institute of Technology, 2024. (in Chinese)
- [262] MU Y, CHEN T X, PENG S J, et al. RoboTwin: Dual-arm robot benchmark with generative digital twins (early version)[EB/OL]. (2024-09-04)[2024-09-30]. <https://arXiv.org/abs/2409.02920>.
- [263] LIU S M, WU L X, LI B G, et al. RDT-1B: A diffusion foundation model for bimanual manipulation[EB/OL]. (2025-03-01)[2025-05-02]. <https://arXiv.org/abs/2410.07864>.
- [264] KIM M J, PERTSCH K, KARAMCHETI S, et al. OpenVLA: An open-source vision-language-action model [EB/OL]. (2024-09-05)[2024-09-30]. <https://arXiv.org/abs/2406.09246>.
- [265] TEAM O M, GHOSH D, WALKE H, et al. Octo: An open-source generalist robot policy[EB/OL]. (2024-05-26)[2024-05-30]. <https://arXiv.org/abs/2405.12213>.
- [266] BROHAN A, BROWN N, CARBAJAL J, et al. RT-1: Robotics transformer for real-world control at scale[EB/OL]. (2023-08-11)[2024-05-30]. <https://arXiv.org/abs/2212.06817>.
- [267] LICĂ M, SHIREKAR O, COLLE B, et al. MindForge: Empowering embodied agents with theory of mind for lifelong cultural learning[EB/OL]. (2024-11-20) [2024-11-30]. <https://arXiv.org/abs/2411.12977>.
- [268] ZHAO Z H, CHAI W H, WANG X, et al. See and think: Embodied agent in virtual environment[M]//Computer Vision - ECCV 2024. Cham: Springer Nature Switzerland, 2024: 187-204.
- [269] HE C Y, CHENG S X, QIU Z H, et al. Continual egocentric activity recognition with foreseeable-generalized visual-IMU representations[J]. IEEE Sensors Journal, 2024, 24(8): 12934-12945.
- [270] GRAUMAN K, WESTBURY A, BYRNE E, et al. Ego4D: Around the world in 3 000 hours of egocentric video[EB/OL]. (2022-03-11)[2024-05-30]. <https://arxiv.org/abs/2110.07058>.
- [271] SAPOVAL N, AGHAZADEH A, NUTE M G, et al. Current progress and open challenges for applying deep learning across the biosciences[J]. Nature Communications, 2022, 13: 1728.
- [272] WEBB S. Deep Learning for Biology[M]. California: O'Reilly Media, Inc., 2018.
- [273] ZENI C, PINSLER R, ZÜGNER D, et al. MatterGen: A generative model for inorganic materials design[EB/OL]. (2024-01-29)[2024-05-30]. <https://arXiv.org/abs/2312.03687>.
- [274] MERCHANT A, BATZNER S, SCHOENHOLZ S S, et al. Scaling deep learning for materials discovery[J]. Nature, 2023, 624(7990): 80-85.
- [275] BI K F, XIE L X, ZHANG H H, et al. Accurate medium-range global weather forecasting with 3D neural networks[J]. Nature, 2023, 619(7970): 533-538.
- [276] JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold[J]. Nature, 2021, 596(7873): 583-589.
- [277] SINGH V, SHRIVASTAVA S, SINGH S K, et al. Multi-scale temporal convolutional networks and continual learning based in silico discovery of alternative antibiotics to combat multi-drug resistance[J]. Expert Systems with Applications, 2023, 215: 119295.
- [278] ZHANG Y X, LI X Y, CHEN H M, et al. Continual learning for Abdominal multi-organ and Tumor segmentation[C]//Medical Image Computing and Computer Assisted Intervention - MICCAI 2023. Cham: Springer, 2023: 35-45.
- [279] BRINGAS S, DUQUE R, LAGE C, et al. CLADSI: Deep continual learning for Alzheimer's disease stage identification using accelerometer data[J]. IEEE Journal of Biomedical and Health Informatics, 2024, 28(6): 3401-3410.
- [280] QAZI M A, ALMAKKY I, HASHMI A U R, et al. DynaMMo: Dynamic model merging for efficient class incremental learning for medical images[M]//Medical Image Understanding and Analysis. Cham: Springer Nature Switzerland, 2024: 245-257.
- [281] KUMARI P, CHAUHAN J, BOZORGPOUR A, et al. Continual learning in medical image analysis: A comprehensive review of recent advancements and future prospects[EB/OL]. (2023-12-28)[2024-05-30]. <https://arXiv.org/abs/2312.06817>.

org/abs/2312.17004.

- [282] WEI B Q, CHEN J J, TSENG Y C, et al. Representative data selection for efficient medical incremental learning[J]. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2023: 1-4.
- [283] LI C H, JHA N K. DOCTOR: A multi-disease detection continual learning framework based on wearable medi-

cal sensors[J]. ACM Transactions on Embedded Computing Systems, 2024, 23(5): 1-33.

- [284] ZHU X Y, JIANG Z G, WU K, et al. Lifelong histopathology whole slide image retrieval via distance consistency rehearsal[M]//Medical Image Computing and Computer Assisted Intervention - MICCAI 2024. Cham: Springer Nature Switzerland, 2024: 274-284.

## 作者简介



**方岩** 男, 2000年8月出生于安徽省宣城市. 现为北京交通大学信息科学研究所电子科技博士生. 主要研究方向包括计算机视觉与深度学习基础模型.

E-mail: yanfang@bjtu.edu.cn



**左旺孟** 男, 1977年6月出生于河南省新乡市. 2007年毕业于哈尔滨工业大学, 获工学博士学位. 现为哈尔滨工业大学计算机学院特聘教授、博士生导师. 主要研究方向包括元学习、迁移学习和生成式对抗网络模型方法及其在底层视觉、图像生成、视觉跟踪、物体检测和图像分类等领域的应用.

E-mail: cswmzuo@gmail.com



**魏云超** 男, 1987年8月出生于河北省邢台市. 2016年毕业于北京交通大学, 获工学博士学位. 现为北京交通大学计算机科学与技术学院教授、博士生导师. 主要研究方向包括面向非完美数据的视觉认知、图像/视频的分割和物体检测、多模态数据建模、生成式人工智能等. 中国电子学会会员编号: E190130784M.

E-mail: yunchao.wei@bjtu.edu.cn



**赵耀** 男, 1967年12月出生于江苏省徐州市. 1996年获北京交通大学工学博士学位. 现为北京交通大学计算机科学与技术学院教授、博士生导师. 主要研究方向包括数字媒体信息处理与智能分析、人工智能、计算机视觉、AIGC、AI视频编码. 中国电子学会会员编号: E190004729S.

E-mail: yzhao@bjtu.edu.cn



**丛润民** 男, 1989年9月出生于山东省潍坊市. 2019年毕业于天津大学信息与通信工程专业, 获工学博士学位. 现为山东大学控制科学与工程学院教授、博士生导师. 主要研究方向包括计算机视觉、人工智能、多媒体信息处理、视觉显著性计算、遥感影像解译、水下环境感知、智慧医疗、深度学习等. 中国电子学会会员编号: E190025265S.

E-mail: rmcong@sdu.edu.cn