

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2025)08-2633-12

论文引用格式: Zhou Y F, Du K L, Lyu F, Hu F Y and Liu G C. 2025. Class activation map replay and minimum entropy sampling-based multilabel class-incremental learning. *Journal of Image and Graphics*, 30(8):2633-2644(周怡凡, 杜凯乐, 吕凡, 胡伏原, 刘光灿. 2025. 类激活图回放和最小熵采样的多标签类增量学习. *中国图象图形学报*, 30(8):2633-2644)[DOI:10.11834/jig.240643]

类激活图回放和最小熵采样的多标签类增量学习

周怡凡¹, 杜凯乐¹, 吕凡², 胡伏原³, 刘光灿^{1*}

1. 东南大学自动化学院, 南京 210096; 2. 中国科学院自动化研究所, 北京 100190;

3. 苏州科技大学电子与信息工程学院, 苏州 215009

摘要: 目的 多标签类增量学习旨在学习多标签分类任务时,能够持续学习新任务中的新标签信息,同时保持对之前学习任务的良好性能。不同于单标签连续学习中的标签互斥性假设,多标签连续学习中的跨任务标签无法被补全。因此,样本的大量标签缺失使得跨任务之间的关联被完全阻断,模型无法区别新类与旧类,导致相似类别的混淆。**方法** 为了解决类别混淆问题,提出一种类激活图回放和最小熵采样的多标签类增量学习,以实现跨任务的监督信息传递。具体地,首先在传统经验回放策略的基础上,开辟一块新的存储区用于存储样本正类的类激活图。其次,基于交叉熵的数值对当前任务样本进行采样,以挑选那些拥有精确类激活图的样本放入存储区作为回放样本。最后,在后续任务中回放样本,对当前模型的正类激活图输出进行正向监督,并将存储的正类的激活图作为其他类别激活图的反向监督,使其显著区域不重合。**结果** 在PASCAL VOC(pattern analysis, statistical modeling and computational learning visual object classes)数据集上进行广泛实验,验证了所提出方法的有效性和鲁棒性。对比实验结果表明,类激活图监督回放方法显著提升了模型在多标签连续学习中的性能。**结论** 本文所提出的类激活图监督回放法为跨任务类别补充了监督信息,缓解了多标签类增量学习中的标签缺失问题。

关键词: 类增量学习(CIL);多标签分类;多标签类增量学习(MLCIL);类激活图;最小化熵

Class activation map replay and minimum entropy sampling-based multilabel class-incremental learning

Zhou Yifan¹, Du Kaile¹, Lyu Fan², Hu Fuyuan³, Liu Guangcan^{1*}

1. School of Automation, Southeast University, Nanjing 210096, China;

2. Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China;

3. School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China

Abstract: Objective Multilabel class-incremental learning seeks to enable models to learn new label information continuously while ensuring robust performance on previously acquired tasks within the context of multilabel classification. This task diverges from the label exclusivity assumption inherent in single-label problems because multilabel classification is effectively modeled as multiple binary classification tasks. In this framework, each sample can be associated with multiple labels simultaneously, reflecting the complexity of real-world scenarios. Moreover, the task setting in class-incremental learning imposes major constraints on how labels are assigned. Specifically, samples can only receive positive labels corre-

收稿日期:2024-11-05;修回日期:2025-03-25;预印本日期:2025-04-02

* 通信作者: 刘光灿 guangcanliu@seu.edu.cn

基金项目: 国家自然科学基金项目(U21B2027)

Supported by: National Natural Science Foundation of China(U21B2027)

sponding to their respective tasks, while the labels for categories pertaining to other tasks remain unknown. Consequently, this limitation results in a substantial loss of negative samples across categories because many potential negative examples are not represented in the available data. The implications of this scenario are profound. Not only do the various classes suffer from a considerable deficit of negative samples, but the learning process is further complicated by the emergence of numerous unknown mappings between classes. This lack of clarity can lead to severe label confusion, wherein the model struggles to distinguish between similar categories. As a result, the effectiveness of multilabel class-incremental learning is severely hindered, necessitating the development of novel methods to address these challenges.

Method To address the issue of label confusion, this study introduces a novel bidirectional class activation mapping replay method, designed to facilitate the effective transfer of supervisory information across tasks. The primary innovation of our model lies in the establishment of a new storage area integrated within the traditional replay strategy. This dedicated storage area is specifically allocated to retain the class activation maps of the positive samples, thereby preserving critical information about the features associated with each class. Class activation mapping is a simple yet effective weakly supervised method for object localization that is capable of generating pixel-level masks for foreground objects. In an ideal scenario, an accurate class activation map can effectively represent the precise location of the target object within the image. Furthermore, leveraging the inherent spatial exclusivity of different objects within a single image enables the realization of cross-task information transfer. In the next step, we implement a cross-entropy-based sampling method aimed at selecting samples with highly accurate class activation maps. This careful selection process ensures that only the most representative samples are stored as replay samples, thereby enhancing the quality of the supervisory information available for future tasks. During the execution of subsequent tasks, we utilize these stored samples in a replay mechanism. This replay provides not only standard supervision for the positive class activation map output via the current network but also reverse supervision for the other class. Such reverse supervision ensures that the prominent regions of the class activation maps for the other classes do not overlap with the positive class. Reverse supervision for potentially confusing categories does not interfere with the classification of other positive classes across different tasks. Clearly, different objects occupy distinct regions within an image. If we assume that the localization of the class activation maps is sufficiently accurate, then the prominent regions of the class activation maps related to one class will not overlap with those belonging to other classes, as reflected in our reverse supervision loss function, which would yield a very low loss value. Consequently, the method we designed can correct the localization and output of the network for the confusing categories without disrupting the ability to learn the positive labels and features.

Result This study selects a range of widely recognized and representative class-incremental comparison methods based on recent advancements in the field, including classic regularization-based methods, replay-based methods, and the top-performing replay and regularization methods from the single-label class-incremental learning, along with methods specifically designed for multilabel class-incremental learning. In simple task settings, our model, while slightly lagging behind the state-of-the-art in terms of mean average precision, demonstrates remarkable advantages in terms of the F1 score. Importantly, as the number of tasks increases, our method exhibits substantial advantages. Furthermore, our experiments demonstrate that our model effectively reduces the false positive rate in network predictions. Particularly in more challenging task settings, our model shows a considerable advantage, with a nearly 30% reduction in false positive rate compared with other methods. This result indicates that our model considerably alleviates the confusion associated with multilabel class-incremental learning. Finally, we provide extensive ablation experiments to validate the effectiveness of each module within our model.

Conclusion This study posits that the label confusion caused by label absence in multilabel class-incremental learning is as crucial as the problem of forgetting. Class activation maps reveal the prominent regions that the network focuses on, ideally corresponding to the locations of labeled objects. Given the practical exclusivity of object positions, we propose a bidirectional class activation mapping replay method that stores the class activation maps of the positive samples during the class-incremental learning process. This method leverages these stored class activation maps as supervisory information in future tasks. Moreover, a class activation map exclusivity loss function is designed to implement the reverse supervision, thereby distinguishing the new categories across tasks from the prominent regions of stored positive classes, thereby alleviating confusion. Furthermore, we introduce a cross-entropy-based sampling method to select samples with reliable class activation maps. Experimental results demonstrate that the proposed bidirectional class activa-

tion mapping replay method remarkably enhances the model accuracy in the context of multilabel class-incremental learning.

Key words: class incremental learning (CIL); multi-label classification; multi-label class-incremental learning (MLCIL); class activation mapping; minimizing entropy

0 引言

类增量学习(class-incremental learning, CIL)(Yan等, 2021; Douillard等, 2022; Buzzega等, 2020)旨在使模型在不断学习新类别的同时保持旧类别的分类能力。现有的类增量学习主要关注单标签任务情境,然而在实际应用中,不同类别可能同时出现,例如一幅图像可能包含多个对象(如,牧羊犬在牧羊)。对于实际应用而言,多标签类增量学习(multi-label class-incremental learning, MLCIL)(Dong等, 2023; Du等, 2024)是一个更具挑战性但更符合现实场景的任务。

单标签类增量学习存在标签互斥假设(Boschini等, 2023),可认为新类样本不包括旧类标签,如图1所示。这种负标签补全为旧类补充了负样本,使其可以作为模型区分新旧类的重要信息。例如,“狼”与“狗”的特征十分相似,新类“狼”的出现会加重类别混淆。但因为单标签的互斥假设,“狼”样本关于“狗”的标签被补全为无,让网络拥有了知道“狼不是狗”的能力。因此,跨任务样本之间具有潜在信息补充,只需关注如何消除灾难性遗忘(McCloskey和Cohen, 1989)。

然而,多标签类增量学习中的标签不具有标签互斥假设,带来了严重的标签混淆问题。具体而言,不同于单标签问题中的标签互斥假设,多标签分类可被表示为多个二分类问题。与此同时,类增量学习的任务设定要求样本只能得到其所属任务下类别的真实标签,而对于其他任务的类别,样本的真实值未知。因此,多标签类增量学习缺乏标签补全能力,不仅各类别损失了负样本,还使学习过程中出现大量未知映射关系,产生了标签混淆问题。在多标签的问题设定下回顾“狼”与“狗”的例子:新类“狼”的样本关于跨任务旧类“狗”的标签完全缺失,即网络无法得知“狼不是狗”。因此,对于“狼”的样本,网络将有可能在“狗”的分类头上输出过高置信度。该现象称为多标签连续学习下的潜在混淆。这种混淆现

象表现为网络输出的假阳性(false positive, FP),且随着网络学习的类别不断增加,这种错误积累会越来越严重(Du等, 2024)。

虽然多标签问题中不存在标签互斥假设,但一幅图像中不同物体所在像素是互斥的。如果能得到各物体的位置信息,利用这种像素互斥性就可以实现跨任务的信息补充。类激活图(class activation mapping, CAM)(Zhou等, 2016)是一种简单而有效的弱监督对象定位方法,该方法能够可视化模型关注的显著区域。如图1所示,对于学习充分、分类准确的类别,类激活图的定位即能代表该类别物体所处的像素区域。

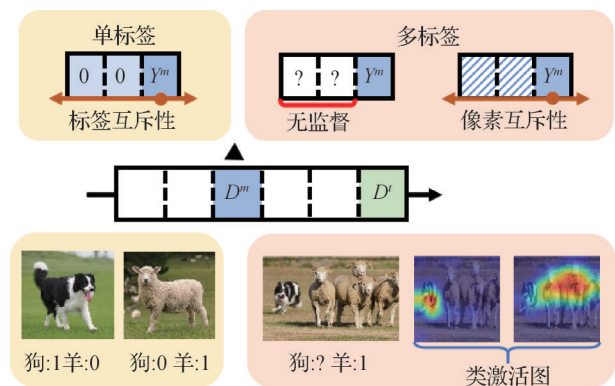


图1 类增量学习在单标签与多标签问题设定下的对比

Fig. 1 Comparison of single-label class-incremental learning and multi-label class-incremental learning

为了解决标签缺失带来的潜在混淆,本文提出一种类激活图回放和最小化熵采样的多标签类增量学习方法。该方法由类激活图回放和最小化熵采样两个模块构成。具体而言,对于类激活图回放模块,提出在记忆存储区中开辟新的区域存储正类的类激活图,从而保存网络的注意力信息。然后在连续学习的过程中,利用存储的信息设计了3种类型的监督来优化网络,以缓解灾难性遗忘和潜在混淆:1)遵循传统经验回放策略进行标签监督,通过校正网络在连续任务流中对过去任务的预测值防止遗忘;2)利用存储的正类激活图,对当前模型的正类激活图输出进行正向监督,通过对网络推理过程中关注

区域的约束来防止遗忘;3)取出潜在混淆类别,使用存储的类激活图对这些类别进行反向监督,使潜在混淆类别的显著区域不与正类的显著区域重合。这实现了跨任务的信息传递,补充了各类别的负样本信息。此外,还设计了一个最小化熵采样模块,该模块根据网络对于训练集样本的预测表现,排除不可靠的噪声样本,从而选择更具有特征代表性的样本进行后续的回放,进一步保证了存储的类激活图的可信性。

本文的贡献主要有:1)提出类激活图回放模块,设计了类激活图正反向监督损失函数。通过引入双向监督机制,减轻遗忘的同时确保新类别的显著区域不会与旧任务中的正类显著区域发生重叠,补充了负样本的缺失。2)为了确保存储的样本和类激活图的代表性和可信性,引入基于最小化熵的样本选择模块。通过衡量样本的特征代表性,优先选择那些在当前任务中对类区分有贡献的样本进行存储和回放。3)进行广泛实验,验证所提出方法的有效性和鲁棒性。实验结果表明本文方法显著提升了模型在多标签类增量学习中的性能。

1 相关工作

1.1 单标签类增量学习

1)基于正则化的方法。这类方法通过在损失函数中引入正则项,使网络更新的参数保留旧的知识。如,保留历史任务中的重要参数(Kirkpatrick等,2017;刘展阳和刘进锋,2024)。还有一些方法,例如LwF(learning without forgetting)(Li和Hoiem,2017)和AGCN(augmented graph convolutional network)(Du等,2024)采用知识蒸馏技术缓解遗忘问题,即通过软标签引导模型在学习新任务时保持对旧任务的记忆。

2)基于回放的方法。这类方法通过存储一组旧任务的示例来代表旧数据,并在训练当前任务时与新数据一起使用。早期的回放方法,如ER(experience replay)(Rolnick等,2019)通过构建一个内存缓冲区保存旧任务的样本,并与新数据一起进行训练。在此基础上,DER++(dark experience replay++) (Buzzega等,2020)引入了对缓冲区中存储数据的知识蒸馏惩罚,从而进一步减少遗忘。此外,iCaRL(incremental classifier and representation learning)

(Rebuffi等,2017)通过使用herding技术选择示例并设计不同的蒸馏损失用于防止遗忘。

3)基于架构的方法。这些方法为每个任务提供独立的参数,从而防止遗忘。大多数基于架构的方法需要额外的任务标签,并且通常仅限于任务增量场景。Yan等人(2021)和Douillard等人(2022)将基于架构的方法与排练方法相结合,达到了更好的防遗忘效果。但这些方法需要动态扩展网络参数以适应新数据,可扩展性有限。

1.2 多标签类增量学习

1)多标签分类。近年来,多标签分类引起了广泛关注(张晶等,2020;邢素霞等,2023)。早期的方法使用循环神经网络(recurrent neural network, RNN)(Wang等,2016)或长短期记忆网络(long short-term memory network, LSTM)(Wang等,2017)来依次建模标签与语义之间的关系,而当前最先进的方法则通过空间信息或图结构来增强性能。例如,Chen等人(2022b)利用先验知识指导类别间的自适应信息传播。Chen等人(2023)提出基于图卷积网络的多标签分类模型。Zhao等人(2021)通过构建结构和语义相关图探索多标签分类。然而,标签关系图的建立依赖于图像完整的标签标注,而在多标签类增量的任务中,严重的标签缺失限制了这些方法的应用。

2)多标签增量学习。随着单标签类增量学习和多标签分类的不断研究,多标签增量学习已成为一个新的领域。Kim等人(2020)提出一个新的采样策略,用于缓解多标签数据集中长尾分布对基于回放方法的影响。KRT(knowledge restore and transfer)(Dong等,2023)提出一种知识恢复和迁移框架,缓解了标签缺失的问题。AGCN(Du等,2024)利用伪标签构建固定的统计标签关系,并缓解灾难性遗忘。然而,这些方法都没有意识到多标签类增量学习中潜在混淆的影响。

1.3 类激活图

类激活图(CAM)(Zhou等,2016)是一种简单但有效的弱监督目标定位方法,能够生成前景对象的像素级掩码。为了改进CAM的性能,一系列高级CAM变体方法应运而生,如Grad-CAM(Selvaraju等,2017)、ReCAM(Chen等,2022a)和AdvCAM(Lee等,2021)。这些变体通过结合梯度或其他信息,提高了定位的准确性和模型的可解释性。

在非连续学习的研究领域, Grad-CAM已被证明可以编码信息以学习新类, 如弱监督语义分割方法大多采用类激活图定位物体(白雪飞等, 2024)。近年来, 有部分研究采用类激活图指导连续学习。例如, Luo等人(2023)利用类激活图来减少内存消耗并取得了有潜力的结果, 但它主要侧重于数据压缩。与本文最相似的方法是 RRR (remembering for the right reasons) (Ebrahimi等, 2021), 其也将类激活图存储在内存中, 并在回放过程中使用来防止遗忘。然而, 其没有认识到类激活图的负类补充作用。本文方法首次尝试将回放类激活图扩展到多标签连续学习领域, 证明了其在减轻假阳性预测和混淆问题方面的有效性。

2 理论基础

2.1 问题描述

假设共有 T 个多标签图像分类任务, 表示为 $D^1, D^2, \dots, D^T$ 。每个任务的数据集包含图像 $\mathbf{x}$ 和对应任务的类别标签 $\mathbf{y}$, 表示为 $D^t = \{\mathbf{x}_i^t, \mathbf{y}_i^t\}$ 。式中, 类别标签 $\mathbf{y}_i^t$ 表示对应于第 i 幅图像第 t 个任务的真实标签, 且可能包含不止一个类别, 即满足 $\mathbf{y}_i^t \in \mathbf{K}^t$ 且 $1 \leq |\mathbf{y}_i^t| \leq |\mathbf{K}^t|$ 。属于不同任务的标签空间 $\mathbf{K}$ 彼此不重叠, 即 $\forall m, n (m \neq n), \mathbf{K}^m \cap \mathbf{K}^n = \emptyset$ 。根据连续学习的任务要求, 每个任务的数据集划分为训练集和测试集 $D^t = \{\mathbf{X}^t, \mathbf{Z}^t\}$, 在第 t 个任务的训练阶段, 只有当前任务的训练集数据 $\mathbf{X}^t$ 和一小部分存储在记忆区 $\mathbf{M}$ 中的过去任务样本可用。在当前任务学习完毕后, 模型被允许取出有限数量的样本存储在 $\mathbf{M}$ 中, 以便后续任务中进行回放。在测试阶段, 网络需要在测试集的组合上进行评估, 即 $\mathbf{Z}^{1-t} = \mathbf{Z}^1 \cup \mathbf{Z}^2 \cup \dots \cup \mathbf{Z}^t$, 并期望识别出学过的所有类。

2.2 类激活图生成

本文使用梯度加权类激活映射 (Grad-CAM) 来完成类激活图生成任务。该方法不依赖网络结构的设计而且是可微分的, 便于进行端到端的网络训练。

Grad-CAM 是一种白盒可解释性技术, 其使用梯度确定特征图激活对预测的影响。由于已知卷积神经网络中的后续层对更高级别的语义进行编码, 因此根据模型输出相对于这些特征图激活的梯度可以

发现哪些高级语义对预测很重要。当使用 Grad-CAM 可视化图像分类时, 对最后一个卷积层采用目标类别预测的梯度进行加权求和, 会突出显示图像中的目标类别区分区域 (例如在识别鸟类时鸟类的翅膀)。

假设已知图像分类输出中某个类别 c 的预测分数为 y_c , 给定一个具有 K 个特征图的卷积层, 其中单个特征图记为 $\mathbf{A}_k \in \mathbf{R}^{u \times v}$ 。Grad-CAM 对每个特征图 $\mathbf{A}_k$ 取 y_c 的导数并对宽和高维度进行全局平均池化, 得到权重 $a_k = \frac{\partial y_c}{\partial \mathbf{A}_k}$, 表明特征图 $\mathbf{A}_k$ 对预测 y_c 的重要性。接下来, 将这些重要性作为权重计算特征图的线性组合, 最后经过 ReLU (rectified linear unit) 激活函数得到网络对于类别 c 的激活图 $\mathbf{Q}_c$, 可表示为

$$\mathbf{Q}_c = f_{\text{ReLU}}(\boldsymbol{\alpha}^T \mathbf{A}) \quad (1)$$

式中, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_K]$, $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_K]$ 。

3 方法

本文方法的总体框架如图 2(a) 所示, 在每个任务的学习阶段, 网络通过当前任务的训练数据学习新的类别, 同时使用记忆存储区 $\mathbf{M}$ 中保存的图像和知识, 使用类激活图回放法对网络进行防遗忘和防混淆监督。当前任务的图像数据训练完毕后, 使用最小化熵采样策略选择最优的样本并生成对应类别的类激活图存储到 $\mathbf{M}$ 中。3.1 节和 3.2 节分别对类激活图回放法和最小化熵采样策略进行了详细的介绍。

3.1 类激活图回放

如图 2(b) 所示, 以存储第 m 个任务为例。与传统的经验回放只存储样本 $\mathbf{x}$ 和标签 $\mathbf{y}$ 不同, 本文将记忆区 $\mathbf{M}$ 划分为 $\mathbf{M}_x$, $\mathbf{M}_y$ 和 $\mathbf{M}_{\text{cam}}$ 。其中, $\mathbf{M}_x$ 存储需要回放的样本 $\mathbf{x}$, $\mathbf{M}_y$ 存储各样本 $\mathbf{x}$, $\mathbf{x} \in \mathbf{M}_x$ 在任务 m 下各类别的标签 $\mathbf{y}^m$, $\mathbf{M}_{\text{cam}}$ 存储各样本 $\mathbf{x}$, $\mathbf{x} \in \mathbf{M}_x$ 的正类 k_+ , $\mathbf{y}_{k_+}^m = 1$ 对应的类激活图 $\mathbf{Q}_{k_+}^m$ 。 $\mathbf{M}_y$ 中存储的标签用于回放时监督样本在其任务下类别的分类结果。 $\mathbf{M}_{\text{cam}}$ 中存储的正类对应的激活图用于监督所有类别的激活图。

在任务 $t (t > m)$ 阶段, 从记忆区取出在第 m 个任务存入的样本 $\mathbf{x}$, $\mathbf{x} \in \mathbf{M}_x$ 输入到当前任务 t 的网络中, 得到各任务下类别 $\mathbf{K}^{1-t} = \{k^1, k^2, \dots, k^t\}$ 的预测

结果 $\hat{Y}^{1-t} = \{\hat{y}^1, \hat{y}^2, \dots, \hat{y}^t\}$ 。此外,根据之前介绍过的类激活图计算方法可以得到关于各任务的类激活图 $\hat{Q}^{1-t} = \{\hat{Q}^1, \hat{Q}^2, \dots, \hat{Q}^t\}$ 。

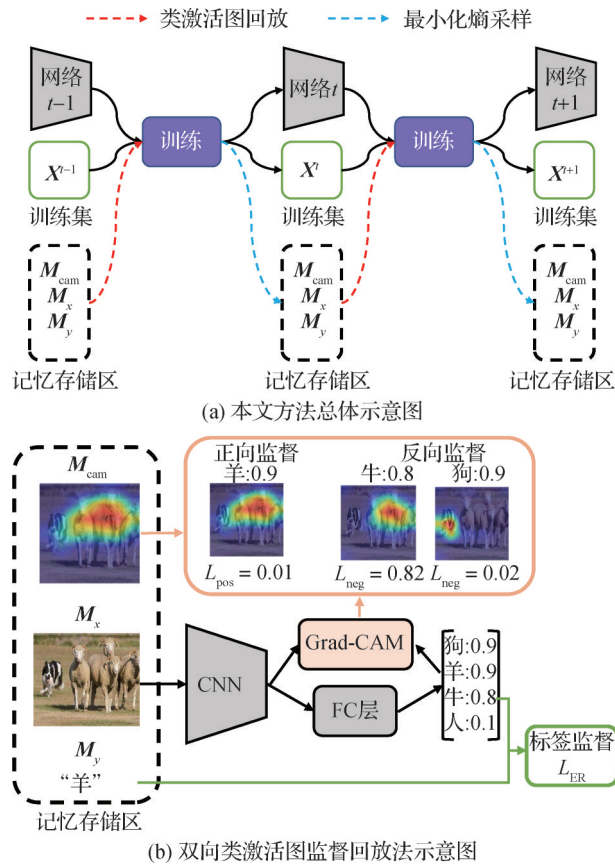


图2 本文方法总体示意图及双向类激活图监督回放法示意图

Fig. 2 Schematic diagram of the overall method and the bidirectional class activation mapping replay method ((a) the overall method; (b) the bidirectional class activation mapping replay method)

对于任务 m 下类别 k^m , 本文利用 M_y 和 M_{cam} 存储的信息设计了3种类型的监督来优化网络以减轻灾难性遗忘和潜在混淆。

1) 标签监督。利用在 M_y 中存储的标签 y^m 监督预测值 $\hat{y}^m$, 得到对于任务 m 下类别 k^m 的分类损失, 具体为

$$L_{ER} = \frac{1}{N} \sum_{i=1}^N (y_i^m \log \hat{y}_i^m + (1 - y_i^m) \log (1 - \hat{y}_i^m)) \quad (2)$$

式中, N 表示样本总数。

该损失函数与传统的经验回放方法类似, 区别在于使用了适用于多标签分类的二元交叉熵损失。基于 M_y 的回放本质上是利用存储的真实标签, 校正网络在连续任务流中对于过去任务的预测值以防止

遗忘。

2) 正向监督。利用在 M_{cam} 中存储的正类激活图 $Q_{k_i}^m$ 监督当前网络对于这些类的激活图 $\hat{Q}_{k_i}^m$, 具体为

$$L_{pos} = \mathbb{E}_{(x, y_{k_i}^m, Q_{k_i}^m) \sim (M_x, M_y, M_{cam})} \|\hat{Q}_{k_i}^m - Q_{k_i}^m\|_1 \quad (3)$$

式中, $\mathbb{E}$ 表示期望损失函数采用 L_1 范数。基于 M_{cam} 的正向监督利用存储的正类激活图, 校正网络在连续任务流中对于这些类的关注区域, 减少遗忘带来的偏移。换句话说, 正向监督通过对网络推理过程中关注区域的约束来防止遗忘。

3) 反向监督。对于 k_+ 之外的类别 K^{1-t}/k_+ , 本文设计了一个反向激活图监督策略, 减轻多标签类增量学习中负类缺失带来的混淆。首先, 本文设置一个阈值 θ 来选取潜在混淆类别。具体地, 当前网络预测值大于阈值 θ 的类别被视为潜在混淆类别, 即网络偏向于将其视做当前样本的正类 k_θ 。值得注意的是, 由于本文要解决的任务是多标签分类, 网络输出的正类 k_θ 允许有多个, 如图2(b)所示的“狗”和“牛”。然后, 使用 M_{cam} 中正类对应的激活图对当前网络关于潜在混淆类的激活图进行反向监督, 具体为

$$L_{neg} = \mathbb{E}_{(x, y_{k_i}^m, Q_{k_i}^m) \sim (M_x, M_y, M_{cam})} \sum_{k_\theta} (\exp(Q_{k_i}^m \cdot \hat{Q}_{k_\theta}^m) - 1) \quad (4)$$

利用图像中的像素互斥性, 本文使用点乘计算两个 $u \times v$ 的特征图之间的相似程度。通过在点积结果上应用 $\exp(\cdot) - 1$, 放大了相似度高的特征, 使模型更为敏感, 同时保证了损失函数的最低值为0, 即当特征图完全不相似时损失最小。该损失函数从特征层面惩罚了潜在混淆类别与已知的正类所对应的激活图重合区域, 从而约束了网络对于类别 K^{1-t}/k_+ 的关注区域。基于 M_{cam} 的反向监督利用在过去任务中充分训练得到的正类激活图, 限制其他类别的关注区域, 减轻了负类缺失带来的特征混淆现象。

值得注意的是, 对于潜在混淆类别的反向监督不会干扰不同任务下的其他正类的分类。显然, 不同物体在图像上所处的区域不同, 如图2(b)中“狗”和“羊”。如果“羊”的定位足够准确, 那么关于它的类激活图的显著区域不与“狗”的重叠, 表现在反向监督损失函数上会是一个很低的损失值。因此, 本文方法可以在不干扰网络学习正确标签和特征的基础上, 纠正网络对于假阳性混淆类别的定位和输出,

从而起到了任务之间的补充作用。

模型最终的损失函数设计为

$$L = L_{\text{task}} + \lambda(L_{\text{ER}} + L_{\text{pos}} + L_{\text{neg}}) \quad (5)$$

式中, L_{task} 指常规的针对当前任务训练样本的交叉熵分类损失, λ 作为超参数调整类激活图回放所占的比重。超参数的取值在本文消融实验部分做了详细的讨论。

3.2 最小化熵采样

本文将在 M_{cam} 中存储的类激活图作为监督信息, 作用于后续的所有任务中。因此在理想情况下, 所存储的类激活图不仅应该能表示网络在样本图像上对于目标类别的关注位置, 还应该相对精准才能作为监督信息存储网络的知识。然而, 由于训练集中的图像可能含有噪声, 网络并不能对所有训练集图像进行准确的类别预测和定位。因此, 本文提出最小化熵样本选择策略, 该模块根据网络对于训练集样本的输出结果, 排除不可靠的噪声样本, 从而最大程度上保证了类激活图的准确性和可靠性。具体而言, 本文定义了一个样本选择指标, 计算为

$$S(X^t) = \underset{x_i^t \in X^t}{\text{minimize}} \left(y_i^t \log \hat{y}_i^t + (1 - y_i^t) \log (1 - \hat{y}_i^t) \right) \quad (6)$$

在第 t 个任务训练完毕后, 该模块对于当前任务的每个类, 采样训练集 X^t 中损失最小的前 n 个样本加入到记忆区 M 中。该模块以二元交叉熵损失最小为评价标准, 保证了类激活图的准确度, 在本文消融实验中证明了这一点。

总的来说, 算法流程如下:

输入: 依次给定当前任务 t ($t = 1, 2, \dots, T$) 的训练集 X^t , 包含当前所有任务的测试集 $Z^{1-t} = Z^1 \cup Z^2 \cup \dots \cup Z^t$, 给定记忆存储区 M 。

输出: 模型在测试集上的评价指标。

- 1) 计算当前任务 t 训练样本 X^t 的分类损失 L_{task} 。
- 2) 根据式(2)—式(4)计算记忆区中样本的类激活图回放损失。
- 3) 根据式(5)计算总损失函数并反向传播更新网络参数。
- 4) 根据式(6)选择样本, 并将其与对应的类激活图和标签存储到记忆存储区 M 。
- 5) 计算模型在测试集 Z^{1-t} 上的测试指标。
- 6) 重复步骤1)—5), 直到 $t = T$ 。

4 实验

4.1 实验设定

本文使用 PASCAL VOC (pattern analysis, statistical modeling and computational learning visual object classes) (Everingham 等, 2010) 数据集评估方法的有效性。PASCAL VOC 数据集共包含 20 个物体类别的 9 963 幅图像, 其中, 5 011 幅图像用于训练, 4 952 幅图像用于测试。平均每幅图像的标签数为 2.4。遵循现有的连续学习工作 (Yan 等, 2021; Luo 等, 2023), 使用两种任务设定: 从零开始学习 (learning from start, LFS) 和从一半学习 (learning from half, LFH)。在 LFS 中, 模型在所有 N 个阶段观察相同数量的类, 其中 N 为 5 或 10。在 LFH 中, 模型第 1 阶段在前一半类上训练。然后, 模型在随后的 N 个阶段中均匀地学习剩余的类。

对于模型学习表现的评估, 采用 3 个指标, 即: 类平均 F1 测度 (CF1)、总体 F1 测度 (OF1) 和平均精度均值 (mean average precision, mAP)。这 3 个指标在多标签分类领域和多标签类增量学习领域中得到广泛应用。本文采用最后一个任务训练完毕后, 模型在所有类别图像上的测试结果的指标作为方法的评价指标。

遵循现有的多标签类增量学习的研究 (Dong 等, 2023), 采用 ImageNet-21k 预训练的 TRResNetM (Ridnik 等, 2021) 作为骨干网络 (所有比较的方法都使用相同的骨干网络), 使用 Adam 优化器和权重衰减为 $1\text{E-}4$ 的 One CycleLR 调度器对模型进行 20 个 epoch 的训练, 批次大小设置为 64, 学习率设置为 $4\text{E-}5$ 。在对比实验部分, 所有方法统一使用每个类固定数目的内存设置, 遵循现有研究方法的惯例, 每个类存储 2 个训练样本, 即 $n = 2$ 。此外, 每个实验进行 3 次, 并报告平均结果。

4.2 对比实验

4.2.1 准确率比较

根据最近的几项类增量学习工作选择广泛认可且有代表性的类增量对比方法, 其中 FT (fine-tuning) 表示没有采用任何连续学习技术的微调方法, 作为所有对比方法的下界。LwF (Li 和 Hoiem, 2018) 是基于正则化的代表性工作。iCaRL (Rebuffi 等, 2017) 和 ER (Rolnick 等, 2019) 是经典的基于回

放的方法。DER++ (dark experience replay++) (Buzzega 等, 2020)和PODNet(pooled outputs distillation network)(Douillard 等, 2020)分别是在单标签类增量领域性能最佳的基于回放和基于正则化的方法。AGCN(Du 等, 2024)和KRT(Dong 等, 2023)是专门针对多标签类增量设计的方法。表1展示了不同方法在4种任务设定上的实验结果。可以得到以下观察结果:在较为简单的任务中,如LFS任务设定中 $N=5$ 的情况下,本文方法尽管在mAP指标上落后于KRT方法,但在F1指标上有较大的优

势。值得注意的是,随着任务难度的不断提升,即任务数量 N 增大,本文方法展现出了强大的优势。与此同时,基于回放的方法,如ER、DER++逐渐超过基于正则化的LwF、KRT等方法。这表明,随着任务数量的增加,每个任务训练的类别减少,类别之间的混淆愈加明显。而基于正则化的方法尽管在较为简单的任务上能够更好地保留网络的能力,克服灾难性遗忘,但是在困难任务上会加重错误积累,从而加重混淆使表现变差,这也印证了本文之前的分析。

表1 不同连续学习算法在LFS和LFH两种设定下的对比结果
Table 1 Comparison results of different continuous learning algorithms under LFS and LFH settings

方法	LFS						LFH					
	$N=5$			$N=10$			$N=5$			$N=10$		
	CF1	OF1	mAP	CF1	OF1	mAP	CF1	OF1	mAP	CF1	OF1	mAP
FT	40.6	43.9	62.5	20.6	23.9	27.7	19.1	29.8	45.8	5.9	12.6	13.1
LwF	53.1	53.8	71.1	32.2	32.4	38.9	39.7	37.9	57.5	7.6	15.5	24.8
iCaRL	52.3	51.9	67.0	27.5	29.6	41.6	38.6	36.4	61.5	8.1	15.3	31.8
ER	47.2	46.3	71.5	29.9	30.3	44.7	44.7	41.6	68.6	8.6	14.9	34.1
PODNet	52.0	50.3	76.6	30.2	31.2	40.2	49.3	50.8	59.1	10.7	18.6	31.9
DER++	59.8	56.4	74.4	<u>37.6</u>	<u>38.8</u>	<u>50.4</u>	<u>52.8</u>	44.2	64.4	<u>38.9</u>	<u>25.8</u>	<u>48.9</u>
AGCN	58.8	51.2	73.4	34.1	34.5	40.9	44.6	43.8	65.1	12.3	23.2	44.6
KRT	<u>62.1</u>	<u>59.6</u>	84.2	24.1	30.3	34.0	36.4	<u>44.6</u>	<u>68.9</u>	3.7	25.6	41.9
本文	69.3	71.1	<u>78.2</u>	47.8	41.0	60.1	60.5	59.8	74.1	47.8	39.9	57.8

注:加粗、下划线字体表示各列最优、次优结果。

4.2.2 假阳性率

图3展示了本文提出的模型与对比方法在LFH两种任务设定上的假阳性率比较结果。图中每个柱体顶端的“工”字型符号表示误差棒,反映了多次重复实验中数据的不确定性,展示了每个数据点的标准误差范围。从图3可以看出,本文提出的模型可以有效降低网络预测的假阳性率,尤其是在 $N=10$ 的较为困难的任务设定上,模型有较大的优势,与其他的方法相比,假阳性率下降了近30%。这表明本文方法可以有效减轻由多标签类增量学习带来的混淆现象。

4.3 消融实验

4.3.1 模块消融

表2展示了在LFS($N=5$)的实验设定上的模块消融实验结果。由于本文是基于回放的方法,因此

采用传统的经验回放(ER)方法作为本文的基线方法。模型1和模型2分别是在基线方法上添加类激活图回放模块和最小化熵采样模块的结果。模型3是本文方法的最终结果。从表2可以看出,在基线的基础上添加两个模块均能带来较大的提升。其中,模型1的提升更大,说明类激活图回放模块可以通过避免混淆有效地提升模型分类准确度。而本文的最终模型3相比于模型1有了进一步提升,这说明最小化熵采样模块可以有效地选择更适合回放的样本,充分证明了本文方法的有效性。

4.3.2 采样策略

图4展示了采用不同采样策略所选择的样本对比。从图中可以看出,本文提出的样本采样策略,即最小化熵,与最大化熵相比有较大优势。如,在“飞

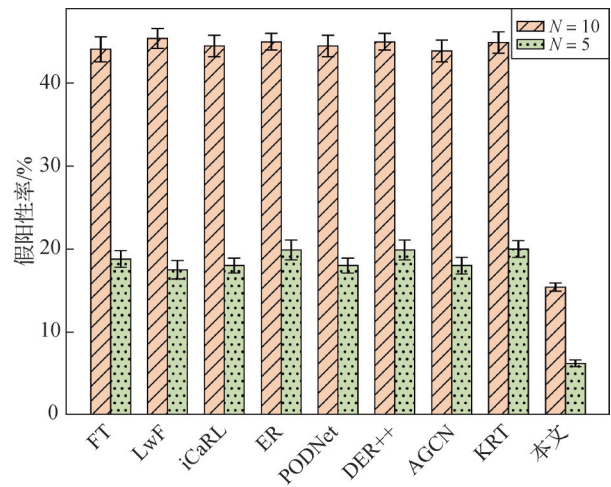


图 3 假阳性率对比
Fig. 3 Comparison of false positive rates

表 2 模块消融实验结果

Table 2 Module ablation experiment results

方法	CF1	OF1	mAP	平均 mAP
基线方法	47.2	46.3	71.5	85.7
模型 1	64.3	66.6	75.1	87.2
模型 2	59.3	55.9	73.3	86.8
模型 3	69.3	71.1	78.2	88.1

机”和“狗”这两个类别的例子中,本文的采样策略可以选择主体更为突出的样本,而最大化熵则会选择含有背景噪声的样本。在“自行车”这个类别的示例

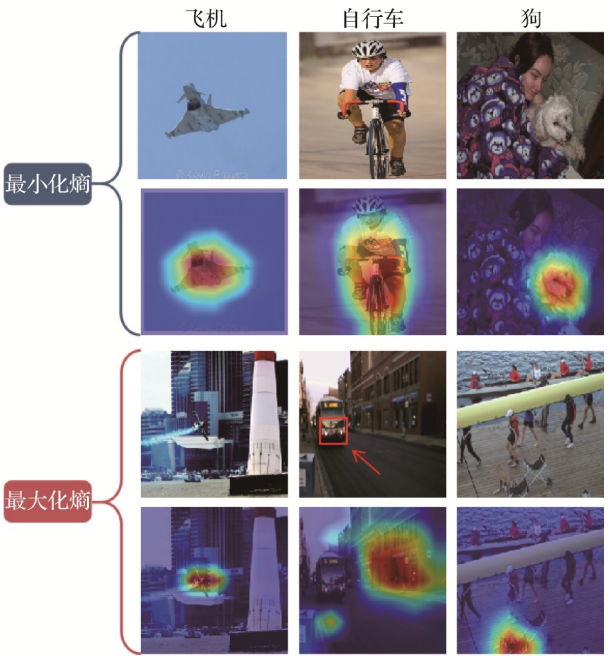


图 4 不同采样方法所选样本对比
Fig. 4 Comparison of samples selected by different methods

中,最大化熵策略不仅选择了噪声样本,且类激活图的定位并不准确,对网络学习产生负面效果。

表 3 展示了每个样本采样个数 n 的消融实验结果。从表 3 可以看出, n 的取值增大时,即每个类别保存的样本数目增多,基线方法(ER)和本文方法的准确率都有显著提高。与此同时,在 n 取值相同时,本文方法显著优于基线方法,这体现出了本文方法的鲁棒性和有效性。

表 3 参数 n 的消融实验结果

Table 3 Results of ablation experiment of parameter n

方法	$n = 2$				$n = 5$				$n = 10$			
	CF1	OF1	mAP	平均 mAP	CF1	OF1	mAP	平均 mAP	CF1	OF1	mAP	平均 mAP
基线	47.2	46.3	71.5	85.7	49.6	53.5	75.2	86.0	56.8	59.5	79.3	88.9
本文	69.3	71.1	78.2	88.1	70.6	72.0	80.8	89.6	71.2	72.3	85.5	90.6

4.3.3 超参数消融

为了验证本文方法的鲁棒性,超参数消融实验分别调整了超参数 λ 和 θ 。图 5 展示了模型在不同参数下的表现,任务设定与模块消融相同。从图中可以看出,本文方法在参数为 $\lambda = 10, \theta = 0.5$ 的时候表现最好。在不同的参数条件下,模型准确率变化不大,说明本文方法有较高的鲁棒性。

4.4 可视化实验

为了验证本文方法在解决多标签类增量学习中对潜在混淆的有效性,可视化实验结果如图 6 所示。图 6 中展示的是在所有类增量任务学习结束后进行测试的结果。可以明显看出,经过本文方法的修正之后,原本混淆的类别的类激活图定位不再与真实类别相近。与此同时,对于混淆类别的预测值减小,

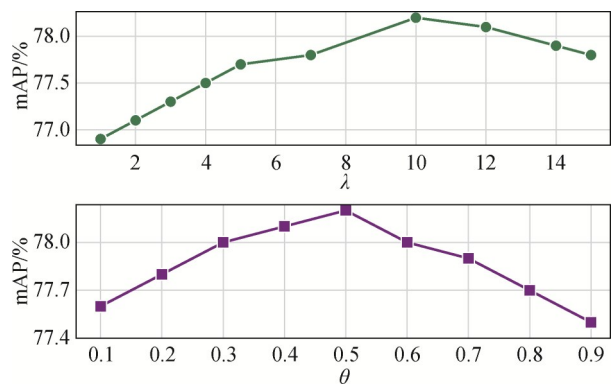


图5 超参数消融实验结果

Fig. 5 Results of hyper-parameter ablation experiment

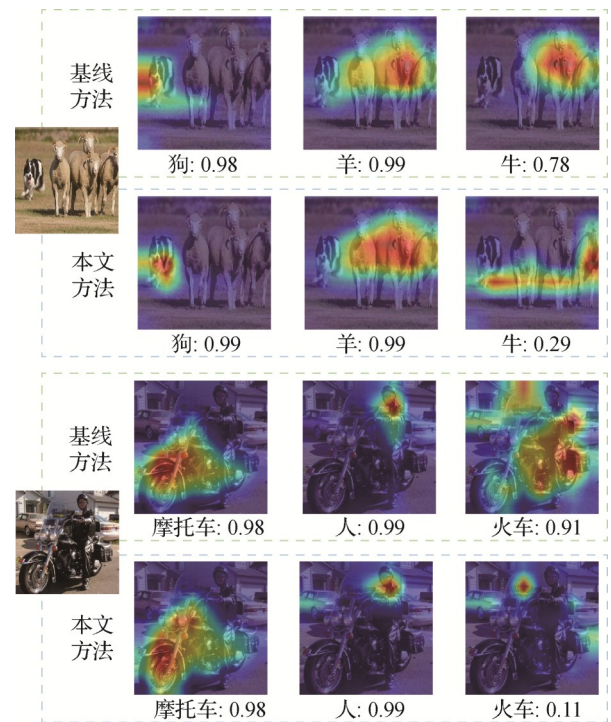


图6 可视化实验结果与分类预测值

Fig. 6 Visualization of experimental results and classification predictions

说明本文方法有效减轻了混淆现象带来的假阳性问题。

5 结 论

本文认为在多标签类增量学习中,标签缺失带来的混淆问题与遗忘问题同样显著。类激活图揭示了网络关注的显著区域,在理想情况下即标签物体所在区域。由于一幅图像中不同物体所在像素具有互斥性,本文提出在连续学习的过程中存储样本正类的类激活图,并在未来任务作为监督信息的类激

活图监督回放法:通过标签监督、正向监督和设计类激活图互斥损失函数实现反向监督,区分跨任务新类别与存储正类的显著区域,以此减轻混淆。此外,本文还提出一个基于交叉熵的采样方法,用于挑选拥有可信类激活图的样本。实验结果表明,本文提出的基于类激活图监督回放的方法显著提升了在多标签连续学习任务下模型的准确率。由可视化实验结果可见,原本混淆的标签被区分,进一步验证了所提出方法在缓解多标签连续学习潜在混淆问题上的能力。

参考文献(References)

Bai X F, Lu L B and Wang W J. 2024. Saliency guided object complementary hiding for weakly supervised semantic segmentation. *Journal of Image and Graphics*, 29(4): 1041-1055 (白雪飞, 卢立彬, 王文剑. 2024. 显著性引导的目标互补隐藏弱监督语义分割. *中国图象图形学报*, 29(4): 1041-1055) [DOI: 10.11834/jig.230156]

Boschini M, Bonicelli L, Buzzega P, Porrello A and Calderara S. 2023. Class-incremental continual learning into the extended der-verse. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5): 5497-5512 [DOI: 10.1109/tpami.2022.3206549]

Buzzega P, Boschini M, Porrello A, Abati D and Calderara S. 2020. Dark experience for general continual learning: a strong, simple baseline//*Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc.: 15920-15930

Chen T S, Lin L, Chen R Q, Hui X L and Wu H F. 2022b. Knowledge-guided multi-label few-shot learning for general image recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3): 1371-1384 [DOI: 10.1109/tpami.2020.3025814]

Chen Z M, Wei X S, Wang P and Guo Y W. 2023. Learning graph convolutional networks for multi-label recognition and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6): 6969-6983 [DOI: 10.1109/TPAMI.2021.3063496]

Chen Z Z, Wang T, Wu X W, Hua X S, Zhang H W and Sun Q R. 2022a. Class re-activation maps for weakly-supervised semantic segmentation//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 959-968 [DOI: 10.1109/cvpr52688.2022.00104]

Dong S L, Luo H Y, He Y H, Wei X, Cheng J and Gong Y H. 2023. Knowledge restore and transfer for multi-label class-incremental learning//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 18665-18674 [DOI: 10.1109/iccv51070.2023.01715]

Douillard A, Cord M, Ollion C, Robert T and Valle E. 2020. PODNet:

- pooled outputs distillation for small-tasks incremental learning//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 86-102 [DOI: 10.1007/978-3-030-58565-5_6]
- Douillard A, Ramé A, Couairon G and Cord M. 2022. Dytox: transformers for continual learning with dynamic TOken expansion//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 9275-9285 [DOI: 10.1109/cvpr52688.2022.00907]
- Du K L, Lyu F, Li L Y, Hu F Y, Feng W, Xu F L, Xi X F and Cheng H J. 2024. Multi-label continual learning using augmented graph convolutional network. *IEEE Transactions on Multimedia*, 26: 2978-2992 [DOI: 10.1109/tmm.2023.3305871]
- Ebrahimi S, Petryk S, Gokul A, Gan W, Gonzalez J E, Rohrbach M and Darrell T. 2021. Remembering for the right reasons: explanations reduce catastrophic forgetting. *Applied AI Letters*, 2(4): #e44 [DOI: 10.1002/ail.2.44]
- Everingham M, Van Gool L, Williams C K I, Winn J and Zisserman A. 2010. The PASCAL visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2): 303-338 [DOI: 10.1007/s11263-009-0275-4]
- Kim C D, Jeong J and Kim G. 2020. Imbalanced continual learning with partitioning reservoir sampling//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 411-428 [DOI: 10.1007/978-3-030-58601-0_25]
- 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]
- Lee J, Kim E and Yoon S. 2021. Anti-adversarially manipulated attributions for weakly and semi-supervised semantic segmentation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 4070-4078 [DOI: 10.1109/cvpr46437.2021.00406]
- 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]
- Liu Z Y and Liu J F. 2024. Data-free class incremental learning based on knowledge distillation. *Journal of Computer Applications*, 44(S2): 12-17 (刘展阳, 刘进锋. 2024. 基于知识蒸馏的不存储旧数据的类增量学习. *计算机应用*, 44(S2): 12-17) [DOI: 10.11772/j.issn.1001-9081.2023111716]
- Luo Z L, Liu Y Y, Schiele B and Sun Q R. 2023. Class-incremental exemplar compression for class-incremental learning//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 11371-11380 [DOI: 10.1109/cvpr52729.2023.01094]
- McCloskey M and Cohen N J. 1989. Catastrophic interference in connectionist networks: the sequential learning problem. *Psychology of Learning and Motivation*, 24: 109-165 [DOI: 10.1016/s0079-7421(08)60536-8]
- 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. Honolulu, USA: IEEE: 5533-5542 [DOI: 10.1109/cvpr.2017.587]
- Ridnik T, Lawen H, Noy A, Baruch E B, Sharir G and Friedman I. 2021. TRResNet: high performance GPU-dedicated architecture//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 1399-1408 [DOI: 10.1109/wacv48630.2021.00144]
- Rolnick D, Ahuja A, Schwarz J, Lillicrap T P and Wayne G. 2019. Experience replay for continual learning//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #32
- Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 618-626 [DOI: 10.1109/iccv.2017.74]
- Wang J, Yang Y, Mao J H, Huang Z H, Huang C and Xu W. 2016. CNN-RNN: a unified framework for multi-label image classification//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2285-2294 [DOI: 10.1109/cvpr.2016.251]
- Wang Z X, Chen T S, Li G B, Xu R J and Lin L. 2017. Multi-label image recognition by recurrently discovering attentional regions//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 464-472 [DOI: 10.1109/iccv.2017.58]
- Xing S X, Ju Z H, Liu Z J, Wang Y and Fan F Q. 2023. Multi-label classification of chest X-ray images with pre-trained vision Transformer model. *Journal of Image and Graphics*, 28(4): 1186-1197 (邢素霞, 鞠子涵, 刘子骄, 王瑜, 范福强. 2023. 视觉Transformer预训练模型的胸腔X线影像多标签分类. *中国图象图形学报*, 28(4): 1186-1197) [DOI: 10.11834/jig.220284]
- Yan S P, Xie J W and He X M. 2021. DER: dynamically expandable representation for class incremental learning//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 3013-3022 [DOI: 10.1109/cvpr46437.2021.00303]
- Zhang J, Wang Y B and Fang S. 2020. Multi-label hyperspectral image classification. *Journal of Image and Graphics*, 25(3): 568-578 (张晶, 王亦斌, 方帅. 2020. 多标签高光谱图像地物分类. *中国图象图形学报*, 25(3): 568-578) [DOI: 10.11834/jig.190285]
- Zhao J W, Yan K, Zhao Y F, Guo X W, Huang F Y and Li J. 2021.

Transformer-based dual relation graph for multi-label image recognition//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 163-172 [DOI: 10.1109/iccv48922.2021.00023]

Zhou B L, Khosla A, Lapedriza A, Oliva A and Torralba A. 2016. Learning deep features for discriminative localization//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2921-2929 [DOI: 10.1109/cvpr.2016.319]

作者简介

周怡凡,女,硕士研究生,主要研究方向为多标签类增量连续学习。E-mail: 220232050@seu.edu.cn

刘光灿,通信作者,男,教授,主要研究方向为低维模型与优化在机器学习、模式识别、计算机视觉、信号处理、序列预测领域的应用。E-mail: guangcanliu@seu.edu.cn

杜凯乐,男,博士研究生,主要研究方向为连续学习、多标签学习和图表示学习。E-mail: kailedu@seu.edu.cn

吕凡,男,博士后,主要研究方向为连续学习、视觉—语言理解、小样本学习、可持续人工智能和交互式人工智能。E-mail: fan.lyu@cripac.ia.ac.cn

胡伏原,男,教授,主要研究方向为图像处理、模式识别、人工智能以及智能规划技术的应用研究。E-mail: fuyuanhu@mail.usts.edu.cn