

# 基于主动学习的图像分类技术:现状与未来

刘 颖<sup>1,2</sup>, 庞羽良<sup>1</sup>, 张伟东<sup>1,2</sup>, 李大湘<sup>1,2</sup>, 许志杰<sup>2,3</sup>

(1. 西安邮电大学图像与信息处理研究所, 陕西西安 710121; 2. 无线通信与信息处理技术国际联合研究中心, 陕西西安 710121; 3. 英国哈德斯菲尔德大学, 西约克郡 HD13DH)

**摘 要:** 图像分类作为计算机视觉领域中的重要研究方向之一, 应用领域非常广泛. 基于深度学习的图像分类技术取得的成功, 依赖大量的已标注数据, 然而数据的标注成本往往是昂贵的. 主动学习作为一种机器学习方法, 旨在以尽可能少的高质量标注数据达到期望的模型性能, 缓解监督学习任务中存在的标注成本高、标注信息难以大量获取的问题. 主动学习图像分类算法根据样本选择策略, 从未标记样本数据集中选择出信息量丰富, 对分类模型训练贡献更高的样本进行标注, 以更新已标注训练数据池, 如此循环直至满足给定的停止条件或模型标注预算耗尽. 本文对近年来提出的主动学习图像分类算法进行了详细综述, 并根据所用样本数据处理及模型优化方案, 将现有算法分为三类: 基于数据增强的算法, 包括利用图像增广来扩充训练数据, 或者根据图像特征插值后的差异性来选择高质量的训练数据; 基于数据分布信息的算法, 根据数据分布的特点来优化样本选择策略; 优化模型预测的算法, 包括优化获取和利用深度模型预测信息的方法、基于生成对抗网络和强化学习来优化预测模型的结构, 以及基于 Transformer 结构提升模型预测性能, 以确保模型预测结果的可靠性. 此外, 本文还对各类主动学习图像分类算法下的重要学术工作进行了实验对比, 并对各算法在不同规模数据集上的性能和适应性进行了分析. 另外, 本文探讨了主动学习图像分类技术所面临的挑战, 并指出了未来研究的方向.

**关键词:** 图像分类; 主动学习; 数据增强; 数据分布; 模型预测信息; 模型结构优化

**基金项目:** 国家自然科学基金青年项目 (No.62106195)

**中图分类号:** TP751; TP391

**文献标识码:** A

**文章编号:** 0372-2112(2023)10-2960-25

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

**DOI:** 10.12263/DZXB.20230397

## Active Learning-Based Image Classification Technology: Status and Future

LIU Ying<sup>1,2</sup>, PANG Yu-liang<sup>1</sup>, ZHANG Wei-dong<sup>1,2</sup>, LI Da-xiang<sup>1,2</sup>, XU Zhi-jie<sup>2,3</sup>

(1. Center for Image and Information Processing, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China;

2. International Joint-Research Center for Wireless Communication and Information Processing, Xi'an, Shaanxi 710121, China;

3. Huddersfield University, West Yorkshire HD13DH, United Kingdom of Great Britain and Northern Ireland)

**Abstract:** As one of the important research directions in the field of computer vision, image classification has a wide range of applications. The success of deep learning-based image classification techniques depends on a large amount of annotated data. However, the cost of data annotation is often expensive. Active learning is a machine learning method that aims to achieve the expected model performance with as few high-quality annotated data as possible, and it can alleviate the problem of high annotation costs and difficulty in obtaining a large amount of annotation information in supervised learning tasks. Based on a sample selection strategy, active learning for image classification selects samples from the unlabeled dataset which are informative and thus contribute more to the training of the classification model, in order to update the annotated training data pool. This process is repeated until a given stopping condition is met or the model annotation budget is exhausted. This paper provides a comprehensive survey of the active learning image classification algorithms published in recent years. According the strategies applied in sample data processing and model structure optimization, existing algorithms are classified into three categories: algorithms based on data augmentation, including those using image augmentation to expand the scale of training data or using the differences in image feature interpolation to select high-quality training data; algorithms based on data distribution information, which optimize sample selection strategies based on the characteris-

tics of data distribution; algorithms for optimizing model predictions, including methods for optimizing the acquisition and utilization of deep model prediction information, improving the predictive model structure through the use of generative adversarial networks and reinforcement learning, as well as enhancing model prediction performance based on the Transformer architecture to ensure the reliability of model predictions. In addition, this paper also conducts experimental comparisons on important academic work under various types of active learning image classification algorithms, and analyzes the performance and adaptability of each algorithm on datasets of different scales. Furthermore, this paper discusses the challenges faced by active learning image classification technology and points out future research directions.

**Key words:** image classification; active learning; data augmentation; data distribution; model prediction information; model structure optimization

**Foundation Item(s):** National Natural Science Foundation of China (Youth Program) (No.62106195)

## 1 引言

图像分类是计算机视觉领域中的一大基本任务。图像分类任务的核心在于图像特征提取和分类器的设计。随着深度学习(Deep Learning, DL)<sup>[1]</sup>技术的不断发展,基于卷积神经网络(Convolutional Neural Networks, CNN)<sup>[2]</sup>的图像特征提取技术取得了巨大的成就。卷积神经网络可以通过组合简单特征形成更复杂和抽象的特征,从而提高图像分类任务的准确性和鲁棒性。作为一种数据表示学习的方法,深度学习可以通过迭代更新深度网络层级参数来训练和优化模型,从而使结果更加接近真实值。常用于图像分类的深度网络包括 LeNet<sup>[3]</sup>, GoogLeNet<sup>[4]</sup>, AlexNet<sup>[5]</sup>, VGGNet<sup>[6]</sup>, ResNet<sup>[7]</sup>等。然而,在图像分类领域,为了得到高精度的分类器,深度学习模型很大程度上依赖大量已标注数据来优化模型参数。特别是在需要高水平专业知识的领域,如医学图像<sup>[8]</sup>、遥感图像<sup>[9]</sup>等,获取大量的高质量已标注数据集需要消耗大量的人力。

主动学习(Active Learning, AL)<sup>[10]</sup>作为一种能够降低样本标注成本的学习方法,正逐渐受到越来越多的关注。主动学习作为监督式机器学习中的一种范式,旨在标注尽可能少的样本,同时最大化模型的性能增益。具体来讲,主动学习根据样本选择策略从未标记的数据集中选择信息丰富的样本,交由 Oracle 进行标注,以降低模型所需数据量、计算资源和存储资源的需求,同时保持分类器性能。Oracle 是一个能够提供准确标签的信息源,可以是人类专家或自动化系统。样本选择策略决定了算法选择哪些样本以获得最大的模型性能提升。目前,主动学习已被应用于分类与检索<sup>[11]</sup>、图像分割<sup>[12]</sup>、目标检测<sup>[13]</sup>等多种图像处理任务。

在早期研究中,文献[10]将目前主动学习方法定义为三种基本框架:基于成员查询的主动学习、基于流的选择性采样和基于池的主动学习。基于成员查询的主动学习方法是指学习器可以请求查询输入空间中任何未标记样本的标签,包括学习器生成的样本。基于流的选择性采样是指每次从未标记数据源中提取一个样

本数据,学习器必须决定是查询标签还是丢弃该数据。基于池的主动学习框架则维护一个未标注数据集,由样本选择策略从未标记集中选择要标注的样本。

目前,基于池的主动学习框架更适用于图像分类任务中。该框架能同时处理批量数据,从未标记数据集中选出对模型训练最有帮助的数据进行标注,提高标注数据效率,降低成本。此外,该框架适用于数据集规模较大、标注数据较少的情况,符合多数图像分类技术场景。相比之下,基于成员查询的主动学习算法需要逐个查询成员并进行标注,不适用于大规模的数据集。基于流的选择性采样在处理流数据时,对每个数据点进行快速分类来实现快速标注,从而处理大量的数据流。但在图像分类中,每个数据点都是一个独立的图像,对每个图像进行分类和标注可能会带来更多的标注成本,因此该方法在图像分类中使用较少。该方法主要适用于需要时效性的小型移动设备的应用场景,因为这些小型设备通常具有有限的存储和计算能力。图 1 展示了基于池的主动学习的基本框架。

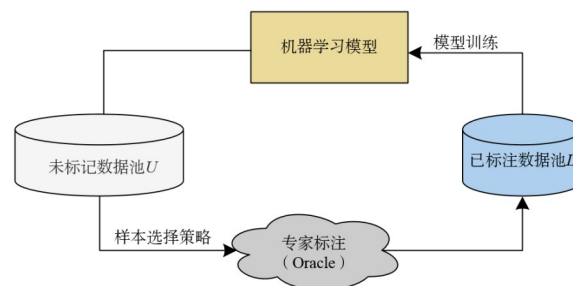


图 1 基于池的主动学习框架<sup>[10]</sup>

在初始状态下,从未标记数据池  $U$  中随机选择样本,交由 Oracle 查询标签以获得标记的数据集  $L$ 。然后,使用监督学习算法在  $L$  上训练模型。随后,根据新样本中获取的信息选择要查询的样本,由 Oracle 标注后添加到  $L$  中,并进行下一步模型训练。如此循环迭代,直到标签预算耗尽或达到预定义终止条件。

近年来,在主动学习图像分类领域中,一些研究者探索了如何将深度模型和主动学习策略结合起来,以

提高图像分类的效率和准确性,如基于核心集<sup>[14]</sup>、基于贝叶斯卷积神经网络<sup>[15]</sup>等,利用模型的代表性或不确定性来指导样本的选择.此外,部分研究者探索了如何利用强化学习来优化主动学习的过程,以提高图像分类的性能和稳定性<sup>[16]</sup>,使模型快速适应不同的数据分布和任务.部分研究者则利用生成对抗网络来增强主动学习的能力,以提高图像分类的泛化性和鲁棒性.如基于生成对抗网络<sup>[17]</sup>、基于条件生成对抗网络<sup>[18]</sup>等,利用生成器产生新的样本,并使用判别器作为主动学习策略来选择最有信息量或多样性的样本进行标注.最近,部分研究者针对数据分布情况,探索了数据不平衡问题对主动学习的影响,以提高图像分类的鲁棒性,如基于类平衡<sup>[19]</sup>,利用类平衡因素避免偏向于某些类别的样本.

尽管已有的综述工作总结了近年来主动学习算法在算法改进、计算机视觉任务(目标检测、图像分割、视频处理)和自然语言处理等领域中的应用<sup>[20-22]</sup>,但尚未对图像分类这一特定任务展开更详细深入的介绍.鉴于此,本文基于近十年来国内外公开发表的重要学术工作,对现有的主动学习图像分类算法进行了详细综述.此外,考虑到当前研究工作中,一些研究者在评估算法性能时使用不同的模型标注预算,本文通过实验比较和分析了不同类别代表性算法在相同标注预算下的性能,并对算法的优缺点进行了探讨.此外,针对目前主动学习图像分类算法所面临的挑战,本文提出了几个具有潜力的研究方向.

如何有效利用数据进行模型训练,以及如何优化主动学习图像分类算法架构,是影响主动学习图像分类算法性能的关键因素.因此,本文从两个方面详细总结近年来提出的主动学习图像分类算法.一方面,鉴于主动学习图像分类算法基于有限的已标注数据来进行模型训练,本文首先从最直观的数据处理角度出发,来总结近年来基于数据增强的主动学习图像分类算法.通过数据增强的手段,算法能够扩充有限的标注数据,从而提高算法的性能.此外,在数据处理的过程中,考虑到数据分布情况对主动学习样本选择策略的影响,本文详细总结了基于数据分布信息的主动学习图像分类算法.利用数据分布的特征,针对不同的数据分布情况来设计相应的主动学习样本选择策略,以实现更加高效和准确的样本选择.另一方面,随着近年来深度学习与主动学习图像分类算法的逐步融合,众多研究者通过优化模型架构以及对模型训练过程的改进,来提升深度模型预测性能.例如,优化深度模型的预测信息、基于生成对抗网络、基于强化学习策略和基于Transformer结构来提升主动学习模型的预测效果.故本文还对优化模型预测的主动学习图像分类算法进行

了详尽的总结.综上,本文根据主动学习图像分类算法所用样本数据处理及模型优化方案,将现有算法分为三大类:基于数据增强的主动学习图像分类算法、基于数据分布信息的主动学习图像分类算法以及优化模型预测的主动学习图像分类算法.

本文的结构如下:第2节介绍了基于主动学习图像分类算法的基本框架;第3节根据所用样本数据处理及模型优化方案,将现有主动学习图像分类算法分为基于数据增强、基于数据分布信息以及优化模型预测三大类,并进行详细介绍;第4节通过实验数据对比分析了各类典型算法的性能;第5节讨论了主动学习图像分类技术所面临的技术挑战,并指出了未来研究趋势.第6节对本文工作进行了总结.

## 2 主动学习图像分类介绍

### 2.1 主动学习图像分类算法框架

主动学习图像分类方法根据样本选择策略,从未标记的样本数据集中选择出对模型训练贡献更大的样本数据,以更新已标注训练数据集.具体工作模式是抽样迭代训练的过程<sup>[21]</sup>.首先,使用初始已标注数据集 $L$ 训练分类器模型 $C$ .然后,通过样本选择策略 $Q$ 从未标注数据集中选择部分高质量数据,并由Oracle对这些选中的样本进行标注.标注的新样本将被添加到标注样本集 $L$ 中,形成新的训练集,以参与下一次分类器训练.该步骤为循环过程,迭代进行分类器训练和样本选择标注.算法流程如图2所示.

#### 2.1.1 数据预处理

对数据进行预处理,可提高模型的鲁棒性和泛化能力,使模型更好地适应不同的输入数据.本文将主动学习图像分类算法的数据预处理方式分为以下2种.

一是对数据集的量级和数据集模式进行扩增,使经过样本查询函数选择出的少量标注图像数据包含更多的语义信息.在早期研究中,数据增强主要采用传统的方法,如旋转、平移、缩放等操作,但是这些方法存在局限性,不能满足复杂场景下的需求.目前,基于深度学习的方法已经成为数据增强的主流技术,例如使用生成网络和变分自编码器进行数据扩充和样本合成等操作.随着深度学习与主动学习的深度融合,数据处理的方式也愈加复杂,在本文的后续内容中将进行更深入的介绍.

二是对数据集进行处理,使主动学习模型适应不同的数据分布情况,进而样本选择策略能够选择更有价值的样本<sup>[23]</sup>.例如样本难度评估和样本平衡处理等.在某些应用场景下,数据集往往是不均衡的,一些样本可能比其他样本更具有代表性和难度.因此,样本难度评估可以帮助选择具有代表性的样本来增强模型的泛

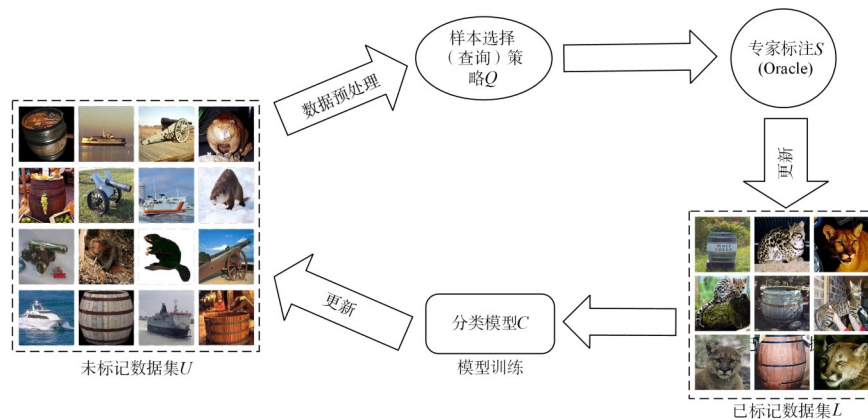


图2 主动学习图像分类算法流程

化能力。早期样本难度评估方法主要基于样本的统计信息和特征分布,如使用欧氏距离等测量方法来判断样本之间的相似性,从而进行样本选择和样本加权等操作。随着深度学习和主动学习等技术的发展,目前的方法主要基于模型输出的置信度、熵、梯度等方法来度量样本的不确定性和难度,以及基于生成模型的样本难度评估方法等。此外,在样本平衡处理方面,除了传统的欠采样和过采样方法,为了更有效地平衡数据集中类别之间的数量和质量差异,目前学者们采用了基于生成对抗网络的样本生成方法和基于辅助任务的样本扩增等方法。

### 2.1.2 常见样本选择策略

如前所述,深度学习基于大量的已标注数据来训练模型。与深度学习不同,主动学习从数据集开始,主要通过设计复杂的样本选择策略,从未标记的数据集中选择最佳样本并查询其标签。因此,样本选择策略的设计对主动学习的性能至关重要,相关研究也相当丰富。例如,在一组给定的未标记数据集中,主要的选择策略包括基于不确定性的方法<sup>[24]</sup>、基于代表性的方法<sup>[14]</sup>以及基于多样性的方法<sup>[25]</sup>等。

基于不确定性的方法根据模型预测的概率分布或分类边界等指标,选择模型预测结果最不确定的样本作为下一轮的训练数据。基于代表性的方法根据当前已有的样本分布或特征分布等指标,选择出能够代表未标记数据分布的样本作为下一轮的训练数据。基于多样性的方法通常会优先选择距离已有标注样本最远的样本或者选择与已有标注样本差异性最大的样本,以保证被选择出的样本的多样性。由于基于不确定性的抽样方法通常会导致抽样偏差,因此当前选择的样本难以更好地代表未标记数据集的分布。另外,只考虑促进抽样多样性的策略可能会导致标注成本增加,因为可能会选择大量信息含量较低的样本。因此,近年来许多研究者还研究了混合选择策略<sup>[26,27]</sup>,并试图在多

种选择策略之间找到平衡。

在早期的主动学习图像分类任务中,常见的样本选择方法如表1所示。近年来,越来越多的工作利用深度模型来学习如何评估样本的重要性,以改进样本选择策略。该深度模型可以是一个分类模型或一个生成模型等。例如,利用深度卷积神经网络的特征表达能力和预测概率来评估样本的不确定性、多样性和代表性等指标,从而选择最有利于模型学习的样本;或者将主动学习的样本选择策略与生成模型结合,可以实现从数据空间中合成最有信息量的样本;或者从已有的样本中提取最有信息量的部分,从而提高标注效率和模型性能。利用深度模型的强大表达能力,可使样本选择策略获得更高的精度和鲁棒性,从而适应复杂场景下应用。

此外,在主动学习图像分类中,选择合适的样本选择策略需考虑多方面因素,如任务特点、分类器性能、标注成本等。例如,对于大规模数据集,可使用不确定性采样策略最小化标注成本,确保分类器性能;对于具有复杂结构的数据集,可采用多样性策略提高样本多样性,避免选择相似样本;还可结合分类器进行样本选择,如使用置信度量、边缘度量等方法选择最具信息量的样本。综合考虑任务特点和分类器性能,选择合适的样本选择策略是主动学习图像分类中关键一步。第3节将详细介绍不同算法中样本选择策略的工作原理。

### 2.1.3 分类器

分类器根据数据的特征度量进行数据分类。传统的机器学习图像分类算法已被广泛应用于主动学习模型中。在主动学习任务中,常用一些分类算法包括支持向量机<sup>[35]</sup>、K近邻算法<sup>[36,37]</sup>和余弦相似度<sup>[38]</sup>等。

近年来,深度学习图像分类算法表现优异,但仍面临训练数据标注困难和高维数据分类复杂度高挑战。为解决这些问题,一些学者尝试将深度分类模型与主动学习相结合。例如,2017年Feng等人<sup>[39]</sup>在主动学

表 1 样本选择策略函数总结

样本选择(查询)策略函数	原理	样本选择偏好	计算方法
随机采样	从未标记样本中随机采样	—	—
最大熵采样 <sup>[28]</sup>	选择信息熵最大的样本	不确定性	$H(x) = - \sum_{i=1}^m p(y_i x, L) \log_2 p(y_i x, L)$
DeepGini <sup>[29]</sup>	选择被错误分类可能性最大的样本	不确定性	$\operatorname{argmax}_{x \in u} \left( 1 - \sum_{i \in [c]} p^2(y = i x; \theta) \right)$
BALD <sup>[15]</sup>	选择贝叶斯神经网络预测不一致性最大的样本	不确定性+代表性	$\operatorname{argmax}_{x \in u} \left( H(y x; u) - E_{\theta \sim p_{\theta, u}} [H(y x; \theta)] \right)$
Dropout-Entropy <sup>[30]</sup>	选择在多个 DNN 上不确定性最大的样本	不确定性	$H(y x; h) = - \sum_{i \in [c]} p(y = i x; h) \log(y = i x; h)$
最小置信度 <sup>[31]</sup>	选择预测置信度最低的样本	不确定性	$\operatorname{argmax}_{x \in u} (1 - p(y = y' x; \theta))$
Margin 采样 <sup>[32]</sup>	选择最高置信标签和第二置信标签之间差异较小的样本	不确定性	$\operatorname{argmin}_{\substack{x \in [C] \\ x \in u}} \left( \operatorname{argmax}_{x \in [C]} (p(y = i x; q)) - \operatorname{argmax}_{x \in [C] \setminus y'} (p(y = i x; q)) \right)$
多边界聚类 and 优先排序 <sup>[33]</sup>	从优先级较高的数据集中选择样本	多样性	—
期望梯度长度 <sup>[34]</sup>	选择使梯度下降最大的样本	不确定性	$\operatorname{argmax}_{x \in u} \left( \sum_{i \in [C]} p(y = i x; \theta) \ \nabla J(\theta, x, y = i)\  \right)$

习框架下,将深度残差网络用于图像缺陷检测和分类. 2018年 Ahmed等<sup>[40]</sup>使用VGG16模型实现了一个用于人脸表情识别分类的增量式主动学习框架. 2018年 Haut等<sup>[41]</sup>将贝叶斯卷积神经网络与主动学习样本选择策略相结合,提出了一种用于高光谱图像分类的算法,取得了良好的分类性能. 在医学图像分类领域,2018年 Sayantan等<sup>[42]</sup>基于深度置信网络来学习图像的特征表示,有效提升了模型的性能. 这些研究表明,将深度分类网络与主动学习相结合能够有效地提高分类性能,为后续研究提供了有益的参考.

### 3 主动学习图像分类算法

基于主动学习的图像分类方法旨在通过一定的样本选择策略,选择对模型训练提供更多贡献的样本,在节省大量数据标注成本的情况下,得到较高性能的分类器. 在标注数据有限的背景下,对图像数据进行处理,以充分利用选择出的高质量样本或直接生成高质量的训练样本,以及针对数据的分布信息来提高主动学习算法中样本选择策略的适应性,是提高主动学习图像分类模型性能的一种直观方法. 此外,随着深度学习技术的发展,如何有效利用深度模型的信息来评估样本的价值,并优化主动学习算法的模型结构,已成为一个研究热点. 早期的主动学习图像分类算法通过结合熵、置信度等来设计样本选择策略. 这些方法易于适应各种任务,且大多只涉及数据选择过程,较少优化网络训练过程;并且样本选择过程可能会引入选择偏差,导致算法选择某些易于分类的数据而忽略了一些重要

的难以分类的数据. 因此,更好的主动学习方法需要综合考虑数据选择和模型训练过程的改进,以提高其性能和鲁棒性.

因此,在目前的主动学习图像分类任务中,算法的改进主要分为数据驱动和模型驱动两方面<sup>[22]</sup>. 在现有算法中,数据驱动方面主要包括图像增广以及对图像特征插值处理等,以及在算法改进的过程中将数据的分布信息考虑其中. 模型驱动方面主要包括附加额外网络、修改损失函数、集成生成对抗网络、集成强化学习方法和基于Transformer结构等. 本节旨在从基于数据增强、基于数据的分布信息以及优化模型预测的角度出发,介绍近年来主动学习图像分类领域的研究成果. 现有算法分类归纳如图3所示.

#### 3.1 基于数据增强的主动学习图像分类算法

基于主动学习的图像分类算法依赖少量信息量丰富的已标注数据来进行模型训练,同时包含大量的未标注数据. 本节从数据增强的角度出发,介绍现有的主动学习图像分类方法. 例如,通过图像增广来扩充训练数据,或通过对图像特征进行插值来判断图像数据的信息丰富性,从而选择需要标注的数据等.

##### 3.1.1 基于图像增广

由于数据标注的成本较高或标注数据不足,采用图像增广<sup>[43]</sup>处理技术可最大限度地利用已有的标注数据. 图像增广通过随机改变训练样本,可以降低模型对某些属性的依赖,从而提高模型的泛化能力. 例如,简单的图像增广方法包括对图像进行不同方式的旋转和裁剪,使感兴趣的物体出现在不同位置,从而减轻模型

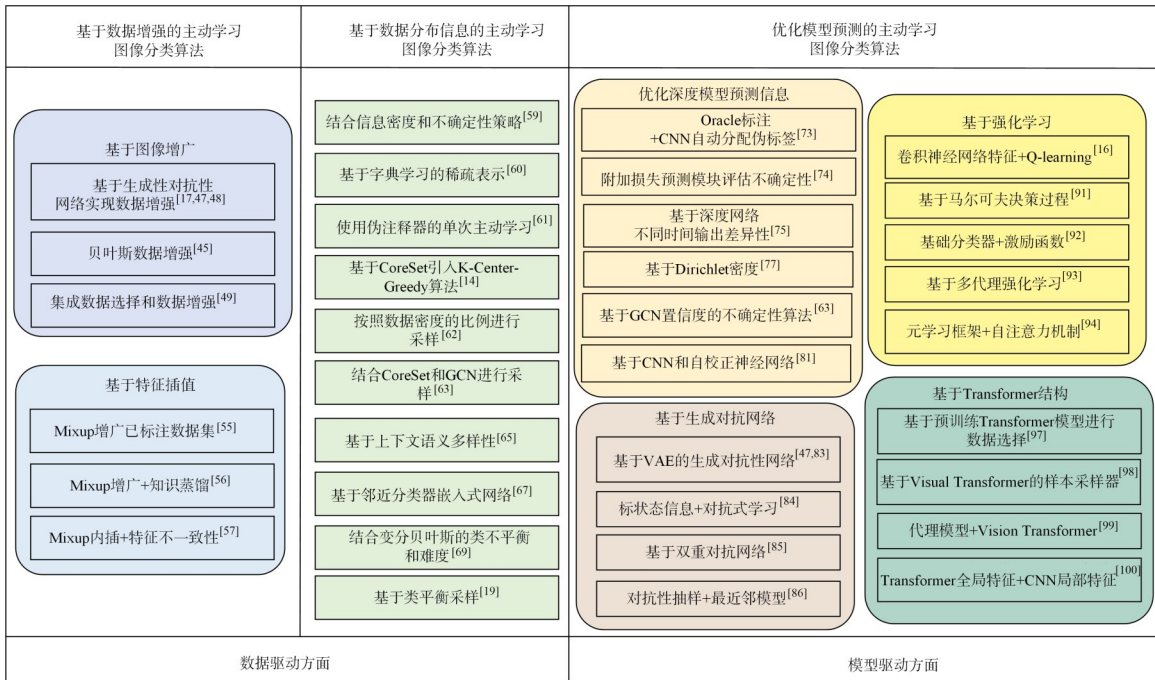


图3 主动学习图像分类算法分类

对物体出现位置的依赖性,也可通过调整亮度、色彩等因素来降低模型对色彩的敏感度,如图4所示的处理过程.

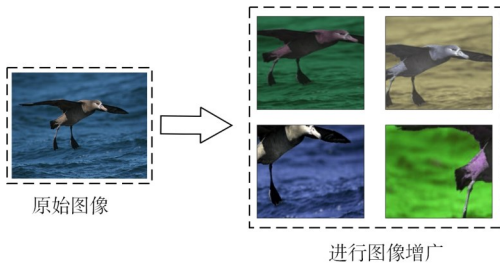


图4 利用图像增广方法处理图像数据

然而,传统的图像增广技术处理图像方式有限,且扩充后的图像质量难以保证.部分研究者考虑通过利用生成对抗性网络(Generative Adversarial Net, GAN)<sup>[44]</sup>来生成可靠性更强的训练数据.2017年,Zhu等<sup>[17]</sup>提出的生成对抗主动学习(Generative Adversarial Active Learning, GAAL)首次将GAN引入样本查询方法中.GAAL的目标是使用生成学习来生成比原始数据集包含更多信息的样本.GAAL通过GAN构造出靠近分类边界的样本,使生成的新样本具有较高不确定性.然而,随机数据扩增并不能保证生成的样本比原始数据中包含更多的信息,从而造成计算资源的浪费<sup>[21]</sup>.2019年,TRAN等<sup>[45]</sup>基于贝叶斯数据增强提出了贝叶斯生成主动深度学习算法(Bayesian Generative Active Deep Learning, BGADL).该算法在GAAL上进行了进一步的

扩展,结合辅助分类器生成对抗网络和变分自编码器(Variational Auto Encoder, VAE)<sup>[46]</sup>等方法,目的是生成属于不同类别的不同区域的样本.该方法通过辅助分类器生成对抗网络和贝叶斯数据增强来产生与所选择样本信息丰富程度一样的新样本.

此外,2019年提出的变分对抗主动学习(Variational Adversarial Active Learning, VAAL)<sup>[47]</sup>和对抗表示主动学习(Adversarial Representation Active Learning, ARAL)<sup>[48]</sup>不仅将生成性对抗学习引入网络体系结构实现数据增强,而且使用已标记和未标记的数据集来训练分类网络.在此过程中,鉴别器尝试区分重构图像和原始图像之间的差异,从而帮助选择最具信息量的未标记图像进行标记.VAAL通过减少基于不确定性的批量查询策略的依赖来解决批量查询策略容易受到离群值干扰的问题.ARAL对VAAL进行了扩展,以尽量减少使用人工标注样本,在充分利用现有或生成的数据信息的同时提高模型学习能力.ARAL额外使用深度生成网络产生的样本来联合训练整个模型,并通过共享鉴别器的特征来训练分类器.这种方法不仅可以提高学习到的表示的质量,而且可以进一步提高分类性能.

为了进一步保证生成数据拥有丰富的信息量.2021年, Kim等<sup>[49]</sup>提出了一种“前瞻数据采集”(Look-Ahead Data Acquisition, LADA)的算法,旨在集成数据选择和数据增强.该算法在进行数据选择之前考虑数据增强的效果,并综合考虑数据增强所产生的非标记数据和虚拟数据以进行数据选择.传统的样本选择策略不考虑数据增强的潜在增益,而LADA则通过将

数据增强集成到采集过程中,来考虑虚拟数据的信息量.此外,LADA 还通过优化数据增强策略,以最大化预测获取分数来增强虚拟数据实例的信息量.

然而,以上方法在对样本图像进行增强时,并未充分考虑原始图像关键特征的完整性.针对这一问题,Gong 等人<sup>[50]</sup>结合 KeepAugment 数据增强方法对每个循环过程中所选择出的高质量样本进行数据增强.该算

法首先基于显著图来检测原始图像上的重要区域,并在数据增强期间保留这些重要的信息区域,这种信息保留策略允许生成更可靠的训练样本,并将一种低计算量 SpinalNet<sup>[51]</sup>的深度网络模型改进分类网络,算法框架如图 5 所示.该算法进一步提升了基线算法的性能,尤其在分类类别数较少的数据集上,显示出了更先进的性能.

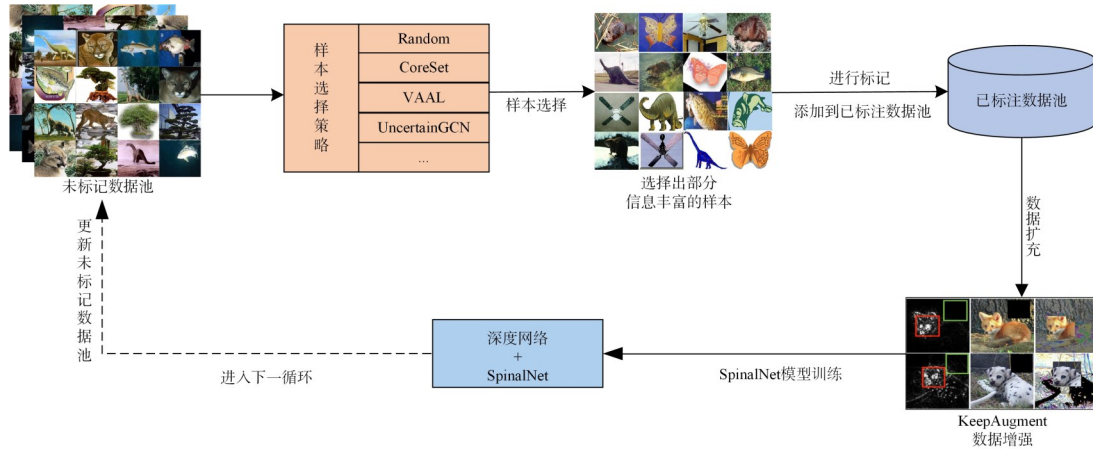


图5 结合数据增强和SpinalNet的主动学习图像分类算法<sup>[50]</sup>

数据增强技术在提高数据多样性和改善模型泛化性能方面具有明显优势.但在应用数据增强时,需考虑现实世界任务可能存在的问题.例如,仅生成未标记样本可能会导致数据增强生成不自然或人类难以解释的实例,从而降低模型的可解释性<sup>[50]</sup>.此外,生成对抗网络可能会生成与原始数据集中的样本不同的样本,给模型训练带来挑战<sup>[52]</sup>.总体而言,数据增强为近年的研究提供了有效的方法,且这种对数据利用技巧的探索也是必不可少的.

### 3.1.2 基于特征插值

与传统的图像增强和生成网络不同,通过对图像特征进行插值处理以实现数据增强,而无需生成额外数据,在目前的主动学习图像分领域中取得了显著的成果.特别是,Zhang 等人<sup>[53]</sup>在 2017 年提出的 Mixup 算法,近年来受到一些主动学习算法研究者的重视<sup>[54]</sup>.

Mixup 是一种通过线性内插构造新的训练样本及其对应的标签的方法.该方法首先从原始训练集中随机选取两个样本-标注对,然后对这两个样本-标注对的特征向量进行线性内插.最后,获得一个新的样本-标注对 $(\tilde{x}, \tilde{y})$ ,数学定义如式(1)所示.

$$\begin{aligned} \lambda &\sim \text{Beta}(a, a) \\ \tilde{x} &= \lambda x_i + (1 - \lambda)x_j \\ \tilde{y} &= \lambda y_i + (1 - \lambda)y_j \end{aligned} \quad (1)$$

其中,  $\lambda \in [0, 1]$ ,  $\lambda \sim \text{Beta}(a, a)$  表示  $\lambda$  服从参数为  $a$  的 Beta 分

布.内插后的图像示例如图 6 所示.这种线性建模减少了在预测训练样本之外的数据时的不兼容性,提升了模型的泛化性.

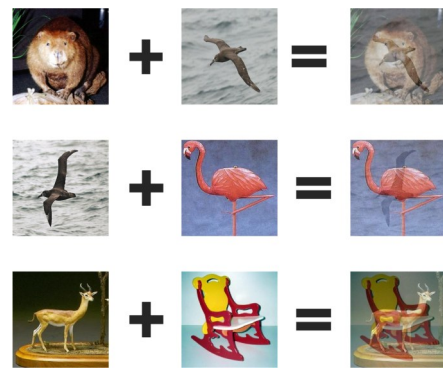


图6 Mixup 增强示例

2020 年, Ma 等人<sup>[55]</sup>基于 Mixup 数据增强方法,首先提出了一种结合数据增强的主动学习图像分类算法.该算法在每轮的迭代过程中,根据主动学习的样本选择策略进行样本选择,由 Oracle 对待标记的样本进行标记,以更新标记数据集,并对更新后的已标记数据集进行 Mixup 数据增强,随后训练分类模型.进一步地,2020 年 Wang 等人<sup>[56]</sup>基于知识蒸馏模型提出了一种与 Mixup 结合的主动学习算法.该算法首先使用 Mixup 合成一批图像,然后使用主动学习算法从中选择最有价值的子集来查询教师模型.查询到教师模型的输出后,将其视为查询图像的真实标签信息,并使用这些标签

来训练学生神经网络,以减少对大规模数据集的依赖,从而训练出高性能的分类模型.此外,2022年Parvaneh等人<sup>[57]</sup>提出的ALFA-Mix算法通过结合Mixup寻找对其表示信息进行干扰而导致的预测不一致,来判定未标记样本的信息丰富性,以进一步提升未标记样本的信息利用率.在有标注和无标注样本的特征表示之间构造内插以形成样本的扰动版本,然后检验预测的标签.通过评估样本扰动版本预测的标签的可变性来识别信息量最大的未标记样本.具体来讲,将未标记的样本集合划分为多个子集,每个子集对应一个特征空间的子空间.对于每个子空间,使用K-Means算法将其内部的样本聚类成若干个簇.对于每个簇,选择距离其质心最近的样本进行标记,这些样本被称为代表性样本.

将所有代表性样本标记后,与已标记的样本特征构建内插,重新训练模型并预测未标记样本的标签.通过将未标记样本的特征与已标记样本相结合,有效地探索其周围邻域,从而选择最有价值的样本进行标注.该算法基本框架如图7所示.

与基于图像增广的算法相比,该方法在不生成新数据的情况下,只在原始数据之间衡量样本信息丰富性,能够有效地判断样本信息丰富性.然而在进行特征插值融合时,若融合的样本之间存在很大的差异,可能会导致融合后的样本信息不准确<sup>[58]</sup>,且现有方法没有充分利用训练数据中丰富的信息,如目标显著性、相对排列等方面的信息.此外,在高维空间中进行插值操作,可能会产生异常值.

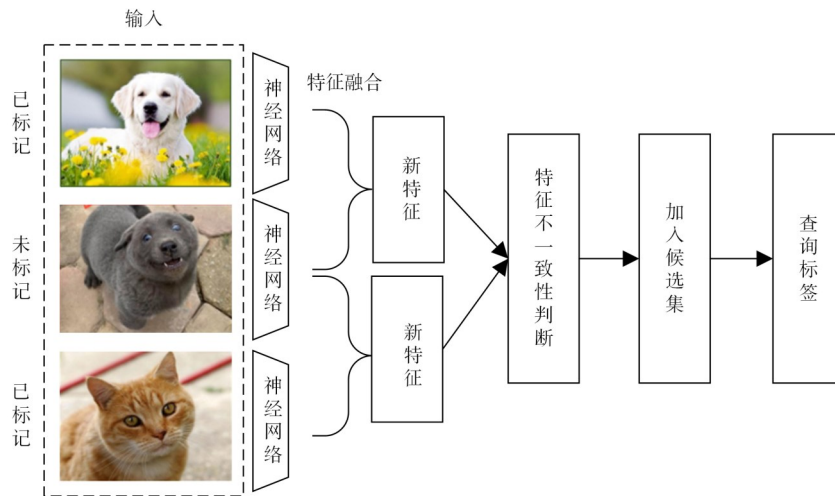


图7 ALFA-Mix算法基本框架<sup>[57]</sup>

### 3.2 基于数据分布信息的主动学习图像分类

数据样本的分布是数据集的内在特征.样本在几何分布中的位置及其与邻域样本的关系,决定了该样本在模型训练中的重要性.

为了使模型更好地适应数据分布,部分研究者在早期的研究中提出了一些自适应样本分布的主动学习方法.2013年,Li等人<sup>[59]</sup>提出了一种结合信息密度和不确定性策略的主动学习方法,用于主动学习图像分类.该方法通过监测训练数据的分布,动态地调整选择策略,以选择更能代表数据分布情况的样本.正如前文所述,基于代表性的算法通常会考虑样本的分布信息和特征分布情况等,且数据通常具有冗余性,故使用更具代表性的样本来训练深度分类模型在直观上是很好的选择.Liu等人<sup>[60]</sup>使用字典学习的稀疏表示来搜索代表性样本.该算法旨在选择训练集中最具代表性和不确定性的样本,并在遥感图像和高光谱图像分类任务中获得了良好的性能.此外,Yang等人<sup>[61]</sup>提出了使用伪注释器的单次主动学

习,其中伪注释器可以作为一种寻找最具代表性样本的特殊方法.

在基于代表性方法的研究中,2017年Sener等人<sup>[14]</sup>首次提出一种基于CoreSet的算法,通过使用全局训练集的代表性替代子集的局部几何特征来提高学习效果.该方法将K-Center-Greedy引入主动学习框架,以选择核心集进行训练.选择中心点的过程是通过最小化数据重复点与其最近中心之间的最大距离来实现的.类似于聚类算法,该方法对远点和离散点不太敏感.然而,基于CoreSet的方法往往只是查询数据点,以尽量覆盖数据流上的所有点,而不考虑样本的密度.这导致查询的数据点过度代表稀疏区域的样本点.针对这一问题,2019年Gissin等人<sup>[62]</sup>提出的判别式主动学习(Discriminative Active Learning, DAL)将主动学习图像分类任务视为一个二进制分类任务,目的是使进一步使查询到的标记数据集与未标记数据集难以区分.DAL的关键优势在于,可以按照数据密度的比例从未标记的数据集中采样,而不会使稀疏域中的样本点产生偏差.

此外, DAL 不限于分类任务, 能够容易地应用于到其他任务中. 进一步, 2021 年 Caramalau 等人<sup>[63]</sup> 基于 CoreSet 算法和图卷积网络 (Graph Convolutional Network, GCN)<sup>[64]</sup> 提出了 CoreGCN 算法. 该算法利用 GCN 学习图像特征之间的关系并结合 CoreSet 算法来选择最具代表性的未标记示例.

从空域分布角度出发, 2020 年 Agarwal 等人<sup>[65]</sup> 根据不同类的图像在同一空域分布的差异性, 提出了一种用于主动学习的语义多样性方法 (Contextual Diversity for Active Learning, CDAL). 该方法有助于在不同的上下文和背景中选择具有不同示例对象的样本. 语义多样性取决于一个重要的观察结果, 即 CNN 预测的感兴趣区域的概率向量通常包含来自更大感受野的信息. 基于此, 作者在 CoreSet 基础上提出了 CDAL-CS 算法, 该算法不会受到维度诅咒的影响. 进一步作者基于强化学习策略 (Reinforcement Learning, RL)<sup>[66]</sup> 提出了 CDAL-RL 算法, 采用了特定任务的状态表示, 并使用了基于上下文多样性的激励, 该激励以无监督的方式结合不确定性和多样性来优化样本选择策略.

一方面, 若已标记样本与未标记样本的分布存在显著偏差, 则可能会影响选择策略的性能. 特别是, 大多数基于不确定性/多样性的方法通常基于 Softmax 分类器的预测来计算. 然而, 这仅在训练的特征和分类器能推广到未标记数据集的前提下才成立<sup>[67]</sup>. 2021 年, Wan 等人<sup>[67]</sup> 针对以上问题提出了邻近分类器嵌入式网络 (Neighbor Classifier Embedded Network, NCE-Net). NCE-Net 利用一个 Soft 邻近分类器<sup>[68]</sup>, 在“拒绝”或“混淆”置信度的指导下进行样本选择, 选择出远离分类决策边界且具有丰富信息性的样本进行标注, 进行下一步训练.

另一方面, 为了进一步探索已标记样本与未标记样本的分布差异性, 以更好地适应现实应用场景下的数据分布情况, 一些研究者开始关注现实世界中存在的数据集不平衡问题, 即数据的长尾分布<sup>[69]</sup>. 2021 年 Choi 等人<sup>[70]</sup> 提出结合变分贝叶斯的类不平衡和难度的算法, 该算法基于贝叶斯规则, 将类不平衡性纳入主动学习框架, 如图 8 所示. 当评估分类器在给定样本上出错的概率时, 同时考虑三个方面: (1) 错误标记类别的概率, (2) 给定预测类别的数据的可能性, (3) 预测类别丰度的先验概率. 通过训练 VAE 并将其与分类器联系, 使用分类器的深度特征表示作为 VAE 的输入, 促进 VAE 训练. 该算法通过考虑所有三种概率, 特别是数据的不平衡性, 在数据不平衡数据集上显示出优异性能. 此外, 2022 年 Javad 等人<sup>[19]</sup> 提出了用于图像分类的类平衡主动学习算法 (Class-Balanced Active Learning, CBAL), 目标是使选择的样本更偏向于均

匀分布. 该算法纠正了未标记数据池中出现的类不平衡问题, 以缓解采样偏差和数据集不平衡带来的问题. 该方法具有较强通用性, 可与常用的不确定性和代表性方法结合.

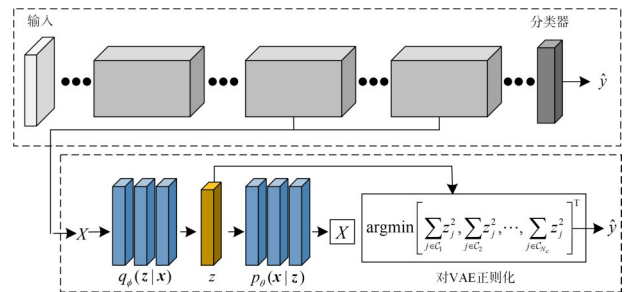


图8 结合变分贝叶斯的类不平衡算法框架<sup>[70]</sup>

通过充分利用数据分布信息, 可以设计更加准确和具有针对性的样本选择策略, 以提高主动学习的效率和选择样本的准确性. 结合数据分布信息可以有效改善模型的鲁棒性, 使其能够更好地应对数据集中的偏差、噪声和离群点. 此外, 数据分布信息的应用还能够有效解决类别不平衡问题, 从而使模型能够更好地适用于现实应用场景. 然而, 需要注意的是, 在不同规模的数据集上, 基于数据分布信息的算法可能呈现出不同的性能表现, 且模型的稳定性难以得到充分保证. 此外在实际应用中, 若数据分布与真实分布存在较大差异, 模型可能会出现过拟合或欠拟合等问题<sup>[71]</sup>.

### 3.3 优化模型预测的主动学习图像分类算法

主动学习图像分类算法旨在通过少量已标注数据获取高性能分类器, 这需要充分优化获取和利用模型信息的途径或优化模型结构来保证模型预测结果的可靠性. 例如, 可通过结合深度学习模型的学习能力来提高算法的性能; 利用深度模型在不同层和不同时间输出之间的差异作为选择需标记数据的依据, 或者利用图卷积神经网络生成更高阶的特征表示; 结合生成对抗网络来进一步优化算法架构, 提高样本选择策略的可靠性; 在算法结构中引入强化学习策略, 以动态优化分类器的训练过程, 并根据不同的任务环境优化样本选择策略; 基于最新提出的视觉 Transformer 结构来提升模型的预测性能.

#### 3.3.1 优化深度模型预测信息

深度学习在高维数据处理和自动特征提取的背景下具有较强的学习能力, 而主动学习在降低标注成本方面具有显著的潜力. 因此, 将深度学习和主动学习的结合是一个直接的方法, 有利更好地扩大两者的应用潜力. 通过结合两者的优势, 部分研究者提出了 DeepAL<sup>[21]</sup> 方法. 图 9 展示了 DeepAL 图像分类算法框架. 深度模型在标记的训练集  $L_0$  上初始化或预训练, 而未

标记的池  $U$  的样本用于通过深度模型提取特征. 接下来, 根据相应的选择策略选择样本, 并由 Oracle 进行标注, 形成新的已标注训练集  $L$ , 然后在  $L$  上训练深度模型, 同时更新  $U$ . 重复此过程, 直到标签预算耗尽或达

到预定义的终止条件. 从图 9 中的 DeepAL 框架示例中, 可大致将 DeepAL 框架分为两部分: 在未标记数据集上施行主动学习算法的样本选择策略和在深度学习模型上的训练.

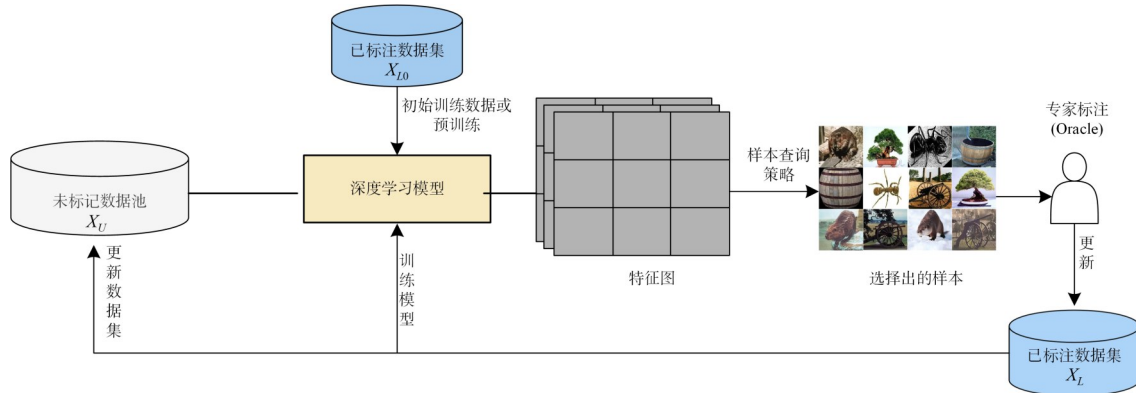


图 9 DeepAL 图像分类算法结构<sup>[21]</sup>

主动学习算法和深度学习算法之间存在处理通道不一致的问题, 即大多数主动学习图像分类算法主要关注分类器的训练, 主要使用基于固定特征表示的查询策略<sup>[21]</sup>. 然而, 在深度学习中, 特征学习和分类器训练共同优化. 简言之, 简单地将主动学习和深度学习结合起来作为两个独立的问题来处理, 可能会导致一定的歧义<sup>[72]</sup>. 2017 年 Wang 等人<sup>[73]</sup>提出的具有成本效益的主动学习 (Cost-Effective Active Learning, CEAL) 算法是首批将主动学习和深度学习结合解决深度图像分类问题的作品之一. 该算法将少量不确定性样本由 Oracle 标记, 而大量高置信度样本则由 CNN 自动分配伪标签, 从而有效降低标注成本. 两种类型的样本随后用于微调 CNN, 并重复更新过程.

测量不同时间输出的差异性 (Temporal Output Discrepancy, TOD)<sup>[76]</sup>, 评估模型在不同优化步骤给出的输出差异来估计样本损失, 即较高的差异对应较高的样本损失, 如图 11 所示. 具体来讲, 在每次迭代中, 训练模型会根据每个样本的损失向后传播误差, 而具有高损失的样本通常会给训练模型的参数带来信息更新. 当真实标注样本不可用时, TOD 可以测量仅依赖训练模型的样本的潜在损失, 从而降低了累积样本损失.

另一问题在于深度模型和浅层模型的学习模式不同, 即深度模型由特征提取阶段和任务学习阶段组成, 传统的基于不确定性的选择策略难以直接应用于深度模型. 仅使用深度模型最后一层的输出来评估样本预测的不确定性是不准确的, 因为深度模型的不确定性实际上由两个阶段的不确定性信息组成. 针对以上问题, 2019 年 Yoo 等人<sup>[74]</sup>提出了用于主动学习的学习损失 (Learning Loss for Active Learning, LLAL) 框架. 该框架将深度模型中间不同隐藏层的特征视为多视图数据, 考虑了目标模型不同网络层之间的不确定性, 使不确定度的评估更加准确, 如图 10 所示. 学习损失预测模块预测未标记数据集的目标损失, 使用 TOP-K 策略选择查询样本. LLAL 方法已适用于当前任务范围较广的深度网络. 除了通过附加网络模块来获取深度模型各层信息之外, 2021 年 Huang 等人<sup>[75]</sup>通过模型在不同时间段输出的差异性来利用深度模型各阶段信息, 并提出了一种新的 DeepAL 方法. 该方法的核心是

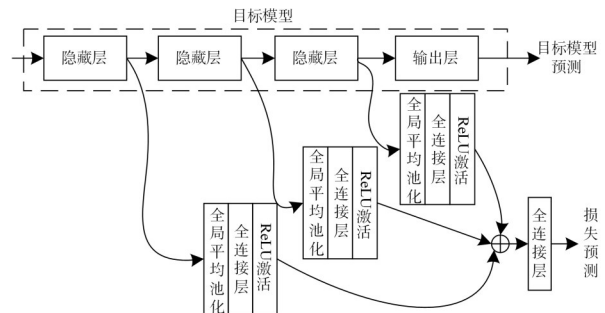


图 10 用于主动学习的学习损失预测模块<sup>[74]</sup>

为了进一步解决深度模型末层输出信息无法有效且准确地评估样本的信息丰富性的问题, 2022 年 Patrick 等人<sup>[77]</sup>提出了深度证据主动学习算法 (Deep Evidential Active Learning, DEAL). 该算法通过将 CNN 的 Softmax 标准输出替换为 Dirichlet 密度<sup>[78]</sup>的参数, 这使模型的输出是 Dirichlet 分布而不是概率分布. 该算法使用 CNN 作为模型, 并使用贝叶斯框架中的证据理论来计算预测的不确定性. 在每次迭代中, DEAL 算法选择最小边缘作为度量标准, 并使用该度量标准选择最具信息量和代表性的未标记数据实例进行标注.

此外, 面对深度模型使用的大规模数据集, 为了保

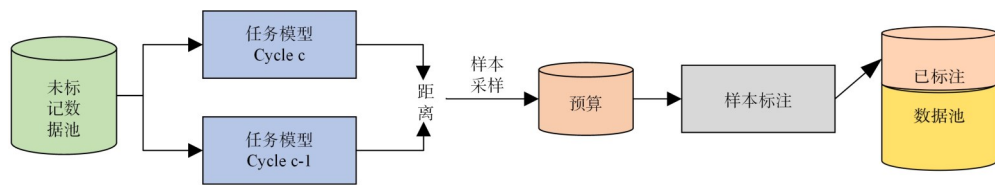


图 11 基于 TOD 的单次迭代步骤<sup>[75]</sup>

证从大型数据集中选择出的训练数据的质量,2021年 Caramalau 等人<sup>[63]</sup>基于 GCN 来获取更高阶的特征表示,并提出了一种用于主动学习的循序图卷积网络(Sequential Graph Convolutional Network, SGCN),如图 12 所示. 图网络中的节点代表数据池中的图像的特征,图网络中的边来编码特征之间的相似性. 作者基于不确定性选择策略提出了 UncertainGCN,该算法利用 GCN 模型对未标记节点进行预测,根据置信度得分选择不确定性最高的节点进行标注. 图卷积神经网络在面对大规模数据集时,需建立更多的图节点来更好地学习图像特征之间的相似性,这无疑会消耗更多的计算资源,提高标注成本<sup>[79]</sup>. 鉴于此,2022 年 Ilić 等人通过结合 CNN 和自校正神经网络(Self-Correcting Neural Network, SCN)<sup>[80]</sup>,提出了一种基于自我修正神经网络的主动学习算法(Active Learning Using a Self-Correcting Neural Network, ALSCN)<sup>[81]</sup>. 该算法中, CNN 仅使用手动标记的数据进行训练,并对未标记的数据进行预测. 而 SCN 使用所有可用数据进行训练,其中部分由手动标记,其余使用网络自动标记. ALSCN 算法的特性使其在处理大规模数据集时,能够通过仅标记数据集的一部分样本,从而显著降低样本标注的成本. 同时,该算法将有差异的样本选出进行手动标记,从而提高训练数据集的质量.

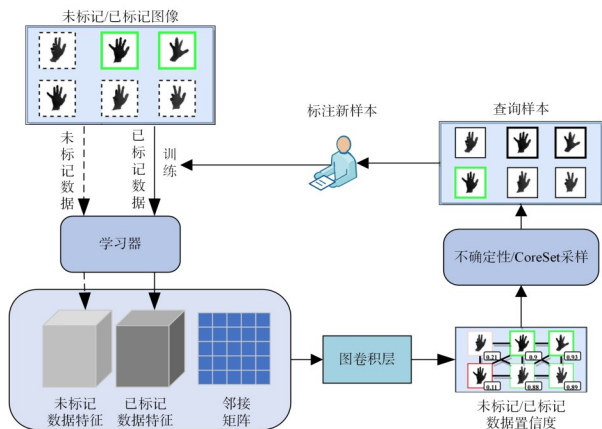


图 12 用于主动学习的循序图卷积网络(SGCN)基本架构<sup>[63]</sup>

在深度学习和主动学习融合中,深度学习主要负责特征信息提取处理,主动学习负责样本选择查询. 深度网络模型的各个隐含层和阶段信息为高价值样本选

择提供了更多依据,深度模型强大的特征表示能力能够有效提高主动学习图像分类算法性能. 然而,主动学习与深度学习处理流程不一致,使两者难以有效结合;且深度学习需要大量标注数据,而主动学习的标注数据较少,容易引起样本分布偏差;在确定哪些样本需要进行标注时,大多基于不确定性的样本策略可能仅仅基于样本的不确定性排名来选择样本,因此在考虑多个样本属性时可能会忽略样本之间的相关性,从而使更新后的已标注训练集具有冗余性<sup>[21]</sup>.

### 3.3.2 基于生成对抗网络

如前所述,在 3.1.1 节中,GAN 网络已被应用于数据增强领域来减少标注图像过程中的成本. 进一步,本节将从优化模型结构的角度来总结近年来的研究成果.

2019 年, Sinha 等人<sup>[47]</sup>依据 VAE 学习隐空间的强大能力提出了变分对抗式主动学习(Variational Adversarial Active Learning, VAAL). 在此方法中,样本选择由对抗性网络执行,该网络判别样本属于已标记池或未标记池,如图 13 所示. VAE 和对抗性网络中的鉴别器被构建为类似于 GAN 的 Mini-Max 博弈<sup>[82]</sup>,使 VAE 被训练来学习特征空间,而鉴别器学习如何选择不确定较高的样本来进行标注. 进一步,在 2021 年 Kim 等人<sup>[83]</sup>对 VAAL 算法进行了改进,并提出了基于任务感知的变分对抗式主动学习网络(Task-Aware Variational Adversarial Active Learning, TA-VAAL). 该算法考虑已标注和未标记数据分布,并使用排序条件生成对抗网络在 VAAL 上嵌入归一化排序损失信息,去除预测输入样本间相对距离. 通过对真实损失信息排序来重塑隐空间,来选择具有较高真实损失值的样本.

另外,已标注数据池处于持续更新过程中,且早期训练中已标注数据池通常很小,这限制了对抗性训练模型的性能. 因此,一些研究者使用样本的状态信息来指示样本是否被标记,该状态信息可直接用作主动学习算法的监督信息. 未标记数据池中不同样本对目标任务具有不同重要性,且未标记样本与标记池中样本越相似,其被标记的优先级越低. 2020 年, Zhang 等人<sup>[84]</sup>结合样本的状态信息提出了一种状态重新标记的对抗式主动学习模型(State Relabeling Adversarial Active Learning, SRAAL),其由表示生成器和状态鉴别器组成. 该生成器利用标注信息生成样本的统一表示,将语义嵌入整个数据表示中. 鉴别器中设计了一个在线

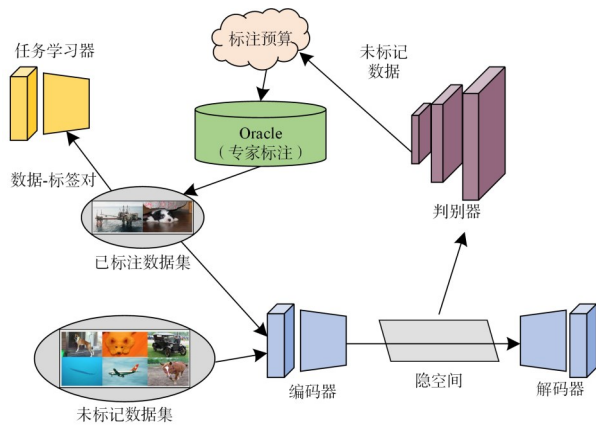


图 13 基于任务感知的变分对抗性主动学习网络基本结构<sup>[47]</sup>

不确定度指示器,指示器计算每个未标记样本的不确定性得分,作为其新的状态标签.因此,可以根据鉴别器的预测状态选择信息量最大的样本.

进一步,2020年Wang等人<sup>[85]</sup>在基于单一GAN算法的基础上提出了一种创新性的算法——用于深度主动学习的双重对抗网络(Dual Adversarial network for deep Active Learning, DAAL).该算法同时考虑了不确定性和代表性两种基本的样本选择策略.与以往需要多阶段数据选择的混合主动学习方法不同,DAAL算法使用不同的采集函数逐步评估不确定性和代表性.这种结构能够在主动学习阶段中选择不确定性最高和最具代表性的数据点,从而在未标记池中准确地选择信息量最大的数据点.

然而,由于GAN模型高度复杂和计算规模较大,现有大多数基于GAN的算法需更高的训练成本.为此,在2020年Mayer等人<sup>[86]</sup>提出一种新的对抗式主动学习方法——对抗性抽样(Adversarial Sampling for Active Learning, ASAL).该算法使用GAN生成高熵样本,并使用特征提取器和最近邻模型从池中检索相似的真实样本.该方法避免了在整个数据集上进行不确定性采样所需的昂贵计算,并且可以在较短的时间内找到最相关的真实样本.因此,ASAL具有比传统不确定性采样方法更低的运行复杂度.

将GAN与主动学习算法结合,不仅可以通过数据增强解决标注不足的问题,而且对抗式的学习方法能够有效提升样本选择策略判别样本信息丰富性的能力.然而,GAN的训练过程可能不稳定,特别是主动学习算法初始迭代阶段,生成器和判别器之间的动态平衡可能很难实现,导致模型难以收敛或难以获得良好的分类性能<sup>[87]</sup>.此外,GAN的训练通常需要更多的计算资源和时间成本.

### 3.3.3 基于强化学习

传统的DeepAL算法由深度学习和主动学习两部

分组成,手工设计这两部分需要大量成本,而且受限于研究人员的经验.并且在传统主动学习流程中,样本选择策略通常被视为固定先验,只有在标签预算耗尽后才能评估其适用性.这使研究者难以动态地调整样本选择策略.因此,一种合理的选择是利用强化学习方法来实现对样本选择策略的动态调整.

鉴于此,2017年Fang等人<sup>[88]</sup>将启发式主动学习算法重新定义为一个强化学习问题.随后,2019年Haussmann等人<sup>[89]</sup>提出强化主动学习(Reinforced Active Learning, RAL)算法,该算法使用贝叶斯神经网络作为样本选择策略的学习预测器.贝叶斯神经网络预测器综合考虑了提供的所有概率信息,并形成全面的概率分布.随后,该概率分布将被传递给贝叶斯神经网络概率策略网络.在每一轮的标注中,贝叶斯神经网络通过接受来自Oracle的反馈进行强化学习.这种反馈被用于微调样本选择策略,以持续提升其性能.2019年,Liu等人<sup>[90]</sup>提出的深度强化主动学习(Deep Reinforcement Active Learning, DRAL)采用了类似的思想.对于每个查询锚点(探针),代理(强化主动学习器)在主动学习算法流程中从数据池中顺序选择实例,并将其交给Oracle以获得带有二进制反馈(正/负)的手动标注信息.状态评估所有实例之间的相似性关系,并根据Oracle反馈计算激励以调整代理查询.

在深度强化主动学习基础上,2019年Sun等人<sup>[16]</sup>将深度卷积神经网络提取图像的特征作为强化学习算法的“状态”,并使用深度Q-learning算法来训练一个Q-网络,根据Q网络的输出来决定是否对数据进行标注.同时,为了进一步优化强化主动学习算法中分类器的动态训练过程,2020年Wang等人<sup>[91]</sup>将主动学习建模为马尔可夫决策过程,并基于Actor-Critic架构的强化学习算法,使用深度确定性策略梯度算法来训练模型.此外,2022年Cui等人<sup>[92]</sup>通过使用一个持续更新的基础分类器和一个激励函数,并使用分类器的后验概率作为其置信度,来决定应该为哪些数据样本进行标注,有效地提升了分类器的性能.进一步,针对深度强化主动学习面临的大规模数据集问题,2022年Zhang等人<sup>[93]</sup>通过将批量主动学习问题定义为一种协作的多代理强化学习问题,提出了一种新颖的批量模式的强化主动学习算法框架.该算法基于图神经网络的批量主动学习设置,其中学习代理可以一次获取多个样本的标签;同时引入了一种值分解方法,将总的Q值分解为单个Q值的平均值,以避免多代理机制可能引发的组合爆炸问题.此外,2023年Chen等人<sup>[94]</sup>基于元框架,将自注意力机制与激励函数整合到深度强化学习结构中,以解决主动学习算法面临的数据相关性和数据不平衡的问题.

强化学习可以使主动学习图像分类算法更具自主

决策能力. 学习代理可以通过与环境的交互, 根据不同的学习任务 and 环境的反馈, 自适应地调整其标注样本选择策略以及动态优化分类器的训练过程. 这种自主决策能力使主动学习图像分类算法可以更好地探索标注样本的空间, 找到对学习任务更有益的样本.

### 3.3.4 基于 Transformer 结构

Transformer 最早应被用于自然语言处理领域, 是一种主要基于自注意机制的深度神经网络. 其由于强大的表示能力, 目前已被广泛应用于计算机视觉任务中<sup>[95, 96]</sup>. 在各种视觉基准测试中, 基于 Transformer 的模型能够表现出比卷积神经网络和循环神经网络等其他类型的网络相当或更好的性能. 尤其, 2020 年 Google 提出的 Vision Transformer 模型在图像分类任务显示出的卓越性能, 引起了更多的研究者的关注. 目前, 部分研究者已将 Transformer 应用于主动学习图像分类任务中.

2021 年, Xie 等人<sup>[97]</sup>提出了一种通用和高效的主动学习算法 (General and Efficient Active Learning, GEAL), 该算法基于预训练 Transformer 模型来进行数据选择. 基于 Transformer 模型的强大表示能力和可迁移性, 该算法可以在不需要额外训练或监督的情况下, 使用单次推理从不同数据集中选择数据. 该方法具有较高的通用性和效率, 能有效提升样本选策略的效率. 然而该方法仅使用一般预训练的 Visual Transformer 来提取图像的特征. 为了进一步利用 Transformer 来提取主动学习模型中的信息, 2021 年 Caramalau 等人<sup>[98]</sup>将 Visual Transformer 作为主动学习流程中的采样器. Visual Transformer 模拟了标记和未标记样本之间的非局部视觉概念依赖关系, 这对识别具有影响力的未标记样本至关重要. 此外, 2022 年 Khan 等人<sup>[99]</sup>通过引入代理模型提出了一种代理模式主动学习 (Proxy Model Ac-

tive Learning, PMAL), 使 Vision Transformer 在主动学习算法中具备更好的适应性. 具体来讲, 该算法使用未标记的数据对 Vision Transformer 进行预训练, 以使其能够更好地理解数据; 并使用代理模型对未标记数据进行分类, 选择不确定性最高的样本进行标记; 使用已标记数据对 Vision Transformer 进行微调, 以使其能够更好地适应主动学习任务.

主动学习图像分类算法基于少量的已标注数据进行模型训练, 而基于卷积神经网络的主动学习图像分类模型在处理相似性较高的图像时, 其分类性能受到一定限制. 针对以上问题, 2023 年 Tang 等人<sup>[100]</sup>提出一种具有学习全局特征的 Transformer, 该方法能够结合卷积神经网络学习局部特征的优势, 从而实现更准确的预测. 尤其是在与主动学习算法相结合的情况下, 该模型甚至在仅使用初始训练集的 30% 的情况下, 显示出了与大多数同类模型在完整训练集上相当的性能水平.

对于主动学习图像分类任务, Transformer 可以有效地捕捉图像中的关键信息和特征, 并将其编码成高质量的表示向量. 这种强大的表示学习能力使 Transformer 能够更好地捕捉样本之间的特征依赖关系, 以及图像中的上下文信息, 从而提高主动学习图像分类算法的性能. 此外, Transformer 的预训练和微调机制使其能够充分利用大规模数据的信息, 并应用到主动学习图像分类任务中, 提升了算法的性能.

### 3.4 分析与总结

为了更好地训练分类模型, 样本选择策略需确定哪些未标记样本应被选择以获最大信息增益. 因此, 本节总结了各算法采用的样本选择策略, 并对这些算法使用的分类器和图像分类数据集进行了总结, 如表 2 所示.

表 2 主动学习图像分类算法总结

算法类别		年份	样本选择策略	分类器	数据集	
基于数据增强的主动学习图像分类算法	基于图像增广	GAAL <sup>[17]</sup>	2017	查询合成+不确定性采样	线性 SVM	CIFAR10, MNIST
		BGADL <sup>[45]</sup>	2019	BALD + 不确定性采样	ResNet18	CIFAR10, CIFAR100, SVHN, MNIST
		ARAL <sup>[48]</sup>	2019	预测的不确定性+代表性采样	VGG16	CIFAR10, SVHN, ImageNet, MNIST
		LADA <sup>[49]</sup>	2021	代表性+不确定性采样	ResNet18	FashionMNIST, SVHN, CIFAR10, CIFAR100
	基于特征插值	文献[55]	2020	不确定性采样	ResNet18	Digits, CIFAR10
		文献[56]	2020	不确定性采样	ResNet18	CIFAR10, MNIST, FashionMNIST, Places365-Standard
		ALFA-Mix <sup>[57]</sup>	2022	代表性采样+不确定性采样	MLP, ResNet-18, DenseNet121, Vision Transformers	MNIST, EMNIST, CIFAR10, CIFAR100, Mini ImageNet

续表

算法类别		年份	样本选择策略	分类器	数据集	
基于数据分布信息的主动学习图像分类算法	文献[59]	2013	信息密度+不确定性采样	逻辑回归	Natural Scene Categories, Caltech101	
	文献[60]	2016	代表性+不确定性采样	深度置信网络	PAVIAU	
	CoreSet <sup>[14]</sup>	2017	数据几何分布+代表性采样	CNN	CIFAR10,CIFAR100,SVHN	
	文献[61]	2017	代表性采样	ResNet50	MNIST,Caltech101,ImageNet	
	DAL <sup>[62]</sup>	2019	代表性采样	LeNet	CIFAR-10,MNIST	
	CDAL <sup>[65]</sup>	2020	不确定性+代表性采样	ResNet18	CIFAR-10,CIFAR-100	
	CoreGCN <sup>[63]</sup>	2021	GCN+代表性采样	ResNet18	CIFAR-10,CIFAR-100,FashionMNIST,SVHN	
	NCE-Net <sup>[67]</sup>	2021	NCE-Net+不确定性采样	ResNet18	CIFAR10,CIFAR100	
	文献[70]	2021	贝叶斯+类不平衡性	ResNet18	CIFAR-10,CIFAR-100	
CBAL <sup>[19]</sup>	2022	均匀分布+不确定性采样	ResNet18	CIFAR-10,CIFAR-100,Tiny ImageNet		
优化模型预测的主动学习图像分类算法	优化深度模型预测信息	CEAL <sup>[73]</sup>	2016	计算标注成本+不确定性采样	AlexNet	Caltech-256,CACD
		LLAL <sup>[74]</sup>	2019	不确定性采样	ResNet18	CIFAR-10
		TOD <sup>[75]</sup>	2021	时间输出差异+不确定性采样	ResNet18	CIFAR-10,CIFAR-100,SVHN
		UncertainGCN <sup>[63]</sup>	2021	GCN+不确定性采样	ResNet18	CIFAR-10,CIFAR-100,Fashion MNIST,SVHN
		ALSCN <sup>[81]</sup>	2022	SCN+不确定性采样	CNN	CIFAR-10,MNIST,Fashion MNIST
		DEAL <sup>[77]</sup>	2022	不确定性+代表性采样	LeNet,ResNet18	CIFAR-10,MNIST
	基于生成对抗网络	VAAL <sup>[47]</sup>	2019	VAE+不确定性采样	VGG16	CIFAR10,CIFAR100,Caltech-256,ImageNet
		TA-VAAL <sup>[83]</sup>	2021	Ranking Loss+VAE+不确定性采样	ResNet18	CIFAR10,CIFAR100 Fashion MNIST,SVHN
		SRAAL <sup>[84]</sup>	2020	状态标记+不确定性采样	ResNet18	CIFAR10,CIFAR100,Caltech-101
		DAAL <sup>[85]</sup>	2020	不确定性采样+代表性采样	Wide-ResNet	CIFAR10,CIFAR100
		ASAL <sup>[86]</sup>	2020	不确定性采样	CNN	CIFAR10,MNIST,SVHN
	基于强化学习	文献[16]	2019	不确定性采样	VGG16	CIFAR10,CIFAR100,SVHN
		文献[91]	2020	不确定性采样	ResNet50	chestCT,Retinopathy
		文献[92]	2022	不确定性采样	随机森林	Opportunity,PAMAP2,SKODA
		文献[93]	2022	不确定性采样	GNN\GCN	Cora,Citeseer,Pubmed,Reddit
		文献[94]	2023	元学习框架+不确定性采样	—	—
	基于Trans-former结构	GEAL <sup>[97]</sup>	2021	多样性采样+代表性采样	ResNet18	CIFAR10,CIFAR100
文献[98]		2021	不确定性采样	VGG16\ ResNet18	CIFAR10,CIFAR100,FashionMNIST	
PMAL <sup>[99]</sup>		2022	不确定性采样	Vit,VGG11,VGG16	Caltech101,Caltech256,CIFAR-10	
文献[100]		2023	不确定性采样	Transformers+CNN	Polyp,Cancer,CVC-ClinicDB	

如前文所述,主动学习图像分类算法是基于少量信息量大的已标注数据进行训练的.数据集的规模和复杂性限制了该方法的有效性.为了克服这些限制,数据增强技术被广泛应用于图像分类算法中,以扩展数据集并减少标注的数据样本数量.尤其是生成性网络可用于生成更多高质量的训练数据.此外,Mixup提供了一种不同模式的图像增广方法.Mixup不仅可以合成新的训练样本,还可利用未标注样本和已标注样本之间的特征信息差异来评估样本信息丰富性,且无需产生额外的计算成本.

在样本选择策略的设计中,评估样本数据的分布情况可有效计算样本的信息量.可以根据数据分布情况调整选择策略,使算法更关注具有代表性的样本.例如,在处理长尾分布数据时,采用类平衡的选择方法可使待标记数据集包含来自各类别的数据.此外,在一些基于难度的策略中,与不确定性采样、信息熵采样等方面的结合,可使算法更关注更难以分类的样本,从而提高算法性能.

深度学习模型与主动学习算法的融合为图像分类任务中的主动学习提供了新的优化方向.在这种融合中,采用附加的网络模块学习深度模型各层之间的损失,进一步优化深度模型的输出结构,以评估样本的不确定性.此外,采用图卷积网络学习标注和未标注样本特征之间的关系,可更好地优化模型的预测结构.特别地,基于GAN的主动学习算法通过生成器的生成能力和判别器的鉴别能力,不仅可对训练数据进行扩充,还可有效地预测待标记样本的信息量.进一步,基于强化学习的主动学习算法可以通过与环境交互,选择最具信息量的样本进行标记,实现算法在不同环境下的动态适应性,从而降低标记数据的需求量.另外,视觉Transformer的提出,使主动学习图像分类模型能够凭借其强大的表示能力来有效地提升模型的预测效果.这些方法为主动学习图像分类算法的进一步研究提供了重要的理论支持.

## 4 实验对比分析

当前研究工作中,对于算法性能评价呈现多样性,即不同的研究者采用不同的模型标注预算.且现有的总结性工作中尚未对算法性能进行实验测试和分析.故本节选取了不同类别下的重要学术工作进行实验对比分析,共在四种公共数据集上测试了算法的性能.此外,实验模拟了现实应用场景下类别数据分布不平衡的情况,构造出类别数据分布不平衡数据集,并在该数据集上进行了测试和分析.

### 4.1 实验设置

为确保实验设置的普适性,本节实验参考了在主

动学习图像分类领域具有重要影响力的学术论文中所描述的对比实验方法<sup>[14,47,63,74,83]</sup>.这些论文的实验设置得到了广泛的认可,为其他研究人员提供了可复现实验和可比较结果的基准.基于表2的数据集统计信息,本文在四个具有广泛代表性的公共数据集上进行实验,包括三个RGB图像数据集(CIFAR10, CIFAR100, SVHN),以及一个灰度数据集(FashionMNIST).本文测试的主动学习图像分类算法包括基于样本随机采样的主动学习算法(Random)、基于数据增强和生成对抗式网络的VAAL<sup>[47]</sup>算法、基于数据分布信息的CoreSet<sup>[14]</sup>和CoreGCN<sup>[63]</sup>算法以及优化模型预测信息的LLAL<sup>[74]</sup>和UncertainGCN<sup>[63]</sup>算法.

对于各算法,将整个训练集视为一个未标记池.随机抽取一个小子集并查询该子集的标签作为冷启动.本次实验共进行了10个周期的子实验,以充分呈现各算法的性能表现.对于CIFAR-10, SVHN和FashionMnist数据集各子周期的样本标记预算为1 000,对于类别较多的CIFAR-100数据集则设置为2 000.均采用ResNet-18作为图像分类网络,使用3次实验的均值作为最终测试结果,以消除实验中的随机性.

### 4.2 主动学习图像分类公用数据集介绍

近年来,现有文献中用于主动学习图像分类的公用数据集主要包括FashionMnist<sup>[101]</sup>, CIFAR10<sup>[102]</sup>, CIFAR100<sup>[103]</sup>及SVHN<sup>[104]</sup>等,如表2所示.表3总结了以上数据集的数据情况.图14展示了各数据集样本示例.

表3 主动学习图像分类常用数据集

数据集	图像尺寸	图像数量	类别数	类内样本数量	训练数量	测试数量
Fashion-Mnist	28×28	70 000	10	7 000	60 000	10 000
CIFAR10	32×32	60 000	10	6 000	50 000	10 000
CIFAR100	32×32	60 000	100	600	50 000	10 000
SVHN	32×32	100 000	10	10 000	73 257	26 032

FashionMnist由德国电子商务公司Zalando提供,包含10个类别的图像,如T恤/上衣、裤子、套头衫、连衣裙、外套、凉鞋、衬衫、运动鞋、包和靴子等,是主动学习图像分类任务中常用的一种灰度图像数据集.

CIFAR10数据集包含10个类别的图像数据.数据均来自现实世界中真实的物体,且同一类别物体的特征和比例都不尽相同,这为分类识别带来很大困难.相对于CIFAR10, CIFAR100数据集则包含100个类别数据,类内样本数据更少且数据模式更为复杂,在分类任务中更具挑战性.

SVHN中的数据来源于谷歌街景图像中的门牌号,实现对0~9的数字识别. SVHN包含了数量级更多的标记数据,并且来自一个非常困难、未解决的现实世界问题——识别自然场景图像中的数字.

在生物医学图像分类领域中,常用的数据集包括 Erie County<sup>[105]</sup>, EEG<sup>[106]</sup>, BreaKHis<sup>[107]</sup>, SVEB 和 SVDB<sup>[42]</sup> 等. 在高光谱图像分类识别,常用的数据集包括 PaviaC、PaviaU, Salinas Valley, Indian Pines<sup>[108]</sup>, Washington DC Mall 和 Urban<sup>[109]</sup>等.

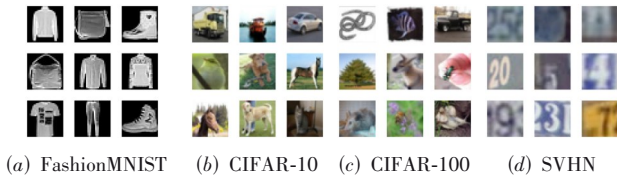


图 14 各数据集样本示例

此外,西安邮电大学图像与信息处理研究所依托与公安部门合作的平台所自建的轮胎花纹图像数据<sup>[110]</sup>,为主动学习图像分类的现实应用研究工作提供了数据支持. 该数据集是目前公开用于学术研究的最大的轮胎花纹数据集,包含轮胎表面花纹数据和轮胎压痕花纹数据各 80 类,每类 30 张不同亮度不同尺度和不同旋转角度的图片,如图 15 所示.



图 15 轮胎花纹数据集样本示例

### 4.3 评估方法

在单标签任务中,分类任务最直观的指标是 Accuracy,即准确率. 其计算公式如式(2)所示. 其中 TP (True Positive)为预测正确,实际为正类; FN (False Negative)为预测错误,实际为负类; FP (False Positive)为预测错误,实际为正类; TN (True Negative)为预测正确,实际为负类<sup>[111]</sup>.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

在此基础上,在多类分类问题中使用 TOP-1 准确率作为评价指标. TOP-1 准确率是指模型在预测中最有可能的类别与真实类别完全匹配的比率,也称为分类准确率. 记样本的类别为  $x_i$ , 测试样本总数为  $N$ , 样本类别标签为  $y_i$ , 预测类别函数为  $f, f(x_i)$  即为预测概率的最大值,则 TOP-1 Accuracy 计算方法如式(3)所示. 本文采用 TOP-1 精度作测试算法的性能指标.

$$TOP-1 Accuracy = \frac{\sum_{i=0}^{N-1} (f(x_i) == y_i)}{N} \quad (3)$$

此外, TOP-5 准确率也是图像分类领域中一种常用的性能评价指标. TOP-5 准确率是指模型的前 5 个最高

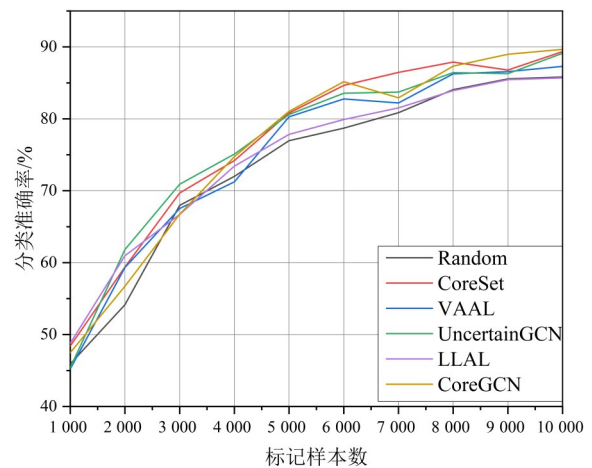
概率答案中的任何一个与预期答案匹配,即模型预测的前 5 个置信度最高的类别中包含真实类别,则认为模型的预测是正确的,如式(4)所示. 相比于 TOP-1 准确率, TOP-5 准确率提供了更宽松的评估. TOP-5 准确率通常用于评估数据规模大且具有多个类别的图像分类任务,如 ImageNet 数据集. 由于主动学习图像分类领域中常用的数据集分类类别数相对较少,故本文后续实验测试中采用 TOP-1 准确率作为性能评价指标.

$$TOP-5 Accuracy = \frac{\sum_{i=0}^{N-1} (f(x_i) \text{ in } y_{i\text{set}})}{N} \quad (4)$$

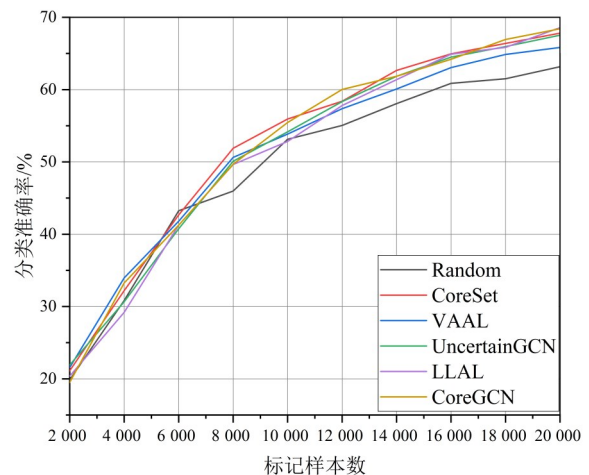
其中,  $y_{i\text{set}}$  表示由模型预测的前 5 个置信度最高的类别标签组成的集合.

### 4.4 在公共数据集上的表现

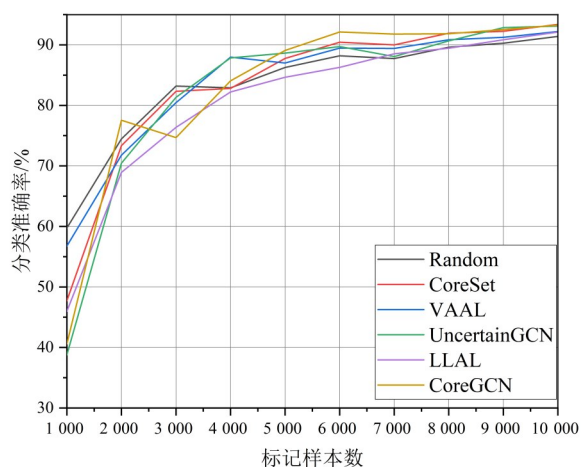
如前所述,随机采样通常用于各种类型算法的性能对比. 其性能表现往往作为对比实验中的下限. 如图 16 所示,各实验方法的最终实验效果均优于随机采样方法.



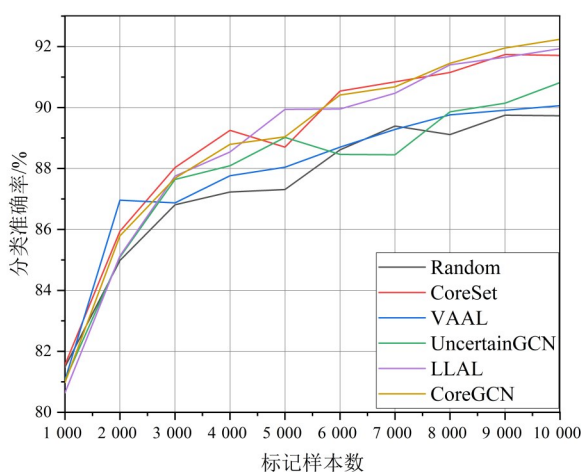
(a) 在 CIFAR10 数据集上的测试结果



(b) 在 CIFAR100 数据集上的测试结果



(c) 在SVHN数据集上的测试结果



(d) 在FashionMnist数据集上的测试结果

图 16 公共数据集测试结果

首先,基于数据分布信息的 CoreSet 和 CoreGCN 算法显示出了强大竞争性. CoreSet 通过定义一个无标签的核心集合选择问题,并提供了一种基于数据点几何形状的严格界限来解决这个问题.在主动学习图像分类算法中,CoreSet 试图选择一个子集,使这个界限最小化.通过选择最具代表性的样本进行标记,CoreSet 方法可以大大降低标记成本,因而在固定标注预算的前提下,选择出的样本能够有效提升分类模型的性能.特别是基于 CoreSet 以及利用 GCN 获取高阶特征表示信息的 CoreGCN 算法,在部分数据集上达到了最优结果.其次,在优化深度模型的预测方面,LLAL 算法通过附加损失预测模块来连接到目标模型的多个层次,以考虑多个网络层的知识进行损失预测来进行不确定性评估.而 VAAL 则利用数据增强和对抗式学习方法为样本不确定性评估提供了更有效的方法,在部分数据集上,性能均优于 LLAL 算法.此外,结合 GCN 的不确定

性算法 UncertainGCN,通过学习图像特征之间的相似性来评估样本的不确定性,其性能更是优于 VAAL 算法.

在不同模式的数据集上,不同算法呈现出不同的效果.然而,结合样本分布信息的算法在各数据集中表现出了较强的竞争性.为了深入分析,在第 4.5 节中,本文将进一步在类别分布不平衡的数据集上进行测试.

#### 4.5 在类别不平衡数据集上的表现

上节实验中,尽管在样本选择之前,本文将未标记的样本随机分配到一个子集,但数据集中各类别中的数据呈现均匀分布,即每个类别可用的图像数量相等.然而,在没有与数据分布相关的先验信息的情况下,这种情况在实际应用场景中并不常见,现实应用场景中的数据资源存在严重不平衡的情况.并且,研究在此场景下的应用对主动学习领域算法的研究具有重大意义.本文在后续的实验中模拟了此场景中的数据,并对各算法的性能进行了分析.

文献[63]提供了一种实验方法设置,在实验过程中将 CIFAR10 数据集中各类别中的数据进行处理从而构成一个不平衡版本的 CIFAR-10 数据集(CIFAR-10im).上节实验中,默认 50 000 个训练样本是未标记的,且给定的 10 个类别中每个类别均有 5 000 个样本.在 CIFAR-10im 数据集中,10 个类中别的 5 个类别包含 10% 的原始数据.因此,新的初始未标记池由 27 500 个图像组成.本实验的参数设置与其他实验保持一致.实验结果如图 17 所示.

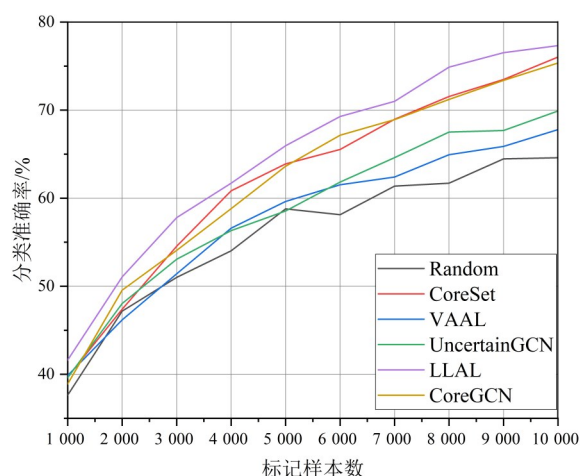


图 17 在类别数据不平衡 CIFAR-10 数据集上的表现

在类别数据分布不平衡的情况下,需要有效评估待标记样本的信息丰富性,并根据数据分布的特征来优化主动学习图像分类模型.如图 17 所示,LLAL 算法结合深度模型不同层级之间的信息,能够更可靠地评

估带标记样本的信息价值,并在不平衡的数据集上取得最佳性能.此外,考虑样本分布的 CoreSet 算法和 CoreGCN 算法也表现出较高的性能.因此,在数据集不平衡的情况下,结合样本分布信息,也是一种有效的方法.

#### 4.6 测试算法对比分析

本节根据各算法在公共数据集 CIFAR10, CIFAR100, SVHN, FashionMnist 以及模拟的类别不平衡数据集 CIFAR10im 上的性能表现,对各类算法的特点

进行了进一步的总结,并以 Random 算法为基准对比其他算法在节约标注成本方面的性能表现,具体如表 4 所示.

根据表 4 中的统计数据,不同类型的算法在不同规模的数据集中表现出显著的差异.因此,在研究过程中,需要针对数据的特性设计有效的方法.特别是在类别数较多或数据分布不平衡的情况下,算法的性能受到限制.因此,研究者需要特别关注这些情况下算法的表现,并可能需要采取适当的样本选择策略以提高算法的性能.

表 4 各测试算法对比

单位:%

算法	原理	性能(节约标注成本比例)				
		CIFAR10	CIFAR100	SVHN	FashionMnist	CIFAR10im
Random	随机选择样本进行标注	—	—	—	—	—
CoreSet	数据点的几何形状来表征选定子集的性能	35	30	25	55	40
VAAL	生成对抗性网络实现数据增强,VAE 学习图像特征的潜在表示,鉴别器选择样本	25	20	15	30	20
LLAL	深度模型不同层级的信息	15	25	10	55	40
UncertainGCN	GCN 构建已标记/未标记样本之间关系+节点熵信息	25	25	15	35	30
CoreGCN	GCN 构建已标记/未标记样本之间关系+CoreSet 采样	25	25	50	45	45

## 5 技术挑战与未来研究趋势

目前,基于主动学习的图像分类算法在分类类别较少、模式较为简单的数据集上表现出较好的分类效果,而面对分类类别较大的数据集,其效果仍不理想.一定的数据处理和模型优化方法在一定程度上能够提升分类器的性能,但对于模式较为复杂的数据集则需要更高的数据标注预算.基于此,本节将介绍目前基于主动学习的图像分类算法面临的挑战,同时针对目前面临的挑战对未来的研究趋势进行讨论.

### 5.1 基于主动学习图像分类面临的挑战

#### 5.1.1 模型的任务无关性问题

目前大多数基于主动学习的学习模型,均为特定的任务而设定.在模型的通用性和可移植性等方面存在一定的不足之处.如第 4 节所示,对于分类类别数目较少的数据集,基于主动学习的图像分类算法表现出了更好性能.对于分类类别数目较多或者数据集存在样本分布不均衡的情况,基于主动学习的图像分类算法的性能将会受到一定的限制.此时,需对以上情况来设定特定的模型,而设计出一个完整的且基于特定的任务模型往往需要较高的成本,也不易于用于其他任务<sup>[112]</sup>.故针对不同类型的数据,如何选择合适的样本选择策略和图像分类器是目前基于主动学习图像分类算法面临的问题之一.

#### 5.1.2 样本选择策略的融合问题

基于不确定性的样本选择策略所选择的样本,其

分布往往存在偏差.而基于多样性的样本选择策略,其成本较高且获得的样本信息量可能较低,无法更高效地提高分类器的性能.故如何将多种样本选择策略进行融合,从而构成更高效、更具普遍适用性的混合策略,是目前研究过程的一大挑战<sup>[113]</sup>.

#### 5.1.3 模型训练模式的问题

根据对本文所介绍算法的深入研究,大部分研究工作每轮样本选择后,均采用重新训练分类模型的策略.然而,从计算资源的角度而言,每个周期内从头开始训练深度模型是难以接受的做法,尤其是针对大规模数据集或复杂的模型架构,这无疑会显著增加训练时间成本.此外,每轮重新训练分类模型还可能导致之前轮次已经获得的知识与信息丧失.因此,如何优化目前主动学习图像分类算法的训练模式成为当前亟待解决的一项重要挑战.

### 5.2 基于主动学习图像分类未来的研究方向

#### 5.2.1 针对模型的任务无关性问题

(1)结合自注意力机制.近年来,Transformer 在计算机视觉领域取得了巨大的成功,其性能已能与 CNN 方法媲美.Transformer 的核心是自注意力机制,并被应用到不同的计算机视觉任务中,如高分辨率图像合成<sup>[114]</sup>、目标跟踪<sup>[115]</sup>、目标检测<sup>[116]</sup>、分类<sup>[117]</sup>、分割<sup>[118]</sup>以及目标识别<sup>[119]</sup>等.随着 ViT<sup>[120]</sup>,BoTNet<sup>[121]</sup>,Swin Transformer<sup>[122]</sup>等架构相继被提出,分类模型的性能得到不断突破.如前所述,已有部分研究者将视觉 Transformer

应用于主动学习图像分类任务中。相较于传统的卷积神经网络,Transformer架构展现出更大的灵活性和可扩展性。不同于依赖固定的卷积操作,Transformer架构采用自注意力机制和前馈神经网络层进行特征提取和表示。这种架构的灵活性使Transformer能够更好地适应各种不同的任务和输入类型,从而提高了其任务无关性。然而,尽管这一架构潜力巨大,但目前在主动学习领域的研究工作还相对较少。未来的研究可以进一步基于Transformer架构,使主动学习模型能够在不同场景的任务中得到更广泛的应用。

(2)结合元学习方法。结合元学习同样能够在一定程度上提高模型的任务无关性。元学习是一种学习如何学习的方法,可以帮助模型适应不同的任务,并且不需要重新训练模型。该学习方法不仅可以提高模型的泛化能力,使其能够更好地适应新任务,还可减少数据量,提高模型的训练效率。未来,将元学习方法与主动学习图像分类算法进一步结合,可有效提升算法的性能。如基于MAML(Model-Agnostic Meta-Learning)<sup>[123]</sup>来学习样本选择策略。MAML可以通过在多个任务上进行元训练来学习一个通用的模型,然后在新任务上进行微调。在每个任务中,MAML都会学习一个新的样本选择策略,以便在该任务上获得最佳性能。基于元学习策略可快速适应新任务,提升模型架构的泛化性能,使样本选择策略能够适应不同任务和数据分布情况等,提升主动学习图像分类算法的任务无关性,使模型在面对新的图像分类任务时能够更加灵活、高效地学习和适应。

### 5.2.2 针对样本选择策略的融合问题

强化样本间的关系。2021年Caramalau等<sup>[63]</sup>通过GCN和建立起图像特征的相似性关系并融合不同样本选择策略,在图像分类任务中显示出了优异的性能表现。未来,可结合注意力机制和图网络来充分学习样本之间的关系,从而对样本数据的分布情况进行更深入的分析<sup>[124,125]</sup>。通过结合图网络的预测信息(基于不确定性)和图网络构建的样本之间的关系(基于代表性和多样性),可将不同样本选择策略进行融合,以获得更优的主动学习效果。

### 5.2.3 针对模型训练模式的问题

基于增量训练模式。目前主流的主动学习算法,均在每个子周期实验中重新训练深度分类模型,这在一定程度上浪费了计算资源。针对主动学习模型训练模式的问题,增量训练法<sup>[126]</sup>在原有模型的基础上,不断增加新的数据进行训练,以更新原有模型的参数和结构,以及增加新的类别,这符合主动学习算法的训练模式。尽管简单的增量训练可能引入模型参数的偏差,但该方法能有效地降低模型的训练成本。目前,与

此相关的研究仍然较为有限,因此该领域仍然是一个具有重要意义的研究方向。

## 6 结束语

本文从主动学习的基本概念出发介绍了基于主动学习的图像分类算法,对主动学习算法中常用的样本选择策略、数据集进行了介绍,并将现有主动学习图像分类算法分为基于数据增强、基于数据分布信息以及优化模型预测的主动学习图像分类算法三大类。其中基于数据增强的主动学习图像分类算法主要通过图像增广来扩充训练数据,并通过对图像特征进行插值处理来进行样本选择。基于数据分布信息的主动学习图像分类算法从未标注/标注数据的分布角度出发,来衡量样本的信息丰富性。优化模型预测的算法充分利用深度模型的结构信息、上下文信息以及时间输出差异信息来有效评估样本的价值性。同时,结合生成对抗网络的特点,对主动学习图像分类算法架构进行优化,以提高模型预测的鲁棒性。另外,基于强化学习策略与环境的交互,强化学习能够指导算法在决策过程中做出更合理、准确和稳定的选择。进一步,基于Transformer模型来捕获更准确和丰富的特征表示,从而改善主动学习图像分类算法的性能。此外,本文通过实验分析总结了不同主动学习图像分类算法的性能。最后,讨论了目前主动学习图像分类算法面临的挑战,并指出了该领域的未来研究趋势。

## 参考文献

- [1] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [2] GU J X, WANG Z H, KUEN J, et al. Recent advances in convolutional neural networks[J]. *Pattern Recognition*, 2018, 77: 354-377.
- [3] EL-SAWY A, EL-BAKRY H, LOEY M. CNN for handwritten Arabic digits recognition based on LeNet-5[C]//International Conference on Advanced Intelligent Systems and Informatics. Cham: Springer, 2017: 566-575.
- [4] BALLESTER P, ARAUJO R M. On the performance of GoogLeNet and AlexNet applied to sketches[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. New York: ACM, 2016: 1124-1128.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2023-05-04]. <https://arxiv.org/abs/1409.1556>.

- [7] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.
- [8] KOTHAWADE S, SAVARKAR A, IYER V, et al. CLINICAL: Targeted active learning for imbalanced medical image classification[C]//Workshop on Medical Image Learning with Limited and Noisy Data. Cham: Springer, 2022: 119-129.
- [9] TUIA D, VOLPI M, COPA L, et al. A survey of active learning algorithms for supervised remote sensing image classification[J]. IEEE Journal of Selected Topics in Signal Processing, 2011, 5(3): 606-617.
- [10] SETTLES B. Active learning literature survey[EB/OL]. (2010-01-26)[2023-05-04]. <http://digital.library.wisc.edu/1793/60660>.
- [11] BELUCH W H, GENEWEIN T, NURNBERGER A, et al. The power of ensembles for active learning in image classification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 9368-9377.
- [12] MACKOWIAK R, LENZ P, GHORI O, et al. CEREALS - cost-effective REgion-based active learning for semantic segmentation[EB/OL]. (2018-10-23) [2023-05-04]. <https://arxiv.org/abs/1810.09726>.
- [13] BRUST C A, KÄDING C, DENZLER J. Active learning for deep object detection[EB/OL]. (2018-09-26)[2023-05-04]. <https://arxiv.org/abs/1809.09875>.
- [14] SENERO O, SAVARESE S. Active learning for convolutional neural networks: A core-set approach[EB/OL]. (2017-08-01)[2023-05-04]. <https://arxiv.org/abs/1708.00489>.
- [15] GAL Y, ISLAM R, GHAHRAMANI Z. Deep Bayesian active learning with image data[C]//Proceedings of the 34th International Conference on Machine Learning. New York: ACM, 2017: 1183-1192.
- [16] SUN L, GONG Y H. Active learning for image classification: A deep reinforcement learning approach[C]//2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI). Piscataway: IEEE, 2019: 71-76.
- [17] ZHU J J, BENTO J. Generative adversarial active learning[EB/OL]. (2017-02-25)[2023-05-04]. <https://arxiv.org/abs/1702.07956>.
- [18] MAHAPATRA D, BOZORGTABAR B, THIRAN J P, et al. Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2018: 580-588.
- [19] BENGAR J Z, VAN DE WEIJER J, FUENTES L L, et al. Class-balanced active learning for image classification [C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2022: 3707-3716.
- [20] LIU P, WANG L Z, RANJAN R, et al. A survey on active deep learning: From model driven to data driven[J]. ACM Computing Surveys, 54(10s): 221.
- [21] REN P Z, XIAO Y, CHANG X J, et al. A survey of deep active learning[J]. ACM Computing Surveys, 54(9): 180.
- [22] ZHAN X Y, WANG Q Z, HUANG K H, et al. A comparative survey of deep active learning[EB/OL]. (2022-03-25)[2023-05-04]. <https://arxiv.org/abs/2203.13450>.
- [23] WANG M, MIN F, ZHANG Z H, et al. Active learning through density clustering[J]. Expert Systems With Applications, 2017, 85: 305-317.
- [24] GORRIZ M, CARLIER A, FAURE E, et al. Cost-effective active learning for melanoma segmentation[EB/OL]. (2017-11-24)[2023-05-04]. <https://arxiv.org/abs/1711.09168>.
- [25] ZHDANOV F. Diverse mini-batch active learning [EB/OL]. (2019-01-17)[2023-05-04]. <https://arxiv.org/abs/1901.05954>.
- [26] NGUYEN H T, SMEULDERS A. Active learning using pre-clustering[C]//Proceedings of the Twenty-First International Conference on Machine Learning. New York: ACM, 2004, DOI: 10.1145/1015330.1015349.
- [27] YANG L, ZHANG Y Z, CHEN J X, et al. Suggestive annotation: A deep active learning framework for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2017: 399-407.
- [28] NÚÑEZ J A, CINCOTTA P M, WACHLIN F C. Information entropy[C]//Chaos in Gravitational N-Body Systems. Dordrecht: Springer, 1996: 43-53.
- [29] FENG Y, SHI Q K, GAO X Y, et al. DeepGini: Prioritizing massive tests to enhance the robustness of deep neural networks[C]//Proceedings of the 29th ACM SIGSOFT International Symposium on Software Testing and Analysis. New York: ACM, 2020: 177-188.
- [30] ZHANG X C, CHEN F L, LU C T, et al. Mitigating uncertainty in document classification[EB/OL]. (2019-07-17)[2023-05-04]. <https://arxiv.org/abs/1907.07590>.
- [31] JAIN P, KAPOOR A. Active learning for large multi-class problems[C]//2009 IEEE Conference on Computer

- Vision and Pattern Recognition. Piscataway: IEEE, 2009: 762-769.
- [32] ZHOU J, SUN S L. Improved margin sampling for active learning[C]//Chinese Conference on Pattern Recognition. Berlin: Springer, 2014: 120-129.
- [33] SHEN W J, LI Y H, CHEN L, et al. Multiple-boundary clustering and prioritization to promote neural network re-training[C]//Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering. New York: ACM, 2020: 410-422.
- [34] YOU X G, WANG R X, TAO D C. Diverse expected gradient active learning for relative attributes[J]. IEEE Transactions on Image Processing, 2014, 23(7): 3203-3217.
- [35] GUO H S, WANG W J. An active learning-based SVM multi-class classification model[J]. Pattern Recognition, 2015, 48(5): 1577-1597.
- [36] LENG Y, QI G H, XU X Y, et al. A new SVM active learning algorithm based on KNN[J]. Advanced Materials Research, 2014, 926/927/928/929/930: 2906-2909.
- [37] YAN L, XU X Y, SUN C L, et al. A SVM active learning method based on confidence, KNN and diversity[C]//2015 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE, 2015: 1-6.
- [38] YE C, WU J, SHENG V S, et al. Multi-label active learning with label correlation for image classification[C]//2015 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2015: 3437-3441.
- [39] FENG C, LIU M Y, KAO C C, et al. Deep active learning for civil infrastructure defect detection and classification [C]//Computing in Civil Engineering 2017. Reston: American Society of Civil Engineers, 2017: 298-306.
- [40] AHMED M U, WOO K J, HYEON K Y, et al. Wild facial expression recognition based on incremental active learning[J]. Cognitive Systems Research, 2018, 52: 212-222.
- [41] HAUT J M, PAOLETTI M E, PLAZA J, et al. Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(11): 6440-6461.
- [42] SAYANTAN G, KIEN P T, KADAMBARI K V. Classification of ECG beats using deep belief network and active learning[J]. Medical & Biological Engineering & Computing, 2018, 56(10): 1887-1898.
- [43] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 1-48.
- [44] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [45] TRAN T, DO T T, REID I, et al. Bayesian generative active deep learning[C]//Proceedings of the 36th International Conference on Machine Learning. Long Beach: ICML, 2019: 6295-6304.
- [46] PU Y C, GAN Z, HENAO R, et al. Variational autoencoder for deep learning of images, labels and captions[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2016: 2360-2368.
- [47] SINHA S, EBRAHIMI S, DARRELL T. Variational adversarial active learning[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 5971-5980.
- [48] MOTTAGHI A, YEUNG S. Adversarial representation active learning[EB/OL]. (2019-12-20)[2023-05-04]. <https://arxiv.org/abs/1912.09720>.
- [49] KIM Y-Y, SONG K, JANG J, I-C MOON. Lada: Look-ahead data acquisition via augmentation for deep active learning[C]//Advances in Neural Information Processing Systems, Virtual Conference: NIPS Foundation, 2021: 22919-22930.
- [50] GONG C Y, WANG D L, LI M, et al. KeepAugment: A simple information-preserving data augmentation approach[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 1055-1064.
- [51] DIPU KABIR H M, ABDAR M, KHOSRAVI A, et al. SpinalNet: Deep neural network with gradual input[J]. IEEE Transactions on Artificial Intelligence, 2022. DOI: 10.1109/TAI.2022.3185179.
- [52] QIAN Y M, HU H, TAN T. Data augmentation using generative adversarial networks for robust speech recognition [J]. Speech Communication, 2019, 114: 1-9.
- [53] ZHANG H Y, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization[EB/OL]. (2017-10-25)[2023-05-04]. <https://arxiv.org/abs/1710.09412>.
- [54] VENKATARAMANAN S, KIJAK E, AMSALEG L, et al. AlignMixup: Improving representations by interpolating aligned features[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 19152-19161.
- [55] MA Y, LU S X, XU E Y, et al. Combining active learn-

- ing and data augmentation for image classification[C]// Proceedings of the 3rd International Conference on Big Data Technologies. New York: ACM, 2020: 58-62.
- [56] WANG D D, LI Y D, WANG L Q, et al. Neural networks are more productive teachers than human raters: Active mixup for data-efficient knowledge distillation from a blackbox model[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 1495-1504.
- [57] PARVANEH A, ABBASNEJAD E, TENEY D, et al. Active learning by feature mixing[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 12227-12236.
- [58] KIM J H, CHOO W, JEONG H, et al. Co-mixup: Saliency guided joint mixup with supermodular diversity [EB/OL]. (2021-02-05)[2023-05-04]. <https://arxiv.org/abs/2102.03065>.
- [59] LI X, GUO Y H. Adaptive active learning for image classification[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 859-866.
- [60] LIU P, ZHANG H, EOM K B. Active deep learning for classification of hyperspectral images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(2): 712-724.
- [61] YANG Y Z, LOOG M. Single shot active learning using pseudo annotators[J]. Pattern Recognition, 2019, 89: 22-31.
- [62] GISSIN D, SHALEV-SHWARTZ S. Discriminative active learning[EB/OL]. (2019-07-15)[2023-05-04]. <https://arxiv.org/abs/1907.06347>.
- [63] CARAMALAU R, BHATTARAI B, KIM T K. Sequential graph convolutional network for active learning[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 9578-9587.
- [64] ZHANG S, TONG H H, XU J J, et al. Graph convolutional networks: A comprehensive review[J]. Computational Social Networks, 2019, 6(1): 1-23.
- [65] AGARWAL S, ARORA H, ANAND S, et al. Contextual diversity for active learning[C]//European Conference on Computer Vision. Cham: Springer, 2020: 137-153.
- [66] WANG X, WANG S, LIANG X X, et al. Deep reinforcement learning: A survey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, DOI: 10.1109/TNNLS.2022.3207346.
- [67] WAN F, YUAN T N, FU M Y, et al. Nearest neighbor classifier embedded network for active learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(11): 10041-10048.
- [68] FROSST N, PAPERNOT N, HINTON G. Analyzing and improving representations with the soft nearest neighbor loss[C]//Proceedings of the 36 th International Conference on Machine Learning. Long Beach: ICML, 2019: 2012-2020.
- [69] OUYANG W L, WANG X G, ZHANG C, et al. Factors in finetuning deep model for object detection with long-tail distribution[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 864-873.
- [70] CHOI J, YI K M, KIM J, et al. VaB-AL: Incorporating class imbalance and difficulty with variational Bayes for active learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 6745-6754.
- [71] OVADIA Y, FERTIG E, REN J J, et al. Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2019: 14003-14014.
- [72] GUO Y J, HU Q, CORDY M, et al. Robust active learning: Sample-efficient training of robust deep learning models[C]//Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI. New York: ACM, 2022: 41-42.
- [73] WANG K Z, ZHANG D Y, LI Y, et al. Cost-effective active learning for deep image classification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(12): 2591-2600.
- [74] YOO D, KWEON I S. Learning loss for active learning [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 93-102.
- [75] HUANG S Y, WANG T Y, XIONG H Y, et al. Semi-supervised active learning with temporal output discrepancy [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2022: 3427-3436.
- [76] MEHTA R, SHUI C J, NICHYPORUK B, et al. Information gain sampling for active learning in medical image classification[C]//International Workshop on Uncertainty

- for Safe Utilization of Machine Learning in Medical Imaging. Cham: Springer, 2022: 135-145.
- [77] HEMMER P, KÜHL N, SCHÖFFER J. DEAL: Deep evidential active learning for image classification[M]//Deep Learning Applications, Volume 3. Singapore: Springer, 2022: 171-192.
- [78] TEH Y W. Dirichlet process[M]//Encyclopedia of Machine Learning. New York: Springer, 2010: 280-287.
- [79] WU Z H, PAN S R, CHEN F W, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.
- [80] KALUZA P, VINGRON M, MIKHAILOV A S. Self-correcting networks: Function, robustness, and motif distributions in biological signal processing[J]. Chaos, 2008, 18(2): 026113.
- [81] ILIĆ V, TADIĆ J. Active learning using a self-correcting neural network (ALSCN)[J]. Applied Intelligence, 2022, 52(2): 1956-1968.
- [82] BERNARD S, BAS P, KLEIN J, et al. Explicit optimization of min max steganographic game[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 812-823.
- [83] KIM K, PARK D, KIM K I, et al. Task-aware variational adversarial active learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 8162-8171.
- [84] ZHANG B C, LI L, YANG S J, et al. State-relabeling adversarial active learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 8753-8762.
- [85] WANG S, LI Y X, MA K, et al. Dual adversarial network for deep active learning[C]//European Conference on Computer Vision. Cham: Springer, 2020: 680-696.
- [86] MAYER C, TIMOFTE R. Adversarial sampling for active learning[C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2020: 3060-3068.
- [87] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [EB/OL]. (2018-02-16)[2023-05-04]. <https://arxiv.org/abs/1802.05957>.
- [88] FANG M, LI Y, COHN T. Learning how to active learn: A deep reinforcement learning approach[EB/OL]. (2017-08-08)[2023-05-04]. <https://arxiv.org/abs/1708.02383>.
- [89] HAUSSMANN M, HAMPRECHT F A, KANDEMIR M. Deep active learning with adaptive acquisition [EB/OL]. (2019-06-27)[2023-05-04]. <https://arxiv.org/abs/1906.11471>.
- [90] LIU Z M, WANG J Y, GONG S G, et al. Deep reinforcement active learning for human-in-the-loop person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 6121-6130.
- [91] WANG J W, YAN Y G, ZHANG Y B, et al. Deep reinforcement active learning for medical image classification [C]//Medical Image Computing and Computer Assisted Intervention - MICCAI 2020. Cham: Springer, 2020: 33-42.
- [92] CUI Y L, HIREMATH S K, PLOETZ T. Reinforcement learning based online active learning for human activity recognition[C]//Proceedings of the 2022 ACM International Symposium on Wearable Computers. New York: ACM, 2022: 23-27.
- [93] ZHANG Y H, TONG H H, XIA Y L, et al. Batch active learning with graph neural networks via multi-agent deep reinforcement learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(8): 9118-9126.
- [94] CHEN S Y, WANG R, LU J. A meta-framework for multi-label active learning based on deep reinforcement learning[J]. Neural Networks, 2023, 162: 258-270.
- [95] TOUVRON H, CORD M, DOUZE M, et al. Training data-efficient image transformers & distillation through attention[EB/OL]. (2020-12-23)[2023-05-04]. <https://arxiv.org/abs/2012.12877>.
- [96] CHEN C F R, FAN Q F, PANDA R. CrossViT: cross-attention multi-scale vision transformer for image classification[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2022: 347-356.
- [97] XIE Y C, TOMIZUKA M, ZHAN W. Towards general and efficient active learning[EB/OL]. (2021-10-15)[2023-05-04]. <https://arxiv.org/abs/2112.07963>.
- [98] CARAMALAU R, BHATTARAI B, KIM T K. Visual transformer for task-aware active learning[EB/OL]. (2021-06-07)[2023-05-04]. <https://arxiv.org/abs/2106.03801>.
- [99] KHAN A, HAQ I U, HUSSAIN T, et al. Pmal: A proxy model active learning approach for vision based industrial applications[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2022, 18(2s): 123.
- [100] TANG S G, YU X Y, CHEANG C F, et al. Transformer-

- based multi-task learning for classification and segmentation of gastrointestinal tract endoscopic images[J]. *Computers in Biology and Medicine*, 2023, 157: 106723.
- [101] XIAO H, RASUL K, VOLLGRAF R. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms[EB/OL]. (2017-08-25)[2023-05-04]. <https://arxiv.org/abs/1708.07747>.
- [102] DOON R, KUMAR RAWAT T, GAUTAM S. Cifar-10 classification using deep convolutional neural network [C]//2018 IEEE Punecon. Piscataway: IEEE, 2019: 1-5.
- [103] KRIZHEVSKY A. Learning Multiple Layers of Features From Tiny Images[D]. Toronto: University of Toronto, 2009.
- [104] NETZER Y, WANG T, COATES A, et al. The Street View House Numbers (SVHN) Dataset[D/OL]. [2023-05-04]. <http://ufldl.stanford.edu/housenumbers/>.
- [105] FOLMSBEE J, LIU X L, BRANDWEIN-WEBER M, et al. Active deep learning: Improved training efficiency of convolutional neural networks for tissue classification in oral cavity cancer[C]//2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). Piscataway: IEEE, 2018: 770-773.
- [106] ALAHMARI S S, GOLDFOG D, HALL L O, et al. Automatic cell counting using active deep learning and unbiased stereology[C]//2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). Piscataway: IEEE, 2019: 1708-1713.
- [107] SPANHOL F A, OLIVEIRA L S, PETITJEAN C, et al. A dataset for breast cancer histopathological image classification[J]. *IEEE Transactions on Biomedical Engineering*, 2016, 63(7): 1455-1462.
- [108] DENG C, XUE Y M, LIU X L, et al. Active transfer learning network: A unified deep joint spectral-spatial feature learning model for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(3): 1741-1754.
- [109] LIN J Z, ZHAO L, LI S Y, et al. Active-learning-incorporated deep transfer learning for hyperspectral image classification[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(11): 4048-4062.
- [110] LIU Y, LEI Y B, RASHID S F. Graph convolution network with node feature optimization using cross attention for few-shot learning[C]//Proceedings of the 2nd ACM International Conference on Multimedia in Asia. New York: ACM, 2021: 1-7.
- [111] DENG X Y, LIU Q, DENG Y, et al. An improved method to construct basic probability assignment based on the confusion matrix for classification problem[J]. *Information Sciences*, 2016, 340/341: 250-261.
- [112] DASGUPTA S, HSU D J, MONTELEONI C. A general agnostic active learning algorithm[C]//Proceedings of the 20th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2007: 353-360.
- [113] COHN D A, GHAHRAMANI Z, JORDAN M I. Active learning with mixture models[M]//Multiple Model Approaches to Modelling and Control. Boca Raton: CRC Press, 1997: 167-183.
- [114] DALMAZO, YURT M, ÇUKUR T. ResViT: Residual vision transformers for multi-modal medical image synthesis[EB/OL]. (2021-06-30)[2023-05-04]. <https://arxiv.org/abs/2106.16031>.
- [115] MEINHARDT T, KIRILLOV A, LEAL-TAIXÉ L, et al. TrackFormer: Multi-object tracking with transformers [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 8834-8844.
- [116] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//European Conference on Computer Vision. Cham: Springer, 2020: 213-229.
- [117] DEY D, CHATTERJEE B, CHAKRAVORTI S, et al. Rough-granular approach for impulse fault classification of transformers using cross-wavelet transform[J]. *IEEE Transactions on Dielectrics and Electrical Insulation*, 2008, 15(5): 1297-1304.
- [118] ZHENG S X, LU J C, ZHAO H S, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 6877-6886.
- [119] LI M, LIU J, ZHENG C, et al. Exploiting multi-view part-wise correlation via an efficient transformer for vehicle re-identification[J]. *IEEE Transactions on Multimedia*, 2023, 25: 919-929.
- [120] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. (2020-10-22)[2023-05-04]. <https://arxiv.org/abs/2010.11929>.
- [121] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition[C]//2021 IEEE/CVF

Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 16514-16524.

- [122] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2022: 9992-10002.
- [123] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//Proceedings of the 34th International Conference on Machine Learning - Volume 70. New York: ACM, 2017: 1126-1135.
- [124] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2017-10-30) [2023-05-04]. <https://arxiv.org/abs/1710.10903>.
- [125] ZHOU J, CUI G Q, HU S D, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57-81.
- [126] WU Y, CHEN Y P, WANG L J, et al. Large scale incremental learning[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 374-382.



**李大湘** 男,副教授,硕士生导师.现为西安邮电大学图像与信息处理研究所成员、无线通信与信息处理技术国际联合研究中心成员.主要研究方向为图像分类识别.

E-mail: [www.ldx@163.com](mailto:www.ldx@163.com)



**许志杰** 男,现任英国哈德斯菲尔德大学(University of Huddersfield)工程与计算机学院教授,视觉计算与虚拟技术研究中心主任,无线通信与信息处理技术国际联合研究中心成员.主要研究方向为图形图像处理.

E-mail: [z.xu@hud.ac.uk](mailto:z.xu@hud.ac.uk)

## 作者简介



**刘颖** 女,2007年获得澳大利亚莫纳什大学博士学位.现为西安邮电大学教授、硕士生导师、图像与信息处理研究所成员、无线通信与信息处理技术国际联合研究中心成员.主要研究方向为图像理解与识别.中国电子学会会员编号:E190091672M.

E-mail: [liuying\\_ciip@163.com](mailto:liuying_ciip@163.com)



**庞羽良(通讯作者)** 男,2021年西安邮电大学通信与信息工程学院硕士研究生.主要研究方向为主动学习图像分类.

E-mail: [yuliangpang0315@163.com](mailto:yuliangpang0315@163.com)



**张伟东** 男,生物医学工程专业博士,副教授,硕士生导师.现为西安邮电大学图像与信息处理研究所成员、无线通信与信息处理技术国际联合研究中心成员.主要研究方向为室内场景理解.

E-mail: [chluzhre@126.com](mailto:chluzhre@126.com)