E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP181　　**文献标识码**: A　　**文章编号**: 1006-8961(2025)06-1690-27

论文引用格式:Fu H, Feng Q, Tu J H, Zhao H B, Zhang C, Du X and Qian H. 2025. Advances in incremental learning research. Journal of Image and Graphics, 30(6):1690-1716(付浩, 冯前, 涂嘉航, 赵涵斌, 张超, 杜歆, 钱徽. 2025. 增量学习研究进展. 中国图象图形学报, 30(6):1690-1716)[DOI:10.11834/jig.240790]

增量学习研究进展

付浩[1],冯前[1],涂嘉航[1],赵涵斌[1*],张超[1],杜歆[2],钱徽[1]

1. 浙江大学计算机科学与技术学院,杭州　310027; 2. 浙江大学信息与电子工程学院,杭州　310027

摘　要:海量的数据和计算机强大的计算能力使深度模型在众多单模态和多模态任务上取得优异性能。当前高性能深度模型通常在静态的学习场景中训练,模型只在全部的数据集上进行一次联合训练。然而,在实际应用中,数据是不断产生的,任务以多批次的形式持续地到达,在这种环境下,深度模型面临动态的学习场景,即增量学习场景。由于无法同时访问所有的旧任务数据,增量学习场景中的深度模型在训练时会面临灾难性遗忘问题。如何缓解灾难性遗忘问题是增量学习领域的重要研究目标。本文围绕增量学习领域的研究进展,从增量学习问题定义、评估指标、增量学习范式和增量学习挑战总结增量学习相关背景;从模型参数正则化、样本重放、模型结构化和预训练模型微调4个角度汇总增量学习最新方法;从语义分割、图像生成和文本生成3个单模态领域应用以及视觉—语言、音频—视觉两个多模态领域应用归纳增量学习应用及相关方法。从国内和国外两个角度对比增量学习领域的科研投入和发展情况,并对增量学习领域的未来发展进行展望。本文可为研究人员和从业人员提供增量学习领域的最新进展。

关键词:增量学习(IL);持续学习;灾难性遗忘(CF);机器学习;深度学习

Advances in incremental learning research

Fu Hao[1], Feng Qian[1], Tu Jiahang[1], Zhao Hanbin[1*], Zhang Chao[1], Du Xin[2], Qian Hui[1]

1. College of Computer Science and Technology, Zhejiang University, Hangzhou　310027, China;

2. College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou　310027, China

Abstract: The rapid development of information technology has led to the explosive growth of multimedia data, including text, images, and videos. The vast amount of data and the powerful computational capabilities have significantly driven advances in deep learning, allowing models to achieve exceptional performance in single-modal tasks, such as object detection and semantic segmentation, as well as multimodal tasks, such as cross-modal retrieval and question answering. High-performance deep learning models are typically developed in static learning scenarios in which they can access all training data simultaneously and update the model using the entire dataset. However, in real-world applications, deep models always learn from task streams that dynamically receive new data over time. In these scenarios, deep models must simultaneously retain old task data and maintain performance on old and new tasks through repeated joint training. This approach of unrestricted joint training and continuous dataset expansion to maintain high performance incurs substantial time and financial costs. Furthermore, the prolonged runtime of high-power devices during joint training leads to significant carbon emissions, thereby contributing to environmental pollution. Such costs are even higher in multimodal scenarios due to the

收稿日期:2024-12-30;**修回日期**:2025-01-13;**预印本日期**:2025-01-20

*通信作者:赵涵斌　zhaohanbin@zju.edu.cn

基金项目:国家自然科学基金项目(62402430, 62206248, 62476238)

Supported by: National Natural Science Foundation of China (62402430, 62206248, 62476238)

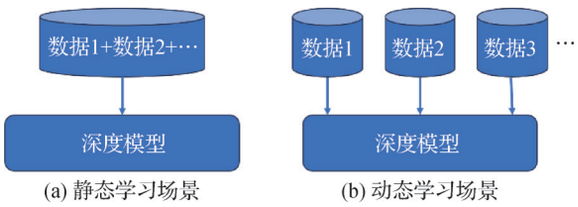
vast amounts of multimodal data and the large number of model parameters. This learning strategy, which is misaligned with the sustainability goals of artificial intelligence (AI), is impractical for real-world applications. Therefore, deep models must be capable of adapting to new tasks in dynamic environments, commonly known as incremental learning scenarios. However, developing a high-performance incremental learning model in incremental learning scenarios remains a challenging task, because the model must update its knowledge base each time it receives new data, without access to previously learned data. Due to the unavailability of previously learned data, models face the problem of catastrophic forgetting (CF) in which they tend to forget previously acquired knowledge when learning new tasks. CF degrades model performance, especially in scenarios with high privacy and security requirements, such as personal data in medical image processing. Therefore, incremental learning methods must possess the ability to acquire new knowledge (plasticity) while simultaneously retaining previously learned knowledge (stability). Enhancing stability reduces plasticity, while increasing plasticity causes instability in old tasks. These conflicting demands form the stability-plasticity dilemma, and resolving it is the primary challenge for researchers specializing in the field of incremental learning. This paper categorizes incremental learning methods into four perspectives: regularization-, replay-, and architecture-based methods, as well as methods based on fine-tuning pre-trained model. Regularization-based incremental learning methods mitigate catastrophic forgetting by adding regularization terms that optimize model parameters, adjusting the update of key parameters related to previous tasks while learning new tasks. Such an approach can be divided into two types: parameter and output regularization. The former adds regularization terms to the model's parameters, while the latter applies them to the model's outputs. Replay-based incremental learning methods transfer key knowledge from representative old samples to mitigate catastrophic forgetting. Replay-based methods are further divided into generation- and experience-based replay wherein the former generates old representative samples using generative models, while the latter retains actual old representative samples for training. Structure-based methods mitigate CF by maintaining task-specific model parameters. These methods are subdivided into neuron-expansion and parameter-isolation approaches. Neuron-expansion approaches allocate new model parameters for each task by expanding the network, while parameter-isolation approaches freeze key parameters of old tasks to ensure that learning new tasks does not interfere with previous ones. The fine-tuning of pre-trained model methods apply fine-tuning strategies to large models to learn new tasks directly, while leveraging their generalization ability to maintain performance on old tasks. Such methods can be divided into prompt- and representation-based fine-tuning approaches. The former introduces prompts to fine-tune the pre-trained model, thereby balancing old and new knowledge. In comparison, the latter directly leverages the model's generalization ability to build classifiers, utilizing high-quality feature representations to handle new tasks without significant changes to the model architecture. This paper also provides a mathematical definition of incremental learning scenarios and the objective function for optimizing incremental learning models. In particular, it summarizes six key evaluation metrics used in the incremental learning field: average accuracy (AA), average incremental accuracy (AIA), average forgetting rate (AF), forward transfer rate (FTR), backward transfer rate (BTR), and learning plasticity (LP). Based on the tasks stream pattern in incremental scenarios, the need for task identification, and the number of classification heads, this paper divides incremental learning into three subfields: task-incremental learning (TIL), domain-incremental learning (DIL), and class-incremental learning (CIL), along with detailed explanations and mathematical definitions. This paper summarizes the latest research progress in single-modal fields, including semantic segmentation, image generation, and large language models, as well as multimodal fields like vision-language and vision-audio multimodal incremental learning, based on the application scenarios of incremental learning. Furthermore, this paper surveys the publication status of papers in leading English journals and major conferences in the field to compare the research levels of incremental learning between domestic and international scholars. The comparison analyzes research investments and progress in incremental learning, both domestically and internationally, based on the total number of published papers and the average citation count per paper. Finally, this article anticipates three future directions for the development of incremental learning: large multimodal model incremental learning, incremental learning based on novel deep network architectures, and continual forgetting of knowledge for AI security.

Key words: incremental learning (IL); continual learning; catastrophic forgetting (CF); machine learning; deep learning

0 引 言

近年来，海量的数据和强大的计算能力促进了深度学习的发展，使深度神经网络（深度模型）不仅在图像识别、目标检测、图像去噪和机器翻译等单模态任务中取得了优异的性能（Floridi 和 Chiriatti，2020），还在机器人交互、社交媒体内容分析、跨模态信息检索以及语音问答等多模态领域表现出较高的准确性和效率（Bayoudh 等，2022）。深度模型通常面临静态的学习场景，即模型在初始化阶段能取得全部训练数据，并在全部训练数据上更新模型（Delange 等，2021）。随着互联网的快速发展，图像、视频和文本等数据在实际应用中呈现出持续增长和高频更新的特点。例如，人脸识别的普及使得移动客户端产生大量新的个体面部信息；短视频平台创作者数量不断增加产生大量的新视频；海量的待分类新闻资讯带来推荐问题；道路数量增加和路面情况千变万化使自动驾驶模型的学习愈发困难；大模型的快速发展需要不断应对日益增加的多模态数据和不断丰富的模态种类。上述应用需要深度模型动态地学习新任务，而在这些动态的学习场景中，深度模型仍缺乏持续学习的能力。人类的学习是一个循序渐进的知识迁移（knowledge transfer）过程，需要不断地获取新知识，并在整合新旧知识的过程中对旧知识进行补充和修正（French，1999）。为了使得深度模型在动态的学习场景中也能接近或达到人类的智能水平，模型也需要不断地积累知识，这种动态的学习场景称为增量学习（incremental learning，IL）场景（Delange 等，2021）。如图 1 所示，在增量学习场景中，深度模型的学习目标是一个任务流，其在各个特定的阶段只能访问当前阶段的任务的训练数据，并利用这些训练数据更新模型（Golab 和 Özsu，2003）以不断获取新知识，同时要求模型保留从旧任务中已经学习得到的旧知识。

增量学习的核心问题是灾难性遗忘（catastrophic forgetting，CF）问题，即由于隐私安全（Chamikara 等，2018；Delange 等，2021）和存储受限（Krempl 等，2014；Gaber，2012），旧任务的训练数据通常无法被全部保留，模型在学习新任务时只能使用新任务的训练数据，进而导致模型在旧任务上的性能出现大幅下降。在多模态相关应用中，深度模型需要学习多个模态的知识，同时还要克服模态间的距离，更加剧了灾难性遗忘。为了缓解灾难性遗忘问题，深度模型需要同时具有可塑性（plasticity）和稳定性（stability）（Delange 等，2021）。可塑性要求模型具有从新数据中提炼新知识的能力；稳定性指模型在学习新知识时，对已经获取的旧知识的保留能力。然而，这两种性质在增量学习场景中通常呈现此消彼长的关系，引发了稳定性—可塑性困境（stability-plasticity dilemma）。为了缓解稳定性—可塑性困境，研究者们提出众多增量学习的方法以平衡两者的关系。本文根据知识迁移的不同角度，将增量学习方法总结为 4 类：基于正则化的增量学习、基于样本重放的增量学习、基于结构化的增量学习和基于预训练模型微调的增量学习。



(a) 静态学习场景　(b) 动态学习场景

图 1　动态学习场景和静态学习场景对比

Fig. 1　Comparison of dynamic and static learning scenarios ((a) static learning scenario; (b) dynamic learning scenario)

1）基于正则化的增量学习，从模型优化的角度迁移旧模型知识。通过引入正则项实现新旧知识的平衡，根据正则项添加的位置具体分为基于模型参数的正则化和基于网络输出的正则化。（1）基于模型参数的正则化在模型参数层面增加正则项，根据新旧任务的重要性，对模型参数分配不同的更新权重；（2）基于网络输出的正则化在模型的中间层或最后一层的输出中利用知识蒸馏（knowledge distillation，KD）等将旧模型中的重要知识有导向性的蒸馏到新模型中来保留模型对旧任务的性能。

2）基于样本重放的增量学习，从样本信息的角度迁移旧模型知识。这类方法在训练新数据时会储存部分旧样本信息，并在学习下一个任务时，将这些样本和新数据一起联合训练实现新旧知识的平衡。根据旧样本信息的不同来源，可以细分为基于生成样本的重放和基于经验样本的重放。（1）基于生成样本的重放在学习新任务的同时利用新数据训练一个生成模型，在学习下一个任务时通过该生成模型逆向生成伪样本（Shin 等，2017）模拟旧任务数据；（2）

基于经验样本的重放则在每次训练新任务后从训练数据中随机选取或有目的性挑选出一部分真实数据，或保存旧模型某一层的旧样本特征。

3）基于结构化的增量学习，从模型结构的角度迁移旧模型知识。通过为每一个任务维护特定的模型参数实现新旧知识的平衡。参数维护方法可以细分为动态架构和参数隔离。（1）基于动态架构的增量学习方法通过压缩、扩展神经网络的方式更新模型以适应新任务并保留模型对于旧任务的性能；（2）基于参数隔离的方法在学习新任务时专注于通过路径模块（Fernando 等，2017）、门控模块（Abati 等，2020）等对旧任务的重要参数进行冻结，或通过剪枝等方式消除旧任务的冗余参数以释放参数空间学习新任务参数。

4）基于预训练模型微调的增量学习，从大模型泛化能力的角度迁移旧模型知识。将大模型作为基础，通过新任务数据定向更新模型参数。基于预训练模型微调的增量学习可以细分为基于提示的预训练模型微调和基于表征的预训练模型微调。（1）基于提示的预训练模型微调冻结预训练模型参数，并维护一个提示池，在增量学习模型推理时，从提示池中选取提示信息来引导模型的推理过程；（2）基于表征的预训练模型微调方法利用预训练模型的泛化能力来适配下游任务。这类方法利用预训练模型的高质量特征表征能力，通过构造分类器进行新任务的分类或回归，不需要在模型架构上做出大的改动。

本文围绕增量学习场景中的深度模型，详细探讨其背景、发展与挑战。

1 增量学习基础

1.1 研究背景

近年来，机器学习的下游应用例如图像识别、目标检测、人机交互以及跨模态检索等各种技术的需求日益增加，持续产生的新数据、超百亿的模型参数以及不断丰富的模态种类使得深度网络的训练代价越来越大，这引发了研究者重新审视传统深度网络训练的合理性——在训练传统深度网络时，深度模型通常面临静态的学习场景，在模型初始化阶段能够获取全部训练数据，并且在学习结束后不再对当前模型进行任何更新，对新数据的学习只能通过重新训练的途径。这种静态的学习场景不仅带来了大量的时间和能源的消耗（Narayanan 等，2021）、数据隐私问题，并面临保存旧任务数据带来持续增长的存储开销问题。如图 1（a）所示，静态环境下的深度模型在整个训练数据集上学习，并不具有连续学习新任务的能力，这种处理新任务的局限性催生了动态环境下的增量学习模型。图 1（b）表示动态场景下增量学习模型的学习过程。目标任务不再同时被模型学习，而是以任务流的方式依次到达。增量学习模型在每学习一组任务后都会更新，因而具有可以连续处理新任务的特点。一个高性能的增量学习模型可以节省大量的时间成本和物理成本，并且不再依赖于对旧数据的长期保存。

在当前大规模动态环境下，增量学习模型的应用范围和适配场景更广泛，这是由于增量学习模型具有 3 个特点。1）缓解存储限制。对于大多数深度模型，联合训练（joint training）是模型的性能上限。但是联合训练需要大规格的数据库储存所有已经出现过的样本信息，同时还需要持续的扩容数据库规模来应对大量新数据的出现。通过无限制的对数据库的扩容保持模型的高性能需要高昂的储存开销。在多模态场景中，由于训练数据量大，模型参数多的原因，联合训练成本更是成倍增加，在实际应用中是不可取的。增量学习模型采取非联合训练的方式，通过牺牲少量的模型性能缓解存储问题，从而打破静态环境下的深度学习被存储所限制的壁垒。2）保护数据隐私。储存所有旧任务样本在实际应用中通常也是不可取的，因为很多数据并不被允许长时间的保存。例如，在医疗应用中，未经允许长时间保存病人的个人信息是被严格禁止的（McClure 等，2018）。政府也不能在未经许可的情况下收集和使用个人信息（Mittelstadt 等，2016）。增量学习具有在不访问所有旧任务样本的情况下更新模型的特点，因此，增量学习可以在一定程度上避免静态环境下的机器学习因数据隐私带来的困扰。3）人工智能的可持续性。通过长时间高功率运行的计算机对超过数百亿个参数的大模型进行一次联合训练会产生大量的碳排放。GPT-3（generative pre-trained Transformer 3）具有 1 750 亿个参数，对 GPT-3 模型进行一次联合训练需要 1 024 块 NVIDIA A100 GPU 连续不停的高负荷运作 34 天（Narayanan 等，2021），同时会消耗 1 287 兆瓦时的电能，并产生 502 吨的碳排放，相当于 112 辆车行驶一年排出的尾气（Patterson

等,2021)。这不符合可持续人工智能的要求。增量学习模型得益于可持续学习的特点,大大减少了大模型联合训练的次数,进而避免因训练而带来的碳排放。

1.2 问题定义

本节从数学角度定义增量学习场景。对于动态场景下,具有k个子任务的任务流的定义为

$$\boldsymbol{T}=[(\boldsymbol{D}^1,\boldsymbol{C}^1),(\boldsymbol{D}^2,\boldsymbol{C}^2),\cdots,(\boldsymbol{D}^k,\boldsymbol{C}^k)] \tag{1}$$

式中,$\boldsymbol{C}^t$表示第t个任务的样本真实标签集合,$\boldsymbol{C}^t=\{c_1^t,c_2^t,\cdots,c_{n^t}^t\}$,定义$\hat{\boldsymbol{C}}^t=\bigcup_{i=1}^{t}$,$\boldsymbol{C}^t$表示增量学习$t$个任务时所有的类别;$\boldsymbol{D}^t$表示第$t$个任务的训练集,$\boldsymbol{D}^t=\{(x_1^t,y_1^t),(x_2^t,y_2^t),\cdots,(x_{m^t}^t,y_{m^t}^t)\}$,其中$x$是训练时样本的输入数据,可以是单模态或多模态,$y$是样本的真实标签,有$y^t\in\boldsymbol{C}^t$,真实标签$y$以独热编码的形式表示,即$y_i\in\{0,1\}^{\hat{C}^t}$。模型在学习第$t$个任务期间只能访问当前任务的数据$\boldsymbol{D}^t$和标签集合$\boldsymbol{C}^t$。增量学习任务要优化的期望风险为

$$f^*=\arg\min_{f\in\mathcal{H}}\mathop{\mathbb{E}}_{(x,y)\in\boldsymbol{D}^1\cup\boldsymbol{D}^2\cup\cdots\cup\boldsymbol{D}^t}\ell(f(x),y) \tag{2}$$

式中,$\mathcal{H}$是函数假设空间,ℓ为交叉熵函数,衡量预测值和真实标签的差距,(x,y)为新旧任务的训练样本,在实际应用中,旧数据往往是不可用的,因此,可以使用符合新旧任务数据分布的伪样本$(\tilde{x},\tilde{y})$代替真实训练样本,具体为

$$f^*=\arg\min_{f\in\mathcal{H}}\mathop{\mathbb{E}}_{(\tilde{x},\tilde{y})\sim\boldsymbol{D}^1\cup\boldsymbol{D}^2\cup\cdots\cup\boldsymbol{D}^t}\ell(f(\tilde{x}),\tilde{y}) \tag{3}$$

增量学习的优化目标是在函数假设空间中找出使所有的$(\tilde{x},\tilde{y})$总精准度最高的函数f^*。

1.3 评估指标

增量学习模型性能受多种因素的影响,例如模型大小、模型结构、数据集规模以及任务流构成等。本节从准确率、遗忘率、迁移率和模型学习可塑性4个角度总结了领域内6种比较常用的增量学习模型评估指标(Lopez-Paz和Ranzato,2017)。

首先定义p_m^n为模型学习过m个任务之后,对于第n个旧任务在测试集上的准确率($n\leqslant m$).

1)平均准确率(average accuracy,AA)表示模型当前的全局准确率,可以当做衡量模型整体性能的指标。AA_m表示模型在学习第m个新任务后,针对于所有任务的准确率。平均准确率的定义为

$$AA_m=\frac{1}{m}\sum_{n=1}^{m}p_m^n \tag{4}$$

平均准确率的优点是衡量当前模型的时候考虑了所有的旧任务。

2)平均增量准确率(average incremental accuracy,AIA)衡量所有新旧模型的平均准确率,定义为

$$AIA_m=\frac{1}{m}\sum_{n=1}^{m}AA_n \tag{5}$$

平均增量准确率的优点是考虑了模型的历史变化,更能体现出增量学习模型的鲁棒性。

3)平均遗忘率(average forgetting,AF)表示模型对于旧任务的遗忘程度。f_m^n是在增量学习的前$m-1$次迭代中,第n个任务准确率最高的数值和第m次迭代时其准确率的差值,具体为

$$f_m^n=\max_{i\in\{n,n+1,\cdots,m-1\}}p_i^n-p_m^n \tag{6}$$

平均遗忘率以F_m表示,具体为

$$AF_m=\frac{1}{m-1}\sum_{n=1}^{m-1}f_m^n \tag{7}$$

平均遗忘率的优点是它衡量了增量学习模型在学习新任务时,对于旧任务的稳定性。此外,初始任务遗忘率表示模型在增量学习过程中,增量学习模型在第1个任务上的性能差值。具体可以表示为

$$F_m^1=p_1^1-p_m^1 \tag{8}$$

初始任务遗忘率相对于平均遗忘率更加简易,但在一定程度上也可以反映增量学习模型对旧任务的遗忘程度。

4)前向迁移率(forward transfer forgetting,FWT)延伸到模型整体的遗忘率,即在学习第m个任务时,考虑所有前$m-1$个任务,定义为

$$FWT_m=\frac{1}{m-1}\sum_{n=2}^{m}\left(p_n^n-\kappa_n\right) \tag{9}$$

式中,κ_n是随机初始化模型通过第n个任务数据上训练的模型精度。较大的FWT表明模型不仅能够吸收新知识,还能通过之前学习的旧知识加深对新知识的理解。

5)后向迁移率(backward transfer,BWT)用于测量在学习新任务时对先前任务的影响,其定义为

$$BWT_m=\frac{1}{m-1}\sum_{n=1}^{m-1}\left(p_m^n-p_n^n\right) \tag{10}$$

BWT可以是正向的(帮助)或负向的(遗忘)。BWT的值越高,增量学习过程对旧任务的帮助就越多。BWT的优点是可以直观给出新知识的学习对于旧知识的影响。

6)学习可塑性(learning plasticity, LP)衡量模型

对新任务的学习性能。衡量标准为模型联合训练性能与增量学习性能的差距。联合训练要求所有已学习任务的数据$\boldsymbol{D}=\bigcup_{n=1}^{m}\boldsymbol{D}_m$保持可用状态。联合训练模型对于新任务$m$的性能为$p_m^*$，增量学习模型对于新任务的性能为$p_m^m$，增量学习模型的学习可塑性定义为两者的差值：

$$LP_m = p_m^* - p_m^m \tag{11}$$

学习可塑性只针对当前任务计算遗忘率，主要衡量增量学习模型性能和联合训练性能的差距。

除了以上提及到的衡量指标外，还有一些指标可以衡量增量—学习的性能，如最小类增量准确率（Konaté 等，2023）、海森矩阵最大特征值（Kong 等，2023）和准确率曲线下面积（Koh 等，2022）等。

1.4 增量学习挑战

增量学习的核心挑战是克服灾难性遗忘问题。该问题通常出现在基于反向传播的深度网络中，这是因为当模型学习新任务时，网络会以遗忘旧样本特征为代价更新参数以适应新的数据。模型可能会在新任务上表现很好，但是对于旧任务会出现不可逆转的性能下降。

图 2 展示了灾难性遗忘的过程。任务 1 由类别 1 和类别 2 构成。任务 2 由类别 3 和 4 构成。模型在学习任务 1 之后，生成了对应的决策边界，如图 2（a）所示，决策边界准确地区分了任务 1 中的样本。模型在此基础上直接学习任务 2 后，会生成新的决策边界，如图 2（b）所示。新决策边界可以准确地区分新任务的两个类别，但是新样本的引入导致旧决策边界发生偏移（图中橙色部分为偏移量）而发生灾难性遗忘，进而导致模型对旧样本的预测发生误差。图 2（c）展示了理想状态下的增量学习结果，即新任务的学习并不会影响模型在旧任务上的性能。这种结果可以通过联合训练的方法实现，因此联合训练也是增量学习的理论最优解。

图 3 从模型参数优化角度展示灾难性遗忘的过程。灰色区域表示初始化模型在学习第 1 个任务之后返回的以θ_A为中心点的低误差权重区间。淡黄色区域表示初始化模型直接学习第 2 个任务返回的以θ_B为中心点的低误差参数区间。当模型在第 1 个任务上以无干涉的增量学习方式直接学习第 2 个任务，则参数会以逼近θ_B的方向偏移出第 1 个任务的低误差区间，导致模型在第 1 个任务上出现不可逆

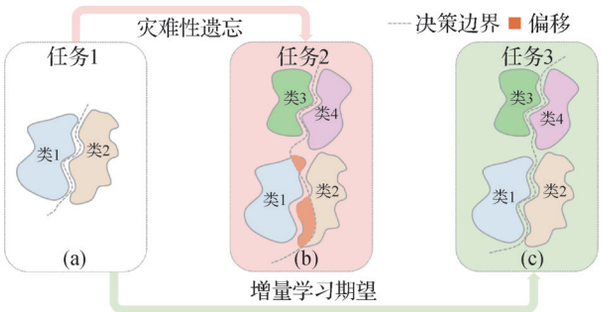


图 2　灾难性遗忘示意图（一）

Fig. 2　Illustration of catastrophic forgetting (1)

转的性能下降。增量学习的期望是使模型的误差方向朝着两个任务的低误差权重区间的交集方向移动（橙色箭头方向），使模型对新旧任务的处理结果都能保持在较低的误差空间内。在不访问任何旧数据的情况下，大多数增量学习模型都会产生灾难性遗忘。

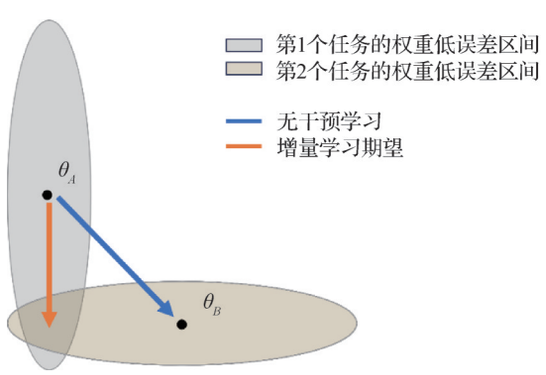


图 3　灾难性遗忘示意图（二）

Fig. 3　Illustration of catastrophic forgetting (2)

如何最大程度地减少灾难性遗忘是增量学习领域的主要研究目的。但增量学习不能仅以减少灾难性遗忘为导向，模型对新任务的学习性能、模型训练时间、模型空间复杂度以及训练所需存储大小也是衡量增量学习模型的重要指标。增量学习为解决灾难性遗忘问题面临的挑战可以具体化为可塑性—稳定性困境。模型对新任务的学习能力称为模型可塑性，可塑性要求模型具有从新任务中整合知识的能力，学习新任务的能力越强，模型的可塑性就越高；模型在学习新任务后对旧任务的性能维持能力称为模型稳定性，稳定性要求模型在学习新学习的知识时，减少对旧知识的显著干扰，新任务的学习对旧任务的负影响越大，模型的稳定性就越低。在同一个模型中，稳定性的提升意味着可塑性的下降，而可塑

性的上升会导致模型在旧任务上的不稳定。这两个互相冲突的需求构成了稳定性—可塑性困境。同时，模型内部的网络还存在泛化性(generalizability)，在大多数情况下，更高的泛化性可以同时提高模型的稳定性和可塑性，如图4所示，模型可以在牺牲模型结构复杂性和模型参数量的条件下提高模型泛化性，达到稳定性和可塑性兼顾。如何在模型的稳定性、可塑性和泛化性之间衡量，找到模型的最佳状态是增量学习的重要挑战。

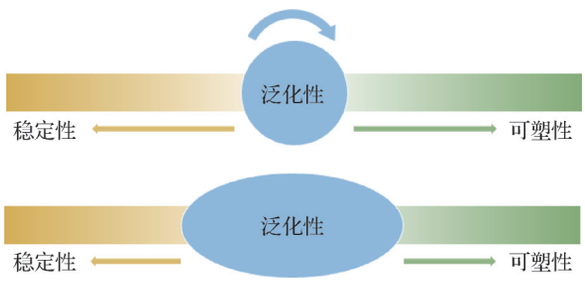


图4　稳定性—可塑性困境示意图

Fig. 4　Illustration of stability-plasticity dilemma

2　增量学习范式

增量学习在动态场景中的应用逐渐增加，越来越多的研究对增量学习的不同范式也进行了深入的探索和细致的分析。本节将增量学习根据不同的学习范式分成任务增量学习(task-incremental learning, TIL)、域增量学习(domain-incremental learning, DIL)和类增量学习(class-incremental learning, CIL)，其学习困难程度依次递增。

2.1　任务增量学习

任务增量学习(Hsu等，2018；Van de Ven和Tolias，2019)最重要的特点是模型在测试时可以获得训练样本的相关任务标识，即模型总是知道在执行哪一个任务。在1.2节给出定义基础上，还要求对于不同的任务i和任务j，其任务的标签空间无交叉，即$\boldsymbol{C}^i \cap \boldsymbol{C}^j = \varnothing$。任务增量学习在测试阶段，对于任一待测试样本，必须指定具体任务t对应的真实样本标签集合$\boldsymbol{C}^t$，模型的目标是在标签集合$\boldsymbol{C}^t$中推断出样本的标签。任务增量学习场景中，模型的输出应该是多头的，并且任务头的数量为总任务个数k。如图5所示，模型在任务增量学习中甚至可以为每个任务构造一个完全独立的网络(Jiang和Celiktutan，2023)，在这种情况下，完全不存在灾难性遗忘的问题。但是任务增量学习不应该只以减少灾难性遗忘为导向，还要注重模型复杂度和模型训练时间。因此，任务增量学习的核心问题是如何使增量学习模型发掘并维持跨任务共享参数。此外，任务增量学习还需要“专家”对每个测试样本人工添加任务标识，所以在实际应用中并不常用。对比来说，域增量学习和类增量学习则不需要对每个样本附加任务标识。

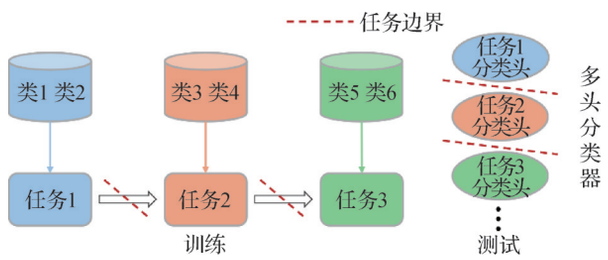


图5　任务增量学习示意图

Fig. 5　Illustration of task-incremental learning

2.2　域增量学习

域增量学习(Hsu等，2018；Van de Ven和Tolias，2019)中，不同时刻到达的数据属于同一个任务的不同域别。在本文1.2中定义的基础上，还要求对不同的任务i和任务j，其分类标签空间重合，即$\boldsymbol{C}^i = \boldsymbol{C}^j$。如图6所示，域增量学习的分类器是单头分类器，因此在测试阶段，不要求明确待推断样本的真实标签集合。域增量学习的训练样本在分布上有着明显区别，但在训练和测试阶段，样本都不需要附带任务标识。尽管不需要样本的任务标识，在结果中仍然可以分析出样本具体属于哪一个模态空间。域增量学习相对于任务增量学习，不需要“专家”对每个测试样本添加任务标识，其训练难度相较于任务增量学习更加困难。

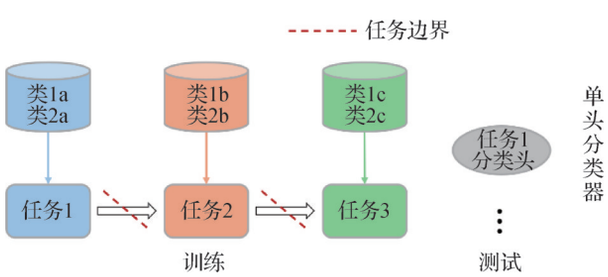


图6　域增量学习示意图

Fig. 6　Illustration of domain-incremental learning

2.3 类增量学习

类增量学习(Hsu 等,2018;Van de Ven 和 Tolias,2019,Zhao 等,2021a)是增量学习场景中探索最为广泛的一个范式。在类增量学习场景中,深度模型必须能够在不需要任务标识的情况下分辨到目前为止学习过的所有任务。在推断时,深度模型使用单头分类器,如图 7 所示,类增量学习与任务增量学习相同,同样要求任务间类别没有重叠,引用本文 1.2 节的定义参数,对于不同的任务 i 和任务 j,有 $\boldsymbol{C}^i \cap \boldsymbol{C}^j = \varnothing$。类增量学习在测试阶段需要在无任务标识的情况下在所有分类 $\hat{\boldsymbol{C}}^t$ 中判断正确分类。因此前后数据之间的互相干扰更大,学习难度更高。相对于任务增量学习,类增量学习不具有任务标识;相对于域增量学习,类增量学习的类别更多,因此新旧任务数据之间的互相干扰更大,学习难度更高。

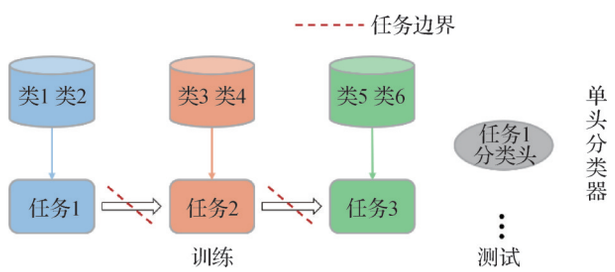


图 7 类增量学习示意图

Fig. 7 Illustration of class-incremental learning

除了上述 3 种增量学习范式之外,还有无任务标识增量学习(task-free incremental learning)(Aljundi 等,2019a)、在线增量学习(online incremental learning)(Aljundi 等,2019a)以及模糊边界增量学习(blurred boundary incremental learning)(Bang 等,2021)等,在此不做介绍。

3 增量学习方法

本节通过 4 个角度来维护增量学习深度模型的可塑性和稳定性,分别为模型优化、数据知识迁移、模型结构知识迁移和预训练模型知识迁移。对应的主流方法如图 8 所示:1)基于正则化的方法,这类方法从模型优化的角度,增加正则项;2)基于重放的方法,这类方法从数据集的角度,存储旧样本的相关信息;3)基于结构化的方法,这类方法从模型结构的角度,维护特定模型参数;4)基于预训练模型微调的方法,这类方法从预训练模型先验知识的角度,引入模型微调模块。图 9 展示了增量学习相关方法。

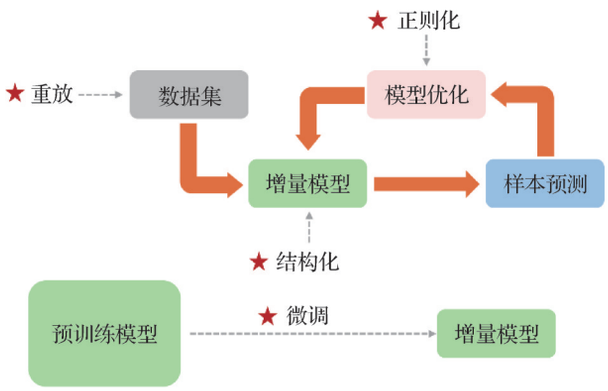


图 8 增量学习方法分类示意图

Fig. 8 Illustration of classification of incremental learning methods

3.1 基于正则化的方法

基于正则化的方法通过在训练过程中引入正则项以实现增量学习模型对新旧知识之间的平衡。根据正则项添加的位置具体可分为模型参数正则化和网络输出正则化。模型参数正则化在模型参数层面添加正则项,而网络输出正则化则在网络输出部分增加正则项。

3.1.1 模型参数正则化

模型参数正则化的方法在模型参数层面增加正则项,首先衡量深度模型中可学习参数对旧任务的重要性,并根据重要性控制可学习参数的更新,从而最大限度地减少在学习新任务时对重要性较高的旧模型参数的影响。模型参数正则化的方法通常包括两步操作:1)在学习新任务后估计模型中每个参数(假设参数是独立的)对任务的重要性;2)在训练下一个任务时,将上一个任务中每个参数的重要性作为权重系数来界定对参数本身的更新幅度。模型参数正则化的损失函数中,除交叉熵分类损失外,还引入一个额外的损失,该损失结合参数权重衡量模型每一个可学习参数的偏移量。针对参数权重,Kirkpatrick 等人(2017)提出的 EWC(elastic weight consolidation)方法采用费歇耳(Fisher)信息矩阵的对角线近似参数权重,在增量学习过程中有效地平衡新旧知识,保证对旧知识尤为重要的参数不被新任务过度修改。然而该方法只在学习完每个任务后衡量模型参数的重要性,忽略了在参数空间中不同的学习轨迹对模型参数重要性的影响。Liu 等人(2018)

改进了EWC方法，将模型的参数空间旋转到一个特定的空间，模型在该空间中能够提供更好的费歇耳信息矩阵，使得EWC在处理增量学习任务时更加鲁棒。然而该方法在训练过程中需要保持模型的参数数量固定，进而在较小的模型上产生更多的计算成本。Lee等人（2020）同样致力于对EWC中费歇耳信息矩阵的近似方法进行改进，他们扩展了克罗内克（Kronecker）因式分解技术，用于对费歇耳信息矩阵进行块对角近似，更准确地捕捉参数之间的依赖性，从而在增量学习场景中更有效地保护对旧模型比较重要的参数。相对于关注参数权重的计算，Zenke等人（2017）提出通过计算每个模型参数在训练过程中的“路径积分”衡量参数的重要性。同时，Zenke等人（2017）还指出批量更新模型参数可能会导致对参数重要性的高估，而从预训练模型开始更新则可能会导致对某些参数的重要性低估。为了解决这个问题，Aljundi 等人（2018）提出 MAS（memory aware synapses）方法，该方法通过积累所学函数的敏感度（梯度大小）在线计算参数权重。Chaudhry 等人（2018）提出RWalk方法，该方法将费歇耳信息矩阵近似和沿学习轨迹积分两个方法融合在一个算法中，并使用代表性旧样本进一步改进结果。Batra和Clark（2024）通过结合EWC和变分持续学习两个正则项共同抑制灾难性遗忘。其中，EWC方法通过费歇耳信息矩阵代替后验分布，而变分持续学习则基于先验分布，通过变分推断近似后验分布。此外，先前的方法虽然设计了复杂的正则项，但往往会陷入到新旧知识的平衡问题中。Wen等人（2025）提出应

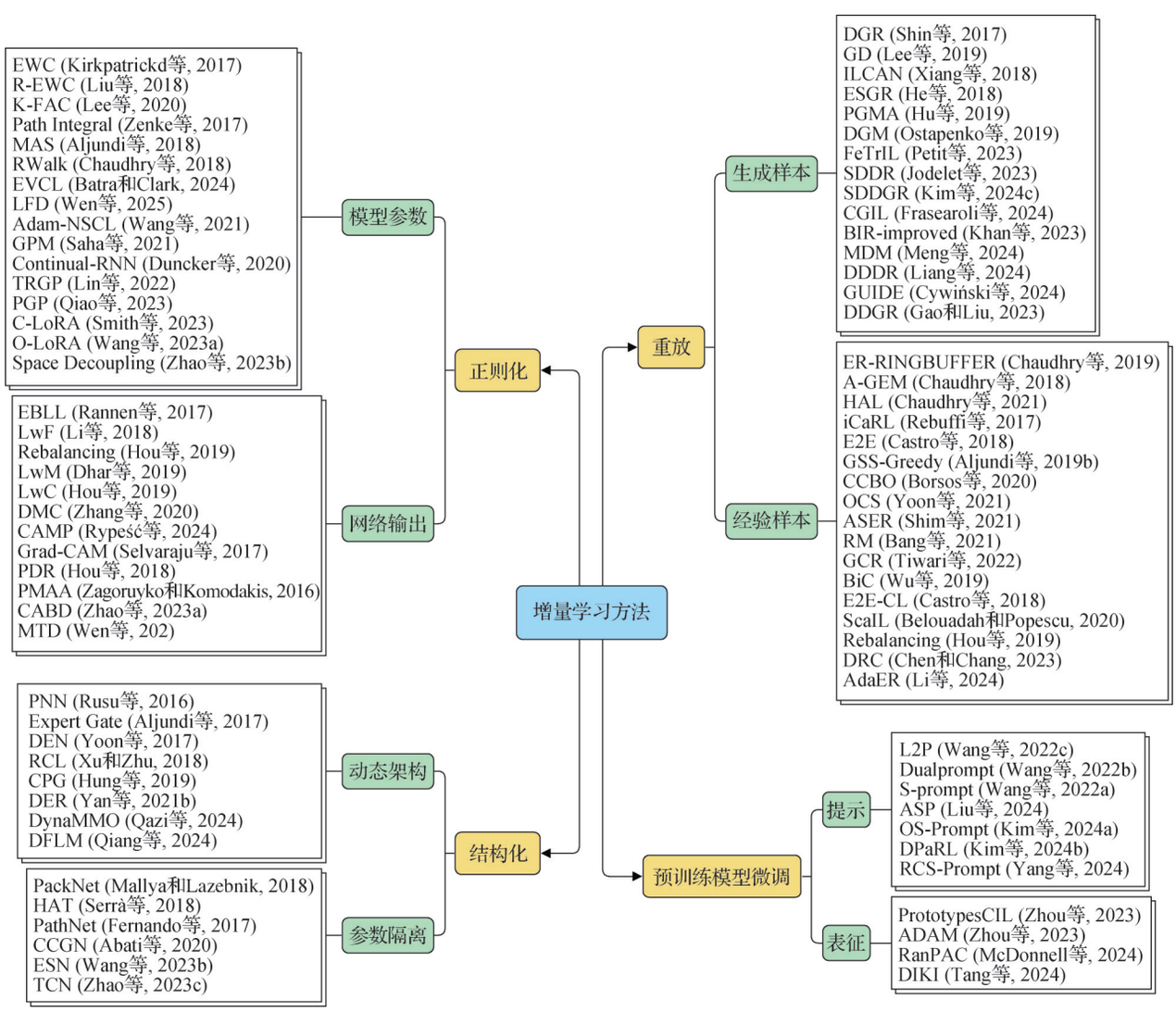


图9　增量学习相关方法总览

Fig. 9　Overview of incremental learning methods

该将对新参数约束方向接近旧任务的遗忘较少的方向，而不是单组旧参数，并提出一个后处理过程来找到这一间隔中的平衡点，更好地平衡新任务与旧任务之间的关系。近些年，研究者们还从模型参数梯度优化的方向控制优化过程减少灾难性遗忘。Saha等人（2021）和Wang等人（2021）在学习每一个任务后建模特征子空间，在学习下一个任务时利用特征子空间对新任务学习的梯度进行修正以实现对旧知识的影响最小化。Duncker等人（2020）针对循环神经网络（recurrent neural network，RNN）也提出类似的设计。随后，Lin等人（2022）发现从梯度角度设计的正则项对新任务的学习约束过强，因此从前向促进知识迁移的角度提出置信域引导的梯度正则化设计。Qiao等人（2023）最先提出将梯度正则化引入到基于预训练模型的增量学习方法中，通过对微调参数和输入特征进行特征分解，高效地对学习新任务的梯度进行调整。Smith等人（2023）和Wang等人（2023a）将梯度正则化方法引入到针对不同任务学习的低秩适配器中。虽然基于梯度的正则化设计能够有效减少模型学习新任务时对于旧知识的灾难性遗忘，但对于新任务梯度的强约束往往会使得新任务的学习能力严重受损。对此，Zhao等人（2023b）提出将特征空间解耦为松弛空间和平稳空间，并在相应的子空间对梯度施加不同程度的正则化约束，从而实现新旧知识间合理的平衡。

3.1.2　网络输出正则化

网络输出正则化方法在深度模型的中间层或最后一层的输出添加正则项来平衡模型对新旧任务的关注度，通常与知识蒸馏（Hinton等，2015）相关。知识蒸馏最初是为了从较大的教师模型中学习出一个更紧凑的学生模型。在此基础上，Li和Hoiem（2018）提出的LwF（learning without forgetting）方法首次将知识蒸馏运用在增量学习中。具体地，LwF在学习新任务时额外增加一个与蒸馏相关的损失以减弱旧模型参数的偏移量，并引入温度系数解决真实标签的概率过高的问题。LwF最初是一种针对任务增量学习的方法，后续也逐渐成为类增量学习的关键组成部分。Rannen等人（2017）通过优化自动编码器（autoencoder）扩展LwF，从而将特征投影到维度更少的流形。模型会对每一个增量任务都会学习一个自动编码器，虽然会使得计算和存储成本呈线性增长，但自动编码器与总模型相比很小，因此可以同时兼顾资源的消耗和模型的性能。Hou等人（2018）在LwF的基础上提出保留一小部分信息量比较大的旧任务数据，并利用这部分数据作为蒸馏的训练样本以减少模型对旧知识的灾难性遗忘，同时额外使用新任务的数据训练出一个专家网络，并利用该网络的输出作为软标签进行蒸馏以巩固模型对新知识的学习。在面对新旧数据之间的不平衡导致模型的训练偏差问题时，Hou等人（2019）在LwF方法的基础上提出LwC（learning without forgetting-multi-class）方法。该方法首先对新旧模型的输出结果进行L2归一化，以消除不同任务之间输出结果尺度差异的影响。接着，计算余弦相似度（cosine similarity）以约束新旧模型的输出。该正则化方式通过比较标准化后的向量降低模型对新旧数据的不平衡的敏感性。Zagoruyko和Komodakis（2016）提出利用教师模型的注意力引导学生模型的思想。在此基础上，Dhar等人（2019）提出LwM（learning without memorizing）方法，在训练新任务时通过注意力蒸馏损失保持旧任务上训练的模型对输入数据的注意力区域不变。有研究者注意到新任务具有明确且强的监督，而旧任务的监督较弱并且只能通过知识蒸馏的方式进行约束的问题。为了消除这种不对称性，Zhang等人（2020）提出DMC（deep model consolidation）方法，专注于训练时旧任务和新任务之间的不对称性。DMC方法引入辅助分类样本，并且对旧模型和新模型应用双重蒸馏损失。Zhao等人（2021b）提出的CIVC（class-incremental video classification）方法在蒸馏之前将重要信息进行分解，而不是作为一个整体，有效地继承了旧模型中的关键知识。Rypeść等人（2024）在LwF的基础上，提出将可学习的投影器与特征蒸馏相结合，从而增强模型的适应性。此外，从单一旧模型中蒸馏得到的知识并不足够将大部分重要知识过滤到新模型中，而对旧模型进行多角度的增强可以蒸馏出更多重要知识。对此，Zhao等人（2023a）提出同时对两个教师模型进行蒸馏。一个是在所有任务数据上训练的模型，具有丰富的通用知识，可以用于减轻模型对当前新任务数据的过拟合；另一个是在上一个旧任务上训练的旧模型，包含了对旧任务的知识，用于减轻其遗忘。通过从两个互补的旧模型中同时蒸馏，可以平衡模型对新旧任务的性能。此外，Wen等人（2025）提出采用权重置换、特征扰动和多样性正则化技术对旧

模型进行增广，确保其多样性，再通过对多个增广的旧模型进行蒸馏来保证新模型可以学习到更为全面的旧任务知识。

3.2 基于重放的方法

基于样本重放的增量学习方法通过存储有关旧任务的样本信息，在训练新任务时同时复用来平衡模型对新旧知识的性能。根据旧信息存储形式的不同，基于样本重放的增量学习方法具体可以细分为基于生成样本的重放和基于经验样本的重放。

3.2.1 生成样本重放

基于生成样本重放的增量学习方法存储旧任务模型的关键信息如分布特征或模型参数，并利用这些信息通过生成模型逆向生成旧任务伪样本储存旧任务知识，这种方法也称为伪样本重放。随后，该类方法通过将新任务的训练样本与生成的旧任务伪样本合并进行联合训练减少模型对旧知识的灾难性遗忘。常见的生成模型包括生成对抗网络（generative adversarial network，GAN）（Gui 等，2021）、变分自编码器（variational autoencoder，VAE）（Elasri 等，2022）、流模型（flow）（Elasri 等，2022）和扩散模型（diffusion models，DMs）（Ho 等，2020）等。生成样本与真实样本相比，虽然在视觉表征上相似，但在高维特征空间中的差异通常较明显。生成模型的性能可以决定生成样本的质量，采用不合适的生成模型生成的伪样本训练增量学习模型会导致其在大规模数据集上无法达到满意的结果（Lavda 等，2018；Van de Ven 和 Tolias，2018）。因此，如何训练一个高质量的伪样本生成器就是基于生成样本重放的增量学习方法最重要的目标。Hu 和 Rostami（2023）使用变分自编码器生成旧任务伪样本，并引入了时间感知的正则化方法动态调整监督学习、潜在正则化和数据重建这 3 个训练目标项的权重。Khan 等人（2023）提出在基于 VAE 的生成重放模型中，生成的特征在映射到潜在空间时与原始特征相差甚远，因此可以在当前模型和旧模型之间的潜在空间中引入蒸馏操作，以减少特征漂移。Gulrajani 等人（2017）首次将生成对抗网络应用到增量学习任务中，采用 WGAN-GP（Wasserstein generative adversarial network with gradient penalty）在简单的数据集（Lecun 等，1998）上实现了不错的性能。但该方法在大规模图像生成任务上面临的挑战依然严峻（Shmelkov 等，2018）。Lee 等人（2019）提出 GD（global distillation）方法，利用生成模型生成旧任务的辅助数据，这些数据连同新任务数据一起用于训练新模型。Xiang 等人（2019）提出一种基于条件对抗网络的生成样本重放方法，将带标签信息的嵌入向量作为生成对抗网络的条件，生成伪样本代替重放的旧样本。然而，使用生成对抗网络生成旧任务代表性样本可能存在生成器训练的不稳定性或生成数据质量低下的困扰，而扩散网络生成的旧任务伪样本的质量更高，与真实样本更接近。因此，研究者们提出可以使用扩散模型代替生成对抗网络。Meng 等人（2024）采用多分布匹配扩散模型提高生成样本的质量，并集成了选择性合成图像增强扩展训练数据的分布。Liang 等人（2024）提出一种基于扩散模型的数据重放方法。该方法使用预训练的条件扩散模型（conditional diffusion models，CDMs）逆向生成每个类别的伪样本，加强了生成样本的表征能力。该方法还通过对比学习（contrastive learning）进一步增强分类器在生成样本和真实样本之间的泛化能力，显著减少计算资源和时间消耗，同时确保旧任务伪样本的有效生成。Cywiński 等人（2024）使用一种新颖的引导机制调整和优化扩散模型在扩散过程中的概率分布，并采用动态调整的学习策略以优化学习效率。此外，当前基于生成重放的增量学习方法通常多次使用生成的样本更新生成器，导致生成的伪样本偏离先前任务样本的分布。对此，Gao 和 Liu（2023）提出通过分类器指导扩散模型的生成过程。基于生成样本重放的增量学习方法虽然可以减少模型对旧任务知识的灾难性遗忘，但在每次学习新任务时仍需要训练生成模型以适配不断增加的新数据，这个过程会耗费额外的存储和计算资源。

3.2.2 经验样本重放

基于经验样本重放的增量学习方法直接从旧任务的训练样本中选择具有代表性的样本，这些样本通常称为范例（exemplar）。该类方法在学习新任务时，会将训练样本与范例合并在一起进行联合训练以减少模型对旧任务知识的灾难性遗忘。在存储空间受限的背景下，如何合理选择具有代表性的旧样本是该类方法研究的关键挑战。早期研究通常采用较为简单的样本采样策略，例如从旧任务数据中随机采样。Rebuffi 等人（2017）首次提出通过保留少量具有代表性的旧样本缓解遗忘问题。在模型学习新任务时，该方法采用 Herding 排序法（Welling，

2009)对每个样本和对应类别中心的距离进行排序，选取离类别中心更近的样本作为范例。除了Herding排序法，还有K-Means(Chaudhry等，2019)、Plane Distance(Riemer等，2018)和Entropy(Riemer等，2018)等采样方法，但这些方法对模型效果的提升较为一般。更高级的采样方法则是基于梯度或可优化的准则，例如Aljundi等人(2019b)提出的GSS(gradient based sample selection)方法通过参数梯度最大化范例多样性，使得模型能够覆盖更多的旧任务特征。Borsos等人(2020)提出一种基数约束的多目标优化方法CCBO(cardinality-constrained bilevel optimization)，专注于挑选一组范例提升任务的性能表现。通过双层优化策略，该方法在资源受限的情况下有效平衡了样本存储和模型性能。Zhao等人(2021c)提出使用辅助低保真度样本代替原始范例，并通过双元学习方案减少低保真度样本和真实范例之间的领域偏移，缓解了保存范例所需内存过大的问题。Yoon等人(2021)提出的OCS(online coreset selection)方法在每个小批量数据中，根据梯度信息选择一部分范例，这些范例不仅与之前的旧任务保持高关联性，还包含对当前新任务的主要特征。Shim等人(2021)提出的ASER(adversarial shapley value experience replay)方法通过重新计算夏普利(Shapley)值，选择对模型当前决策最有影响的范例提高模型对旧任务的记忆能力。Bang等人(2021)提出的RM(rainbow memory)方法通过管理范例的多样性增强模型的泛化能力，确保范例能够尽可能广泛地覆盖旧任务特征。Tiwari等人(2022)提出的GCR(gradient coreset replay)方法则通过当前参数与旧任务样本之间的梯度相似性选择与当前模型最相关的旧样本作为范例。Li等人(2024)提出的AdaER(adaptive experience replay)方法通过上下文提示召回策略选择与当前输入数据最不相关的样本进行重放。除此之外，该方法还通过最大化记忆缓冲区的信息熵改进原始抽样策略。然而，尽管储存旧任务范例可以减少灾难性遗忘，但是在学习新任务时，新数据的数量远超旧任务的范例数量，导致模型在训练时更倾向于新数据(Hou等，2019；Rebuffi等，2017；Wu等，2019)。为了缓解这个问题，Wu等人(2019)提出在深度模型的全连接层后添加一个在验证集上优化的线性模型来纠正这种新类偏见。Belouadah和Popescu(2020)提出增大旧任务的分类器权重，使旧任务类别的影响力与新任务类别相匹配。Hou等人(2019)引入基于余弦相似度的分类器减少由数据分布不平衡带来的训练偏差。Castro等人(2018)提出的E2E(end-to-end)方法采用了更均衡的训练方式，在每一个增量任务训练完成后，引入一个额外的阶段，利用每个类别相同数量的样本进行有限次数的二次训练。Chen和Chang(2023)通过残差分类器和分支层合并技术，处理因新旧类样本数量不均导致的模型复杂性增加问题。此外，在新任务的训练过程中增加未标注的辅助训练样本也可以缓解新旧数据间的不平衡问题(Zhang等，2020；Lee等，2019)。

总结来看，基于生成样本重放的增量学习方法严重依赖于生成模型的性能，而基于经验样本重放的增量学习方法则更多受到范例选择策略的影响。然而，即使基于样本重放的增量学习方法相较于基于正则化的增量学习方法取得了更好的性能(Chaudhry等，2021；Buzzega等，2020)，但是其对于旧样本的存储一定程度上违背了增量学习的基本设定，进而受到较多的质疑(Shokri和Shmatikov，2015；Smith等，2021)。

3.3 基于结构化的方法

在基于正则化和重放的增量学习方法中，模型参数通常在所有增量任务之间共享，因此不同增量任务之间的区分较为模糊。相比之下，基于结构化的增量学习方法的目标是为每个增量任务构建专属的模型参数。根据网络结构是否扩展，基于结构化的增量学习方法分为动态架构和参数隔离。

3.3.1 动态架构

基于动态架构的增量学习方法通过压缩、扩展神经网络的方式更新模型以适应新任务并保留模型对于旧任务的性能。Rusu等人(2016)提出的渐进式网络为每个新任务重新构造一个新网络，同时通过横向连接各个网络保留旧样本的特征。Aljundi等人(2017)提出一种基于专家网络的方法。该方法通过一个专家门控模块找寻和新任务相关性最强的先验旧任务，在这个旧模型上结合新数据做进一步的微调。Yoon等人(2017)提出DEN(dynamic expansion network)方法，在对新任务进行训练时，动态决定该任务所需要的网络容量，通过学习任务间的共享知识构造任务共享网络，同时在必要时适当地扩展网络结构。Xu和Zhu(2018)进一步提出一种增量

强化学习的方法，该方法使用演员—评论家的策略，通过基于验证准确性和网络复杂度的奖励机制，动态为每个任务选择神经网络的最优架构。这种方法能够有效地平衡新任务的性能和网络结构的复杂度。Hung等人(2019)提出CPG(compacting, picking, and growing)方法，对每个新任务，CPG首先冻结在旧任务学习到的模型参数，随后，根据新任务的性能需求，选择使用从旧任务中释放的权重空间或扩展当前网络架构。其次，该方法还通过压缩和剪枝去除深度模型在训练新任务时产生的冗余参数，从而提高模型的效率和泛化能力。Yan等人(2021b)将新任务的学习分为两个阶段：表征学习阶段和特征提取器学习阶段。表征学习阶段固定旧模型的表征，同时引入新的可学习特征提取器扩展深度模型网络。在特征提取器学习阶段，通过当前的新任务数据对新的特征提取器进行微调。Qazi等人(2024)提出DynaMMO(dynamic model merging)方法为每一个任务构造一个轻量级可学习模块，并在当前任务学习完毕后将所有可学习模块组合成一个统一模型，在最小化计算开销的同时维持了模型对新任务的可塑性。Qiang等人(2024)提出的DFLM(dynamic feature learning and matching)方法通过学习当前任务与先前任务的特征，并在任务之间共享表征能力，从而在不显著增加计算成本的情况下，有效保留模型对旧任务的性能。

3.3.2 参数隔离

与基于动态架构的增量学习方法不同，基于参数隔离的增量学习方法专注于为不同的任务分配不同的模型参数，以减少任务间的干扰，从而扩大任务之间的知识独立性。Mallya和Lazebnik(2018)提出的PackNet方法利用深度网络中参数的冗余性，通过基于权重的剪枝技术去除模型中的非重要参数，从而为新任务留出参数空间。Serrà等人(2018)提出的一种硬注意力机制，为每一个任务学习一个硬注意力掩码，该掩码利用先前任务的掩码作为条件，通过随机梯度下降进行优化，以减少灾难性遗忘。Fernando等人(2017)提出的PathNet方法在深度网络中嵌入路径模块，并通过遗传算法(genetic algorithm, GA)(Harvey, 2011)识别新旧任务的共享路径。在学习新任务时，固定旧任务路径上的参数并初始化新路径，同时利用遗传算法优化新任务的路径选择。Abati等人(2020)提出为每个卷积层添加任务特定的门控模块，并通过调节通道激活保护旧任务的关键参数，进而保留了模型对旧任务的性能。传统的参数隔离方法通过将不同的模型参数分配给不同的任务减缓灾难性遗忘，但跨任务知识很难被传递且容易被忽视。对此，Zhao和Mu(2023)提出的TCN(twin contrastive networks)方法将增量学习任务流视为一系列单类别分类任务，并训练单独的分类器识别每个类别。同时，为了促进知识传递并充分利用积累的旧任务知识，TCN方法还采用了双网络结构学习不同的特征表示来帮助新任务的学习，并通过固定所有已学习参数减缓灾难性遗忘。Wang等人(2023b)则提出使用参数隔离策略为每个阶段训练独立的分类器，避免不同阶段间的相互干扰，同时通过冻结预训练的骨干网络，仅更新分类器参数来增加模型的稳定性和可塑性。

总结来看，基于正则化的增量学习与基于结构化的增量学习方法相比，后者更倾向于寻找适应新旧增量任务的最佳模型参数。虽然基于结构化的方法几乎可以完全避免灾难性遗忘，但这会影响增量学习模型的可扩展性以及增量学习任务之间的泛化能力。此外，由于这类方法在推理时通常需要任务标识确定应使用的模型参数组，因此很大程度上限制了其在实际应用中的表现和性能。

3.4 基于预训练模型微调的方法

基于预训练模型微调的增量学习方法是指在已有的预训练模型的基础上不断学习新任务，同时尽可能地保持模型对旧任务性能的方法。这类方法在许多增量学习应用中十分有效，因为预训练模型已经通过大量数据训练，具备良好的泛化能力和表征能力。根据对预训练模型微调的方式不同，可以分为基于提示的预训练模型微调增量学习方法和基于表征的预训练模型微调增量学习方法。

3.4.1 基于提示的预训练模型微调

基于提示的预训练模型微调的增量学习方法通过引入提示(prompt)对预训练模型(pre-trained model, PTM)进行微调以实现深度模型对新旧知识的平衡。预训练模型如ViT(vision in Transformer)(Dosovitskiy等，2020)、DeiT(Touvron等，2021)和MoCo v3(Chen等，2021)通常在大规模数据集上训练，具有出色的表征能力。但是在处理下游任务时，预训练模型的训练数据集与下游任务的数据集往往存在领域差异。一些高效的微调技术例如VPT

(vision prompt tuning)(Jia等,2022)、LoRA(low-rank adaptation)(Hu 等 ,2021)和 Adapter(Houlsby 等 ,2019;Zhao等,2021d)等被提出以减少此差异。在增量学习场景中,为了将预训练模型中的先验知识迁移到下游任务,并在增量学习新任务时减少对于旧任务的遗忘,Wang等人(2022c)提出的L2P(learning to prompt)方法首次将提示引入增量学习模型中。该方法通过基于查询—提示池的自注意力机制,利用学习得到的查询向量在提示池(prompt pool)中动态选择与待预测样本最相关的提示进行后续推理。随后,Wang等人(2022b)在L2P的基础上提出DualPrompt方法。该方法采用了相同的提示选择策略,但受互补学习系统(McClelland等,1995;Kumaran等,2016)的启发,延伸出了在增量学习中解耦提示的方法。具体而言,DualPrompt学习了两种提示模块。通用提示模块在每个新任务上进行训练以提取共享特征;任务提示模块则针对当前任务进行优化,从而更好地平衡稳定性和可塑性。Wang等人(2022a)针对域增量学习领域提出S-prompt方法,同样将提示解耦为通用提示和任务提示,不同的是,在测试阶段采用K-means算法进行提示的挑选。在增量学习过程中,固定提示池的大小会导致模型对后续任务的性能大幅度降低,而不断扩展提示池的大小会带来内存增加的问题。对此,Feng等人(2024a)提出的PECTP(parameter-efficient cross-task prompts)方法将固定数量的提示在增量学习过程中持续更新以适配新知识,同时采用正则化的方法对提示的参数施加约束,以减少对旧任务的灾难性遗忘。Feng等人(2024b)提出的LW2G(learn whether to grow)方法从梯度优化的角度解释增量学习过程中新旧任务间的差异,并动态地决定是否需要为新任务增加一组提示。Liu等人(2024a)提出基于注意力的自适应提示机制ASP(attention-aware self-adaptive prompt),该机制可以使模型根据样本标签的注意力分布动态调整提示,以在学习新任务时实现新旧知识的平衡。现有的基于提示的预训练模型微调的方法都需要维护两个前馈阶段,一个是深度模型骨干框架,另一个是用于从提示池内选择提示的专家框架,因此这类模型面临着双倍的计算负担。对于这个问题,Kim等人(2024b)设计了基于提示的单阶段预训练模型微调框架,通过直接使用中间层的标记嵌入作为提示查询向量。这种设计消除了查询提示的前馈阶段,从而降低了训练和推断的计算成本。同时,该方法进一步引入了一个查询池正则化损失调节提示查询模块和提示池之间的关系。此外,当前基于提示的预训练模型微调方法通常为固定的提示机制,在适配开放世界场景的能力方面有所欠缺。对此,Kim等人(2024a)提出的DPaRL(dynamic prompt and representation learner)方法不再依赖传统的静态提示池,而在每个训练阶段学习一个动态提示生成模型,使模型能够在不同任务中灵活调整提示内容,提升模型对新任务的适应能力。传统基于提示的预训练模型微调策略还缺乏一个明确的优化目标建模不同学习阶段之间的类别关系,导致训练样本的类别间隔不清晰。一些旧样本在推断过程中使用新提示也会导致提示—歧义重叠空间中出现新旧类空间重叠问题。对此,Yang等人(2024)提出的RCS-Prompt(rearrange class space)方法通过双向优化提示重新排列类别空间,使其能够区分不同阶段的类别区域。该方法还通过将部分新样本的标签改为旧类别,并利用自定义对称损失进行训练以缓解提示—歧义重叠空间的重叠问题。尽管基于提示的预训练模型微调增量学习方法在平衡预训练知识和下游任务之间取得了一定效果,但随着增量步骤的增加,提示自身可能会面临遗忘问题。此外,存储容量的限制也会约束提示池的大小,从而影响增量学习的整体性能。

3.4.2 基于表征的预训练模型微调

基于表征的预训练模型微调方法利用预训练模型的高表征能力进行新任务的分类或回归,并且不需要在深度模型的架构上做较大改动。这类通过简单的分类器(如线性分类器)的训练实现对新任务的适应更适合在资源受限的环境中使用。Zhou等人(2023)提出的PrototypesCIL通过固定预训练模型的嵌入函数,在增量学习过程中提取每个类别的平均嵌入并将其作为分类条件,在后续任务中通过直接推理实现了良好的性能表现,显著减少了类别增量学习中的训练开销。基于PrototypesCIL,Zhou等人(2023)提出的APER(adapt and merge)方法首先对预训练模型进行适应性训练以缩小其与新任务之间的差距。随后,APER方法将适应后的模型与预训练模型的嵌入函数进行合并,并冻结这个合并后的嵌入函数以确保模型针对旧任务类别的区分能力。McDonnell等人(2024)提出的RanPAC(random pro-

jections)方法通过引入随机投影层减少灾难性遗忘。RanPAC方法在预训练模型的特征表示和输出层之间加入了一个冻结的随机投影层,并使用非线性激活函数捕捉特征之间的交互,同时对特征空间维度进行扩展以增强深度模型的表征能力。Tang等人(2024)提出的DIKI(distribution-aware interference-free knowledge integration)方法构造了一个完全残差机制,将新学习的知识注入冻结的预训练骨干网络中,保持对预训练知识的负面影响最小化。此外,这种残差特性可以明确地控制来自未知分布的测试数据的信息植入过程。

总结来看,基于提示的预训练模型微调可能会受到存储大小的限制,并且提示本身会存在灾难性遗忘问题。而基于表征的预训练模型微调则可能在遇到复杂的新任务时性能受限,需要进行额外的特征工程或结合其他方法以增强表现力。

4 增量学习应用场景

增量学习的快速发展推动了其与各个领域的结合。本节列举了增量学习在单模态和多模态领域中的经典应用。在单模态领域,增量学习常应用于3个下游应用场景:增量语义分割(continual semantic segmentation)场景、增量图像生成(continual image generation)场景和增量大语言模型(continual large language models)场景;在多模态领域,总结了视觉—语言(visual-language)增量学习场景以及音频—视觉(audio-visual)增量学习场景。

4.1 增量语义分割

语义分割是自动驾驶、智能机器人等应用中经常用到的一项技术(Biasetton等,2019;Michieli和Badia,2018)。语义分割增量学习模型可以在非联合训练的情况下持续的接收新样本并归纳已出现的类别。语义分割增量学习场景的主要挑战是样本中并不仅限于一个类别,而是会存在多个已出现的旧类别或未出现的新类别,多个类别之间相互干扰,导致语义分割增量学习相对于传统增量学习会出现更严重的灾难性遗忘。Michieli和Zanuttigh(2019)首先关注到增量学习在语义分割中的灾难性遗忘问题,提出可以通过旧模型对上个任务的输出和中间层特征进行知识蒸馏以缓解灾难性遗忘问题。Douillard等人(2021)提出一种针对特征层面的多尺度池化蒸馏方案Local-POD(local pooling distillation)。Yan等人(2021a)提出在期望最大化(expectation-maximization)框架上开发增量学习策略以平衡语义分割增量学习模型的可塑性和稳定性,并使用自适应的采样方法保留旧样本中信息量较大的训练数据。Huang等人(2021)提出SegInversion方法,该方法使用图像级标签从噪音开始生成图像,并将生成的图像与新类别的真实图像混合当做输入样本到基于蒸馏的模型中训练新的语义分割增量学习模型。Cermelli等人(2022)提出一种端到端的弱增量学习框架,利用图像级标签生成伪样本,使用软标签处理伪样本生成过程中的噪音,并通过引入焦点损失项保持深度模型对不同类别的区分能力,从而减少新类别覆盖旧类别重要特征的风险。研究者观察到在学习新的语义分割任务时,旧类别像素的上下文更倾向于新类别而不是旧类别,从而增加旧类别的灾难性遗忘以及新类别的过拟合。对此,Zhao等人(2022)提出带有校正偏置上下文的增量语义学习框架,并提出偏置上下文无关一致性损失缓解灾难性遗忘问题。Shang等人(2023)提出Incrementer方法,该方法向Transformer的解码器中添加新类别的标记以学习新任务,并提出一种新的知识蒸馏方案,专注于在旧类区域进行蒸馏,从而减少旧模型对新类学习的约束,提高了模型可塑性。Cong等人(2024)提出的Cs2K(class-specific and class-shared knowledge)方法通过类特定知识引导和类共享知识引导两个角度减少灾难性遗忘。其中,类特定知识引导利用旧类别的特征中心引导伪标签的生成,同时维持旧类别的特征中心调整新任务数据的类别分布,减少深度模型对旧任务的遗忘。类共享知识根据旧类别的权重将旧模型与新模型进行整合,强化模型对旧类别的记忆。然而,传统语义分割增量学习模型在学习新任务时通常直接调整新分类器的参数,这可能导致模型对新类别的预测偏差。对此,Xie等人(2024)提出NeST(new classifier pre-tuning)方法,该方法通过在预训练阶段学习一个转换函数,将旧分类器的参数映射到新分类器中,而不是直接调整新分类器的参数。这种方式可以更好地保持深度模型对旧类别的记忆,同时提高对新类别的适应能力。背景转移是语义分割增量学习场景面临的另一个主要问题,即不同的样本中的背景类别不断变化,导致语义分割深度模型在预测背景时不稳定。目前常规

的方法通常采用共享背景分类器判断背景，存在背景预测的稳定性较低，语义分割准确性也随之降低的问题。对此，Park 等人（2024）提出背景—类别分离框架。一方面，采用选择性伪标记和自适应特征蒸馏提取旧任务重要的旧知识；另一方面，通过引入正交性约束，促进背景类与新类之间的分离，并利用标签引导的输出蒸馏进一步提升模型性能。Zhang 和 Gao（2024）则引入背景残差的概念，学习当前样本背景与旧样本背景之间的差异，避免了使用共享背景分类器。同时，提出伪背景二元交叉熵损失和背景适应损失优化模型的背景适应机制，并通过知识蒸馏和背景特征蒸馏两组蒸馏策略减少灾难性遗忘。

4.2 增量图像生成

图像生成是指使用基于数学模型、统计模型或神经网络的计算机算法生成图像的应用。在图像生成领域，比较广泛使用的生成模型是生成对抗网络、变分自编码器和扩散网络等模型。生成模型在学习连续的任务流时也存在灾难性遗忘问题。Wu 等人（2018）在生成对抗模型中融入样本重放，并提出联合训练重放和对齐重放以缓解灾难性遗忘。Zhai 等人（2019）提出的 LifelongGAN 方法通过知识蒸馏生成辅助数据以对齐新旧模型的输出，进而将旧模型知识迁移到新模型中。Zhai 等人（2020）提出 PiggybackGAN 方法，该方法通过旧模型中的卷积层和反卷积层过滤器学习新任务，同时引入一部分未受约束的新过滤器来适应新任务，有效地保持了模型对旧任务的性能，并且减少了学习新任务所需的额外模型参数。Zhai 等人（2021）继续在 LifelongGAN 的基础上提出更新颖、泛化性更高的增量图像生成框架 Hyper-LifelongGAN，该方法将传统的卷积层分解为动态基础过滤器和确定权重矩阵，可以高效地学习每个任务中的细节，并保留旧任务中的重要信息，这个新颖的架构能够以较低的参数成本提高生成模型的性能。Ayromlou 等人（2022）提出的 ClassImpression 方法则使用预训练图像生成模型生成特定类别的图像，同时引入了余弦标准化交叉熵损失、边缘损失以及领域内对比损失减少深度模型对旧任务的灾难性遗忘。Jodelet 等人（2023）提出的 SDDR（stable diffusion for distillation and replay）方法利用文本—图像与训练生成模型模拟后续可能出现的类别样本，进而促进增量学习对新任务的适配能力。此外，SDDR 方法还通过生成旧任务伪样本实现无需范例的增量学习，减少了对储存资源的需求，减轻了深度模型对旧数据的依赖。Seo 等人（2023）针对增量图像生成模型在少样本场景下的低性能问题提出 LFS-GAN（lifelong few-shot GAN）方法，通过可学习分解张量为每个任务学习特定参数，并在不同任务间共享深度模型权重。同时通过引入模式查找损失提高模型在数据量较少的情况下的生成多样性，缓解了少样本场景下增量图像生成模型的灾难性遗忘问题。

4.3 增量大语言模型

大语言模型可以用于解决诸如机器问答、机器翻译和对话系统等曾经具有挑战性的任务，其进展表明了人工智能具有的巨大潜力。但是预训练的大语言模型无法同时满足每个用户的需求，需要进行微调以适配不同的下游应用场景。Winata 等人（2023）通过动态调整大语言模型在增量学习过程中的学习率，结合大语言模型天生的抗遗忘特性减少增量学习过程中的灾难性遗忘。Bai 等人（2022）通过全球原型（global prototypes）对抗负面表征漂移。该方法的核心在于建立和维护一个全球原型集作为不同任务的共享参考点。模型在训练过程中通过对比当前学习的新数据与全球原型之间的差异来保持对旧任务的记忆。Ke 等人（2023）提出基于软掩码机制的大语言模型增量学习方法。该方法通过计算模型中各个单元（如注意力头）对通用知识和领域特定知识的重要性，从而决定在反向传播过程中对哪些单元进行更新，从而实现对深度模型的直接控制。Zhang 等人（2022）提出的增量学习序列生成的方法可以自适应组合多个模块，其中不同模块专注于不同的任务或知识领域。该方法能够动态评估现有模块对新任务的贡献，根据贡献大小决定是重用现有模块还是添加新模块，从而实现对新任务的适应。此外，伪经验回放机制也进一步促进了新旧知识的共享。He 等人（2024）在提出的 KPIG（key-part information gain）方法中引入一个关键部分探测器，该探测器能够动态地从输入数据中计算关键信息增益，增益较高的信息在增量训练过程中会得到重点强调。Wang 等人（2023a）提出的 O-LoRA 方法通过正交低秩子空间投影技术将模型参数分解为共享基和任务特定基。在模型训练过程中，O-LoRA 保持共享基的稳定性，同时允许任务特定基根据新任务灵活

调整，从而确保不同任务间的知识能够独立更新，降低了任务间的相互干扰。在面对大语言模型适配不同领域时存在的高计算需求和适应能差的问题上，Malla等人（2024）提出COPAL方法，利用敏感性分析指导模型的剪枝过程，通过评估模型对新数据引入的扰动的承受能力识别并保留与旧任务数据相关性最高的模型参数，从而减少大语言模型增量学习场景下的灾难性遗忘。

4.4 视觉—语言增量学习

视觉—语言多模态模型通过文本和图像信息建模，并广泛应用于图像标注、视觉问答等任务。比较常见的视觉—语言模型有CLIP（contrastive language-image pre-training）（Radford等，2021）等。视觉—语言多模态模型具有数据差异大、模态对齐困难以及数据规模大的特性，因此训练过程面临着比单模态增量学习模型更多的困难。Srinivasan等人（2022）提出CLiMB（continual learning in multimodality benchmark）基准，并通过研究发现常见的增量学习方法例如iCaRL（incremental classifier and representation learning）等虽然可以减轻多模态增量学习任务中的灾难性遗忘现象，但未能实现多模态场景下跨任务知识的有效迁移。Yan等人（2022）提出一种新颖的视觉—语言预训练增量学习方法，该方法使用生成模型生成与新学习任务相关的负文本样本，用于增强模型在文本和图像任务中的鉴别能力。同时，采用了多模态知识蒸馏技术，确保在引入新知识的同时，减少对旧知识的遗忘。这种负文本重放的策略显著提高了模型在连续学习环境中的稳定性和可塑性。Zheng等人（2023）提出的ZSCL（zero-shot continual learning）方法通过在特征空间和参数空间的双重调节减少灾难性遗忘。具体而言，ZSCL通过一个具有语义多样性的无标签数据集在特征空间内进行知识蒸馏，并通过平均化权重防止参数空间的大幅度变化。ZSCL不依赖于样本重放机制，而是通过模型参数的微调保持旧任务的知识。Zhu等人（2023）提出的CTP（compatible momentum contrast with topology preservation）方法引入了兼容性动量对比和拓扑保持两个策略。兼容性动量通过对比新模型和旧模型的知识，灵活更新文本和视觉模态特征保持不同任务之间的知识连贯性；拓扑保持则在跨任务传递知识时，减少不同任务间的信息损失，确保不同任务之间的特征空间结构一致。Ni等人（2023）提出视觉—语言增量学习框架Mod-X，通过选择性地调整对比矩阵的非对角线信息分布提升模型在新数据领域的适应能力。Yu等人（2024a）提出的DDAS（distribution discriminative auto-selector）方法引入了专家混合适配器（mixture-of-experts，MoE），这些适配器作为模型架构中的插件组件，可以根据输入的特征和任务需求，选择性激活特定的专家网络。这种设计减少了不同任务间的干扰，提高了新知识的吸收能力和旧知识的保持能力。语言—视觉大模型在增量学习过程中面临的问题通常是将旧分类误判为新分类的幻觉，研究者发现在视觉—语言多模态模型中，使用文本特征调整新任务数据的表征能力可以有效减缓增量学习过程中的灾难性遗忘，因此，Huang等人（2024）设计了RAPF（representation adjustment and parameter fusion）方法，在增量学习过程中分别计算新类别和旧类别在文本特种空间的欧几里得距离，并通过一个特定的阈值控制新类别是否属于“易混淆”类别。对于“易混淆”类别，通过文本特征衡量新任务数据对旧任务数据的影响并调整对应参数权重，同时采用分解参数融合的正则项进一步减轻预训练微调过程中的遗忘现象。Xu等人（2024）提出的X-TAIL（cross-domain task-agnostic incremental learning）方法通过岭回归适配器和无需训练的模态融合模块减少CLIP模型在增量学习过程中对初始模型的先验知识以及增量学习的新知识的遗忘，并由初始模型和增量学习之后的模型共同预测样本标签。Yu等人（2024b）针对在CLIP模型上进行增量学习会大幅减少初始模型的零样本学习（zero-shot）能力的问题提出双教师蒸馏网络，该网络在增量学习过程中，不仅对上一个增量学习过程的模型进行蒸馏，还对初始的CLIP模型进行蒸馏。

4.5 音频—视觉增量学习

音频—视觉多模态模型能够同时处理视觉和声音信号，通常应用于自动字幕生成、情感分析、会议摘要、多媒体检索和安全监控任务中。音频—视觉模型相对于单模态模型，在增量学习场景下还需要考虑模态融合的复杂性、更多的资源消耗、同步和时序问题等。Pian等人（2023）提出的D-AVSC（dual-audio-visual similarity constraint）方法在空间维度和时间维度分别进行知识蒸馏以保留旧知识的关键信息，并通过融合音频和视觉数据的特征，将样本对在

实例和分类层面上分别对齐来学习新知识，在保证深度模型对新任务的学习性能的情况下有效地减少了对旧任务知识的灾难性遗忘。Mo 等人（2023）提出的 CIGN（class-incremental grouping network）方法构造了一种新颖的网络架构，该架构的核心通过一个创新的分组机制处理连续的音视频数据流。具体而言，CIGN 方法将音频和视觉特征分别编码，并在隐空间中动态组合，对每个新类别引入专门的分组模块以捕捉其独特特征。Zuo 等人（2024）提出一种结合层次化数据增强的方法，利用多层次的数据增强策略，通过模拟新类数据的特征分布，以增强深度模型对新类的泛化能力。同时，利用知识蒸馏有效地将旧模型中的知识迁移到新模型。Lee 等人（2024）提出的 STELLA（spatio-temporal localized alignment）方法采用了空间时间局部对齐机制，减少必要重训练的数据量。Yue 等人（2024）提出的 MMAL（multi-modal analytic learning）方法通过深度融合音频与视觉数据，利用这两种模态的互补性增强音频—视觉多模态增量学习模型的泛化能力。此外，MMAL 还采用了模块化的知识蒸馏技术，在每次新类别引入时动态调整网络结构，从而保证模型对旧类别知识的稳定性。

5　国内外研究进展比较

为了对比国内外在增量学习领域的科研进展，本文对该领域在英文核心期刊和重要学术会议上的论文发表情况进行调研分析。

图 10 展示了增量学习领域 2000—2024 年国内外的论文发表总体发展趋势。从图中可以看出，论文发表数量整体呈现持续增长的趋势，尤其是从 2018 年开始，论文发表数量呈现出明显的增长态势，并在 2023 年达到高峰。此外，2024 年 1 月至 2024 年 9 月期间，国内外在增量学习领域的相关论文发表数量达到了 2 670 篇，且相较于前一年，月均发表的论文数量依然保持了上升的态势，表明增量学习研究热度不断增加。

图 11 进一步展示了同一时期不同国家在英文核心期刊及重要会议上的论文发表量。结果显示，21 世纪以来，我国在增量学习领域的论文发表数量位居全球首位，表明我国在增量学习领域的科研投入和产出已处于世界领先地位。

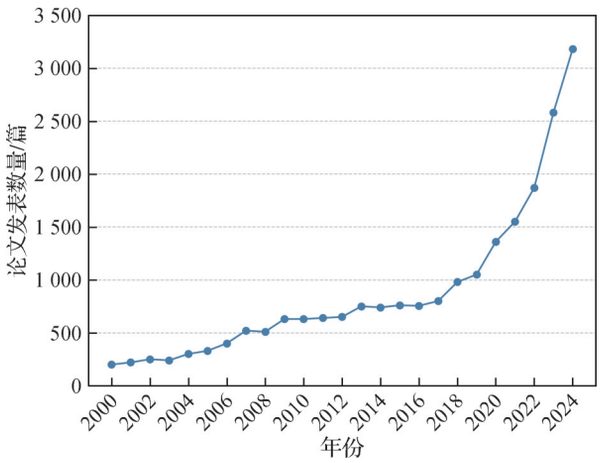


图 10　增量学习相关论文数量（统计数据源自 Scopus）

Fig. 10　Number of papers related to incremental learning (data is sourced from Scopus)

图 12 统计了过去 20 年各国论文发表数量与相对引用量的关系（统计数据源自知因），数据显示尽管我国在论文发表总量上处于世界第一，但论文相对引用量与美国和英国相比具有一定差距，表明我国在增量学习领域的研究影响力方面仍有提升空间。

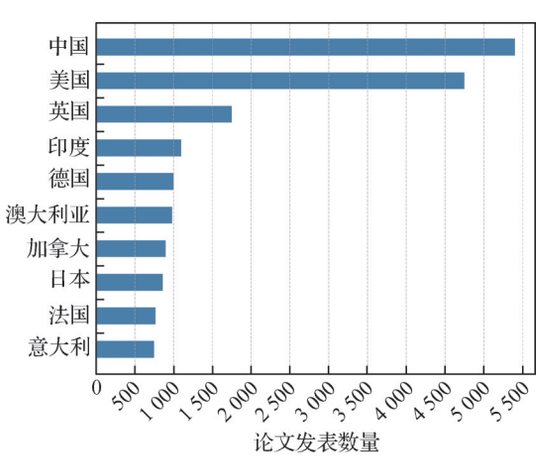


图 11　不同国家增量学习领域论文数量（统计数据源自 Scopus）

Fig. 11　Number of papers in incremental learning field in different countries (data is sourced from Scopus)

本文还调研了发表于人工智能与计算机视觉领域 7 类顶级会议（NIPS，CVPR，AAAI，ICML，ICLR，ICCV，ECCV）中增量学习相关的较新的文章，共计 498 篇。以国内研究机构人员为第一作者的文章为 189 篇，占比超过 37. 95%，远超其他国家。其中，北京大学团队发表相关论文 16 篇，清华大学 8 篇，华南理工大学 7 篇，浙江大学 6 篇，南开大学 6 篇等。由此可见，国内高校科研团队在增量学习领域投入了

大量研究，并取得了显著成果。另一方面，以国外研究机构人员为第一作者发表的文章共309篇，其中，韩国首尔大学和英国约克大学各12篇，新加坡国立大学6篇。与国外相比，国内在增量学习领域的研究团队数量更多，且在顶级学术会议上的论文发表量保持较高水准。此外，华为阿里巴巴、英特尔、苹果等企业也在增量学习领域发表了相关论文。本文还调研了上述7类会议中增量学习应用层面的相关文章发表情况。在498篇增量学习相关论文中，涉及实际应用层面的论文共108篇，包括增量语义分割30篇，增量目标检测22篇，增量动作识别4篇等。在这些文章中，国内研究机构发表45篇，占比超过41.66%，表明国内更注重增量学习在实际应用中的创新研究。

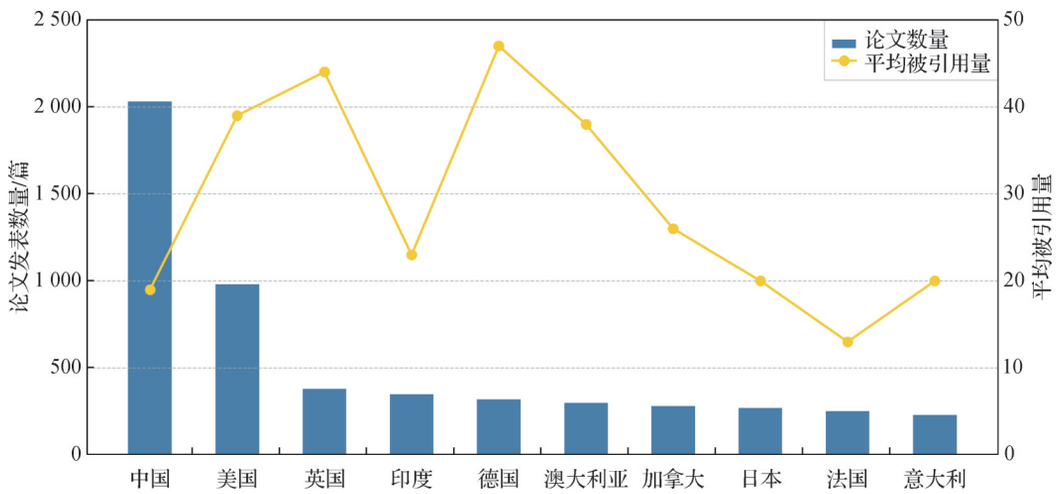


图12　国内外增量学习领域论文统计示意图（统计数据源自知因）

Fig. 12　Domestic and international incremental learning field paper statistics illustration (data is sourced from Aminer)

6　结　语

近年来，增量学习领域的研究取得了显著的进展。随着人工智能领域新兴应用的不断涌现，增量学习领域中新的研究方向也不断扩展。本节介绍未来值得研究的3个增量学习任务，分别为面向多模态大模型的增量学习、基于新型深度神经网络结构的增量学习以及面向人工智能安全的知识持续遗忘。

6.1　面向多模态大模型的增量学习

随着大语言模型能力的不断提升，诸如视觉—语言多模态大模型（large vision language model，LVLM）、语音—语言多模态大模型（speech language model，SpeechLM）等多模态大模型在计算机视觉、语音翻译等传统任务上取得了优异的性能。在实际应用中，多模态大模型同样面临对多个任务序列学习的增量学习场景。最近，有研究者针对视觉—语言多模态大模型探究了持续指令微调策略。He等人（2023）首次将正则化、网络扩展和经验重放等传统的增量学习方法整合到多模态大模型中。有工作（Anonymous authors，2024）通过不同的任务流分布将多模态大模型增量学习进一步应用在域增量学习、能力增量学习和数据集增量学习场景中。Zheng等人（2024）描述了多模态大模型在增量学习任务中的灾难性遗忘和负前向迁移挑战，对持续视觉语言问答、持续视觉语言预训练和持续指令微调三者的差异进行了分析。然而，目前多模态大模型领域内的增量学习方法仅限于视觉—语言大模型，并且由于多模态大模型的参数量规模较大，现有的多模态大模型增量学习方法训练开销仍然巨大；其次，多模态数据集的高度多样化和增量任务边界的模糊性导致此类方法在实际增量学习场景中性能欠佳。因此，多模态大模型增量学习领域的未来发展可以着重在模态多样性、多模态数据集、多模态增量任务边界以及模型参数量的研究上。

6.2　基于新型深度神经网络架构的增量学习

深度神经网络的架构设计经历了重要的演变，每个阶段都衍生了相应的增量学习方法。早期的卷积神经网络（convolutional neural network，CNN）通过局部连接和权重共享机制，在保持较低参数量的同时提高了深度模型的训练效率，并成功应用于图

像分类、目标检测等多种视觉任务。研究者们基于卷积神经网络提出正则化（Wen 等，2025；Zhao 等，2023a）、重放（Meng 等，2024；Chen 和 Chang，2023）和结构化（Qiang 等，2024；Wang 等，2023b）等增量学习的方法，解决了卷积神经网络在动态环境中面临任务流时的灾难性遗忘问题。然而，卷积神经网络及其增量学习方法在处理复杂序列任务时表现不佳。与卷积神经网络不同，Transformer 利用自注意力（self-attention）机制动态捕捉数据序列中任意位置之间的相关性，使得模型在面向语音翻译、机器人多模态任务等更复杂的任务时能够更加灵活。在传统增量学习方法之外，研究者针对 Transformer 预训练模型提出基于提示的预训练模型微调（Kim 等，2024b）以及基于表征的预训练模型微调（Tang 等，2024）的增量学习策略，极大程度地减少了 Transformer 及其相关模型在面临动态任务流时的灾难性遗忘。Transformer 及其衍生出的增量学习方法虽然在处理语言数据和序列任务上有出色表现，但在处理图像数据和其他高维度数据时存在计算复杂度较高、对大规模数据依赖较强等问题。在此背景下，基于状态空间模型（state-space models，SSM）的新型深度学习架构 Mamba 应运而生（Gu 和 Dao，2023）。Mamba 深度模型使用 SSM（state space models）参数作为输入的函数，并取代 Transformer 架构中的自注意力机制，通过硬件感知的并行算法构造端到端的神经网络架构。Mamba 架构相较于卷积神经网络和 Transformer 架构，在处理长序列及高维度数据时具有更高效且更具扩展性的特点。在增量学习场景下，Mamba 架构也存在灾难性遗忘问题。然而，目前增量学习领域的方法主要针对卷积神经网络及 Transformer 架构，并没有针对 Mamba 架构进行特殊设计的方法。因此，增量学习领域未来的研究方向可以重点关注 Mamba 等新型深度网络架构在增量学习场景下的解决方案。

6.3 面向人工智能安全的知识持续遗忘

越来越多的开源预训练大模型得到工业界和学术界的研究者使用，并在各个应用任务中取得了优异的性能。然而，在一些对隐私性或安全性要求较高的应用场景，例如面向儿童的问答系统或基于人脸识别的访客系统，预训练模型中的不合法或非必要的知识应当被擦除或遗忘。其中，不合法知识包括预训练图像生成模型中色情和暴力有关的危险概念，非必要知识包括已经不再使用的人脸信息等。在实际应用中，预训练模型需要擦除的概念或遗忘的知识通常是序列化的任务，这类对预训练模型的部分知识进行持续擦除的场景为知识持续遗忘（incremental forgetting）场景。知识持续遗忘任务的主要目标是在有效遗忘不合法或非必要知识的同时保留预训练模型中的其他知识。目前已经有研究者针对预训练的目标检测、人脸识别和图像分类等视觉感知模型展开了知识持续遗忘技术的研究（Chatterjee 等，2024；Zhao 等，2024）。但在文本生成、多模态检索、多模态感知以及图像和视频生成等领域的知识持续遗忘技术仍存在空白。有效的知识持续遗忘技术能够提升预训练模型在实际应用中的安全性，并减少模型的冗余表征能力，可以作为增量学习领域未来的研究方向。

致谢：本文由中国图象图形学学会视觉感知智能系统专业委员会组织撰写，该专业委员会链接为 https://www.csig.org.cn/16/201704/49325.html，在此表示感谢。

参考文献（Reference）

Abati D，Tomczak J，Blankevoort T，Calderara S，Cucchiara R and Bejnordi B E. 2020. Conditional channel gated networks for task-aware continual learning//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：3930-3939［DOI：10.1109/CVPR42600.2020.00399］

Aljundi R，Babiloni F，Elhoseiny M，Rohrbach M and Tuytelaars T. 2018. Memory aware synapses：learning what（not）to forget//Proceedings of the 15th European Conference on Computer Vision —— ECCV 2018. Munich，Germany：Springer：144-161［DOI：10.1007/978-3-030-01219-9_9］

Aljundi R，Chakravarty P and Tuytelaars T. 2017. Expert gate：lifelong learning with a network of experts//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu，USA：IEEE：7120-7129［DOI：10.1109/CVPR.2017.753］

Aljundi R，Kelchtermans K and Tuytelaars T. 2019a. Task-free continual learning//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach，USA：IEEE：11246-11255［DOI：10.1109/CVPR.2019.01151］

Aljundi R，Lin M，Goujaud B and Bengio Y. 2019b. Gradient based sample selection for online continual learning［EB/OL］.［2024-09-15］. https://arxiv.org/pdf/1903.08671.pdf

Anonymous authors. 2024. Continual LLaVA：continual instruction tuning in large vision-language models［EB/OL］.［2024-09-15］.

Https://openreview.net/pdf?id=rwmwFnmjAX

Ayromlou S, Abolmaesumi P, Tsang T and Li X X. 2022. Class impression for data-free incremental learning//Proceedings of the 25th International Conference on Medical Image Computing and Computer-Assisted Intervention. Singapore, Singapore: Springer: 320-329 [DOI: 10.1007/978-3-031-16440-8_31]

Bai X Y, Shang J H, Sun Y F and Balasubramanian N. 2022. Enhancing continual learning with global prototypes: counteracting negative representation drift [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2205.12186.pdf

Bang J, Kim H, Yoo Y, Ha J W and Choi J. 2021. Rainbow memory: continual learning with a memory of diverse samples//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 8214-8223 [DOI: 10.1109/CVPR46437.2021.00812]

Batra H and Clark R. 2024. EVCL: elastic variational continual learning with weight consolidation [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2406.15972.pdf

Bayoudh K, Knani R, Hamdaoui F and Mtibaa A. 2022. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. The Visual Computer, 38(8): 2939-2970 [DOI: 10.1007/s00371-021-02166-7]

Belouadah E and Popescu A. 2020. ScaIL: classifier weights scaling for class incremental learning//Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass, USA: IEEE: 1255-1264 [DOI: 10.1109/WACV45572.2020.9093562]

Biasetton M, Michieli U, Agresti G and Zanuttigh P. 2019. Unsupervised domain adaptation for semantic segmentation of urban scenes//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Long Beach, USA: IEEE: 1211-1220 [DOI: 10.1109/CVPRW.2019.00160]

Borsos Z, Mutný M and Krause A. 2020. Coresets via bilevel optimization for continual learning and streaming [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2006.03875.pdf

Buzzega P, Boschini M, Porrello A, Abati D and Calderara S. 2020. Dark experience for general continual learning: a strong, simple baseline [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2004.07211.pdf

Castro F M, Marín-Jiménez M J, Guil N, Schmid C and Alahari K. 2018. End-to-end incremental learning//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 241-257 [DOI: 10.1007/978-3-030-01258-8_15]

Cermelli F, Fontanel D, Tavera A, Ciccone M and Caputo B. 2022. Incremental learning in semantic segmentation from image labels//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 4361-4371 [DOI: 10.1109/CVPR52688.2022.00433]

Chamikara M A P, Bertók P, Liu D, Camtepe S and Khalil I. 2018. Efficient data perturbation for privacy preserving and accurate data stream mining. Pervasive and Mobile Computing, 48: 1-19 [DOI: 10.1016/j.pmcj.2018.05.003]

Chatterjee R, Chundawat V, Tarun A, Mali A A and Mandal M. 2024. A unified framework for continual learning and unlearning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2408.11374.pdf

Chaudhry A, Dokania P K, Ajanthan T and Torr P H. 2018. Riemannian walk for incremental learning: understanding forgetting and intransigence//Proceedings of the 15th European Conference on Computer Vision – ECCV 2018. Munich, Germany: Springer: 556-572 [DOI: 10.1007/978-3-030-01252-6_33]

Chaudhry A, Gordo A, Dokania P, Torr P and Lopez-Paz D. 2021. Using hindsight to anchor past knowledge in continual learning//Proceedings of the 35th AAAI Conference on Artificial Intelligence. Virtually: AAAI: 6993-7001 [DOI: 10.1609/aaai.v35i8.16861]

Chaudhry A, Rohrbach M, Elhoseiny M, Ajanthan T, Dokania P K, Torr P H S and Ranzato M A. 2019. On tiny episodic memories in continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1902.10486.pdf

Chen X L, Xie S N and He K M. 2021. An empirical study of training self-supervised vision transformers//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 9620-9629 [DOI: 10.1109/ICCV48922.2021.00950]

Chen X W and Chang X B. 2023. Dynamic residual classifier for class incremental learning//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 18697-18706 [DOI: 10.1109/ICCV51070.2023.01718]

Cong W, Cong Y, Liu Y Y and Sun G. 2024. Cs2K: class-specific and class-shared knowledge guidance for incremental semantic segmentation [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2407.09047.pdf

Cywiński B, Deja K, Trzciński T, Twardowski B and Kuciński Ł. 2024. GUIDE: guidance-based incremental learning with diffusion models [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2403.03938.pdf

Delange M, Aljundi R, Masana M, Parisot S, Jia X, Leonardis A, Slabaugh G and Tuytelaars T. 2021. A continual learning survey: defying forgetting in classification tasks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(7): 3366-3385 [DOI: 10.1109/TPAMI.2021.3057446]

Dhar P, Singh R V, Peng K C, Wu Z Y and Chellappa R. 2019. Learning without memorizing//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 5133-5141 [DOI: 10.1109/CVPR.2019.00528]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N. 2020. An image is worth 16x16 words: transformers for image recognition at scale [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2010.11929.pdf

Douillard A, Chen Y F, Dapogny A and Cord M. 2021. PLOP: learning without forgetting for continual semantic segmentation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 4039-4049 [DOI: 10.1109/CVPR46437.2021.00403]

Duncker L, Driscoll L N, Shenoy K V, Sahani M and Sussillo D. 2020. Organizing recurrent network dynamics by task-computation to enable continual learning//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 14387-14397

Elasri M, Elharrouss O, Al-Maadeed S and Tairi H. 2022. Image generation: a review. Neural Processing Letters, 54(5): 4609-4646 [DOI: 10.1007/s11063-022-10777-x]

Feng Q, Zhao H B, Zhang C, Dong J H, Ding H H, Jiang Y G and Qian H. 2024a. PECTP: parameter-efficient cross-task prompts for incremental vision transformer [EB/OL]. [2024-09-15]. https://arxiv.org/pdf /2407.03813.pdf

Feng Q, Zhou D W, Zhao H B, Zhang C and Qian H. 2024b. LW2G: learning whether to grow for prompt-based continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2409.18860.pdf

Fernando C, Banarse D, Blundell C, Zwols Y, Ha D, Rusu A A, Pritzel A and Wierstra D. 2017. PathNet: evolution channels gradient descent in super neural networks [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1701.08734.pdf

Floridi L and Chiriatti M. 2020. GPT-3: its nature, scope, limits, and consequences. Minds and Machines, 30(4): 681-694 [DOI: 10.1007/s11023-020-09548-1]

French R M. 1999. Catastrophic forgetting in connectionist networks. Trends in Cognitive Sciences, 3(4): 128-135 [DOI: 10.1016/S1364-6613(99)01294-2]

Gaber M M. 2012. Advances in data stream mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(1): 79-85 [DOI: 10.1002/widm.52]

Gao R and Liu W W. 2023. DDGR: continual learning with deep diffusion-based generative replay//Proceedings of the 40th International Conference on Machine Learning. Honolulu, USA: JMLR.org: 10744-10763

Golab L and Özsu M T. 2003. Issues in data stream management. ACM SIGMOD Record, 32(2): 5-14 [DOI: 10.1145/776985.776986]

Gu A and Dao T. 2023. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2312.00752.pdf

Gui J, Sun Z N, Wen Y G, Tao D C and Ye J P. 2021. A review on generative adversarial networks: algorithms, theory, and applications. IEEE Transactions on Knowledge and Data Engineering, 35(4): 3313-3332 [DOI: 10.1109/TKDE.2021.3130191]

Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V and Courville A. 2017. Improved training of Wasserstein GANs [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1704.00028.pdf

Harvey I. 2011. The microbial genetic algorithm//Proceedings of the 10th European Conference on Advances in Artificial Life. Budapest, Hungary: Springer: 126-133 [DOI: 10.1007/978-3-642-21314-4_16]

He J H, Guo H Y, Tang M and Wang J Q. 2023. Continual instruction tuning for large multimodal models [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2311.16206.pdf

He Y Q, Huang X C, Tang M H, Meng L X, Li X, Lin W, Zhang W Y and Gao Y F. 2024. Don't half-listen: capturing key-part information in continual instruction tuning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2403.10056.pdf

Hinton G, Vinyals O and Dean J. 2015. Distilling the knowledge in a neural network [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1503.02531.pdf

Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2006.11239.pdf

Hou S H, Pan X Y, Loy C C, Wang Z L and Lin D H. 2018. Lifelong learning via progressive distillation and retrospection//Proceedings of the 15th European Conference on Computer Vision —— ECCV 2018. Munich, Germany: Springer: 452-467 [DOI: 10.1007/978-3-030-01219-9_27]

Hou S H, Pan X Y, Loy C C, Wang Z L and Lin D H. 2019. Learning a unified classifier incrementally via rebalancing//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 831-839 [DOI: 10.1109/CVPR.2019.00092]

Houlsby N, Giurgiu A, Jastrzebski S, Morrone B, De Laroussilhe D, Gesmundo A, Attariyan M and Gelly S. 2019. Parameter-efficient transfer learning for NLP//Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA: [s. n.]: 2790-2799

Hsu Y C, Liu Y C, Ramasamy A and Kira Z. 2018. Re-evaluating continual learning scenarios: a categorization and case for strong baselines [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1810.12488.pdf

Hu E, Shen Y L, Wallis P, Zhu Z A, Li Y Z, Wang S A, Wang L and Chen W Z. 2021. Lora: low-rank adaptation of large language models [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2106.09685.pdf

Hu W P, Lin Z, Liu B, Tao C Y, Tao Z W, Zhao D Y, Ma J W and Yan R. 2019. Overcoming catastrophic forgetting for continual learning via model adaptation//Proceedings of 2019 International Conference on Learning Representations.

Hu Z Z and Rostami M. 2023. Class-incremental learning using generative experience replay based on time-aware regularization [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2310.03898.pdf

Huang L L, Cao X S, Lu H R and Liu X L. 2024. Class-incremental learning with CLIP: adaptive representation adjustment and parameter fusion//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 214-231 [DOI: 10.1007/978-3-031-72949-2_13]

Huang Z L, Hao W T, Wang X G, Tao M Y, Huang J Q, Liu W Y and Hua X S. 2021. Half-real half-fake distillation for class-incremental semantic segmentation [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2104.00875.pdf

Hung S C Y, Tu C H, Wu C E, Chen C H, Chan Y M and Chen C S. 2019. Compacting, picking and growing for unforgetting continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1910.06562.pdf

Jia M J, Tang L M, Chen B C, Cardie C, Belongie S, Hariharan B and Lim S N. 2022. Visual prompt tuning//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 709-727 [DOI: 10.1007/978-3-031-19827-4_41]

Jiang J and Celiktutan O. 2023. Neural weight search for scalable task incremental learning//Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 1390-1399 [DOI: 10.1109/WACV56688.2023.00144]

Jodelet Q, Liu X, Phua Y J and Murata T. 2023. Class-incremental learning using diffusion model for distillation and replay//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision Workshops. Paris, France: IEEE: 3417-3425 [DOI: 10.1109/ICCVW60793.2023.00367]

Ke Z X, Shao Y J, Lin H W, Konishi T, Kim G and Liu B. 2023. Continual pre-training of language models [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2302.03241.pdf

Khan V, Cygert S, Twardowski B and Trzciński T. 2023. Looking through the past: better knowledge retention for generative replay in continual learning//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision Workshops. Paris, France: IEEE: 3488-3492 [DOI: 10.1109/ICCVW60793.2023.00375]

Kim Y, Fang J, Zhang Q, Cai Z W, Shen Y T, Duggal R, Raychaudhuri D S, Tu Z W, Xing Y F and Dabeer O. 2024a. Open-world dynamic prompt and continual visual representation learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2409.05312.pdf

Kim Y, Li Y H and Panda P. 2024b. One-stage prompt-based continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2402.16189.pdf

Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu A A, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, Hassabis D, Clopath C, Kumaran D and Hadsell R. 2017. Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences of the United States of America, 114(13): 3521-3526 [DOI: 10.1073/pnas.1611835114]

Koh H, Kim D, Ha J W and Choi J. 2022. Online continual learning on class incremental blurry task configuration with anytime inference [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2110.10031.pdf

Konaté M A, Yao A F, Chateau T and Bouges P. 2023. Toward industrial use of continual learning: new metrics proposal for class incremental learning//Proceedings of 2023 International Joint Conference on Neural Networks (IJCNN). Gold Coast, Australia: IEEE: 1-7 [DOI: 10.1109/IJCNN54540.2023.10191657]

Kong Y J, Liu L, Chen H H, Kacprzyk J and Tao D C. 2023. Overcoming catastrophic forgetting in continual learning by exploring eigenvalues of hessian matrix. IEEE Transactions on Neural Networks and Learning Systems, 35(11): 16196-16210 [DOI: 10.1109/TNNLS.2023.3292359]

Krempl G, Žliobaite I, Brzeziński D, Hüllermeier E, Last M, Lemaire V, Noack T, Shaker A, Sievi S, Spiliopoulou M and Stefanowski J. 2014. Open challenges for data stream mining research. ACM SIGKDD Explorations Newsletter, 16(1): 1-10 [DOI: 10.1145/2674026.2674028]

Kumaran D, Hassabis D and McClelland J L. 2016. What learning systems do intelligent agents need? Complementary learning systems theory updated. Trends in Cognitive Sciences, 20(7): 512-534 [DOI: 10.1016/j.tics.2016.05.004]

Lavda F, Ramapuram J, Gregorova M and Kalousis A. 2018. Continual classification learning using generative models [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1810.10612.pdf

Lecun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. Proceedings of 1998 IEEE, 86(11): 2278-2324 [DOI: 10.1109/5.726791]

Lee J, Hong H G, Joo D and Kim J. 2020. Continual learning with extended Kronecker-factored approximate curvature//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 8998-9007 [DOI: 10.1109/CVPR42600.2020.00902]

Lee J, Yoon J, Kim W, Kim Y and Hwang S J. 2024. STELLA: continual audio-video pre-training with spatiotemporal localized alignment//Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria: [s.n.]: 27094-27117

Lee K, Lee K, Shin J and Lee H. 2019. Overcoming catastrophic forgetting with unlabeled data in the wild//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 312-321 [DOI: 10.1109/ICCV.2019.00040]

Li X Y, Tang B and Li H F. 2024. AdaER: an adaptive experience replay approach for continual lifelong learning. Neurocomputing, 572: #127204 [DOI: 10.1016/j.neucom.2023.127204]

Li Z Z and Hoiem D. 2018. Learning without forgetting. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(12): 2935-2947 [DOI: 10.1109/TPAMI.2017.2773081]

Liang J L, Zhong J, Gu H L, Lu Z Q, Tang X X, Dai G, Huang S P, Fan L X and Yang Q. 2024. Diffusion-driven data replay: a novel approach to combat forgetting in federated class continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2409.01128.pdf

Lin S, Yang L, Fan D L and Zhang J S. 2022. TRGP: trust region gradient projection for continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2202.02931.pdf

Liu C X, Wang Z Y, Xiong T Y, Chen R B, Wu Y H, Guo J F and Huang H. 2024a. Few-shot class incremental learning with attention-aware self-adaptive prompt [EB/OL]. [2024-09-15].

https://arxiv.org/pdf/2403.09857.pdf

Liu X L, Masana M, Herranz L, Van de Weijer J, Lopez A M and Bagdanov A D. 2018. Rotate your networks: better weight consolidation and less catastrophic forgetting//Proceedings of the 24th International Conference on Pattern Recognition (ICPR). Beijing, China: IEEE: 2262-2268 [DOI: 10.1109/ICPR.2018.8545895]

Lopez-Paz D and Ranzato M A. 2017. Gradient episodic memory for continual learning//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6470-6479

Malla S, Choi J H and Choi C. 2024. COPAL: continual pruning in large language generative models [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2405.02347.pdf

Mallya A and Lazebnik S. 2018. PackNet: adding multiple tasks to a single network by iterative pruning//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7765-7773 [DOI: 10.1109/CVPR.2018.00810]

McClelland J L, McNaughton B L and O'Reilly R C. 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological Review, 102(3): 419-457 [DOI: 10.1037/0033-295x.102.3.419]

McClure P, Zheng C Y, Kaczmarzyk J, Lee J A, Ghosh S S, Nielson D, Bandettini P and Pereira F. 2018. Distributed weight consolidation: a brain segmentation case study [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1805.10863.pdf

McDonnell M D, Gong D, Parvaneh A, Abbasnejad E and van den Hengel A. 2024. RanPAC: random projections and pre-trained models for continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2307.02251.pdf

Meng Z C, Zhang J, Yang C D, Zhan Z, Zhao P and Wang Y Z. 2024. DiffClass: diffusion-based class incremental learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2403.05016.pdf

Michieli U and Badia L. 2018. Game theoretic analysis of road user safety scenarios involving autonomous vehicles//Proceedings of the 29th IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Bologna, Italy: IEEE: 1377-1381 [DOI: 10.1109/PIMRC.2018.8580679]

Michieli U and Zanuttigh P. 2019. Incremental learning techniques for semantic segmentation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul, Korea (South): IEEE: 3205-3212 [DOI: 10.1109/ICCVW.2019.00400]

Mittelstadt B D, Allo P, Taddeo M, Wachter S and Floridi L. 2016. The ethics of algorithms: mapping the debate. Big Data and Society, 3(2): #2053951716679679 [DOI: 10.1177/2053951716679679]

Mo S T, Pian W G and Tian Y P. 2023. Class-incremental grouping network for continual audio-visual learning//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 7754-7764 [DOI: 10.1109/ICCV51070.2023.00716]

Narayanan D, Shoeybi M, Casper J, LeGresley P, Patwary M, Korthikanti V, Vainbrand D, Kashinkunti P, Bernauer J, Catanzaro B, Phanishayee A and Zaharia M. 2021. Efficient large-scale language model training on GPU clusters using Megatron-LM//Proceedings of 2021 International Conference for High Performance Computing, Networking, Storage and Analysis. St. Louis, USA: ACM: #58 [DOI: 10.1145/3458817.3476209]

Ni Z X, Wei L H, Tang S L, Zhuang Y T and Tian Q. 2023. Continual vision-language representation learning with off-diagonal information//Proceedings of the 40th International Conference on Machine Learning. Honolulu, USA: [s.n.]: 26129-26149

Park G, Moon W J, Lee S B, Kim T Y and Heo J P. 2024. Mitigating background shift in class-incremental semantic segmentation [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2407.11859.pdf

Patterson D, Gonzalez J, Le Q, Liang C, Munguia L M, Rothchild D, So D, Texier M and Dean J. 2021. Carbon emissions and large neural network training [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2104.10350.pdf

Pian W G, Mo S T, Guo Y H and Tian Y P. 2023. Audio-visual class-incremental learning//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 7765-7777 [DOI: 10.1109/ICCV51070.2023.00717]

Qazi M A, Almakky I, Hashmi A U R, Sanjeev S and Yaqub M. 2024. DynaMMo: dynamic model merging for efficient class incremental learning for medical images//Proceedings of the 28th Annual Conference on Medical Image Understanding and Analysis. Manchester, UK: Springer: 245-257 [DOI: 10.1007/978-3-031-66955-2_17]

Qiang S Y, Liang Y Y, Wan J and Zhang D. 2024. Dynamic feature learning and matching for class-incremental learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2405.08533.pdf

Qiao J Y, Zhang Z Z, Tan X, Chen C W, Qu Y Y, Peng Y and Xie Y. 2023. Prompt gradient projection for continual learning//Proceedings of the 12th International Conference on Learning Representations.

Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G and Sutskever I. 2021. Learning transferable visual models from natural language supervision [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2103.00020.pdf

Rannen A, Aljundi R, Blaschko M B and Tuytelaars T. 2017. Encoder based lifelong learning//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 1329-1337 [DOI: 10.1109/ICCV.2017.148]

Rebuffi S A, Kolesnikov A, Sperl G and Lampert C H. 2017. ICaRL: incremental classifier and representation learning//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 5533-5542 [DOI: 10.1109/CVPR.2017.587]

Riemer M, Cases I, Ajemian R, Liu M, Rish I, Tu Y H and Tesauro G.

2018. Learning to learn without forgetting by maximizing transfer and minimizing interference [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1810.11910.pdf

Rusu A A, Rabinowitz N C, Desjardins G, Soyer H, Kirkpatrick J, Kavukcuoglu K, Pascanu R and Hadsell R. 2016. Progressive neural networks [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1606.04671.pdf

Rypeść G, Marczak D, Cygert S, Trzciński T and Twardowski B. 2024. Category adaptation meets projected distillation in generalized continual category discovery [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2308.12112.pdf

Saha G, Garg I and Roy K. 2021. Gradient projection memory for continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2103.09762.pdf

Seo J, Kang J S and Park G M. 2023. LFS-GAN: lifelong few-shot image generation//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 11322-11332 [DOI: 10.1109/ICCV51070.2023.01043]

Serrà J, Surís D, Miron M and Karatzoglou A. 2018. Overcoming catastrophic forgetting with hard attention to the task [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1801.01423.pdf

Shang C, Li H L, Meng F M, Wu Q B, Qiu H Q and Wang L X. 2023. Incrementer: transformer for class-incremental semantic segmentation with knowledge distillation focusing on old class//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 7214-7224 [DOI: 10.1109/CVPR52729.2023.00697]

Shim D, Mai Z, Jeong J, Sanner S, Kim H and Jang J. 2021. Online class-incremental continual learning with adversarial Shapley value//Proceedings of the 35th AAAI Conference on Artificial Intelligence. Virtually: AAAI: 9630-9638 [DOI: 10.1609/aaai.v35i11.17159]

Shin H, Lee J K, Kim J and Kim J. 2017. Continual learning with deep generative replay [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1705.08690.pdf

Shmelkov K, Schmid C and Alahari K. 2018. How good is my GAN?//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 218-234 [DOI: 10.1007/978-3-030-01216-8_14]

Shokri R and Shmatikov V. 2015. Privacy-preserving deep learning//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Denver, USA: ACM: 1310-1321 [DOI: 10.1145/2810103.2813687]

Smith J, Balloch J, Hsu Y C and Kira Z. 2021. Memory-efficient semi-supervised continual learning: the world is its own replay buffer//Proceedings of 2021 International Joint Conference on Neural Networks (IJCNN). Shenzhen, China: IEEE: 1-8 [DOI: 10.1109/IJCNN52387.2021.9534361]

Smith J S, Hsu Y C, Zhang L Y, Hua T, Kira Z, Shen Y L and Jin H X. 2023. Continual diffusion: continual customization of text-to-image diffusion with C-LORA [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2304.06027.pdf

Srinivasan T, Chang T Y, Pinto Alva L, Chochlakis G, Rostami M and Thomason J. 2022. CLiMB: a continual learning benchmark for vision-and-language tasks//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: 29440-29453

Tang L X, Tian Z T, Li K, He C M, Zhou H T, Zhao H S, Li X and Jia J Y. 2024. Mind the interference: retaining pre-trained knowledge in parameter efficient continual learning of vision-language models [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2407.05342.pdf

Tiwari R, Killamsetty K, Iyer R and Shenoy P. 2022. GCR: gradient coreset based replay buffer selection for continual learning//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 99-108 [DOI: 10.1109/CVPR52688.2022.00020]

Touvron H, Cord M, Douze M, Massa F, Sablayrolles A and Jégou H. 2021. Training data-efficient image transformers and distillation through attention [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2012.12877.pdf

Van de Ven G M and Tolias A S. 2018. Generative replay with feedback connections as a general strategy for continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1809.10635.pdf

Van de Ven G M and Tolias A S. 2019. Three scenarios for continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1904.07734.pdf

Wang S P, Li X R, Sun J and Xu Z B. 2021. Training networks in null space of feature covariance for continual learning//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 184-193 [DOI: 10.1109/CVPR46437.2021.00025]

Wang X, Chen T Z, Ge Q M, Xia H, Bao R, Zheng R, Zhang Q, Gui T and Huang X J. 2023a. Orthogonal subspace learning for language model continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2310.14152.pdf

Wang Y B, Huang Z W and Hong X P. 2022a. S-prompts learning with pre-trained transformers: an occam's razor for domain incremental learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2207.12819.pdf

Wang Y B, Ma Z H, Huang Z W, Wang Y W, Su Z and Hong X P. 2023b. Isolation and impartial aggregation: a paradigm of incremental learning without interference//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI: 10209-10217 [DOI: 10.1609/aaai.v37i8.26216]

Wang Z F, Zhang Z Z, Ebrahimi S, Sun R X, Zhang H, Lee C Y, Ren X Q, Su G L, Perot V, Dy J and Pfister T. 2022b. DualPrompt: complementary prompting for rehearsal-free continual learning//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 631-648 [DOI: 10.1007/978-3-031-19809-0_36]

Wang Z F, Zhang Z Z, Lee C Y, Zhang H, Sun R X, Ren X Q, Su G L, Perot V, Dy J and Pfister T. 2022c. Learning to prompt for continual learning//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 139-149 [DOI: 10.1109/CVPR52688.2022.00024]

Welling M. 2009. Herding dynamical weights to learn//Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Canada: ACM: 1121-1128 [DOI: 10.1145/1553374.1553517]

Wen H T, Qiu H Q, Wang L X, Cheng H Y and Li H L. 2025. Class incremental learning with less forgetting direction and equilibrium point. IEEE Transactions on Circuits and Systems for Video Technology, 35(2): 1150-1164 [DOI: 10.1109/TCSVT.2024.3477951]

Winata G I, Xie L J, Radhakrishnan K, Wu S J, Jin X S, Cheng P X, Kulkarni M and Preoţiuc-Pietro D. 2023. Overcoming catastrophic forgetting in massively multilingual continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2305.16252.pdf

Wu C S, Herranz L, Liu X L, Van De Weijer J and Raducanu B. 2018. Memory replay GANs: learning to generate images from new categories without forgetting//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc.: 5966-5976

Wu Y, Chen Y P, Wang L J, Ye Y C, Liu Z C, Guo Y D and Fu Y. 2019. Large scale incremental learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1905.13260.pdf

Xiang Y, Fu Y, Ji P and Huang H. 2019. Incremental learning using conditional adversarial networks//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 6618-6627 [DOI: 10.1109/ICCV.2019.00672]

Xie Z Y, Lu H Q, Xiao J W, Wang E G, Zhang L and Liu X L. 2024. Early preparation pays off: new classifier pre-tuning for class incremental semantic segmentation [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2407.14142.pdf

Xu J and Zhu Z X. 2018. Reinforced continual learning//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc.: 907-916

Xu Y C, Chen Y X, Nie J H, Wang Y S, Zhuang H P and Okumura M. 2024. Advancing cross-domain discriminability in continual learning of vision-language models [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2406.18868.pdf

Yan S P, Hong L Q, Xu H, Han J H, Tuytelaars T, Li Z G and He X M. 2022. Generative negative text replay for continual vision-language pretraining [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2210.17322.pdf

Yan S P, Xie J W and He X M. 2021b. DER: dynamically expandable representation for class incremental learning//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 3013-3022 [DOI: 10.1109/CVPR46437.2021.00303]

Yan S P, Zhou J L, Xie J W, Zhang S Y and He X M. 2021a. An EM framework for online incremental learning of semantic segmentation//Proceedings of the 29th ACM International Conference on Multimedia. Virtual Event, China: ACM: 3052-3060 [DOI: 10.1145/3474085.3475443]

Yang L R, Zhao H B, Yu Y L, Zeng X D and Li X. 2024. RCS-Prompt: learning prompt to rearrange class space for prompt-based continual learning//Proceedings of the 18th European Conference on Computer Vision (ECCV). Milan, Italy: Springer: 1-20 [DOI: 10.1007/978-3-031-72970-6_1]

Yoon J, Madaan D, Yang E and Hwang S J. 2021. Online coreset selection for rehearsal-based continual learning [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2106.01085.pdf

Yoon J, Yang E, Lee J and Hwang S J. 2017. Lifelong learning with dynamically expandable networks [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1708.01547.pdf

Yu J Z, Zhuge Y Z, Zhang L, Hu P, Wang D, Lu H C and He Y. 2024a. Boosting continual learning of vision-language models via mixture-of-experts adapters//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 23219-23230 [DOI: 10.1109/CVPR52733.2024.02191]

Yu Y C, Huang C P, Chen J J, Chang K P, Lai Y H, Yang F E, Wang Y C F. 2024b. Select and distill: selective dual-teacher knowledge transfer for continual learning on vision-language models//Proceedings of the 18th European Conference on Computer Vision —— ECCV 2024. Milan, Italy: Springer: 219-236 [DOI: 10.1007/978-3-031-73347-5_13]

Yue X H, Zhang X Y, Chen Y M, Zhang C W, Lao M R, Zhuang H P, Qain X Y and Li H Z. 2024. MMAL: multi-modal analytic learning for exemplar-free audio-visual class incremental tasks//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne, Australia: ACM: 2428-2437 [DOI: 10.1145/3664647.3681607]

Zagoruyko S and Komodakis N. 2016. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1612.03928.pdf

Zenke F, Poole B and Ganguli S. 2017. Continual learning through synaptic intelligence [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1703.04200.pdf

Zhai M Y, Chen L, He J W, Nawhal M, Tung F and Mori G. 2020. Piggyback GAN: efficient lifelong learning for image conditioned generation//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 397-413 [DOI: 10.1007/978-3-030-58589-1_24]

Zhai M Y, Chen L and Mori G. 2021. Hyper-LifelongGAN: scalable lifelong learning for image conditioned generation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 2246-2255 [DOI: 10.1109/

CVPR46437.2021.00228]

Zhai M Y, Chen L, Tung F, He J W, Nawhal M and Mori G. 2019. Lifelong GAN: continual learning for conditional image generation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 2759-2768 [DOI: 10.1109/ICCV.2019.00285]

Zhang A Q and Gao G Y. 2024. Background adaptation with residual modeling for exemplar-free class-incremental semantic segmentation [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2407.09838.pdf

Zhang J T, Zhang J, Ghosh S, Li D W, Tasci S, Heck L, Zhang H M and Kuo C C J. 2020. Class-incremental learning via deep model consolidation [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/1903.07864.pdf

Zhang Y Z, Wang X Z and Yang D Y. 2022. Continual sequence generation with adaptive compositional modules [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2203.10652.pdf

Zhao G Z and Mu K D. 2023. Connection-based knowledge transfer for class incremental learning//Proceedings of 2023 International Joint Conference on Neural Networks (IJCNN). Gold Coast, Australia: IEEE: 1-8 [DOI: 10.1109/IJCNN54540.2023.10191696]

Zhao H B, Fu Y J, Kang M T, Tian Q, Wu F and Li X. 2021a. MgSvF: multi-grained slow versus fast framework for few-shot class-incremental learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(3): 1576-1588 [DOI: 10.1109/TPAMI.2021.3133897]

Zhao H B, Ni B L, Fan J S, Wang Y X, Chen Y T, Meng G F and Zhang Z X. 2024. Continual forgetting for pre-trained vision models//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 28631-28642 [DOI: 10.1109/CVPR52733.2024.02705]

Zhao H B, Qin X, Su S H, Fu Y J, Lin Z B and Li X. 2021b. When video classification meets incremental classes//Proceedings of the 29th ACM International Conference on Multimedia. Virtual Event, China: ACM: 880-889 [DOI: 10.1145/3474085.3475265]

Zhao H B, Wang H, Fu Y J, Wu F and Li X. 2021c. Memory-efficient class-incremental learning for image classification. IEEE Transactions on Neural Networks and Learning Systems, 33(10): 5966-5977 [DOI: 10.1109/TNNLS.2021.3072041]

Zhao H B, Yang F Y, Fu X H and Li X. 2022. RBC: rectifying the biased context in continual semantic segmentation//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 55-72 [DOI: 10.1007/978-3-031-19830-4_4]

Zhao H B, Zeng H, Qin X, Fu Y J, Wang H, Omar B and Li X. 2021d. What and where: learn to plug adapters via NAS for multidomain learning. IEEE Transactions on Neural Networks and Learning Systems, 33(11): 6532-6544 [DOI: 10.1109/TNNLS.2021.3082316]

Zhao L L, Lu J, Xu Y L, Cheng Z Z, Guo D S, Niu Y and Fang X Z. 2023a. Few-shot class-incremental learning via class-aware bilateral distillation//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 11838-11847 [DOI: 10.1109/CVPR52729.2023.01139]

Zhao Z, Zhang Z Z, Tan X, Liu J, Qu Y Y, Xie Y and Ma L Z. 2023b. Rethinking gradient projection continual learning: stability/plasticity feature space decoupling//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 3718-3727 [DOI: 10.1109/CVPR52729.2023.00362]

Zheng J H, Ma Q L, Liu Z, Wu B Q and Feng H W. 2024. Beyond anti-forgetting: multimodal continual instruction tuning with positive forward transfer [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2401.09181.pdf

Zheng Z W, Ma M Y, Wang K, Qin Z H, Yue X Y and You Y. 2023. Preventing zero-shot transfer degradation in continual learning of vision-language models//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 19068-19079 [DOI: 10.1109/ICCV51070.2023.01752]

Zhou D W, Wang Q W, Qi Z H, Ye H J, Zhan D C and Liu Z W. 2023. Class-incremental learning: a survey [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2302.03648.pdf

Zhu H G, Wei Y C, Liang X D, Zhang C J and Zhao Y. 2023. CTP: towards vision-language continual pretraining via compatible momentum contrast and topology preservation [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2308.07146.pdf

Zuo Y K, Yao H T, Zhuang L S and Xu C S. 2024. Hierarchical augmentation and distillation for class incremental audio-visual video recognition [EB/OL]. [2024-09-15]. https://arxiv.org/pdf/2401.06287.pdf

作者简介

付浩，男，博士研究生，主要研究方向为多模态学习和增量学习。E-mail: haof.pizazz@zju.edu.cn

赵涵斌，通信作者，男，助理研究员，主要研究方向为计算机视觉、增量学习和生成模型。E-mail: zhaohanbin@zju.edu.cn

冯前，男，博士研究生，主要研究方向为增量学习和计算机视觉。E-mail: fqzju@zju.edu.cn

涂嘉航，男，博士研究生，主要研究方向为生成模型和计算机视觉。E-mail: tujiahang@zju.edu.cn

张超，男，助理研究员，主要研究方向为机器学习。E-mail: zczju@zju.edu.cn

杜歆，男，副教授，主要研究方向为信息电子。E-mail: duxin@zju.edu.cn

钱徽，男，教授，主要研究方向为人工智能。E-mail: qianhui@zju.edu.cn