基于深度学习的开放集识别研究综述

章 秦[1)] 刘紫琪[1)] 张晓林[2)] 张 鹏[3)] 刘 涵[1)] 陈小军[1)]

[1)](深圳大学计算机与软件学院 广东 深圳 518060)

[2)](山东科技大学电气工程与自动化学院 山东 青岛 266590)

[3)](山东科技大学计算机科学与工程学院 山东 青岛 266590)

摘 要 近年来,机器学习研究不断取得突破,促成了大量智能系统的成熟和落地。然而,当前"深度学习+大规模标注数据+完备先验知识"的机器学习范式过度依赖先验知识的完备性,其应用场景局限于静态封闭的专用系统。现实应用环境具有更多开放性和复杂性,例如现实环境中所包含的类别空间在训练期间无法被完全预知且会有新类别在测试阶段不断出现,这使得实际应用场景下的数据构成和分布都极其复杂,无法通过全局分析来保证模型的有效性。为了打破现有机器学习对完备类别信息的过度依赖,对开放集识别问题的研究已成为一个新的趋势。开放集识别将传统分类问题向开放环境下进行扩展,在保证已知类别准确分类的同时,要求模型还可以有效地识别测试阶段新出现的未知类别样本,避免造成大量误分。本文对近年来开放集识别的研究进行了系统调研,聚焦于基于深度学习的开放集识别方法,对经典模型进行了梳理和介绍,并对其分类效果进行了横向对比。

关键词 开放集识别;深度学习;开放域;分类

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2025.00828

Deep Learning Based Open Set Recognition: A Survey

ZHANG Qin[1)] LIU Zi-Qi[1)] ZHANG Xiao-Lin[2)] ZHANG Peng[3)] LIU Han[1)] CHEN Xiao-Jun[1)]

[1)](College of Computer Science and Software Engineering, ShenZhen University, ShenZhen, Guangdong 518060)

[2)](College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, Shandong 266590)

[3)](College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590)

Abstract In recent years, machine learning research has continuously made breakthroughs, which has led to the maturity and implementation of a large number of intelligent systems. However, the current machine learning paradigm overly relies on complete prior knowledge, and its application scenarios are limited to static closed specialized systems. The emergence of new categories is one of the main challenges in the research of dynamic open environments, to which the conventional closed-world machine learning methods can not handle it well. In order to break the excessive dependence of existing machine learning on complete category information, the research on open set recognition problem has become a new trend, which extends the traditional closed-world classification methods to open-world applications. Open set recognition problem requires the classification models classify the learned known classes correctly, and effectively identify unknown class samples that appear in the testing phase simultaneously, to avoid the large number of misclassifications. This article systematically surveys the

收稿日期:2024-04-10;在线发布日期:2025-01-10。本课题得到国家自然科学基金项目(62206179, 62202280, 62106147)、广东省自然科学基金面上项目(2022A1515010129)、深圳市高等院校稳定支持计划(20220811121315001)、山东省自然科学基金项目(ZR2024QF034, ZR2021QF017)、山东省泰山学者工程青年专家项目(TSQN202312196)、深圳市科技计划项目(ZDSYS20220527171400002)资助。**章 秦**,博士,副教授,中国计算机学会(CCF)会员,主要研究领域为开放集识别、自然语言处理、图数据学习。E-mail: qinzhang@szu.edu.cn。**刘紫琪**,硕士,主要研究领域为开放集识别、图数据学习。**张晓林**,博士,教授,主要研究领域为计算机视觉、数字人技术。**张 鹏**,博士,副教授,中国计算机学会(CCF)会员,主要研究领域为计算机视觉、机器学习、情感计算。**刘 涵**,博士,助理教授,主要研究领域为机器学习、自然语言处理。**陈小军**,博士,研究员,中国计算机学会(CCF)会员,主要研究领域为数据挖掘、机器学习。代码汇总:https://github.com/67catl/open_set_recognition_survey。

research on open set recognition in recent years, focusing on deep learning based open set recognition methods. It introduces the classic models by sorting out them into six categories based on the main ideas they employ, and horizontally compares various research achievements, and give the comparison details of their performances. Related paper and code links are collected and available online.

Keywords　open set recognition; deep learning; open world learning; classification

1　引　言

近年来，机器学习研究不断取得突破，促成了大量智能系统的成熟和落地[1-3]。然而，当前“深度学习＋大规模标注数据＋完备先验知识”的机器学习范式，过度依赖先验知识的完备性，其应用场景局限于静态封闭的专用系统。新类别的不断涌现是复杂开放环境研究的重要挑战之一，例如新形成的网络词汇和新发现的微生物种类等。如果训练阶段的类别信息不全，即使在测试阶段出现了新的类别的样本，大多数现有机器学习模型也无法感知这些新类别，反而“过于自信”地将其强制划分到训练阶段出现的某个类别中[4]，造成大规模误判。因此，大多数现有机器学习模型无法直接应用于复杂的开放环境，也难以直接应用于实际生产任务。

为了打破现有机器学习对完备类别信息的过度依赖，基于实际应用的需求，分类模型从封闭集假设到开放集假设已经成为一个新的研究趋势。开放集识别问题(Open Set Recognition，OSR)[5]将传统分类问题在开放环境下进行扩展，它未能在训练阶段收集，但在测试阶段新出现的一个或多个未知类别视为一个增广类，通过建立模型使其在保证已知类样本准确分类的同时，也能有效识别测试阶段出现的未知类别样本，避免传统分类模型将未知类样本强制划分到已知类所造成的大量误分。

开放集识别问题对机器学习模型提出了更高的要求，要求模型构建更加紧凑、更加清晰的分类边界。图1可视化地对比了传统封闭式环境假设下的分类边界和开放式环境下的开放集分类边界。具体地，左图展示了数据的初始分布状态，包括A、B、C、D四个已知类和E、F两个未知类的数据。A、B、C、D四个已知类别的数据是在训练过程中需要用到的训练数据，在测试阶段也会大量出现。E、F两个未知类别的数据在训练过程中不存在，但在测试阶段可能会出现。第二张图片展示了传统封闭式环境假设下分类模型构建的分类边界，A、B、C、D四个已知类别的数据被分类器的决策边界完全分开，E、F两个未知类别的数据将会被决策边界误分类到已知类别中。未知类E可能被识别为C类或D类；未知类F可能会被识别为D类或A类。图1右图则展示了开放式环境假设下的开放集识别模型所构建的分类边界，它会限制其在A、B、C、D四个已知类别的数据周边并为开放域中的未知类别预留空间，避免开放域中的E、F两个未知类别的数据被误分类到A、B、C、D四类中，而是被准确地标注为“未知类别”样本。理想状态下的开放集分类模型应该符合最大类内特征距离小于最小类间特征距离这个理想特征分类标准[5]，从而最大限度地减小复杂开放环境中未知类别对分类模型的冲击。同时，未知类别的多样性和随机性也给开放集识别任务带来了更大的挑战。

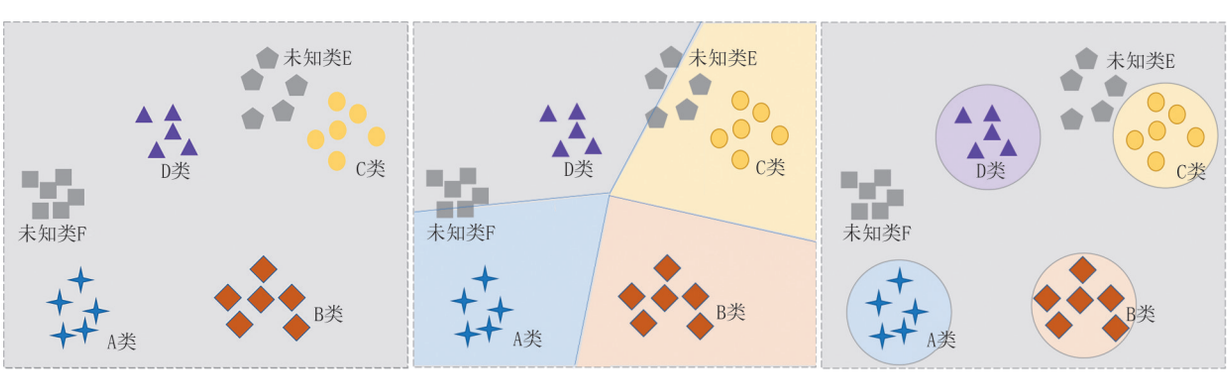

图1　传统分类器与开放式分类器的比较(其中，左图为数据集的初始分布状态，中图为传统封闭式环境假设下的分类边界，右图为开放式环境下的分类边界)

开放集识别问题将传统封闭集分类问题从静态封闭环境扩展到动态开放环境，具有大量的应用景，例如工业领域的零件损毁检测[6]，医疗领域的新药物的发现[7]，信息安全领域的对抗攻击防御[8]，图像处理领域的图像识别[5]、目标检测[9]、图像分割[10]、视频异常检测[11]，自然语言处理领域的新词发现[12]等，都需要开放式方法和技术的支持。以自然语言处理（Natural Language Processing，NLP）中的新词发现问题[13]为例，社会的进步和互联网的普及使得大量未在词典中出现过的新词和新概念频繁出现在网络和生活中，例如“累觉不爱”“人艰不拆”“坑爹”“奥利给”等，这给大量NLP任务带来了挑战。研究表明，中文NLP的最基础问题——中文分词任务中：60%的分词错误是由新词产生的[12]，开放集识别模型可以快速高效地检测出潜在的新词，从而对NLP中的众多基础问题和上层模型提供支持，促进其效果的提升。另一个典型的例子是近年来广受关注的自动驾驶模型[2]。道路驾驶环境作为一种开放式环境，有场景和气候多变、目标种类复杂、物体数量多等特点，要求自动驾驶感知算法能够适应不同的气候和天气变化；识别机动车、非机动车、行人等各类交通参与者；检测和判别障碍物、建筑、坡道、隧道、水域等人工或自然的各类环境；以及各种突然出现的立交系统、施工、拥堵等突发情况。这要求自动驾驶算法具有极强的泛化能力和新事物判别能力，能够快速地应对新出现的各种事物和新特征信息。因此构建适用于开放环境，具有良好分类和辨别性能的开放式学习模型，具有非常重要的现实意义。

基于现实应用的需求，开放集识别问题近年来受到了越来越多研究者的关注，成为机器学习领域研究的热点。在部分研究中，其也被称为增广类学习（Learning with Augmented Classes）[14]，开放世界学习（Open-World Learning）[15]等。该领域也存在一些综述工作，例如Geng等人[16]和Salehi等人[17]从生成式模型（Generative Models）和判别式模型（Discriminative Models）的角度出发对开放集识别方法进行了梳理和介绍；Mahdavi等人[18]从统计模型（Statistical Models）和深度模型（Deep Neural Networks）的角度出发对开放集识别方法进行了梳理和介绍。这三个工作主要聚焦于开放集识别领域2021年及之前的研究工作。随后，高菲等人[19]从模型、度量、增量的角度对开放集识别方法进行了分类和概述。Sun等人[20]从归纳式学习（Inductive Learning）和直推式学习（Transductive Learning）的角度，针对聚焦于图片数据的开放集识别方法做了梳理和归纳，其对深度开放集识别模型的分类思路也沿用了常见的判别式模型和生成式模型的思路。

与已有的综述文章不同，本文聚焦于近期基于深度学习的开放集识别方法，从训练阶段数据使用的角度出发，对该领域的最新进展进行了分类汇总和系统介绍，并对不同类别的方法进行了深入分析。同时，本文还系统介绍了开放集识别问题的评价指标，包括针对该问题的特有评价指标，并对现有主流方法的效果进行了横向对比，展示了详细的实验数据。此外，本文对与开放集识别问题相关和相似的研究问题也进行了简要的介绍和对比分析。具体的，本文的主要内容如图2所示。第2节对开放集识别的问题定义及其关键概念进行介绍；第3节对近年来基于深度学习的开放集识别方法进行分类汇总和系统介绍；第4节介绍开放集识别领域的常用数据集、常用评估指标以及经典方法的效果对比；第5节对与开放集识别相似和相关问题，例如零样本/小样本学习、迁移学习、域自适应学习、主动学习、分布外检测、开放词汇目标检测、开放世界识别等，进行介绍和讨论；第6节对现有研究进行总结，并展望其未来可能的研究方向。

图2 本文组织结构

2 问题定义

传统封闭集分类任务的基本假设是存在充足的有标签训练样本，且训练样本和测试样本独立同分

布(Independently and Identically Distributed，IID)，即样本及其真实类别来自于相同的先验概率分布，如定义1所示。

定义1. 封闭集分类任务。给定一个独立同分布且有标签的训练数据集 $D=\{\boldsymbol{x}_i, y_i\}_{i=1}^{N}, \boldsymbol{x}_i\in X, y_i\in Y=\{1,2,\cdots,K\}$，其中 X 和 Y 分别为样本空间和标签空间。N 为数据集中样本的个数。K 为训练数据集所包含的类别数。传统分类任务的目标即通过学习得到从样本空间到标签空间的映射函数 $f(x): X\xrightarrow{\omega} Y$，其中 w 为映射函数 f 的参数，从而实现对未知数据 $(\boldsymbol{x}_*, y_*), \boldsymbol{x}_*\in X, y_*\in Y$ 的准确预测，即满足

$$y_*=\arg\max_{y} p(y|w,\boldsymbol{x}_*),\ y\in Y \tag{1}$$

开放集识别模型需要对已知类别样本进行正确分类的同时，还可以有效地对测试阶段新出现的未知类别样本进行准确识别。我们用 d 维向量来表示来自不同类别的样本，即样本 $\{(\boldsymbol{x},y), \boldsymbol{x}\in X\subseteq R^d, y\in Y_{tr}\}$，其中 Y_{tr} 包含多个类别，假设共有 K 类。因此样本分类问题转化为在 $(\boldsymbol{x},y)\subset R^{d\times K}$ 空间上的一个概率测度 $P(\boldsymbol{x},y)$。为了简单起见，这里我们假设一个数据样本具有单类特性，即一个数据样本要么属于某个已知类别，要么属于未知类别，不存在既是已知类别样本又是未知类别样本或者既不是已知类别样本也不是未知类别样本的样例。设 $f(\boldsymbol{x}): X\xrightarrow{\omega} Y_{tr}$ 是对已知类别的正样本输入空间的一个可测量的分类函数，将特征向量 x 映射到标签 y。分类的总体目标是寻找一个函数映射 f 能够使得理想风险 R_I 最小化，即：

$$\arg\min_{f} R_I(f):=\int_{R^{d\times K}} L(\boldsymbol{x},y,f(\boldsymbol{x})P(\boldsymbol{x},y)) \tag{2}$$

然而，联合概率分布 $P(\boldsymbol{x},y)$ 是未知的，因此无法直接优化以上目标函数。常用方法是通过经验风险最小化来替代理想风险最小化。值得注意的是，当上述理想风险最小化过渡到开放集识别问题时，模型将开放域中的数据样本标记为任意已知类都是有风险的。因此，我们对开放集识别问题进行如下定义：

定义2. 开放集识别(Open Set Recognition，OSR)。令 $D_{tr}=\{\boldsymbol{x}_i, y_i\}_{i=1}^{N}$ 表示带有标签的训练集，其中 $\boldsymbol{x}_i\in X, X\subseteq R^d$ 表示训练样本，其标签为 $y_i\in Y_{tr}, Y_{tr}=\{1,2,\cdots,K\}$。令 $D_{te}=\{\boldsymbol{x}_i, y_i\}_{i=N+1}^{T}$ 表示测试集，该测试集来自开放环境 $\{R^d, Y_{te}\}, Y_{te}=\{1,2,\cdots,K,K+1,\cdots,M\}$，即测试阶段可能出现训练集中未出现过的未知类别样本。开放集识别的目的是学习一个分类器 $f(\boldsymbol{x}): X\xrightarrow{\omega} Y'$，以最小化期望风险 f^*，其中 $Y'=\{1,2,\cdots,K,unknown\}$。$unknown$ 表示增广类，即所有测试阶段新出现的未知类别 $\{K+1,\cdots,M\}$ 的集合。期望风险 f^* 可定义为

$$f^*=\arg\min_{f\in\mathcal{H}} E_{(x,y)} err(y,f(\boldsymbol{x})) \tag{3}$$

其中，$\mathcal{H}$ 是假设空间，err 是开放集识别学习损失：

$$err(y,f(\boldsymbol{x}))=\begin{cases} I(f(\boldsymbol{x})\neq y), & y\in Y_{tr} \\ I(f(\boldsymbol{x})\neq unknown), & y\notin Y \end{cases} \tag{4}$$

$I(\boldsymbol{x})$ 为指示函数，当 x 为真时，$I(\boldsymbol{x})=1$，否则 $I(\boldsymbol{x})=0$。

考虑开放空间与全空间的比值测度，Sheirer等人[21]进一步定义了开放集风险(Open Set Risk)。

定义3. 开放集风险(Open Set Risk)。假设 Op 表示开放集空间，S_O 表示包括开放集空间和已知类别数据空间的全空间域，那么开放集风险 $R_O(f)$ 定义为

$$R_O(f)=\frac{\int_O f(\boldsymbol{x})\mathrm{d}x}{\int_{S_O} f(\boldsymbol{x})\mathrm{d}x} \tag{5}$$

其中，f 表示可度量的分类函数，当 $f(\boldsymbol{x})=1$ 时，表示测试样本被识别为某个已知类；当 $f(\boldsymbol{x})=0$ 时，表示测试样本不属于任何已知类别。

分类模型 f 将开放空间域中的实例标记为已知类越多，开放空间风险越大。值得注意的是，式(5)中的定义只是开放集风险的理论概率，并没有涉及到损失函数、类条件密度、类先验等具体影响因素。

进一步，针对特定问题或数据领域，Scheirer等人提出了开放度(Openness)概念[5]，如定义4所示。

定义4. 开放度(Openness)。假设 C_{TR} 为训练模型的训练数据类别集合，C_{TE} 为模型测试阶段可能出现的测试数据的类别集合，那么该特定数据领域的开放度(Openness)可定义为

$$Openness=1-\sqrt{\frac{2|C_{TR}|}{|C_{TR}|+|C_{TE}|}} \tag{6}$$

其中，$|\cdot|$ 表示集合中类别的数量。当分类问题的开放度 $Openness=0$ 的时候，该问题退化为封闭集假设下的分类问题。

基于以上定义，我们可以看到对于没有太多约

束的现实应用问题，当测试类别不断增多时，开放集识别问题的开放度会快速增长，无限接近100%。

有了开放集风险和开放度的概念，开放集识别问题也可从下面的角度进行理解：D_{tr}表示训练数据，R_O和R_E分别表示开放集理想风险和经验风险。开放集识别的目标是找到一个可度量的识别函数$f^* \in \mathcal{H}$，可以最小化开放集风险和经验风险，即

$$f^* = \arg\min_{f \in \mathcal{H}} R_O(f) + \lambda R_E(f(D_{tr})) \tag{7}$$

其中，$\mathcal{H}$表示识别函数的假设空间，λ为正则化常数。

3 基于深度学习的开放集识别方法

开放集识别的概念最先在文献[5]中提出，通过基于双平面优化的1-vs-set支持向量机模型来管理开放集风险，同时提出开放集风险是有界的。随着神经网络的发展，基于深度学习的开放集识别模型越来越受到研究者的青睐。现有的针对该领域研究工作的梳理大多聚焦于模型层面，但开放集识别的本质是在数据有限的情况下，如何更好地探索已知类别的精确边界和未知类别所在的开放区域与已知类别区域的联系。在对该任务的探索中，是否有真正的或者模拟的“未知类数据”可利用，对模型的构建思路、训练过程以及整体数据的使用等多方面都有重要的影响。因此，本文从模型训练过程中是否使用或构造辅助数据集的角度出发，将基于深度学习的开放集识别方法分为不依赖辅助数据的开放集识别方法和基于辅助数据的开放集识别方法两大类，如图3所示。

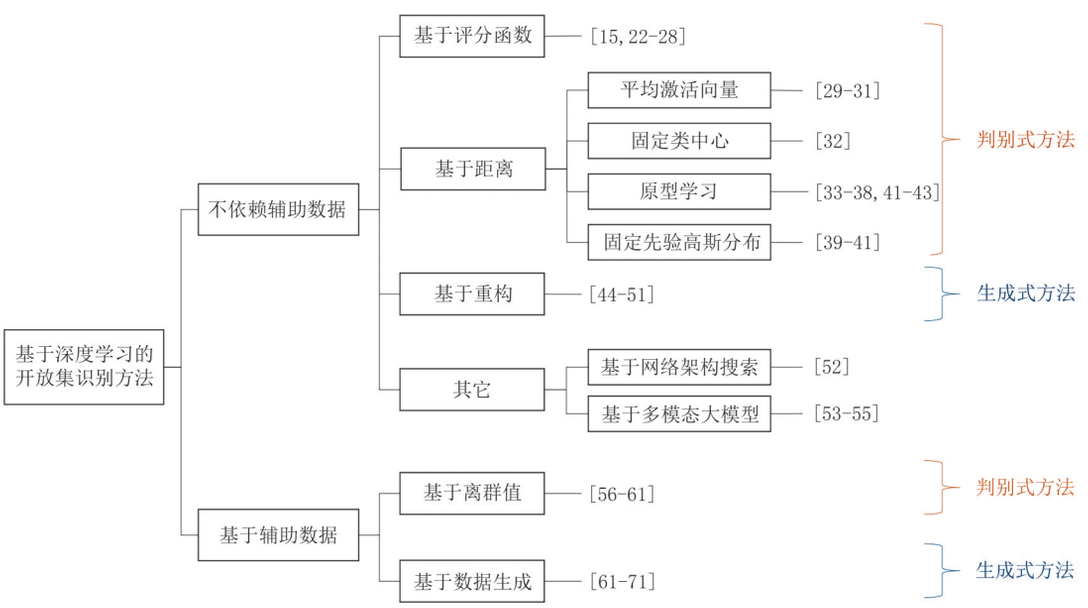

图3 基于深度学习的开放集识别方法分类及其代表方法

在第一类方法中，当前研究又进一步聚焦于以下三个方向：基于评分函数的开放集识别方法[15,22-28]、基于距离的开放集识别方法[29-43]、基于重构的开放集识别方法[44-51]和其他方法[52-55]。在第二类基于辅助数据集的方法中，我们根据辅助数据集的获得方式，进一步将其分为两个子类：基于离群值的开放集识别方法[56-61]和基于数据生成的开放集识别方法[61-71]。其中，基于评分函数、基于距离和基于离群值的开放集识别方法，大多数都是判别式方法；基于重构和基于数据生成的开放集识别方法大多属于生成式方法。判别式方法通常利用判别模型学习和建立划分已知类别和未知类别的决策边界；生成式方法通常利用生成模型对训练数据建模或生成未知类样本，使模型对开放域未知空间具有更加具象的了解。

3.1 不依赖辅助数据的开放集识别方法

在这类方法中，模型仅依赖已知类别的数据，通过对已知类别进行深入细致的表达学习，构建紧凑的类边界和精确的类归属度，实现对已知类别的精准分类和对未知类别样本的准确识别。根据模型构建原理，其又可细分为基于评分函数的开放集识别方法、基于距离的开放集识别方法、基于重构的开放集识别方法以及其它。

3.1.1 基于评分函数的开放集识别

深度学习模型在面对封闭集分类问题时，通常

利用SoftMax函数来预测样本属于各个已知类的概率，并用预测概率值最大的类别标签作为分类结果。在开放集识别问题中，针对来自已知类的测试样本，我们期望模型对其类别的预测具有较大的确定性，即其对应类别的预测概率越大越好；针对来自未知类的测试样本，我们期望其在所有已知类上的概率值低，从而减少其被误分为已知类的概率，这也可以显示模型对该样本具有较大的不确定性。基于这种思想，我们可用SoftMax层输出概率来衡量模型预测的不确定性，并设置合适的阈值实现对未知类别样本的识别。但这个直观的想法在实际中被发现不具有普适性，例如深度网络模型很容易被一些毫无意义的图片愚弄，将无意义的图片以很高的置信度划分到某个已知类别中[4]，即深度模型的过于自信现象。因此，部分研究者尝试对模型预测的置信度增加评分函数，来评估输入样本对类别的归属度，使模型输出的评分和分类准确率能够较好地匹配[22]，或使已知类样本和未知类样本在该分值上有较大的区分度，从而实现更好的开放集识别。

（1）ODIN（Out-of-Distribution Detector for Neural Networks）[23]

ODIN模型[23]的核心设计是利用温度缩放（Temperature Scaling，TS）和测试输入预处理（Input Perturbation，IP）策略加大已知类别样本和未知类别样本的输出置信度之间的差异，使得二分类模型能够更好地对未知类样本进行检测和识别。具体地，该方法在训练阶段学习一个高质量的封闭集分类器。在测试阶段时，不同于常规方法直接输入测试样本来进行预测，ODIN首先在测试样本的特征上加入基于梯度的微小扰动（如公式（8）所示），送入训练好的封闭集模型中。同时，在预测的输出部分，加入温度系数，对原始SoftMax输出进行平滑（如公式（9）所示）。这两个措施可以显著增加测试集中已知类样本和未知类样本的输出置信度之间的差距，进而实现已知类和未知类的区分。

$$\tilde{\boldsymbol{x}}=\boldsymbol{x}-sign\left(-\nabla_{\boldsymbol{x}}\log S_{\hat{y}}(\boldsymbol{x};T)\right) \tag{8}$$

$$S_i(\boldsymbol{x};T)=\frac{\exp\left(f_i(\boldsymbol{x})/T\right)}{\sum_{j=1}^{N}\exp\left(f_j(\boldsymbol{x})/T\right)} \tag{9}$$

其中，$sign(\cdot)$为符号函数，$\exp(\cdot)$为指数函数，$T\in R^{+}$为温度系数，通常在训练阶段设置为1，ε为扰动系数，$S^{\max}(\boldsymbol{x};T)=\max_i S_i(\boldsymbol{x};T)$为SoftMax输出的最大值。

（2）NMD（Neural Mean Discrepancy）[24]

Dong等人[24]经实验发现，小批次的未知类样本和已知类样本的激活向量均值差异明显，足以用于实现未知类样本识别，于是通过结合IPMs（Integral Probability Metrics）[72]提出一个新的度量指标：NMD（Neural Mean Discrepancy）（公式（10）），用于高效地计算模型的平均激活向量，从而实现更好的未知类识别。IPMs是一种分布距离度量指标，可将两组不同的分布利用witness函数映射到新的空间，并使用平均差异作为分布距离。具体地，神经网络f_c^l被用来学习witness函数，并将输入的小批次样本图片$R^{|D_b|\times d'\times d'}$映射到特征图（Activation Maps）$R^{|D_b|\times 1\times d\times d}$。其中，$c$表示第$c$个类别，$l$表示第$l$层网络，$|D_b|$为小批次样本数，$d'$为输入图片的尺寸，$d$为输出特征图的尺寸。

$$\begin{aligned}NMD_c^l(D_b)=&\frac{1}{|D_b|\cdot d^2}\sum_{i=1}^{|D_b|}\sum_{m=1}^{d}\sum_{n=1}^{d}f_c^l(\boldsymbol{x}_i)_{m,n}-\\&\frac{1}{|D_{tr}|\cdot d^2}\sum_{j=1}^{|D_{tr}|}\sum_{m=1}^{d}\sum_{n=1}^{d}f_c^l(\boldsymbol{x}_j)_{m,n}=\\&\mu\left[f_c^l(D_b)\right]-\mu\left[f_c^l(D_{tr})\right]\end{aligned} \tag{10}$$

其中，$\mu\left[f_c^l(D_{tr})\right]$通过每个通道的统计信息批归一化（Batch Normalization，BN）[73]来计算以减少计算开销。最后，将多层的NMD拼接得到$NMD(D_b)=\{NMD_1^1,NMD_2^1,\cdots,NMD_1^l,NMD_2^l,\cdots\}$，用于区分已知类样本和未知类样本。

（3）Energy[25]

Liu等人[25]认为能量（Energy）模型[74]可将每个输入样本映射到单一的标量上（其值被称为能量分数），能量分数可以被理解为未归一化的概率密度的对数，理论上与输入样本的概率密度对齐，能量分数越高说明输入样本在训练过程中出现的可能性越低，因此不太容易出现过度自信的问题，更适合用于识别未知类样本。据此，Liu等人[25]提出基于能量分数的统一框架Energy方法，实现对已知类别和未知类别的区分。假设分类器为$f(\boldsymbol{x})$，其对应第i个类别标签的输出值为$f_i(\boldsymbol{x})$，则能量模型如公式（11）所示

$$Energy(\boldsymbol{x};f)=-T\cdot\log\sum_i^K e^{\frac{f_i(\boldsymbol{x})}{T}} \tag{11}$$

其中，T为温度系数，K为已知类别个数。

能量分数$Energy(\boldsymbol{x};f)$和SoftMax值也有一定联系（如公式（12）所示），最大化SoftMax值等效于同时最小化能量分数$Energy(\boldsymbol{x};f)$和最大化正确标签所对应的输出值$f^{\max}(\boldsymbol{x})$。

$$\log \max_{y} p(y|\boldsymbol{x}) = Energy(\boldsymbol{x}; f) + f^{\max}(\boldsymbol{x}) \quad (12)$$

(4)GEN(Generalized Entropy Score)[26]

当已知类别训练数据保密或封闭集分类器为黑盒模型时,如何仅利用分类器输出的SoftMax值实现对未知类别样本的检测和识别是部分实际应用问题面临的场景和需求。针对这个问题,Liu等人[26]基于广义熵的概念,完全依赖预训练封闭集分类器的SoftMax输出值,提出了一种既简单又有效的分类函数GEN,实现对已知类样本和未知类样本的区分。具体地,假设封闭集分类器在logits空间上的预测分布为$p=\mathrm{SoftMax}(f(\boldsymbol{x}))\in\Delta^K$,广义熵$G$[75]是在logits空间$\Delta^K$上的一个可微非负凹函数。GEN取倒序排序$p_{i1}\geqslant\cdots\geqslant p_{iK}$的前$H$个预测概率来获取广义熵$G_i(p)=-\sum_{j=1}^{H}p_{ij}^{\gamma}(1-p_{ij})^{\gamma}$,并将其作为区分已知类样本和未知类样本的指标,其中$\gamma\in(0,1)$,K为已知类别个数。这里,仅保留前H个预测概率是为了避免广义熵受尾部概率的影响,从而GEN方法可以尽可能放大预测概率与理想的独热编码之间的微小偏差,以简单的方式有效区分已知类别和未知类别样本。

上述方法大多聚焦于计算机视觉(Computer Vision)领域,此外,也有一些针对图学习(Graph Learning)任务的开放集识别方法。例如,Wu等人[15]首次提出开放世界图节点分类任务,并针对该任务提出OpenWGL(Open World Graph Learning)方法。该方法利用变分图自编码器(Variational Graph Autoencoder Network, VGAE)对每个节点进行编码并尝试刻画每个节点的不确定性,通过结合交叉熵损失和类不确定性损失来优化分类效果,增强模型对未知类别的敏感性。Zhang等人[27]使用动态变分自编码器损失(包括KL散度损失和重构损失)和开放世界分类损失来处理结构化序列中的结点分类问题。Zhang等人[28]为解决动态图的开放集结点分类问题提出了DOSSL(Dense Open-world Structured Sequence Learning)方法,通过结合图神经网络和动态变分自编码器捕捉节点的潜在分布,通过结合交叉熵损失、类不确定损失以及原型学习中的密集目标熵损失来学习节点特征的不确定性,辅助开放集识别。

基于评分函数的开放集识别方法将开放集识别任务分解为两个子任务,首先通过二分类器将已知类别样本和未知类别样本区分开,之后将标记为已知类的样本进一步划分到对应的已知类别中。在已有的研究中,基于评分函数的开放集识别方法更多聚焦在第一阶段的二分类任务上,如ODIN方法[23]通过输入前添加基于梯度的小扰动和输入后添加温度系数来区分,NMD方法[24]通过训练好的封闭集分类器的模型平均激活向量作为区分已知类别和未知类别的度量指标,Energy方法[25]通过能量分数实现已知类别和未知类别的区分,GEN方法[26]通过广义熵分数函数来作为区分已知类样本和未知类样本的指标。此类方法通常在第一阶段的二分类任务上寻找更好地区分已知类别和未知类别样本的度量指标,之后引入阈值,将未知类别和已知类别区分开。其潜在的难点除了度量指标的构建和优化之外,阈值的选取也是其难点和弊端之一,当前主要方法大多通过人工进行阈值的选取,如何实现自适应的阈值选取是该类方法可以深入研究的方向之一。

3.1.2 基于距离的开放集识别

在封闭集分类任务中,如图1所示,分类模型会将整个特征空间全部分配给已知类别,导致大量未知类样本的误分。针对这一问题,基于距离的开放集识别模型通过调整样本对已知类别归属度的度量方式对分类空间进行优化,例如通过特殊设计的损失函数、"类原型"的挑选等,增强类间数据的可分离性和类内数据的紧凑性,为未知类别预留足够的空间。具体地,该类方法通过学习每个类别的"类原型"(Class Prototype),例如平均激活向量[29-31]、固定类中心[32]、原型学习[33-38, 41-43]、固定先验高斯分布[39-40]等,并通过样本到类原型的距离来度量样本对该类别的归属度。在模型训练过程中,该类模型不断拉近已知类样本到其对应类别原型的距离,同时拉远该样本到其他类之间的距离,使得学到的已知类别的表达分布和分类界面都更加紧凑,为未知类别提供更多的预留空间。最终,该类模型通过构造K+1类分类器或者通过测试样本到各个类别中心的距离加以合适的阈值来实现对已知类样本的分类和对未知类样本的识别。

(1) OpenMax[29]

针对SoftMax容易出现过于自信的问题,Bendale等人[29]首次提出了基于深度学习的OpenMax方法,期望找到一个特征空间,当模型遇到未知类样本输入时,该未知类样本能在此空间中远离已知类样本的特征分布。具体地,该方法首先计算每个已知类别的平均激活向量作为该类的"类别表示",其激活向量(Activation Vector)为神经网络倒数第二层的输出向量,样本到该类的"类别表

示”的距离则被用来度量该样本到已知类别的归属程度。之后利用EVT极值理论[76]和元识别(Meta-recognition)[77]拟合已知类(K类)的Weibull分布，并利用该分布对原始激活向量进行调整，打破了SoftMax对所有已知类的概率总和为1的限制，直接构建出K+1个类别的OpenMax分数作为最终样本对各个类别的归属度，其中新增加的一个类为未知类。最终，该方法通过结合OpenMax分数和阈值，实现已知类样本的分类和未知类样本的识别。

OpenMax开放集分类器是首个基于深度模型来求解开放集识别问题的方法，具有一定的开创性，但其也存在一些不足，例如OpenMax依赖的激活向量是基于SoftMax和交叉熵损失学习得到的，在特征表示学习模块并未对开放集识别进行适配。

(2) CROSR (Classification-Reconstruction Learning for Open Set Recognition)[30].

对于OpenMax存在的上述问题，Yoshihashi等人[30]在其基础上提出了CROSR方法。该方法同样利用平均激活向量来对类别进行表示，并利用样本到类别中心的距离来拟合Weibull分布，进而实现开放集识别，但该方法在激活向量的学习上进行了优化，使得学习到的激活向量可以更好地辅助开放集识别的实现。具体地，该方法首次结合深度重构网络设计了深度层次重构网络(Deep Hierarchical Reconstruction Net，DHRNet)，使用半监督的方式学习有效的特征表达，通过结合分类损失和重构损失训练网络，提升网络学习特征的能力，让网络能够保留一些在传统分类过程中被丢失，但能够分离已知类和未知类的重要信息。

(3)MetaMax[31].

Lyu等人[31]则认为当数据有限时，OpenMax基于类平均激活向量和各样本到类平均激活向量的距离可能会构造出不准确的Weibull分布，从而提出了一种更加简洁有效的后处理方法MetaMax。MetaMax同样利用极值理论和元识别[76-77]方法来辅助未知类样本的识别。与OpenMax不同的是，MetaMax针对样本的非匹配得分(Non-match Scores)建立Weibull分布，并利用该分布对封闭集分类器的预测进行校准，构建出K+1个类别的分类器。MetaMax不需要计算类平均激活向量，也无需计算样本到类中心的距离，可以更加简单有效地实现开放集识别。

(4)CAC(Class Anchor Clustering)[32].

Miller等人[32]认为OpenMax和CROSR等基于激活向量的方法，虽然相对于传统模型有了显著的性能提升，但其在激活向量空间中的分类界面并不清晰，已知类类边界的可区分性需进一步增强。对此，Miller等人[32]首先利用先验信息将已知类的中心点固定为该类别的独热编码即 $\boldsymbol{c}_i=\left[c_{i0},\cdots,c_{ij},\cdots,c_{iK}\right]$，其中若 $j=i$ 则 $c_{ij}=1$，否则 $c_{ij}=0, i=1,2,\cdots,K$, K 为已知类的类别数。最后利用这些固定类中心构建Class Anchor Clustering损失函数 L_{CAC}：

$$\begin{aligned}L_{CAC}(\boldsymbol{x},y)&=L_T(\boldsymbol{x},y)+\lambda L_A(\boldsymbol{x},y)\\L_T(\boldsymbol{x},y)&=\log\left(1+\sum_{j\neq y}^{N}e^{d_y-d_j}\right)\\L_A(\boldsymbol{x},y)&=d_y=\left\|f(\boldsymbol{x})-\boldsymbol{c}_y\right\|_2\end{aligned}\tag{13}$$

其中，$(\boldsymbol{x},y)$为输入样本，d_y、d_j分别为$\boldsymbol{x}$到类别y和类别j的类中心的距离，λ为平衡损失L_T和L_A的权重参数，L_T为修正的Tuplet损失[78]，促使已知类样本紧密围绕类中心，同时拉大到其他类中心的距离，L_A则进一步直接约束样本到其正确类别类中心之间的距离。

这些策略使得已知类在激活向量空间中的类边界更加紧凑，为未知类预留更多空间，从而使已知类和未知类更易区分，较大地提升了模型的开放集识别能力。Abouzaid等人[79]进一步将其应用在材料表征的介电材料场景中，验证了其在具体应用场景里的有效性和泛化性。

(5)CPN(Convolutional Prototype Network)[33].

Yang等人[80]和Snell等人[81]提出基于卷积原型学习(Convolutional Prototype Learning，CPL)的方法，来解决开放世界的多类识别问题以及分类器的鲁棒性问题。原型学习旨在模拟人类的认知系统，以分类任务为例，每个类别的原型相当于人类大脑中相应类的抽象记忆，输入样本通过神经网络提取深层特征，并与每个已知类别的原型进行匹配，进而获得类别决策。在开放集识别任务中，如果一个测试样本的深层特征和所有已知类别的原型都不能较好地匹配，则其为未知类别样本的可能性更大。

具体地，CPN模型包括一个特征提取器 $f(x;\theta)$(通常为卷积神经网络)和若干个类别原型 $C=\{c_{ij}, i=1,2,\cdots,K, j=1,2,\cdots,H\}$，其中$i$为已知类别的标签下标，$j$为相应类别的原型下标，$K$为已知类别数，$H$为原型个数，其损失函数定义为

$$L_{CPN}=l\big((\boldsymbol{x},y);\theta,C\big)+\lambda pl\big((\boldsymbol{x},y);\theta,C\big)$$

$$l\big((\boldsymbol{x},y);\theta,C\big)=-\log\frac{\sum_{j=1}^{H}e^{-\gamma d\left(f(\boldsymbol{x}),\boldsymbol{c}_{yj}\right)}}{\sum_{i=1}^{K}\sum_{k=1}^{H}e^{-\gamma d\left(f(\boldsymbol{x}),\boldsymbol{c}_{ik}\right)}}$$

$$pl((\boldsymbol{x},y);\theta,C)=\|f(\boldsymbol{x})-\boldsymbol{c}_{yj^*}\|_2^2 \tag{14}$$

其中，$j^*=\arg\min_j\left\|f(\boldsymbol{x};\theta)-\boldsymbol{c}_{yj}\right\|_2^2$。$L_{CPN}$前项是基于样本特征表达$f(\boldsymbol{x})$到原型$\boldsymbol{c}_{ij}$之间距离$d\big(f(\boldsymbol{x}),\boldsymbol{c}_{ij}\big)$的交叉熵损失，可以拉大不同已知类之间的距离；后项则将样本的特征表达拉近到其标签所对应的原型上，使得同类样本分布更加紧凑，减小类内样本之间的距离。二者结合使得模型所学到的特征具有更好的鲁棒性，也可更加准确地拒绝未知类测试样本。CPN[33]方法虽然较好地优化了类内距离，但未考虑特征空间中的反向信息，仍存在较大的开放空间风险。

(6)RPL(Reciprocal Point Learning)[34].

Chen等人[34]则认为仅学习已知类别的表达是不够的，还需要学习反向知识。例如除了学习什么是猫，还需要学习什么不是猫。以此为出发点，提出反向点学习(Reciprocal Point Learning，RPL)方法，通过对每个已知类别的潜在开放空间(反向点)进行建模，间接地将未知类的信息引入只有已知类的网络中。因此，对于类k来说，反向点被用来建模和学习不属于类k的空间。新样本与反向点的距离越大，越可能属于类k，越小则越不可能属于类k。

具体地，RPL方法对每个已知类k构建H个反向点$P^k=\{p_i^k|i=1,2,\cdots,H\}$，并进一步引入可学习边界$\boldsymbol{R}^K$和开放空间风险正则化$L_o$来捕捉基于反向点的未知风险，从而将已知类的类外嵌入空间通过反向点和可学习余量限制在一个有界范围内。其最终损失函数L_{RPL}为

$$L_{RPL}(\boldsymbol{x};\theta,P,\boldsymbol{R})=L_c(\boldsymbol{x};\theta,P)+\lambda L_o(\boldsymbol{x};\theta,P,\boldsymbol{R})$$

$$L_c(x;\theta,P)=-\log\frac{e^{\gamma d\left(f(\boldsymbol{x};\theta),P^k\right)}}{\sum_{i=1}^{K}e^{\gamma d\left(f(\boldsymbol{x};\theta),P^i\right)}}$$

$$L_o(\boldsymbol{x};\theta,P^k,\boldsymbol{R}^k)=\frac{1}{H}\cdot\sum_{i=1}^{H}\left\|d\big(f(\boldsymbol{x};\theta),\boldsymbol{p}_i^k\big)-\boldsymbol{R}^k\right\|_2^2 \tag{15}$$

其中，γ为超参，L_c为分类损失，最大化已知类别的样本与该类反向点之间的距离，L_o是开放空间损失。整个损失函数可以将已知空间推到全局开放空间的外围，然后尽可能地将已知空间和未知空间分开，学习更紧凑的已知类样本表达，减小开放空间风险。

(7) ARPL (Adversarial Reciprocal Points Learning)[35].

Chen等人[35]进一步基于RPL提出了对抗性反向点学习(Adversarial Reciprocal Points Learning，ARPL)方法。在RPL的基础上，采用多个已知类别之间的对抗来降低经验分类风险，利用对抗性的边际约束，限制反向点构建的潜在开放空间来降低开放空间的风险。同时，该方法还引入生成对抗网络来生成未知样本，利用实例的对抗增强，根据反向点和已知类别之间的对抗性机制，提升模型对未知类的可区分性。

(8)PMAL(Prototype Mining and Learning)[36].

上述基于原型学习的开放集识别方法由于对每个类别只学习一个具有判别性的原型，若所学习到的原型在特征空间上更接近低质量的样本，这将会影响开放集识别的性能。Lu等人[36]提出原型挖掘和学习的方法PMAL，旨在学习到具备多样性的原型，以能够代表一个类中各种各样的外观。首先利用已知类数据训练U个分类器$\{\mathcal{C}^u\}_{u=1}^U$(假设$U=2$)，通过马氏距离计算输入样本$\boldsymbol{x}_i$与其它$N$个样本的相对距离$t(\boldsymbol{z}_i)\triangleq\big(d_M(\boldsymbol{z}_i,\boldsymbol{z}_1),\cdots,d_M(\boldsymbol{z}_i,\boldsymbol{z}_N)\big)$，$d_M(\boldsymbol{z}_i,\boldsymbol{z}_j)=\sqrt{(\boldsymbol{z}_i-\boldsymbol{z}_j)\Sigma^{-1}(\boldsymbol{z}_i-\boldsymbol{z}_j)^T}$判断该样本是否可以作为高质量的候选原型即可量化为$r(\boldsymbol{x}_i)\triangleq\exp(-\left\|t(\boldsymbol{z}_i^1)-t(\boldsymbol{z}_i^2)\right\|_2)$，$r(x_i)$越接近1说明该样本的质量越高可以作为候选原型，越低则说明该样本质量越低。选出高质量的候选原型集$C=\{C_i\}_{i=1}^K$之后需要以原型多样性为原则确定最终原型$P=\{P_k\}_{k=1}^K$(公式(16))，避免高冗余。

$$P_k=\bigcup_{i=1}^{H}\{x_i|\max_{x_i\in C_k}\{\min_{x_j\in C_k}\mathrm{d}_M(\boldsymbol{z}_i,\boldsymbol{z}_j)|r(\boldsymbol{x}_j)>r(\boldsymbol{x}_i)\}\} \tag{16}$$

其中，H为原型个数。最后，PMAL通过最小化损失函数L_{PMAL}(公式(17))进行原型学习。

$$L_{PMAL}=L_{cls}+\lambda_p L_p$$

$$L_p=\frac{1}{N}\sum_{i=1}^{N}[\,d(\boldsymbol{z}_i,\boldsymbol{z}(\boldsymbol{P}_m))-d(\boldsymbol{z}_i,\boldsymbol{z}(\boldsymbol{P}_u))+\delta\,]_+ \tag{17}$$

其中，L_{cls}表示SoftMax损失，λ_p为平衡系数，δ为可训练的参数P_u表示其他类别中离P_m最近的原型，$d\big(\boldsymbol{z}_i,\boldsymbol{z}(\boldsymbol{P}_k)\big)=1-\frac{\boldsymbol{z}_i^T\boldsymbol{z}_i^{att}(\boldsymbol{P}_k)}{\|\boldsymbol{z}_i^T\|\|\boldsymbol{z}_i^{att}(\boldsymbol{P}_k)\|}$，$\boldsymbol{z}_i^{att}(\boldsymbol{P}_k)=\mathrm{SoftMax}\left(\frac{\boldsymbol{z}_i^T\boldsymbol{z}(\boldsymbol{P}_k)}{\sqrt{d}}\right)\boldsymbol{z}(\boldsymbol{P}_k)$为Lu等人设计的一种新

的自注意机制(Self-attention)[82]，它考虑到输入样本到类别所有原型之间的相关性，能够更全面地测量点到集合之间的距离。训练结束后通过对SoftMax输出的概率值，或对点到集合的最小距离$\min\{d(z_i, z(P_k))\}$设置阈值拒绝未知类别。

(9)ODL(Orientational Distribution Learning)[37].

上述基于距离的方法大多利用单一特征空间中的无向欧几里得距离构建损失函数并进行优化，但未考虑空间分布的潜在影响。Liu等人[37]针对这一问题提出基于层次空间注意力(Hierachical Spatial Attention)的定向分布学习ODL方法。层次空间注意力可在特征空间中对不同层级的特征进行加权，增强模型对重要特征的关注和学习。具体的，ODL通过以下损失函数对特征空间进行约束，学习更清晰的类别边界：

$$L_{total}=L_{CE}+L_{Intra}+\lambda_{Inter}L_{Inter} \tag{18}$$

$$L_{Intra}=\frac{1}{b}\sum_{i=1}^{b}\|z_i-c_{y_i}\|_2^2+\frac{1}{b}\sum_{i=1}^{b}\sum_{j\neq y_i}\omega_j(1-\cos(\vec{v}_i,\vec{v}_j)) \tag{19}$$

$$L_{Inter}=\frac{1}{K}\sum_{i=1}^{K}\max\left[m-\min_{y\neq i}\|c_i-c_j\|^2,0\right]+\frac{1}{K}\sum_{i=1}^{K}\sum_{o\neq i,j}\omega_o\left(1+\cos\left(\vec{v}_s,\vec{v}_o\right)\right) \tag{20}$$

其中，L_{CE}为交叉熵损失，$C=\{c_y\}_{y=1}^{K}$为K个已知类别的可学习的类中心，$\vec{v}_i=c_{y_i}-z_i$表示样本特征与对应的类中心的空间关系，ω_o，ω_j为权重系数，m，b，λ_{Inter}为超参。L_{Intra}用于约束类内距离，使同一类别的样本在特征空间中更加紧密地聚集在一起；L_{Inter}用于约束不同类别在特征空间中的分离程度，促使不同类别在特征空间中的分离更加明显。通过层次空间注意力机制，ODL模型可以有效地整合不同层级的特征信息，增加开放集识别模型的鲁棒性和泛化性。

(10) JNICS (Jacobian Norm and Inter-Class Separation for Open-Set Recognition)[38].

Park等人[38]探讨了样本表征的雅可比范数$\|\frac{\partial f(x)}{\partial x}\|_F$和类间/类内学习之间的关系，提出了JNICS方法实现开放集识别。该文先从理论进行分析，发现类内学习减少了已知类样本的雅可比范数，类间学习则增加了未知样本的雅可比范数。因此，模型训练的关键在于已知类和未知类之间雅可比范数的差异，这将有利于开放集识别的实现。基于这种理解，作者提出了边际一对多(marginal One vs-Rest，m-OvR)损失函数，通过防止类间原型和类间梯度的崩溃，增强类间的可分离性。综上所述，JNICS利用雅可比范数和边际一对多损失函数来提升模型的开放集识别能力，为开放集识别任务提供了重要的理论基础和方法指导。

(11) CAVECapOSR (Conditional Variational Capsule Network for Open Set Recognition)[39].

胶囊网络[83]摒弃传统卷积神经网络的全连接层，集合一组神经元表示特定对象的各种属性，如位置、大小、方向、纹理等。该方法利用胶囊网络来接收卷积神经网络的特征表示，并且使用K个不同的先验高斯分布对已知类别进行建模，输出一组表示神经元的胶囊向量$C(\mu^{(k)},\sigma^{2(k)})$，其中，整个胶囊网络$C^{(k)}$产生$K$个均值$\{\mu^{(k)}\}_{k=1}^{K}$和$K$个方差$\{\sigma^{2(k)}\}_{k=1}^{K}$。然后通过最小化公式(22)中的损失函数使得已知类别的类间围绕特定的高斯分布中心紧凑分布，并与其他类别的分布中心尽可能地拉开距离，为未知类别留下空间。

$$L_{KL}(x,y)=\frac{1}{K}\sum_{k=1}^{K}D_{KL}[C^{(k)}\|sg[\mathcal{N}_y^{(k)}]]$$

$$L_{contr}(x,y)=\frac{1}{K-1}\sum_{k\neq y}^{K}[m_k-d(sg[C],\mathcal{N}_k)]^+ \tag{21}$$

$$L(x,y)=L_{KL}(x,y)+\alpha L_{contr}(x,y)+\beta L_{rec}(x) \tag{22}$$

其中，$sg[\cdot]$表示不计算梯度符号，N表示K个不同的先验高斯分布，m_k、α、β为超参，$[\cdot]^+$表示返回参数的正数部分。L_{KL}使得概率胶囊C靠近正确类别的分布T_y，L_{contr}使得C尽可能地离其他类别的分布$T_{\neq y}$远，$L_{rec}(x)=\|\hat{x}-x\|_2^2$为重构损失。测试阶段设定阈值，当$\max_k\{d(C,T_k)\}$小于阈值时拒绝其输入；反之，取最大值下标作为已知类标签。

(12) MGPL (Multiple Gaussian Prototypes Learning)[40].

为了学习到更接近真实世界的复杂分布，Liu等人[40]提出了高斯混合原型学习方法MGPL，并且整个框架都是源于贝叶斯推理，为开放集识别任务提供理论支撑。具体地，输入样本x通过编码器f_ϕ获得高斯混合原型$c=(\mu_c,\sigma_c,y)$，其中y为类别标签，μ_c，σ_c为原型高斯分布$\mathcal{N}(\mu_c,\sigma_c)$的均值和方

差，训练中将均值μ_c作为可训练参数，方差σ_c简化为单位矩阵I，每个类别分布由H个高斯原型表示。随后通过解码器f_θ生成重构图像$\hat{\boldsymbol{x}}$，生成式约束如下所示：

$$E_{x,\,y\sim D}[-\log p_\theta(x|y)]=\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2-E_{q(c|\boldsymbol{z},\,y)}\sum_{i=1}^{d}\frac{1}{2}\left[(\mu_i(\boldsymbol{x})-\mu_i^c)^2+\sigma_i^2-\log\sigma_i^2-1\right]-E_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\sum_{c\in C_y}\left[-q\left(c|\boldsymbol{z},\,y\right)\log H\left(q\left(c|\boldsymbol{z},\,y\right)\right)\right] \tag{23}$$

其中，d为潜在特征表示的维度，$q\left(c|\boldsymbol{z},\,y\right)=\frac{\exp\left(-\gamma d\left(\boldsymbol{z},\,c_{yj}\right)\right)}{\sum_{l=1}^{C}\exp\left(-\gamma d\left(\boldsymbol{z},\,c_{yl}\right)\right)}$计算$\boldsymbol{z}$到类别标签$y$的$H$个高斯原型的分配概率，$d(\boldsymbol{z},\,c)=KL(\mathcal{N}(\mu_x,\,\sigma_x)\|\mathcal{N}(\mu_c,\,I))$，$\gamma$为平滑输出概率的温度系数。第一项目的在于通过生成的方式保留更多的原始信息，第二项强制通过KL散度使得潜在特征$\boldsymbol{z}$聚拢到正确标签所对应的原型周围，第三项使得每个类别的原型概率呈均匀分布，更利于未知类检测。然而，生成式约束在网络学习中本质上对判别式特征不利，最小化判别式约束（公式(24)）会缩短输入样本到正确标签对应的原型之间的距离，拉大输入样本到不正确标签对应的原型之间的距离，使得高斯混合原型可以直接用于分类任务而不需要额外训练分类器。

$$q_\phi\left(y|\boldsymbol{x}\right)=\frac{\sum_{j=1}^{H}\exp(-\gamma d(\boldsymbol{z},\,\boldsymbol{c}_{yj}))}{\sum_{k=1}^{K}\sum_{l=1}^{H}\exp\left(-\gamma d(\boldsymbol{z},\boldsymbol{c}_{kl})\right)} \tag{24}$$

MGPL模型的整体损失如公式(25)所示：

$$L_{MGPL}=\lambda E_{x,\,y\sim D}\left[-\log p_\theta\left(\boldsymbol{x}|y\right)\right]+(1-\lambda)E_{x,\,y\sim D}\left[-\log q_\phi\left(y|\boldsymbol{x}\right)\right] \tag{25}$$

其中，λ为平衡两个约束的超参，在测试阶段通过对最小的特征向量到原型的距离$\arg\min_{c_{ij}}\{d(\boldsymbol{z},\,\boldsymbol{c}_{ij})\}$设置阈值实现开放集识别。

除了聚焦于图像数据的开放集识别方法之外，也有一些研究聚焦于图数据或者文本数据上的开放集识别模型。例如，Zhang等人[41]提出了复杂噪声下[84]图数据的鲁棒开放集节点分类模型，ROG_{PL}（Robust Open-Set Graph Learning via Region-Based Prototype Learning）。该方法基于原型学习，通过相似性基础的标签传播纠正噪声标签，并通过学习非重叠区域的开放集原型来解决类内变异和类间混淆问题，从而提高了在开放集场景中的分类性能。在开放集文本识别任务中，Zhang等人[42]采用原型匹配方法来提取视觉特征，并通过可训练的原型来建模已知类别，从而实现已知类别和未知类别的区分。Huang等人[43]通过探索不同的未知类原型生成器，发现基于注意力机制的生成器效果最佳，并且加入语义信息进一步提升开放集识别性能。

基于距离的开放集识别方法通过学习已知类别的类原型，从各个角度对类原型进行刻画。例如，OpenMax[29]、CROSR[30]和MetaMax[31]等利用激活向量，即神经网络倒数第二层的输出向量作为特征空间来区分已知类别和未知类别的开放集识别方法，相比于传统的封闭集分类器有一定的开放集分类效果，但在激活向量空间上的分类边界仍然不够清晰，限制了利用激活向量作为原型的开放集识别方法在性能上的提升。CAC等[32]利用固定类中心作为类原型，相较于利用激活向量作为原型的开放集识别方法，通过设计新的损失函数使得已知类别之间紧凑可分，同时为未知类别留下更多空间，能够有效地在特征空间上提升分类边界的可分性，进而提升开放集识别性能。CPN[33]、RPL[34]、ARPL[35]、PMAL[36]、ODL[37]和JNICS[38]等利用原型学习，学习类似于人脑中对于每个类别的抽象记忆（对应神经网络中对每个类别的抽象特征）作为原型，通过结合对反向知识的学习或原型挖掘等其他方法，使得所学习到的类原型的特征具有更好的鲁棒性，更能够模拟真实世界的情况，是一种更高效可行的基于距离的开放集识别方法。CVAECapOSR[39]和MGPL[40]则利用固定先验高斯分布，学习混合高斯原型，更接近真实世界的复杂分布，并依赖贝叶斯推理为开放集识别方法提供理论依据，使得学习到紧凑可分的特征空间。在该空间中，已知类别围绕先验分布中心紧凑分布，并与其他类别的分布中心尽可能地远离，为未知类别预留了更多空间。综上所述，这些类原型的学习和优化方法，促使特征空间中已知类别围绕类原型分布紧凑，同时加大不同类别原型之间的距离，使得已知类别之间边界清晰，为未知类别留下更多的空间。

在实际应用中，该类模型通过计算输入样例对不同类原型的归属度，并结合阈值来区分已知类别和未知类别，实现开放集识别。例如，OpenMax[29]、CROSR[30]、MetaMax[31]、CAC[32]、CPN[33]、RPL[34]、ARPL[35]等使用一组训练样本和验证数据集上采样的一组未知类样本，对阈值进行网格搜索，选

择效果最好的模型对应的阈值；PMAL[36]、CVAECapOSR[39]等通过交叉验证的方式选择最好的阈值；ODL[37]、MGPL[40]等选择确保95%的训练集合可被识别为已知类时所对应的阈值。该类方法现有的研究也存在着不足，例如大多数现有研究对类原型的设计较为固定和简单，对数据分布的假设较为乐观，面对真实世界的复杂数据分布，该类方法在大量现实应用中的开放集识别性能存在较大的限制。

3.1.3　基于重构的开放集识别

基于重构的开放集识别方法主要通过模型对已知类和未知类样本在重构效果上的差异性，以样本重构质量为依据实现对未知类样本的识别，并通过自监督学习的方式挖掘数据的潜在结构特征，学习嵌入的“不确定性”。通常来说，是在已知类数据上训练的编码器-解码器模型对已知类和未知类中的测试样本具有不同的重构质量，模型对训练过程中已见过类别的样本重构效果要远好于未见过类别样本的重构效果[85]。因而，模型输出的差异可用作识别未知类样本的重要参考指标。

(1)C2AE(Class Conditioned Auto-Encoder)[44]

C2AE[44]通过结合FiLM(Feature-wise Linear Modulation)方法[86]和类别标签对自编码器进行调节，并利用重构效果来增强未知类和已知类之间的差异性。FiLM层由两个神经网络H_γ和H_β组，线性调制输入特征。对于一个随机小批次的输入样本$x=\{x_1,x_2,\cdots,x_b\}$，该方法首先对封闭集分类模型的训练，利用已知类样本和交叉熵损失来训练编码器和分类器；然后固定编码器的模型参数，经过编码器特征提取F后得到潜在特征向量$z=\{z_1,z_2,\cdots,z_b\}$，z和条件向量l_j(如公式(26)所示)通过FiLM层得到处理后的特征向量$z_{l_j}=H_\gamma(l_j)\odot z+H_\beta(l_j)$，将$z_{l_j}$输入到解码器$G$中获得重构后的图片$\hat{\boldsymbol{x}}=G(\boldsymbol{x})$。

$$l_j(y)=\begin{cases}+1, y=j,\\ -1, y\neq j,\end{cases}\quad y,j\in\{1,2,\cdots,k\} \tag{26}$$

基于条件向量的重构损失则如公式(27)所示，

$$L_{C2AE}^{recons}=\frac{\alpha}{b}\sum_{i=1}^{b}\left\|\boldsymbol{x}_i-\widehat{\boldsymbol{x}}_i^{\,m}\right\|_1+\frac{(1-\alpha)}{b}\sum_{i=1}^{b}\left\|\boldsymbol{x}_i^{\,nm}-\widehat{\boldsymbol{x}}_i^{\,nm}\right\|_1 \tag{27}$$

其中，α为超参，$\boldsymbol{x}_i$表示输入的训练样本，$\widehat{\boldsymbol{x}}_i^{\,m}$表示使用正确条件向量重构的样本图片，$\widehat{\boldsymbol{x}}_i^{\,nm}$表示使用不匹配的条件向量重构的样本图片，$\boldsymbol{x}_i^{\,nm}$表示训练数据中的真实样本但给予了额外的不匹配标签。最后利用极值理论对重构损失进行建模，找到最优的区分已知类样本和未知类样本的阈值。

(2)GFROSR(Generative-discriminative Feature Representations for Open-set Recognition)[45]

GFROSR方法利用生成模型和自监督学习来构建一个具有更加丰富特征的深度空间，使得模型可以学到已知类样本更加高级的特征，例如语义和结构属性等，进而增加已知类样本和未知类样本输出概率的差距，实现更好的开放集识别。具体地，该模型首先利用已知类样本训练一个生成模型，对每个样本$\boldsymbol{x}$进行重构，得到重构图片$\hat{\boldsymbol{x}}$；再将输入图片和重构图片拼接得到$\boldsymbol{z}=[\boldsymbol{x},\hat{\boldsymbol{x}}]$，并在扩充后的特征空间中利用交叉熵损失构建分类模型。该生成模型对已知类样本的重构效果会优于未知类样本的重构效果，从而使得在扩充的特征空间中，已知类和未知类的样本具有更大的区分性。此外，该模型还引入了自监督[87]，通过对输入样本进行随机的几何变换，构建自监督子模型来预测变换类型，促使模型学习样本的更多特性，例如形状、结构、语义等。

(3)CGDL(Conditional Gaussian Distribution Learning)[46]

Sun等人[46]利用变分自编码器提出条件高斯分布学习的方法CGDL，该方法框架主要包括编码器F、解码器G、封闭集分类器C和未知类检测器D。编码器F由L层的概率阶梯结构(Probabilistic Ladder Architecture)组成，为保留中间层可能消失的信息以提取高层次抽象的潜在特征，在第l层编码器中输入上一层的输出$\boldsymbol{x}_{l-1}$将产生$\boldsymbol{x}_l$、均值μ_l和方差σ_l^2，最后潜在特征z由第L层编码器输出的均值和方差得到$z=\mu_L+\sigma_L\odot\varepsilon$，其中$\varepsilon\sim\mathcal{N}(0,I)$，所得到的潜在特征$z$将输入到封闭集分类器、解码器和未知类检测器中。解码器接收潜在特征z后输出重构的$\hat{\boldsymbol{x}}$。损失函数如公式(28)所示：

$$L_{CGDL}=-\left(L_{cls}+\beta L_{KL}+\lambda L_{rec}\right) \tag{28}$$

$$L_{KL}=\frac{1}{L}\left[D_{KL}\left(q_\phi(z|\boldsymbol{x},y)\|p_\theta^{(y)}(z)\right)+\sum_{l=1}^{L-1}D_{KL}\left(q_\theta(\hat{\boldsymbol{x}}_l|\hat{\boldsymbol{x}}_{l+1},\boldsymbol{x})\|q_\theta(\hat{\boldsymbol{x}}_l|\hat{\boldsymbol{x}}_{l+1})\right)\right] \tag{29}$$

其中，β在训练阶段由0到1线性增加，λ为常数，L_{cls}表示SoftMax损失，L_{rec}表示使用L_1距离的重构损失$\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_1$，$L_{KL}$如公式(29)所示，第一项表示强制条件

后验概率分布$q_\phi(z|\boldsymbol{x},y)$接近不同的$K$个多元高斯模型$p_\theta^{(y)}(\boldsymbol{z})=\mathcal{N}(\boldsymbol{z};\mu_y,I)$,$\boldsymbol{\mu}_y$通过全连接层将输入的第$k$标签的独热编码(One-hot Label)映射到潜在空间所得到的;第二项表示在中间层也使用KL散度。

最后测试阶段计算输入样本到各分布的概率以及重构损失,假定潜在空间向量$\boldsymbol{z}$的维度为d,第k类别的多元高斯分布模型为$f_k(\boldsymbol{z})=\mathcal{N}(\boldsymbol{z};\mu_k,\sigma_k^2)$,$\mu_k$和$\sigma_k^2$为第$k$类别所有正确分布的样本在特征空间上的均值和方差,则输入样本到各分布的概率计算公式为$P_k(\boldsymbol{z})=1-\int_{\mu_0-|z_0-\mu_0|}^{\mu_0+|z_0-\mu_0|}\cdots\int_{\mu_d-|z_d-\mu_d|}^{\mu_n+|z_n-\mu_n|}f_k(t)\mathrm{d}t$,最后当$P_k(\boldsymbol{z})<\tau_l$或者$R>\tau_r$时拒绝为未知类,否则输出封闭集分类器概率最高所对应的类别标签。

(4)OpenHybrid[47]

基于流的模型(Flow-based Model)允许神经网络可逆,可以无监督地通过最大似然估计拟合训练样本的概率分布,根据输入样本的概率密度能够预测这个样本属于已知类还是未知类别。OpenHybrid方法使用基于流的生成模型进行异常检测,利用分类器和密度估计器以端对端的方式学习联合特征空间,确保内部分类不受异常检测影响的同时,提升识别未知类别样本的能力。具体而言,OpenHybrid包括一个编码器,用于将输入样本编码到联合嵌入空间中;一个分类器,用于对已知类样本进行细分;以及一个基于流的模型作为密度估计器,用于检测样本是否属于未知类别。在训练过程中,OpenHybrid通过最小化分类器和密度估计器的损失函数来联合训练模型,并使用负样本采样技术来平衡正负样本的数量,进一步提高模型性能。

(5)GMVAE(Gaussian Mixture Variational Autoencoder)[48]

该方法通过直接构建$(K+1)$类分类器,其中第$K+1$类用于识别测试阶段可能会出现的所有未知类样本实现开放集识别。具体地,该方法基于前面介绍的CROSR[30]方法中构建的深度层次重构网络DHRNet,引入高斯混合变分自编码器(Gaussian Mixture Variational Autoencoder, GMVAE),在嵌入空间中同时进行样本重构和基于类别的相似度学习,扩展变分自编码器的无监督学习框架,为每个类别引入高斯混合先验,并在类别内进行子聚类,以提供更灵活的重构策略,从而实现了更加准确和稳健的开放集分类。综上所述,GMVAE方法充分挖掘了数据的潜在结构特性,利用对嵌入学习的“不确定性”和不同类别中子集数量等信息,获得了更好的分类特征,突破了大量工作对类别嵌入是凸集且只由单个中心点进行表示的局限性。

(6) MOODCAT (Masked Out-of-Distribution Cather)[49]

Yang等人[49]认为未知类样本与已知类样本的语义信息具有较大的差异性,可通过自监督重构的方法来辅助未知类样本识别,并提出模型无关的未知类识别模型MOODCAT。它可以和任意封闭集分类器结合,实现开放集识别。具体地,MOODCAT由随机模块遮盖、生成网络和二类分类器(或图像质量评估模型IQA[88])组成。总体流程分为三个部分,首先对输入样本$\boldsymbol{x}$进行随机遮盖得到$\boldsymbol{x}_m=M(\boldsymbol{x})$;然后利用生成网络$\boldsymbol{G}=E\bullet D$生成新样本:先经过编码器获得特征向量$\boldsymbol{z}=E(\boldsymbol{x}_m)$,使用KL散度对潜在特征向量$\boldsymbol{z}$进行优化,并用高斯分布对其进行重采样,即$\boldsymbol{z}=\mu(\boldsymbol{x}_m)+\Sigma(\boldsymbol{x}_m)\bullet\epsilon$,其中$\epsilon\sim N(0,1)$;同时语义标签$y$被用于训练解码器$D$,并对$\boldsymbol{x}$进行重构,得到$\boldsymbol{x}'=D(\boldsymbol{z},y)$;最后使用二类分类器或图像质量评估模型IQA来实现对已知类样本的识别。其中,二类分类器中利用输入样本与正确标签生成的样本作为正样本对,利用输入样本与随机不匹配的标签生成的样本作为负样本对进行训练;图像质量评估模型IQA是对生成图像的质量进行打分,通过设置阈值的方式拒绝生成效果差的图像为未知类样本。

(7)CSSR(Class-Specific Semantic Reconstruction)[50]

与传统基于距离的方法不同,该方法通过将原型学习与自编码器结合,用类定制自编码器学习该类的流形表达获得类原型。具体地,CSSR在骨干网络的顶部为每个类别加入该类别独有的自编码器,对每个已知类别进行建模,基于自编码器对语义特征的重构误差实现样本对类别的隶属度度量,从而指导模型学习更有区分性和更有类代表性的特征信息。此外,与传统的通过自编码器重构原始图像不同,CSSR中的自编码器流形是特定类别的可学习特征,因此重构误差(也就是点到流形的距离)可当作激活向量用于分类。在实现过程中,CSSR模型利用交叉熵损失和自编码器的重构误差对模型进行约束。该方法通过丢弃不必要的信息和重构语义特征,而非原始样本,来解决分类退化的问题;同时通过学习特定类的流形来释放被吞噬的类间区域,减

小了开放集风险。与其它基于距离的方法相比，CSSR模型通过学习特定类的流形很好地处理了类代表不足的问题。这不仅打破了类的高斯假设，而且比使用单个特征点表示类别更能保留类的关键信息。

除了聚焦于图像数据的开放集识别方法之外，也有一些聚焦于文本数据的开放集识别模型。例如，Liu等人[51]通过重建字形图像，可以保留字符的结构信息，增强模型对字符形状的感知能力，从而提高开放集文本识别的性能。

基于重构的开放集识别方法大多利用自编码器实现对样本的重构，并利用已知类别和未知类别的重构效果差异作为区分它们的指标。此类法聚焦的关键问题在于如何学习良好的样本表示和自编码器，使得可以通过重构样本来拉大已知类别和未知类别之间的差异。例如，C2AE[44]利用类别标签数据对自编码器进行调节，结合重构效果来增强已知类别和未知类别之间的差异；GFROR[45]利用重构和自监督学习获得语义、结构属性等更高级的特征，进而增大已知类别和未知类别之间重构效果的差异；CGDL[46]利用变分自编码器和条件高斯分布学习，最后利用重构损失加阈值的方法区分已知类别和未知类别；OpenHybrid[47]利用基于流的生成模型检测未知类别，结合分类器用于已知类别的分类；GMVAE[48]通过结合深度层次重构网络和高斯混合变分自编码器，构造$K+1$类分类器实现开放集识别；MOODCAT[49]对部分输入图片进行遮盖，在标签语义的监督下对输入图片进行重构，最后利用重构损失作为区分已知类别和未知类别的指标；CSSR[50]结合自编码器和原型学习，针对每个已知类别学习一个单独的自编码器作为类原型，接入深度神经网络骨干的顶部，最后通过自编码器的重构损失作为输入样本到各个类别的归属度来区分已知类别和未知类别。综上所述，基于重构的开放集识别方法利用标签数据[44,49]、自监督学习[45,47-49]、条件高斯分布学习[46]、原型学习[50]等，获得更丰富的语义、结构属性等信息，使得测试阶段未知类数据输入所得到的重构效果远远低于已知类别数据输入所得到的重构效果，从而很好地提升了开放集识别的效果。这类方法也有一定的局限性，例如在重构学习过程中，模型容易受到噪声数据或者与类别无关数据的影响，在学习类别信息的同时，也容易学习到与分类无关的信息，从而影响封闭集分类的精度。

3.1.4　其　他

前面所介绍的开放集识别方法大多集中于损失函数的设计、特征空间的构造以及类别隶属度的学习等方面，LNAOSR（Learning Network Architecture for Open-Set Recognition）[52]方法则从神经网络体系结构出发，通过网络架构搜索和变分自编码器对比学习，以紧凑衰减概率模型的理念为基础，寻找最适用于开放集问题的网络。

近年来，随着大规模预训练模型（Large-scale Pre-training Model）的飞速发展，多模态大模型（Vision-Language Model）通过在海量多源异构数据上的训练和学习，获得了强大的知识理解和推理能力，可协助大量下游任务的实现和效果的提升。鉴于预训练任务与下游任务之间普遍存在的数据集和学习目标等差异，提示微调（Prompt Tuning）等技术被进一步引入[89]，以辅助多模态大模型在下游任务中的适配。在开放集识别任务中，一些研究者也对其与大模型的结合进行了探索。例如，CoHOZ（Contrastive multimodal prompt tuning for Hierarchical Open-set Zero-shot recognition）方法[53]通过构建一个全局兼容的层次标签树来识别类别，利用对比连续提示调优来检测未知数据，并通过手工提示从粗粒度到细粒度进行零样本分类，从而实现对未知类别的详细语义检测。R-Tuning方法[54]通过引入来自WordNet的开放词汇，扩展了提示文本的范围，以减轻提示学习中存在的标签偏差和模拟开放集场景，并且该方法通过结合组合调优和测试（Combinatorial Tuning and Testing，CTT）策略，使之在小规模和大规模数据集上有效提升性能。A2Pt方法[55]通过引入跨模态引导激活（Cross-modal Guided Activation，CGA）模块和反关联校准模块，有效消除与已知类别无关的混淆，增强了已知类别与未知类别的区分度。综上所述，基于预训练大模型和多模态大模型，利用大模型微调、对比提示调优、开放词汇扩展和跨模态引导等技术，可以有效提升开放集识别模型对未知类别的检测和区分能力，较大地推动了开放集识别技术的发展，也是此领域当前和未来的研究热点之一。

3.2　基于辅助数据的开放集识别方法

在现实世界中，存在大量来源各异的数据，除了已知类别的数据之外，一些与已知类别“无关”或“无用”的数据被证实可以用来提升开放集识别模型的泛化性[57]，这些更具多样性的数据使得模型可以获得更好的开放集识别效果。因此，一些学者通过选取或者

构造辅助数据对开放集识别原有的已知类别数据进行数据多样性的增强，探索基于辅助数据的开放集识别方法。值得注意的是，基于辅助数据的开放集识别任务与小样本学习和域自适应学习任务都有明显的区别。小样本学习[90]旨在基类（一组具有重组训练样本的类别集）上进行训练，之后在类别不交叉且仅有极少样本的新类数据集上进行适配和使用，新类别的类别信息及少量样本可获得。域自适应学习[91]的任务是先在拥有大量标注样本的源域上进行模型训练，之后将其迁移到较难获得充足标注数据、但与源域有较大相关度的目标域上，且不引起模型性能的大幅下降。模型在训练阶段通常可获得大量无标注的目标域样本。基于辅助数据的开放集识别任务中，未知类的边际信息、相关的域信息以及样本均无法获取，只是尝试在训练时加入与已知类别无关但现实世界中存在的“无关”数据作为辅助数据，协助开放集识别模型对未知区域的探索和理解。

具体地，根据辅助数据的获取方式，已有研究可进一步分为两类：基于离群值的开放集识别方法和基于数据生成的开放集识别方法。基于离群值的方法利用现实世界中已有的非已知类别数据作为辅助数据，基于数据生成的方法直接针对性地生成伪未知类数据或特征作为辅助数据，构建开放集识别分类器。

3.2.1 基于离群值的开放集识别

基于离群值的开放集识别的方法最早由Hendrycks等人[57]提出，又称“离群值暴露（Outlier Exposure，OE）”，旨在通过使用额外的辅助数据集来训练开放集分类器，其中辅助数据集是与测试集中未知类样本无重合、无关联的其他数据。该方法的可行性在于现实应用中总是存在大量无关样本和额外信息，这些样本和信息构成的辅助数据集具有较好的类别和信息的多样性，在辅助已知类特征学习的同时，也可以帮助模型更加深入地认识未知空间，将非已知类的信息泛化到未知类分布，使得模型的鲁棒性和开放集识别性能均得到显著提升。其核心优化函数如公式（30）所示：

$$L_{OE}=\mathbb{E}_{(x,y)\sim D_{in}}\Big[L\big(f(\boldsymbol{x}),y\big)\Big]+\lambda\mathbb{E}_{x'\sim D_{out}^{OE}}\Big[L_{OE}\big(f(\boldsymbol{x}'),f(\boldsymbol{x}),y\big)\Big] \tag{30}$$

其中，D_{in}为已知类数据样本集（In-distribution），D_{out}为未知类数据样本集（Out-of-distribution），D_{out}^{OE}为训练时引入的未知类辅助数据集，与D_{out}^{test}（测试集中的未知类数据）不重合、不相关。$L\big(f(\boldsymbol{x}),y\big)$为已知类的分类损失，$L_{OE}$为未知类识别损失。不同的未知类检测方法对应不同的未知类识别损失，例如L_{OE}可以设计为$f(\boldsymbol{x}')$到均匀分布的交叉熵损失，也可以设计为样本到类别原型距离的均方误差损失。

（1）OECC（Outlier Exposure with Confidence Control）[58]

基于离群值的思路，Papadopoulos等人[58]对Hendrycks等人[57]的工作进行改进，提出了基于置信度控制的离群值暴露方法OECC。该方法对训练集中的已知类采用交叉熵损失进行优化，同时希望训练集中的非已知类数据（离群值）在所有已知类上的概率分布趋于均匀分布，避免未知类数据被误分到某个已知类中。此外，该方法要求在已知类训练数据集上的平均最大预测概率尽可能地靠近封闭集训练的准确率，避免开放集分类器在已知类数据上性能的损耗。整体损失函数如公式（31）所示：

$$\begin{aligned}L_{OECC}=&E_{(\boldsymbol{x},y)\sim D_{in}}\Big[L_{CE}\big(f(\boldsymbol{x};\theta),y\big)\Big]+\\&\lambda_1\sum_{\boldsymbol{x}^{(i)}\sim D_{out}^{OE}}\sum_{l=1}^{K}\left|\frac{1}{K}-e^{z_l}/\sum_{j=1}^{K}e^{z_j}\right|+\\&\lambda_2\left(A_{tr}-E_{\boldsymbol{x}\sim D_{in}}\left[\lim_{l=1,2,\cdots,K}\left(e^{z_l}/\sum_{j=1}^{K}e^{z_j}\right)\right]\right)^2\end{aligned} \tag{31}$$

其中，A_{tr}为神经网络在封闭集训练时的分类准确率，K为已知类类别总数。λ_1,λ_2为权重系数。

（2）Agnostophobia[56]

Agnostophobia意为“the fear of the unknown”，即对未知的恐惧。Dhamjia等人[56]认为开放集识别问题是缓解神经网络和机器学习模型Agnostophobia问题的有效方案，他们将开放集识别问题的类别空间进行了梳理，将其分为两大类(C,U)，三小类(C,B,A)。

① C：模型感兴趣的已知类类别；

② U：模型需要拒绝的所有未知类类别。值得注意的是，U是无限大的；

<1>. $B\subset U$：未知类中的无关类别或容易获得的类别，例如背景、模型不感兴趣的其他已知类等；

<2>. $A=U\subset B$：模型训练阶段完全无法预料的未知类别，大多数只在测试阶段出现。

其中B可用来构造开放集模型训练阶段学习未知类别和开放空间的辅助数据。

为了更好地实现开放集识别，在不改变模型结构的前提下，Agnostophobia方法通过引入两个简单高效的开放集损失函数entropic open-set loss和

objectophere loss，并使用易获得的非已知类样本来提升模型开放集识别的效果。具体地，entropic open-set loss如公式(32)所示，它一方面希望训练集中的未知类样本($\boldsymbol{x}\in D_b$, $b\in B$)的输出分布趋近于均匀分布，使得训练好的网络对于测试阶段出现的任意未知类样本也产生较为均匀的输出，从而不会被误判到某个已知类中；另一方面，对于已知类样本($\boldsymbol{x}\in D_c$, $c\in C$)，它希望其对应类别的输出概率尽可能的高，即分类准确。

$$L_{Entrop}(\boldsymbol{x})=\begin{cases}-\log S_c(\boldsymbol{x}) & \text{if } x\in D_c\\ -\dfrac{1}{C}\displaystyle\sum_{c=1}^{C}\log S_c(\boldsymbol{x}) & \text{if } x\in D_b\end{cases} \tag{32}$$

objectosphere open-set loss则旨在约束特征向量的模长，即让已知类样本的表征在特征空间中有更显著的长度，让未知类样本的特征表示尽可能趋向于0，从而强制让已知类样本和未知类样本有一定的距离。其函数如公式(33)所示。

$$L_{Obs}=\begin{cases}\max(\xi-\|F(\boldsymbol{x})\|,0)^2 & \text{if } x\in D_c\\ \|F(\boldsymbol{x})\|^2 & \text{if } x\in D_b\end{cases} \tag{33}$$

最终，Agnostophobia方法将entropic open-set loss和objectosphere loss结合，作为最终的损失函数，如公式(34)所示。该方法不直接关注于拒绝未知类样本，而是通过寻找对未知类样本更鲁棒的深层特征，使得在此特征空间上已知类样本和未知类样本被有效区分。

$$L_{Agnos}=L_{Obs}+\lambda L_{Entrop} \tag{34}$$

(3)MixOE[59]

MixOE聚焦于细粒度场景下的开放集识别任务，比粗粒度的开放集识别任务更具有挑战性。例如，在鸟类分类器中，细粒度开放集识别任务需要识别出新的、分类器从未见过的子类鸟类物种。由于细粒度场景下未知类样本与已知类样本之间有较强的语义相似性，在特征空间上细粒度的未知类样本也更接近于已知类集群，同时可供使用的离群辅助数据集在训练过程中也容易与细粒度已知类别混淆，从而许多粗粒度场景下的开放集识别方法在细粒度场景下表现并不好[59]，因此，Zhang等人[59]针对这一任务，提出了利用Mixup方法[92]生成虚拟的未知类辅助数据集来覆盖更精确范围的细粒度未知类区域的MixOE方法，为细粒度开放集识别提供有效的解决方案。具体地，MixOE方法将已知类样本x_{in}和未知类样本x_{out}进行简单像素级的混合操作得到虚拟的未知类数据$\tilde{\boldsymbol{x}}=\text{mix}(\boldsymbol{x}_{in},\boldsymbol{x}_{out},\lambda)\sim D_{out}^{virtual}$，$\lambda\in[0,1]$，混合操作有线性混合(即线性插值操作[92])以及剪贴混合(即剪贴-粘贴操作[93])，生成的虚拟未知类数据所对应的标签为$\tilde{y}=\lambda y_{in}+(1-\lambda)U$，其中$U$为已知类别的均匀分布。在训练阶段，MixOE结合已知类训练数据和生成的虚拟未知类数据一起训练，利用联合损失函数(公式(35))进行约束；在测试阶段，模型通过预测置信度和阈值的方式实现已知类别和未知类别的识别。MixOE方法所生成的未知类样本能够高度细化，可以较好地覆盖细粒度场景下的未知类范围，为细粒度场景下的开放集识别任务提供了新的思路。

$$E_{(\boldsymbol{x},y)\sim D_{in}}\Big[L\big(f(\boldsymbol{x}),y\big)\Big]+\beta E_{(\tilde{\boldsymbol{x}},\tilde{y})\sim D_{out}^{virtual}}\Big[L\big(f(\tilde{\boldsymbol{x}}),\tilde{y}\big)\Big] \tag{35}$$

(4)BCROSR (Background-Class Regularization for Open-Set Recognition)[59]

Cho等人[59]提出通过背景类正则化(Background-Class Regularization，BCR)策略和基于距离的分类器来解决开放集识别问题。在训练阶段，BCROSR使用背景类数据来模拟真实的未知类样本，并引入背景类正则化损失来约束模型对已知类和背景类样本的学习，即

$$\begin{aligned}L_{bg}=&E_{\boldsymbol{x}^k\sim D_t}\Big[h\Big(D_E^2\big(g(\boldsymbol{x}^k),\boldsymbol{\mu}\big)\Big)\Big]-\\&E_{\boldsymbol{x}^b\sim D_b}\Big[\log\Big(1-\exp\Big(-h\Big(D_E^2\big(g(\boldsymbol{x}^b),\boldsymbol{\mu}\big)\Big)\Big)\Big)\Big]\end{aligned} \tag{36}$$

其中，D_t和D_b分别表示已知类别和背景类别数据，$D_E^2\big(g(\boldsymbol{x}),\boldsymbol{\mu}\big)$表示输入样本$\boldsymbol{x}$与中心向量$\boldsymbol{\mu}$之间的欧氏距离的平方。$h(\boldsymbol{x})=\sqrt{\boldsymbol{x}+1}-1$可将欧氏距离$\Big(D_E^2\big(g(\boldsymbol{x}),\boldsymbol{\mu}\big)\Big)$缩放到$(0,1]$范围内。通过这个损失函数，BCROSR可以增加背景类样本与中心向量之间的距离，从而迫使背景类样本远离已知类别的特征空间，以实现对未知类别样本的识别。进一步，BCROSR通过结合分类损失来提升已知类的分类效果。

基于离群值的开放集识别方法通过利用系统中易获得的与已知类别无关的辅助数据(其与测试数据集中的未知类样本无重合)来优化开放集分类器，期望借助数据多样性来提升开放集分类器的泛化性。通过迫使开放集分类器对已知类别样本在其所对应的类别标签上的概率尽可能高，对于离群值样本在所有已知类别上的概率分布趋近于均匀分布，

基于离群值的开放集识别方法可以利用置信度控制的方式避免未知类别样本被错误划分到已知类别中，从而学习到更具有区分性的特征空间和特征表达。然而这类方法在开放集识别上的性能很大程度上取决于当前可获得的辅助数据集的质量，同时辅助数据集与已知类别数据集之间的相似性也对最终效果有较大影响，并不是所有的(或者任意的)离群值辅助数据都可以带来开放集识别分类器效果的有效提升。例如，在Agnostophobia[56]的实验中：当CIFAR10数据集被选作辅助数据集时，针对MNIST数据集的类别分类，CIFAR10辅助数据就未能帮助开放集识别模型提升性能；而用NIST Letters作为辅助数据集时，则获得了较好的提升效果。其潜在原因是CIFAR10数据集的图像与MNIST数据集的数字显著不同，机器学习模型可以较为轻易地将其分开，并不会迫使模型去进一步学习已知类样本的细节信息和更精确的边界，而MNIST数据集和NIST Letters数据集具有很强的相似性，机器学习模型需要非常针对性地学习到每个已知类别与这些相似样本之间的区别，并将其分开，从而使得该模型获取到更加紧凑的已知类类边界和更加精细的分类边界。

3.2.2 基于数据生成的开放集识别

利用系统中已有的额外数据集来构造辅助数据集是一个很好的提升数据多样性的思路，但其对数据的选取和质量要求较高，具有较高的成本。基于数据生成的开放集识别方法[61-71]直接自动生成伪未知类样本，在不借助任何测试阶段开放域信息的前提下，将开放集识别任务转化为标准的K+1类分类任务(K为已知类类别数)，通过分类器直接计算新样本属于未知类的概率和属于任意已知类的概率，进行已知类样本分类和未知类样本识别。

(1)G-OpenMax(Generative OpenMax)[62]

Ge等人[62]针对OpenMax算法[29]进行了扩展，提出了新的开放集识别算法G-OpenMax。该方法将OpenMax与生成对抗网络(Generative Adversarial Networks，GAN)结合，使用基于类别标签的条件GAN模型生成新样本，并将生成的样本输入到预训练好的封闭集分类器中，挑选被分类器分错的样本作为未知类样本，与已知类样本一起训练K+1类分类器，并结合OpenMax的方法校准该分类器的SoftMax值，实现良好的开放集识别效果。

(2)OSRCI(Open Set Learning with Counterfactual Images)[63]

Neal等人[63]利用GAN生成未知类样本来约束已知类的边界。为了更好地对已知类边界进行约束，生成的未知类样本要足够靠近已知类样本，但同时也不属于已知类。该方法首先通过编码器E抽取特征，对任意输入的图像$\boldsymbol{x}$获得其嵌入表达$E(\boldsymbol{x})$；然后在嵌入空间中学习和生成靠近已知类边界的未知类样本，如公式(37)所示。其中$\boldsymbol{z}^*$为生成的未知类样本在嵌入空间中的特征表示，$\mathcal{C}_{\mathcal{K}}$为已知类分类器，$G(\boldsymbol{z})$是特征表示$\boldsymbol{z}$经过解码器$G$生成的未知类样本。公式(37)的第一项希望生成的未知类样本足够靠近已知类，公式的第二项希望生成的未知类样本不属于已知类中的任意一类，即它在已知类上的预测概率分布趋于均匀分布。获得足够的未知类样本后，生成的未知类样本与已知类样本混合，共同训练一个包含增广类的$K+1$分类器，实现最终的开放集识别。

$$\boldsymbol{z}^*=\min_{\boldsymbol{z}}\left\|\boldsymbol{z}-E(\boldsymbol{x})\right\|_2^2+\log\left(1+\sum_{i=1}^{K}e^{\mathcal{C}_i(G(\boldsymbol{z}))_i}\right) \tag{37}$$

(3)DIAS(Difficulty-Aware Simulator)[64]

Moon等人[64]认为开放世界中的未知类样本具有极强的多样性，对于分类器来说，判别其是否来自未知类的难度也各不相同。传统基于GAN模型生成的伪未知类样本，由于都以分类器的预测作为标准进行筛选，因而获得的样本都是分类器较易识别的样本类型。因此，Moon等人提出了DIAS框架，它在GAN的基础上进一步提出Copycat模型，通过构造多种不同难度的伪未知类数据来模拟真实开放世界中未知类样本的多样性，提升开放集识别模型的泛化能力。

具体地，DIAS框架由一个生成对抗网络GAN和两个Copycat模型构成。GAN用于生成中等难度的未知类样本，其中一个Copycat模型用于生成高难度和低难度级别的未知类样本图片，另一个Copycat模型作为分类器实现分类功能。作为分类器的Copycat由全连接层$(\boldsymbol{W}_{cls}^1,\boldsymbol{W}_{cls}^2,\cdots,\boldsymbol{W}_{cls}^n,\boldsymbol{W}_{cls}^c)$组成，其中$\boldsymbol{W}_{cls}^c$是带SoftMax的全连接层。中低难度的伪未知类样本的标签被定义为均匀分布标签，即$\tilde{\boldsymbol{y}}=\frac{1}{k}\cdot\boldsymbol{u}$，$\boldsymbol{u}$为全1向量；高难度伪未知类标签被定义为$\tilde{\boldsymbol{y}}=(1-\alpha)\cdot\boldsymbol{y}+\frac{\alpha}{K}\cdot\boldsymbol{u}$，$\boldsymbol{y}$为正确标签的独热编码，$\alpha$为权重参数。整个模型在已知类样本和生成的未知类样本上进行训练，通过交叉熵损失优化分类器，最终实现开放集识别的功能。

（4）PROSER（Placeholders for Open-set Recognition）

Zhou等人[65]提出了基于占位符(placeholder)的开放集识别模型PROSER，分别为数据和分类器分配占位符来提升模型对未知类别识别的效果。具体地，PROSER首先利用Mixup[94]将不同类别样本的嵌入特征通过线性加权混合，生成伪未知类样本，即

$$\tilde{\boldsymbol{x}}=\lambda E(\boldsymbol{x}_i)+(1-\lambda)E(\boldsymbol{x}_j), y_i\neq y_j \quad (38)$$

其中λ为超参，$E(x_i)$为样本x_i的中间层嵌入表达，y_i为其标签。生成的未知类样本$\tilde{\boldsymbol{x}}$的标签为增广类，即$\tilde{y}=K+1$。损失函数如式(39)所示，优化已知类样本的交叉熵损失的同时，希望第$K+1$类成为其正确标签以外的第二大概率。

$$l_{PR1}=\sum_{(x,y)\epsilon D_{tr}} l\left(\hat{f}(\boldsymbol{x}),y\right)+\beta\cdot l\left(\hat{f}(\boldsymbol{x})\backslash y,K+1\right) \quad (39)$$

分类器占位符则将封闭集分类器W从原始的K类扩展到$K+1$类，即由$f(x)=\boldsymbol{W}^T\Phi(\boldsymbol{x})$扩展为$\hat{f}(x)=\left[\boldsymbol{W}^T\Phi(\boldsymbol{x}),\widehat{\boldsymbol{w}^T}\Phi(\boldsymbol{x})\right]$，$\hat{\boldsymbol{w}}$为第$K+1$类对应的网络参数，未知类占位符的交叉熵损失如公式(40)所示。

$$l_{PR2}=\sum_{(\widetilde{\boldsymbol{x}_i},K+1)\in D_{gene}} l\left(\left[\boldsymbol{W},\hat{\boldsymbol{w}}\right]^T\Phi\left(\widetilde{\boldsymbol{x}_i}\right),K+1\right) \quad (40)$$

整个模型通过两个损失进行优化，数据占位符用有限的复杂性样本来模拟未知类样本，将封闭性训练转变为开放性训练；分类器占位符则作为已知类和未知类之间的特定类别边界，更好地分离已知类样本和未知类样本，并通过为新类别保留分类器占位符来矫正过度自信的已知类预测。

（5）GCM-CF（Generative Causal Model Counterfactual-Faithful）[66].

传统的基于数据生成的方法所生成的未知类样本在特征空间上既不属于已知类，也不属于真实的未知类，并不能很好地模拟真实未知类样本，导致模型更加倾向于对已知类分类效果的提升，反而损害模型对未知类样本的识别效果。基于此，Yue等人[66]提出一种基于反事实框架的方法GCM-CF，通过将样本属性和类别属性进行分离，利用反事实推断协助伪未知类样本的生成，促使生成的样本与真实的未知类样本具有较为一致的数据分布，从而解决零样本学习(Zero-Shot Learning，ZSL)和开放集识别的挑战。

具体地，GCM-CF模型的损失函数如式(41)所示。其中，L_Z是针对样本属性的损失，通过最小化重构误差和KL散度来学习样本属性，L_Y是针对类别属性的对比损失，通过最大化样本之间的差异来学习类别属性，L_F是用于提升反事实推断准确性的损失函数，通过训练一个判别器来判断生成的反事实推断样本是否真实，从而保证生成的样本与真实样本的分布一致。这三个损失函数共同作用，合力提升模型的泛化能力。

$$\begin{aligned}
&\mathcal{L}=\mathcal{L}_{\mathcal{Z}}+\lambda_v\mathcal{L}_{\mathcal{Y}}+\lambda_p\mathcal{L}_{\mathcal{F}}\\
&\mathcal{L}_{\mathcal{Z}}=-E_{Q_\phi(Z|X)}\left[P_\theta(X|Z,Y)\right]+\\
&\qquad\beta D_{KL}(Q_\phi(Z|X)\|P(Z))\\
&L_Y=-\log\frac{\exp\left(-dist\left(\boldsymbol{x},\boldsymbol{x}_y\right)\right)}{\sum\limits_{\boldsymbol{x}'\in\tilde{X}\cup\{x_y\}}\exp\left(-dist(\boldsymbol{x},\boldsymbol{x}')\right)}\\
&L_F=E\left[D(\boldsymbol{x},\boldsymbol{y})\right]-E\left[D(\boldsymbol{x}',\boldsymbol{y})\right]-\\
&\qquad\lambda E\left[\left(\left\|\nabla_{\hat{x}}D(\hat{\boldsymbol{x}},\boldsymbol{y})\right\|_2-1\right)^2\right]
\end{aligned} \quad (41)$$

（6）ConOSR（Contrastive Open-set Recognition）[67]

Xu等人[67]从表征学习的视角出发，在SupCon[95]方法的基础上结合Mixup[92]，利用有监督对比学习来提高模型的表征学习能力，进而辅助模型对开放集识别的性能提升。具体地，训练过程分为两阶段：有监督对比学习阶段和分类器训练阶段。在有监督对比学习阶段中，ConOSR使用数据增强技术[96]对原始训练数据进行视图增强，例如利用Mixup方法[92]生成语义模糊的虚拟未知类样本，一起用于训练。在训练过程中，增强后的训练数据首先通过特征提取器$\Phi(\cdot)$和投影网络(ProjectionNetwork)$\Psi(\cdot)$，被投影在表征空间$\boldsymbol{z}_i=\Psi(\Phi(\boldsymbol{x}_i))$中。该阶段主要通过有监督地对比损失进行约束，学习适合于开放集识别任务的样本表征，即

$$L_{con}=-\sum_i\sum_{j\neq i}\frac{s(\boldsymbol{y}_i,\boldsymbol{y}_j)}{\sum\limits_{k\neq i}s(\boldsymbol{y}_i,\boldsymbol{y}_k)}\log\frac{\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_j/\tau)}{\sum\limits_{k\neq i}\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_k/\tau)} \quad (42)$$

其中，τ为超参，$s(\boldsymbol{y}_i,\boldsymbol{y}_j)=\dfrac{\boldsymbol{y}_i\cdot\boldsymbol{y}_j}{\|\boldsymbol{y}_i\|\|\boldsymbol{y}_j\|}$表示平滑后的标签$\boldsymbol{y}_i$和$\boldsymbol{y}_j$之间的余弦相似度。该对比学习中的正样本对选自同一类别中的样本，负样本对选自不同类别的样本。在分类器训练阶段，ConOSR首先冻结特征提取器$\Phi(\cdot)$，之后在特征空间$h_i=\Phi(x_i)$中通过最小化交叉熵损失来优化分类器，利用阈值实现对未知类样本的识别。

(7)OpenMix+[68]

Jiang等人[68]认为大多数基于数据生成的开放集识别方法忽略了开放空间风险与结构风险(Structural Risk)的权衡问题,从而提出了一个新的未知类样本增强策略OpenMix,通过简单的自混合策略生成高质量的未知类样本,为类边界学习提供支撑,并平衡开放空间风险和结构风险。具体地,对每个样本,OpenMix首先选择一个自我混合的对象,这个对象通常是该样本的某个局部,之后对选定的自我混合对象进行变换,扭曲关键特征以打破其类归属度,同时保留其他特征以保持其类相似性,进而生成未知类样本。OpenMix通过自我混合策略生成的未知类样本既具有高相似性又具有低归属度,且该策略简单高效,易于实现,具有较好的应用场景。OpenMix+是OpenMix方法的加强版,它将OpenMix方法与正则化方法结合,实现了更低的开放集结构风险,从而获得更强的开放集识别性能。

(8)IT-OSR(Interative Transductive Open-Set Recognition)[69]

Sun等人[69]基于迭代传导的思想,提出了开放集识别方法IT-OSR。IT-OSR框架包含可信采样(Reliability Sampling Module)、特征生成(Feature Generation Module)和基线更新(Baseline Update Module)三个模块。可信采样模块通过输出空间和潜在特征空间中的伪标签一致性来识别可靠样本;特征生成模块通过设计双对抗生成网络来解决样本不平衡问题,利用特征生成器、真/假鉴别器和已知/未知类别鉴别器综合实现,旨在增加已知类别和未知类别的特征多样性;基线更新模块实现将任意的开放集识别模型与IT-OSR框架的无缝结合,从而实现域迁移学习及开放集识别。

基于数据生成的开放集识别方法利用生成对抗网络和训练集中的已知类数据生成伪未知类样本,利用伪未知类样本探索开放域中的未知空间,并与训练集中的已知类样本共同构建和优化$K+1$类开放集分类器,实现显式的开放集归属度预测。其重点和难点均在于生成策略的设计,构造的伪未知类样本必须有利于开放集分类器对各个类别边界产生更加深入的认识。已有的方法中,G-OpenMax[62]利用训练好的封闭集分类器筛选生成对抗网络的生成样本,并把被分类错误的样本作为未知类数据;OSRCI[63]通过设计损失函数使得所生成的未知类样本与已知类样本足够相似,但不属于已知类样本的任何一类,以此约束已知类样本的边界;DIAS[64]利用生成对抗网络和设计的Copycat网络生成不同难度的未知类样本来模拟开放世界中复杂的未知类样本输入的情况;PROSER[65]利用Mixup生成伪未知类样本,通过分类器占位符扩充封闭集分类器为开放集分类器;GCM-CF[66]通过反事实推断来协助未知类样本的生成,使得生成的未知类样本更真实,提升开放集识别能力。ConOSR[67]利用Mixup方法生成语义模糊的虚拟样本模拟未知类数据,结合标签平滑,利用有监督对比学习策略学习更适合开放集识别任务的样本表征。OpenMix+[68]通过简单的自我混合策略生成既具有高相似性又具有低归属度的未知类样本,平衡开放空间风险和结构风险。IT-OSR[69]采用了迭代式跨领域开放集识别框架,通过可信性采样、特征生成以及基线更新模块,提高模型跨领域开放集识别的性能。在针对图数据的开放集识别任务中,G^2Pxy方法[70]通过生成两种类型的代理未知节点:类间未知代理和类外未知代理,并结合交叉熵损失和补充熵损失,显著提高了未知类检测和已知类分类的效果。在针对文本数据的开放集识别任务中,Zhou等人[65]也基于mix-up方法,通过混合两个类别样本的中间层特征来生成未知类样本,辅助开放集分类器的训练。Ding等人[71]使用生成对抗网络来生成不属于已知类别的伪未知类样本,可以有效提高模型在开放集场景的泛化能力。综上所述,基于数据生成的开放集识别方法可以较好地利用辅助数据集的优势,同时也可以有效解决辅助数据获取成本较高的困难,具有很好的实际应用价值,在公开数据集上的评测效果也证明了这一类方法的有效性(参考表3),然而如何生成更符合真实世界复杂的伪未知类样本以及如何更好地提升生成的伪未知类样本的多样性还需进一步地深入探索。

额外地,也有研究者将基于离群值和基于数据生成的开放集识别方法进行了结合,用于解决开放集识别问题,例如OpenGAN[61]。OpenGAN的提出者Kong等人[61]认为基于离群值的和基于数据生成的开放集识别方法中,前者所得到的模型易对离群值过拟合,后者中常用的GAN模型具有较大的不稳定性,将二者结合则可以很好地弥补这两种方法各自的局限性。因此,在OpenGAN方法中,它既包含真实的其它类数据,同时也利用生成器生成伪开放集数据,与已知类数据一起组成训练数据集,对模型进行训练。其生成器与传统GAN模型不同,它

不直接生成像素级别的开放集图像，而是对特征级别OTS(Off-The-Shelf)进行重建，并利用真实的其他类数据来选择合适的生成器和判别器。

3.3　总结与讨论

在本节中，本文对近年来基于深度学习的开放集识别方法进行了梳理和介绍，根据其是否依赖辅助数据，将现有方法分为了两类：不依赖辅助数据的开放集识别方法和基于辅助数据的开放集识别方法。表1对本章中介绍的主要模型类别从主要原理、优缺点和代表性方法三方面进行了直观展示。

表1　基于深度学习的开放集识别方法分类概括、主要原理优缺点及代表性方法

大类	小类	主要原理	优点	不足与挑战	代表性方法
不依赖辅助数据的开放集识别方法	基于评分函数的开放集识别方法	利用评分函数，促使模型输出的评分分数和分类准确率可以较好地匹配，同时使已知类样本和未知类样本在该分值上有较大区分度	方法简单且高效，能较好地将已知类别和未知类别样本分开	需结合阈值来实现对未知类样本的检测，对阈值的智能选取和自动调节需增强	ODIN[23], NMD[24], Energy[25], GEN[26], OpenWGL[15], OSSC[27], DOSSL[28]
	基于距离的开放集识别方法	通过学习每个类别的“类原型”，利用样本到类原型的距离来度量样本对该类别的归属度，同时使得模型学习的特征空间中类内紧凑类间可分，为未知类别留下更多空间	通过类原型可以有效地在特征空间上描述和学习已知类别和未知类别	对类原型的设计较为简单，对真实世界的复杂性刻画不够	OpenMax[29], CROSR[30], MetaMax[31], CAC[32], CPN[33], RPL[34], ARPL[35], PMAL[36], ODL[37], JNICS[38], MGPL[40], CVAECapOSR[39], ROGPL[41], PMOSR[42], SEMAN-G[43]
	基于重构的开放集识别方法	通过训练自编码器对样本进行重构，利用模型对已知类样本和未知类样本重构质量的显著不同，通过重构效果进行未知类样本识别	有利于学习更丰富的语义信息和样本间的结构信息	容易受到噪声数据或与类别无关的数据的影响	C2AE[44], CGDL[46], CSSR[50], GFROSR[45], OpenHybrid[47], GMVAE[48], MOODCAT[49], OpenSAVR[51]
	基于网络架构搜索的开放集识别方法	从神经网络体系结构出发，通过网络架构搜索和VAE对比学习，寻找最适合开放集任务的网络，实现开放集识别	简单有效且角度新颖	计算时间长，需要大量的人工设计	LNAOSR[52]
	基于多模态大模型的开放集识别方法	利用大模型微调、对比提示调优、开放词汇扩展和跨模态引导等技术，实现对未知类的检测	可有效提升模型对未知类别的检测和区分能力	探索尚处于初期阶段，需更巧妙地与开放集识别相结合	CoHOZ[53], R-Tuning[54], A2Pt[55]
基于辅助数据的开放集识别方法	基于离群值的开放集识别方法	通过使用额外的辅助数据集来训练开放集分类器，与已知样本数据一起构成的数据集具有较好的类别和信息的多样性	使用真实的辅助数据集能够较好地提升开放集分类器的泛化性	对数据选取的质量要求较高	Agnostophobia[56], OECC[58], MixOE[59], BCROSR[59], OpenGAN[61]
	基于数据生成的开放集识别方法	通过直接生成伪未知类数据，在不借助任何测试阶段开放域信息的前提下，将开放集识别任务转化成标准的K+1类分类任务	可有效解决辅助数据获取成本较高的困难	对有效未知类样本生成策略的研究还有较大的优化空间	OpenGAN[61], G-OpenMax[62], OSRCI[63], DIAS[64], PROSER[65], GCM-CF[66], ConOSR[67], OpenMix+[68], IT-OSR[69], DRA[71] G2Pxy[70]

不依赖辅助数据的开放集识别方法是指模型在训练过程中仅使用已知类别的训练数据实现开放集识别，根据其建模思路，又可将其细分为基于评分函数的、基于距离的、基于重构的和其他四个子类别。其中，基于评分函数的开放集识别方法通过对传统SoftMax的预测置信度进行校正，探索更好地区分已知类样本和未知类样本的置信度评估指标，实现对未知类样本更准确地识别。但该类方法通常需要结合阈值来实现对未知类样本的检测，当前对阈值的智能选取或调节是该类方法可进一步优化的方向之一。基于距离的开放集识别方法通过学习已知类别的类原型，利用度量学习使得特征空间中已知类别类内紧凑、类间可分，为未知类别预留了更多可用空间。在测试阶段，该类方法通过计算输入样本到

各个类原型之间的距离来实现对已知类别样本和未知类别样本的区分。基于重构的开放集识别方法利用自编码器对已知类别和未知类别的重构效果之间的差异作为区分已知类别和未知类别的重要指标之一，具有较好的可解释性，但该类方法在学习的过程中，容易过拟合到与分类任务无关的信息上，从而对封闭集分类的精度造成一定损害。

依赖辅助数据的开放集识别方法是指直接利用系统中的“无关”数据或利用生成方法间接获得伪未知类样本作为辅助数据，结合已知类数据进行开放集识别优化的方法。依据辅助数据的获取方式，其又可划分为基于离群值的和基于数据生成的开放集识别方法两个子类别。基于离群值的开放集识别方法使用训练过程中系统中可获得的、与测试集中未知类不重合的未知类别数据作为辅助数据集来训练一个$K+1$类的开放集分类器，实现对样本数据未知类概率的显性评估，为开放集识别提供了新的思路。该类方法的关键点之一是对辅助数据的选取，辅助数据质量不高或与已知类数据不适配等情况将直接影响开放集识别的最终性能，这也提升了辅助数据选取的难度。基于数据生成的开放集识别方法不依赖系统中已有的辅助数据，直接利用生成对抗网络等生成模型对伪未知类别进行样本生成，形成人工辅助数据，帮助K+1类开放集分类器的训练，具有更好的灵活性和可控性。其中，如何生成更接近真实数据的复杂伪未知类样本，以及生成更具有多样性的伪未知类样本，是该类问题的研究重点。

此外，在实际应用中，对于基础神经网络模型的选取也是影响开放集识别模型的重要因素。大多数现有的开放集识别算法都采用了在计算机视觉领域较为经典且性能良好的神经网络模型作为基本模型，使用频率较高的网络框架包括AlexNet[97]、VGGNet[98]、ResNet[99]、Wide ResNet[100]、DenseNet[101]、ImprovedGAN[102]等。其中，AlexNet[97]、VGGNet[98]被常用于小规模数据集上的开放集识别[29-30,46,56,83]，如MNIST，CIFAR10，CIFAR100等数据集。ResNet[99]、Wide ResNet[100]、Improved GAN[102]、DenseNet[101]则更多用于大规模数据集上的开放集识别[23-25,31-32,34-36,39-40,45,49-50,56,58-59,63,65]，例如SVHN，TinyImageNet等数据集。在大量的开放集识别实验中，不同的基本模型被发现具有各自不同的优势，例如NMD开放集识别算法[24]在实验中证明使用ResNet-34在开放集识别上的平均性能最好，具有较高的鲁棒性；CROSR开放集识别算法[30]展示了在CIFAR10数据集上分别使用VGGNet和DenseNet的推理运行时间，在开放集识别准确率相当的情况下，VGGNet比DenseNet的推理速度快5—7倍。因此，对基本网络的选择不仅取决于开放集模型自身的需求，也取决于实际应用对性能和效率的需求和取舍。此外，Vaze等人[103]通过大量实验证明了模型在开放集上的识别能力与其在封闭集分类上的性能高度相关，且这种关系适用于多种损失函数和各种模型架构，还进一步证明了这种相关性在开放集识别的小规模数据集、大规模数据集以及其它相关子领域（例如分布外检测）的评估中都成立。

除了针对开放集识别这一基础问题的研究之外，鉴于开放集识别模型突破了静态封闭环境的局限性，其也被应用于大量的具体场景中，例如Yang等人[104]将开放集识别问题应用于人脸识别任务，Xu等人[105]等人将开放集识别问题应用于类增量学习；Pal等人[106]将开放集识别问题与少样本学习结合；Grcic等人[107]将开放集识别模型应用于密集异常检测场景（Dense Anomaly Detection）；Cai等人[108]将开放集识别模型应用于长尾分布问题等。

4 评 测

本节主要介绍当前开放集识别模型评测相关的内容，例如常用数据集及其设置、常用评估指标，以及经典模型的开放集识别效果和封闭集分类效果等。

4.1 常用数据集

开放集识别任务中常用的数据集和封闭集分类中常用的数据集类似，主要包括MNIST、CIFAR10、CIFAR100、CIFAR+10、CIFAR+50、SVHN、ImageNet、TinyImageNet等小规模数据集，以及ImageNet-100、ImageNet-200、ImageNet-1000、ImageNet-LT、CUB、Stanford Cars、FGVC-Aircraft等大规模数据集。其统计信息如表2所示。这些数据集被用在开放集识别任务中时，通常利用类分离（Class Separation）或者类添加（Outlier Addition）等方式进行扩展（详见表3），用于开放集识别的评测。类分离设定是指在一个标准数据集中随机挑选一些类别作为已知类别，剩余的类别设为未知类别。类添加设定是将某个数据集中的所有类别都作为已知类，在测试阶段加入其他数据集的数据作为未知类别数据。

表 2　常用数据集简介

数据集	介绍	数据集大小/张	类别总数/个	训练集总数/张	验证集总数/张	测试集总数/张
MNIST[109]	MNIST数据集收集了来自250个人的0-9手写数字图片	70 000	10	60 000	-	10 000
CIFAR10[110]	CIFAR10数据集是一个用于识别普适物体的小型数据集	60 000	10	50 000	-	10 000
CIFAR100[110]	CIFAR100数据集与CIFAR10数据集类似，CIFAR100数据集中的100个类分为20个超类，每张图像都带有一个“子类”标签和一个“大类”标签（它所属的类和超类）。例如，该数据集中超类”花”的类别所包含的子类类别标签有兰花、罂粟、玫瑰、向日葵和郁金香等类别	60 000	100	50 000	-	10 000
SVHN[111]	SVHN为街景门牌号码数据集。	630 420	10	73 257＋531 131	-	26 032
TinyImageNet[112]	TinyImageNet数据集是ImageNet数据集[112]的子集	120 000	200	100 000	10 000	10 000
ImageNet[112]	ImageNet数据集是世界上图像分类、图像识别、图像定位等领域最大的数据库	1 431 167	10 000	1 281 167	50 000	100 000
ImageNet-LT[112]	ImageNet-LT是ImageNet数据集[112]的一个长尾版本	115 846	1 000	115 846	20 000	20 000
CUB[103]	鸟类细粒度视觉分类数据集	11 788	200	5 994	-	5 794
StanfordCars[103]	车类细粒度视觉分类数据集	16 185	196	8 144	-	8 041
FGVC-Aircraft[103]	飞机类细粒度视觉分类数据集	10 000	100	6 667	-	3 333

注：验证集总数栏中的“-”表示该数据集官方未明确说明验证集总数，通常需要在训练时对训练集进行进一步划分。

表 3　开放集识别任务中分类数据集的常用设置

数据集	数据集设置	介绍	已知类别数	未知类别数	*Openness*
MNIST[109]	类分离设置	随机选取6个类别作为已知类，其余4个类别作为未知类	6	4	13.39%
MNIST[109]	类添加设置	选取MNIST数据集的全部10个类别作为已知类别，添加其它数据集中的类别作为未知类别:Omniglot[57]、MNIST-noise[36]和Noise等[36]	10	10	18.35%
CIFAR10[110]	类分离设置	随机选取6个类别作为已知类，其余4个类别作为未知类	6	4	13.39%
CIFAR10[110]	类添加设置	选取CIFAR10数据集的全部10个类别作为已知类别，添加其它数据集中的类别作为未知类别: ImageNet-crop、ImageNet-resize、LSUN-crop和LSUN-resize等[46]	10	10	18.35%
CIFAR100[110]	类分离设置	选取20个类别作为已知类，其余80个类别作为未知类	20	80	42.26%
CIFAR＋10	类添加设置	随机选取CIFAR10数据集中4种不属于动物的类别作为已知类，从CIFAR100数据集中随机挑选10种属于动物的类别作为未知类	4	10	33.33%
CIFAR＋50	类添加设置	随机选取CIFAR10数据集中4种不属于动物的类别作为已知类，从CIFAR100数据集中随机挑选50种属于动物的类别作为未知类	4	50	62.86%
SVHN[111]	类分离设置	随机选取6个类别作为已知类，其余4个类别作为未知类	6	4	13.39%
TinyImageNet[112]	类分离设置	随机选取20个类别作为已知类，其余180个类别作为未知类	20	180	57.35%
ImageNet-100	类分离设置	选取ImageNet数据集取前100个类别作为已知类，其余类别作为未知类	100	900	57.36%
ImageNet-200	类分离设置	选取ImageNet数据集取前200个类别作为已知类，其余类别作为未知类	200	800	42.26%
ImageNet-1000	类添加设置	选取ImageNet数据集的全部1000个类别作为已知类别，以ILSVRC2010数据集中出现且未在ILSVRC2012数据集中出现的类别作为未知类别(共360个类别)[113]	1000	360	7.94%
ImageNet-LT[112]	类添加设置	ILSVRC2010数据集[113]中的附加类别作为未知类	1000	360	7.94%
CUB[103]	类分离设置	随机选取100个类别作为已知类，其余100个类别作为未知类	100	32＋34＋34	18.35%
StanfordCars[103]	类分离设置	随机选取98个类别作为已知类，其余98个类别作为未知类	98	76＋0＋22	18.35%
FGVC-Aircraft[103]	类分离设置	随机选取50个类别作为已知类，其余50个类别作为未知类	50	20＋17＋13	18.35%

注：CUB、Stanford Cars、FGCV-Aircraft三个数据集在开放集识别的类分离数据设置下，将未知类别的难易程度进一步分为“简单＋中等＋困难”。

4.2 常用评估指标

开放集识别模型的评价指标与分类模型的类似，常用指标如表4所示，包括准确率（Accuracy）、精确率（Precision）、召回率（Recall）、F1值（F1 score）、AUROC（Area Under the Receiver Operating Characteristic）、真阳率（True Positive Rate）、真阳率达到95%时的假阳率值（FPR@95%TPR）、AUPR（Area Under Precision-Recall Curve），以及开放集识别任务中特有的OSCR（Open-Set Classification Rate）和OpenAUC等评价指标。其中，OSCR（Open-Set Classification Rate）为开放集分类率，是Dhamija等人[56]2018年提出的针对开放集识别任务的专用评估指标。它计算的是正确分类率（Correct Classification Rate，CCR）与错误识别率（False Positive Rate，FPR）曲线下的面积。其中，正确分类率是指被正确分类的样本数占总样本数的比例。错误识别率，即假阳率，是指未知类样本被错误识别为已知类样本的比例。OSCR不容易受到数据集偏差的影响，可用于对类别不平衡应用场景的测试。

表4 开放集识别任务的常用评估指标

测评指标	描述	公式	优点及不足
精确率（Precision）	用于衡量模型被预测为正类的样本的可靠程度	$Precision=\frac{TP}{TP+FP}$	反映模型对于特定类别的分辨能力。在提升其中一项的同时往往另一项会有所降低
召回率（Recall）	用于衡量模型对正类样本的召回能力，是否将正类样本全部预测正确	$Recall=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$	
准确率（Accuracy）	在开放集识别任务中用于计算封闭集分类效果和整体分类效果，为预测正确的样本数与总样本数的比值	$Acc=\frac{TP+TN}{TP+TN+FP+FN}$	计算简单，易于理解，应用广泛。数据不平衡时，结果无法衡量分类器的好坏
Macro-averaged F1值	通过对每个类别的F1值进行普通平均得到	$F1=2\times\frac{Precision\times Recall}{Precision+Recall}$	较好反映模型对尾部类别的识别能力，避免不同类别的测试样本数量对评估结果的影响
FPR@95%TPR	为当TPR为95%时，FPR的值	$TPR=\frac{TP}{TP+FN}$ $FPR=\frac{FP}{FP+TN}$	重点关注高召回率下的假阳性，适用于不平衡数据集。对阈值选择敏感，可能忽视其它性能维度
AUROC	TPR-FPR曲线下的面积	-	这是一个独立于阈值的指标，值越高说明真阳率和假阳率之间有更好的平衡，说明开放集识别性能更好。适用于样本类分布较为均匀的情况
AUPR	精确率-召回率曲线下的面积	-	显示了精确率和召回率之间的变化关系，也是一个独立于阈值的指标
OSCR	正确分类率-错误识别率曲线下的面积	-	不容易受到数据集偏差的影响，可用于对类别不平衡应用场景的测试
OpenAUC	COTPR-OFPR曲线下的面积	见文中公式(43)	能够同时评估模型在封闭集和开放集上的性能，并且具有阈值不敏感性的优点

注：表中TP为正类中预测正确的样本数；FP为负类中预测错误的样本数；TN为负类中预测正确的样本数；FN为正类中预测错误的样本数。

OpenAUC是Wang等人[114]2022年提出的用于评测模型在开放集识别任务中性能的指标。计算的是条件开放集真阳率COTPR（Conditional Open-set True Positive Rate）和开放集假阳率OFPR（Open-set False Positive Rate）曲线下的面积。其中，COTPR表示模型对封闭集已知类样本分类正确的概率，OFPR表示模型将未知类样本误分到已知类中的概率。具体计算公式如下所示：

$$\begin{aligned} OpenAUC&=\int_{-\infty}^{+\infty}COTPR\left(OFPR^{-1}(t)\right)\mathrm{d}t\\ COTPR(t)&=E_{z\sim D_k}\left[1\left[r(\boldsymbol{x})\leqslant t,y=h(\boldsymbol{x})\right]\right]\\ OFPR(t)&=E_{z\sim D_u}\left[1\left[r(\boldsymbol{x})\leqslant t\right]\right] \end{aligned}\tag{43}$$

其中，t表示阈值，D_k表示已知类数据，D_u为未知类数据，$h(\boldsymbol{x})$表示预测所属类别的概率，$r(\boldsymbol{x})$表示样本$\boldsymbol{x}$属于未知类别的概率。OpenAUC能够同时评估模型在封闭集和开放集上的性能，并且具有阈值不敏感性的优点。在满足特定的假设下，作者通过理论证明了最大化OpenAUC指标可以获得更好的开放集识别性能。

4.3 效果对比

开放集识别模型的效果评测主要聚焦在两个方面，一是模型的开放集识别整体效果，即模型在未知类样本识别以及已知类分类上的效果；另一方面，开放集识别模型会进一步关注模型单纯在已知类上的

分类效果(例如封闭集准确率),判断模型的未知类识别能力是否对已知类分类能力有较大的损害。本节对现有经典开放集识别模型的评测结果进行了统计和梳理。根据实验中对数据设置的不同方式,本文分别提供了类分离和类添加(详见4.1节)设置下的开放集识别效果对比,并进一步统计了模型在单纯已知类分类上的效果。

4.3.1　类分离设置下的开放集识别模型效果对比

基于类分离数据设置的模型在开放集识别上的整体效果如表5所示。本文梳理和统计了类分离设置下可获得的开放集识别模型在MNIST、SVHN、CIFAR10、CIFAR+10、CIFAR+50以及TinyImageNet数据集上的AUROC指标(5次随机评测的均值)。可以发现,效果最好的是基于数据生成的IT-OSR-TransP算法[69],在MNIST、CIFAR+10、CIFAR+50以及TinyImageNet数据集上表现都是最优的。究其原因,该方法通过可靠性采样模块和特征生成模块解决了样本不平衡问题,并且IT-OSR-TransP由更复杂的网络结构Swin Transformer作为特征提取器和一个三层感知机作为分类器,模型复杂

表5　开放集识别方法在类分离数据设置下的AUROC结果对比

方法	发表年限	辅助数据集	MNIST	SVHN	CIFAR10	CIFAR+10	CIFAR+50	TinyImageNet
SoftMax+Threshold[22]	-	-	0.978	0.886	0.677	0.816	0.805	0.577
OpenMax[29]	2016	无	0.981	0.894	0.695	0.817	0.796	0.576
CROSR[30]	2019	无	0.991	0.899	-	-	-	0.589
C2AE[44]	2019	无	0.989	0.922	0.895	0.955	0.937	0.748
GFROSR(Plain CNN)[45]	2020	无	-	0.935	0.807	0.928	0.926	0.608
GFROSR(WRN-28-10)[45]	2020	无	-	0.955	0.831	0.915	0.913	0.647
CGDL[46]	2020	无	0.994	0.935	0.903	0.959	0.950	0.762
CPN(OVA+PL, DR)[33]	2020	无	0.990	0.926	0.771	0.839	0.839	0.617
CPN(OVA+PL, PR)[33]	2020	无	0.987	0.924	0.828	0.881	0.879	0.639
RPL[34]	2020	无	0.989	0.934	0.827	0.842	0.832	0.688
RPL++[34]	2020	无	0.993	0.951	0.861	0.856	0.850	0.702
RPL-WRN[34]	2020	无	0.996	0.968	0.901	0.976	0.968	0.809
OpenHybrid[47]	2020	无	0.995	0.947	0.950	0.962	0.955	0.793
ARPL[35]	2021	无	0.997	0.967	0.910	0.971	0.951	0.782
CAC[32]	2021	无	0.991	0.941	0.801	0.877	0.870	0.760
CVAECapOSR[39]	2021	无	0.992	0.956	0.835	0.888	0.889	0.715
GMVAE[48]	2021	无	0.989	0.941	0.896	0.952	0.947	0.782
PMAL[36]	2022	无	0.997	0.970	0.951	0.978	0.969	0.831
MOODCAT[49]	2022	无	-	-	0.895	0.894	0.892	-
CSSR[50]	2022	无	-	0.979	0.913	0.963	0.962	0.823
LNAOSR[52]	2022	无	0.961	0.949	0.843	0.840	0.871	-
MetaMax[31]	2023	无	0.997	**0.997**	0.938	-	-	0.846
MGPL[40]	2023	无	-	0.957	0.840	0.927	0.918	0.730
ODL[37]	2023	无	0.996	0.954	0.885	0.911	0.906	0.746
JNICS[38]	2024	无	-	0.957	0.895	0.962	0.957	0.753
G-OpenMax[62]	2017	有	0.984	0.896	0.675	0.827	0.819	0.580
OSRCI[63]	2018	有	0.988	0.910	0.699	0.838	0.827	0.586
PROSER[65]	2021	有	-	0.943	0.891	0.960	0.953	0.693
GCM-CF[66]	2021	有	0.830	0.708	0.720	0.815	0.817	-
OpenGAN[61]	2021	有	**0.999**	0.988	**0.973**	-	-	0.907
DIAS[64]	2022	有	0.992	0.943	0.850	0.920	0.916	0.731
BCROSR[59]	2022	有	-	0.956	0.948	0.961	0.957	0.785
ConOSR[67]	2023	有	0.997	0.988	0.937	0.979	0.970	0.796
OpenMix+[68]	2023	有	0.981	-	0.869	0.931	0.925	0.751
IT-OSR-ARPL[69]	2024	有	**0.999**	0.982	0.952	0.990	0.991	0.849
IT-OSR-TransP[69]	2024	有	**0.999**	0.983	0.965	**0.991**	**0.993**	**0.943**

度更高。效果次优的是基于离群值和数据生成的OpenGAN算法[61]，该算法不仅借助真实的其它类数据，也借助生成器生成特征级别的伪开放集数据，极大地增强了训练阶段未知类别样本的数据多样性。

4.3.2 类添加设置下的开放集识别模型效果对比

基于类添加数据设置的模型在开放集识别上的整体效果，如表6所示。表6-1中的模型以CIFAR10作为已知类数据集，分别以ImageNet-crop、ImageNet-resize、LSUN-crop、LSUN-resize等数据集作为未知类数据。表6-2的模型以MINST为已知类数据集，分别以Omniglot、MNIST-noise、Noise数据作为未知类数据集，其中，Omniglot是一个来自各种语言字母表的手写字符数据集；Noise是通过对[0,1]上的均匀分布独立采样每个像素值生成随机图像；MNIST-Noise是将MNIST的测试集图像叠加在Noise上得到的。每个数据集均有10,000个测试图像，因此已知类样本数量与未知类样本数量的比例保持为1∶1。

表6-1 开放集识别方法在CIFAR10上基于类添加设置的Macro-F1结果对比

方法	年限	辅助数据集	CIFAR10			
未知类样本来源			ImageNet-crop	ImageNet-resize	LSUN-crop	LSUN-resize
SoftMax[22] FCN/LadderNet/DHRNet	-	-	0.639/0.640/0.645	0.653/0.646/0.649	0.642/0.644/0.650	0.647/0.647/0.649
OpenMax(AlexNet)[29]	2016	无	0.660	0.648	0.657	0.668
OpenMax(LadderNet)[29]	2016	无	0.653	0.670	0.652	0.659
OpenMax(DHRNet)[29]	2016	无	0.655	0.675	0.656	0.664
OSRCI[63]	2018	有	0.636	0.635	0.650	0.648
CROSR(LadderNet)[30]	2019	无	0.621	0.631	0.629	0.630
CROSR(DHRNet)[30]	2019	无	0.721	0.735	0.720	0.749
C2AE[44]	2019	无	0.837	0.826	0.783	0.801
GFROR(Activation)[45]	2020	无	0.757	0.792	0.751	0.805
GFROR(SoftMax)[45]	2020	无	0.821	0.777	0.843	0.784
CGDL[46]	2020	无	0.840	0.832	0.806	0.812
RPL[34]	2020	无	0.811	0.810	0.846	0.820
PROSER[65]	2021	有	0.849	0.824	0.867	0.856
CVAECapOSR[39]	2021	无	0.857	0.834	0.868	0.882
ARPL[35]	2021	无	0.858	0.830	0.845	0.867
CAC[32]	2021	无	0.764	0.752	0.756	0.777
CSSR[50]	2022	无	0.929	0.909	0.941	0.935
BCROSR[59]	2022	有	0.876	0.869	0.880	0.877
MGPL[40]	2023	无	0.862	0.862	0.869	0.868
ODL[37]	2023	有	0.861	0.842	0.871	0.856
ConOSR[67]	2023	有	0.891	0.843	0.912	0.881
OpenMix+[68]	2023	有	0.865	0.887	0.878	0.899
JNICS[38]	2024	无	0.842	0.884	0.851	0.881
IT-OSR[69]	2024	有	**0.959**	**0.973**	**0.971**	**0.971**

可以看到，表6-1效果最好的是基于数据生成的IT-OSR算法[69]，表6-2效果最好的是基于数据生成的DIAS算法[64]。获得效果提升的原因是多方面的，例如DIAS方法[64]不仅使用GAN生成的伪未知类数据辅助开放集分类器的学习，还生成其它不同难度的未知类样本来模拟开放世界中较为复杂的未知类输入的情况，可以更好学习开放集分类器来提升模型在开放式环境中的识别性能。MGPL方法[40]使用高斯混合原型学习模拟真实世界的复杂分布，对比于如CGDL方法[46]仅使用先验高斯分布所学习出的模型，可以较好地提升模型对未知类样本的识别效果。

4.3.3 封闭集分类效果对比

表7展示了开放集识别模型单纯在已知类上的分类效果，用封闭集准确率来衡量。可以看出，这些算法在小规模简单数据集，例如MNIST、SVHN、CIFAR10等，都达到了很好的分类效果（大多都在90%以上），但在TinyImageNet数据集上的效果则较差（大多在52%～84%之间）。其主要原因应为

表 6-2　开放集识别方法在 MNIST 上基于类添加设置的 Macro-F1 结果对比

方法	年限	辅助数据集	MNIST		
未知类样本来源			Omniglot	MNIST-noise	Noise
SoftMax+Threshold[22]	-	-	0.595	0.801	0.829
OpenMax(DHRNet)[29]	2016	无	0.780	0.816	0.826
CROSR[30]	2019	无	0.793	0.827	0.826
CGDL[46]	2020	无	0.850	0.887	0.859
CVAECapOSR[39]	2021	无	0.971	0.982	0.982
PROSER[65]	2021	有	0.862	0.874	0.882
DIAS[64]	2022	有	**0.989**	0.982	**0.989**
MGPL[40]	2023	无	0.981	0.978	0.981
ODL[37]	2023	无	0.982	0.918	0.984
ConOSR[67]	2023	有	0.954	**0.987**	0.988

表 7　开放集识别模型在已知类上的分类准确率对比

方法	年限	辅助数据集	MNIST	SVHN	CIFAR10	CIFAR+	TinyImageNet
SoftMax[22]/OpenMax[29]	2016	无	0.995	0.947	0.801	-	-
G-OpenMax[62]	2017	有	0.996	0.948	0.816	-	-
OSRCI[63]	2018	有	0.996	0.951	0.821	-	-
CROSR[30]	2019	无	0.992	0.945	0.930	-	-
CPN(OVA+PL, DR)[33]	2020	无	0.997	0.967	0.929	-	0.814
RPL-WRN[34]	2020	无	0.996	0.958	0.951	-	0.817
GFROR(Plain CNN)[45]	2020	无	-	0.966	0.928	0.944	0.492
GFROR(WRN-28-10)[45]	2020	无	-	0.973	0.9509	0.974	0.559
CGDL[46]	2020	无	0.996	0.942	0.912	-	-
OpenHybrid[47]	2020	无	0.947	0.929	0.868	-	-
PROSER[65]	2021	有	-	0.964	0.926	-	0.521
ARPL[35]	2021	无	0.995	0.943	0.879	-	0.659
CAC[32]	2021	无	0.998	0.970	0.934	0.952	0.759
GMVAE[48]	2021	无	0.996	0.962	0.946	0.952	0.729
GCM-CF[66]	2021	有	0.830	0.708	0.720	0.815	0.817
PMAL[36]	2022	无	**0.998**	0.971	**0.975**	-	0.847
CSSR[50]	2022	无	-	0.953	-	-	-
BCROSR[59]	2022	有	-	0.974	0.973	0.976	0.802
DIAS[64]	2022	有	0.997	0.970	0.947	0.964	0.700
MGPL[40]	2023	无	0.996	0.967	0.932	-	0.547
ODL[37]	2023	无	0.998	0.969	0.931	0.957	0.735
ConOSR[67]	2023	有	0.817	**0.988**	0.937	0.979	0.796
JNICS[38]	2024	无	0.974	0.964	0.963	0.965	0.782
IT-OSR[69]	2024	有	**0.999**	0.983	0.965	**0.991**	**0.943**

TinyImageNet数据集具有更复杂的数据类别和形态的多样变化，大大增加了其分类难度。

4.3.4　总结和分析

从实验结果上看，目前开放集识别模型在性能上获得的提升与以下几个方面密切相关：(1)学习更接近真实世界的数据分布。真实世界未知类样本的分布往往复杂多样，模拟真实世界的复杂分布有利于提高模型的开放集识别能力。例如，基于距离的MGPL算法[40]使用高斯混合原型学习来模拟数据的复杂分布，较好地提升了模型对未知类样本的识别效果。(2)借助已有的或者生成更复杂更接近真实世界的未知类样本可以显著地提升开放集识别模型对未知类样本的识别能力。例如，OpenGAN算法[61]通过结合真实的其它类数据和生成器生成的伪开放集数据，较好地模拟了真实世界中复杂的未知类别，获得了优异的未知类识别能力。而且从总体

来看，借助了辅助数据集的开放集识别模型相较于未使用辅助数据集的模型获得了更好的效果。(3)更复杂的网络结构。Vaze等人[103]证明封闭集准确度与开放集识别能力高度相关，更复杂的网络结构能够带来更好的封闭集准确度，并且提高开放集识别的能力。例如，IT-OSR-TransP算法[69]使用更复杂的网络结构，在不同数据集及不同设置下均获得不错的性能表现；GFROSR算法[45]使用不同复杂度的网络结构证明了复杂度高的网络结构可以使模型的开放集识别AUROC值提升1%至4%。

5 相似研究问题对比

开放集识别问题打破了传统封闭集分类问题对训练数据和测试数据在类别空间上的一致性要求，它的问题设置与一些其它研究问题具有一定的相似性，例如零样本/小样本学习、迁移学习、域自适应学习、异常检测、主动学习、分布式检测、开放词汇目标检测、开放世界识别等。本文将对这些研究问题进行简单的介绍，并阐述它们与开放集识别问题的主要异同点(详见表8)。

表 8 相似研究问题对比

研究问题	数据设置		学习目标	关注点
	训练集是否提供额外信息	测试集		
零样本学习 小样本学习	新类别的边际信息 少量已标注的新类别样本	存在不可见类	已知类分类+未知类具体类别分类	模型在新类别上的泛化性
迁移学习域自适应学习	少量目标域已标记的样本	目标域数据	目标域上数据分类	模型在目标域上的泛化性
异常检测	否	存在离群点或噪声数据	异常样本检测	识别与大规模数据不相似的样本
主动学习	否	与训练集无不一致要求	未标注数据中信息量大的样本选择	挖掘未标注的富有信息量的样本
分布外检测	否	存在分布外数据	分布外样本识别(二分类)	分布外数据的准确识别
开放词汇目标检测	否	存在不可见类	未知类识别	计算机视觉领域下的开放目标检测
开放世界识别	否	存在不可见类	已知类分类+未知类识别+未知新类别的持续学习	未知类别动态特征的持续学习
开放集识别	否	存在不可见类	已知类分类+未知类识别	对已知类别的准确分类和对分布外数据的准确识别

5.1 零样本/小样本学习

零样本学习(Zero-Shot Learning，ZSL)由Larochell等人[115]首次提出，通过挖掘可见类和未见类之间类别维度的语义联系，使得模型能够识别在训练过程中没出现的未见类[116]。根据测试样本是否包含可见类，零样本学习可以分为传统零样本学习(Conventional ZSL，CZSL)和广义零样本学习(Generalized ZSL，GZSL)。在传统零样本学习中，测试集仅包含未见类；在广义零样本学习中，测试集同时包含可见类和未见类，更具有挑战性。零样本学习在训练的过程中除了已知类的样本信息，还可以获得关于未知类的边际信息(Side Information)[117]，并不是对未知类全无所知。其中，边际信息有多种表示方式，可以是类别的高层语义描述，也可以是类别的文本描述信息；它可以由领域专家给出，也可以用海量的附加数据学习获得，例如利用知识图谱中包含的类别关系作为边际信息来帮助学习。因为可见类和未见类的语义信息在训练时是可见的，如何利用语义信息搭建从可见类到未见类之间的桥梁成为解决零样本学习的关键。更多关于零样本学习的介绍和细节可参考Ren等人[118]的综述文章。

小样本学习(Few-Shot Learning，FSL)[57]是研究利用目标类别少量监督信息来训练机器学习模型的方法。它与零样本学习的类别空间定义一致：在训练阶段可以获得较多的已知类别样本，并获得少量的未见类(目标类别)样本；在测试阶段需对目标类别实现良好的分类效果。当训练阶段目标类别只有一个可获得样本时，小样本学习可具象化为单样本学习(One-Shot Learning)[119]。当前的小样本学习模型主要可以分为基于模型优化、度量学习以及数据增强的三类学习方法[120]。更多关于小样本学习的介绍和细节可参考Song等人的综述文章[121]。

综上，虽然开放集识别和零样本/小样本学习的训练类别空间和测试类别空间都存在不一致性，即在数据设置上都存在训练阶段不可见的位置类别，但它们在学习任务的设置上有较大的差异：(1)开放集识别中没有能够描述未知类别信息的边际信息。零样本学习中假设测试集中出现的新类别有一定的边际信息可在训练阶段获取；小样本学习训练时可直接利用少量已标注的新类别样本。开放集识别在类别空间的设置上更具有一般性，因为更多的未知类别是无法事先获知、被准确描述的，无法获取未知类别的边际信息或者具体的样本，例如新出现的计算机病毒和新发现的物种等。(2)零样本学习和小样本学习的最终分类器除了要实现对已知类的分类，还需要实现对未知类具体类别的分类，而不是像开放集识别模型将所有未知类归纳为一个增广类进行识别。零样本学习和小样本学习对测试阶段出现的未知类具有更加详细和深入的认识。

5.2　迁移学习 & 域自适应学习

迁移学习(Transfer Learning)[58]旨在将数据丰富的源域(Source Domain)信息作为辅助知识，迁移到目标领域(Target Domain)，帮助目标领域训练可靠的决策函数，解决目标领域的应用问题[122]。迁移学习放宽了传统机器学习方法要求训练数据和测试数据服从相同概率分布的限制，只需要源域和目标域之间具有一定的关联。它主要关注跨领域的知识迁移，从而解决目标领域较难获取有标注数据或者仅有少量标注数据时的学习问题，例如对罕见病症的研究[123]，以降低在目标领域收集大规模标注数据所需的昂贵人工成本和时间成本。迁移学习经常借助大规模预训练模型的力量，典型的方法是使用少量目标域样本对预训练模型进行微调。更多关于迁移学习的介绍和细节可参考Zhu等人的综述文章[124]。

域自适应学习(Domain Adaption)[58]是迁移学习的一种特殊情况，是实现迁移学习的重要方法之一。由于迁移学习中训练集和测试集来自不同的分布，容易造成域的偏移(Domain Shift)，这种数据上的域的偏移使得训练好的模型在测试集的效果往往很差。更多关于域自适应学习的介绍和细节可参考Yu等人的综述文章[125]。

虽然迁移学习和域自适应学习与开放集识别学习一样，训练集和测试集的数据分布具有不一致性，但他们之间也存在很多不同之处：(1)迁移学习关注于模型在目标域上的泛化，而开放集识别更关注模型在已知类别上的分类可靠性。(2)迁移学习通常利用预训练模型来进行样本特征提取，并使用少量目标域已标记的样本训练新的分类器，它要求在模型训练之初就对目标域有一定的了解和限制；开放集识别的未知空间更加广阔，对开放集识别方法的泛化性提出了更高要求。

5.3　异常检测

异常检测(Outlier Detection)[126]又叫如异常识别、离群点检测、噪声检测(Anomaly Detection)、偏差检测(Deviation Detection)和例外挖掘(Exception Mining)等，旨在识别不符合“正常模式”的样本。异常检测模型的建模思路与开放集识别模型的建模思路有很多相似的地方，例如利用自编码器对样本进行重构，通常异常样本的重构误差会更大，从而可将重构误差作为区分正常样本和异常样本的参考指标之一。更多关于异常检测的介绍和细节可参考Sikder等人的综述文章[127]。

异常检测和开放集识别都聚焦于识别非目标类别的未知样本，其中目标类别指模型在训练阶段学习的类别。但他们在问题设置上也存在一些差异：(1)异常检测中的离群点或者噪声通常是相对于正常样本的自然偏离，例如偶发性的仪器错误，人工错误或者异常行为等。开放集识别旨在学习一个分类器，使其能在正确分类已知类样本的同时有效识别未知类样本。这里的未知类强调的是一种类别模式，离群点或者噪声通常没有统计意义上所形成的类别模式。(2)在学习目标上，开放集识别需要对所有的类别(包括已知类和增广类)进行分类，异常检测更多聚焦于异常样本的检测，关注异常样本的召回率。

5.4　主动学习

主动学习(Active Learning)[128]是处理缺少类标签数据的有效方法，旨在缓解标注成本过高的问题。它通过选择未标注数据集中信息量最大的样本，将这些样本进行人工标注，并放入训练集进行训练，更新模型，并在这样的反复迭代中以较少的标注成本实现模型性能的显著提升。主动学习和半监督学习都是利用少量有标注样本数据和大量无标注数据提升分类模型性能的方法，半监督学习侧重探索无标注数据中的模型已知部分，而主动学习则尝试挖掘未知的富有信息量的数据，它们具有内在相似性和良好互补性。样本选择策略的好坏直接决定主动学习方法的性能。更多关于主动学习的介绍和细节可参考Ren等人的综述文章[128]。

主动学习与开放集识别的主要差异点在于：(1)主动学习与开放集识别方法的目的不同，主动学习的目的是挑选更好的未标记的样本，通过结合适量的人工标注，使得模型能够以较低的标注成本获得最大程度性能的提升。(2)在训练阶段，开放集识别在训练的时候不出现未知类样本，强调训练集和测试集在类别空间中的不一致性；主动学习对此没有特殊要求。

5.5 分布外检测

分布外检测(Out-of-Distribution Detection，OOD Detection)[129]旨在识别并拒绝与训练数据(In-distribution data，IND数据)不同的数据样本(即Out-of-Distribution Data，OOD数据)。这里的OOD数据可以是来自未知类的样本，也可以是噪声数据或者异常值数据。分布外检测与异常检测的区别在于前者更加侧重于测试阶段出现的未见过的、不同于训练数据的样本，不局限于“异常值”。同时，在分布外检测模型的训练过程中，我们通常设定其在训练阶段无法获得任何分布外数据的相关信息或者样本，而异常检测通常在训练阶段可以有少量的异常样本供模型学习。分布外检测与开放集识别的主要区别在于开放集识别任务通常将问题归纳为$K+1$类分类问题，而分布外检测任务一般归纳为二分类任务，即把输入样本分类为“分布内样本”或者“分布外样本”。如果我们将一个分布外检测模型和一个闭集分类器模型结合起来，它们可以通过二阶段的模式较好地完成开放集识别任务。

对于分布外检测问题的研究，现有方法大致可以分为两类[130]。第一类方法侧重于为已有的DNN模型(例如预训练模型)设计基于阈值的检测函数或者评分函数(Scoring Function)来检测OOD样本。当输入样本的分数小于设定的阈值时，将其判断为OOD样本；第二类方法侧重于重新训练DNN模型，通过获得更好的区分IND样本和OOD样本的样本表征来实现分布外检测。更多关于分布外检测的介绍和细节可参考Cui等人的综述文章[129]。

5.6 开放词汇目标检测

开放词汇目标检测(Open Vocabulary Object Detection)[131]主要研究计算机视觉任务中面向开放词汇下的目标检测任务，其核心目标在于发现并识别开放(无限制)词汇定义下的新对象，也就是在训练过程中未曾出现或者标注的类别。而传统的目标检测任务通常针对一组预定义的对象类别进行训练和检测，对于超出预定义词汇表的对象，传统目标检测模型往往无法准确检测和识别。

开放词汇目标检测最先由Zareian等人[132]在2021年被提出，通过借助于大量的图片-描述(image-caption)数据来覆盖更多的目标检测类别，使目标检测不再受限于带标注数据的少量已知类别，实现了更加泛化的目标检测，可自动识别更多未知物体类别。更多关于开放词汇目标检测的介绍和细节可参考Wu等人的综述文章[131]。

开放词汇目标检测与开放集识别的任务主题有较大差别，前者关注于计算机视觉中的目标检测任务，后者关注于分类任务且不局限于计算机视觉领域。但其共同点都在于打破了传统任务中对类别的封闭式预定义，具有更好的泛化性和鲁棒性，更加适用于真实应用场景。

5.7 开放世界识别

开放世界识别(Open World Recognition)[21]是近几年被提出的新概念，旨在对真实的现实世界进行建模和学习，目前较多地聚焦在自动驾驶汽车[133]、机器人[134]和在线视觉系统[135]等复杂应用场景中。在这些应用场景中，模型需要能够处理未知的输入，能够学习新的类别，并适应不断变化的环境。因此，开放世界识别任务通常需要完成三个子任务：对已知类别的分类、对未知类别的检测以及对未知的新类别的持续学习。

开放世界识别面临的挑战则主要来自于以下三方面：(1)对于训练好的端到端模型来说，更新特征是很困难的；(2)基于现有的类别特征不足以检测未知类别；(3)相同类别之间的实例关系难以利用。针对这些挑战，研究者们利用不同的技术对其进行了研究。更多关于开放世界识别的介绍和细节可参考Parmar等人的综述文章[136]。

开放世界识别相比于开放集识别来说，前者所定义的问题背景更加复杂，预期解决的任务范围更广，它包含了开放集识别问题，不仅需要检测并拒绝未知类样本的输入，还包括了未知类别和动态特征的持续学习和管理。

6 总结和展望

得益于神经网络和深度学习的快速发展，有监督分类模型的性能得到了显著提升且被应用于众多现实场景中。然而，越来越多应用场景面临着新类别的不断涌现，传统机器学习模型对完备类别信息

的过度依赖，难以有效解决这一问题。因此，针对开放动态环境的开放集识别方法成为新的研究热点。它扩展了传统的分类问题，在保证已知类别准确分类的同时，还要求模型能够有效地识别测试阶段新出现的未知类别样本，以避免大量误分类。本文对近年来开放集识别的研究进行了系统调研，聚焦于基于深度学习的开放集识别方法，对经典模型进行了梳理和介绍，对其分类效果进行横向对比，并对与其相似的研究问题，例如零样本/小样本学习、迁移学习、域自适应学习、主动学习、分布外检测、开放词汇目标检测、开放世界识别等学习、主动学习等进行了对比和介绍。

从总体来看，近年来基于深度学习的开放集识别在研究热度和模型性能上都有显著的提升，但也仍存在一些不足。例如，(1)基于评分函数的开放集识别的方法往往在最后检测未知类阶段依赖于阈值，在实际应用中如何获取合适的阈值或者设置一个自适应的阈值仍然具有一定的挑战性；(2)基于距离的方法能够有效学习语义特征并且限制开放集风险，将各个已知类别在特征空间上尽可能紧凑可分，为未知类别留下空间，具有良好的可行性。但当前大多数基于距离的方法假设已知类别的潜在数据分布是固定且唯一的，这与现实场景下的复杂分布有所差异，也影响了其开放集识别的性能；(3)基于重构的方法利用自编码器的重构损失能够有效区分已知类和未知类，但容易学习与分类任务不相关的信息，影响封闭集准确率；(4)基于离群值的方法简单有效，利用与测试集不相关的未知类样本作为辅助数据集，直接训练$K+1$类的分类器，其难点在于辅助数据集的获得与选取，这将直接影响其开放集分类的性能；(5)基于数据生成的方法，利用生成网络生成伪未知类样本加入训练，获得$K+1$类分类器，缓解了前面基于离群值方法中未必可以获得辅助数据的问题，如何生成更符合真实世界的未知类样本是该方法面临的主要挑战。

从另一个角度出发，当前的开放集识别方法也可被归纳为判别式方法和生成式方法两大类。

(1)判别式方法利用判别模型学习并建立能够划分已知类别和未知类别的决策边界，判别模型在决策时需要使用经验设置的阈值，将输入样本分类到某个已知类或做出具体的拒绝响应。但阈值的选择通常取决于已知类的知识，不可避免地会由于缺乏来自未知类的可用信息而产生开放集风险，且需根据数据集的变化进行重新调整。如何获取自适应的、自学习的阈值仍是开放集识别未来的重要研究方向之一。

(2)生成式方法利用生成模型对训练数据建模或生成未知类样本，生成式的增广类识别模型通过对未知类别样本的模拟，使模型对开放域未知空间具有更加具象的了解和更好的解释，为增广类识别提供了另一种思路，但其在未知类样本生成成本方面，当前的生成模型大多设计复杂，训练成本高，同时无法保证生成的未知类样本和已知类样本空间不重合。如何设计更加简洁、高效的伪未知类生成模型是开放集识别的未来的重要研究方向之一。

综合来说，如何精确度量开放集识别模型的预测标签置信度，确定未知类别样本的归属不确定度是开放集识别模型构建的关键，在开放集识别模型的训练模式、建模思路以及多类型数据适配等方面都还有大量的研究空间，例如：

(1)在训练模式方面，开放集识别模型的训练过程中，不同类别的学习速率有较大的不均衡，如何平衡模型训练初期不同类别学习速率上的不均衡性，增强模型初期对未知空间的探索也是开放集识别未来的重要研究方向之一。

(2)在建模思路方面，预训练大模型和多模态大模型的出现带来了新的建模思路。通过对大规模数据中统计规律的学习和对不同模态信息的融合，多模态大模型可极大地促进开放集识别模型在性能上的提升和更多实际应用场景的适配，同时也使模型在面对噪声数据时表现得更加鲁棒。大模型在开放集识别领域的进一步应用和适配是极有价值和前景的研究方向之一。

(3)在数据适配方面，图像数据上的开放集识别方法拥有相对更多的研究基础，许多方法在中小规模数据集以及大规模数据集上均表现出色。然而，对于其他类型的数据，例如文本数据和图数据等，鉴于数据类型的差异，已有的针对图像数据的开放集识别方法并不能直接应用。如何在更多常见数据类型和更多应用场景中实现有效的开放集识别，也是该领域未来的重要研究方向之一。

作者贡献声明　张鹏、刘涵同为通信作者。

参　考　文　献

[1] Muscolo G G, Fiorini P. Force-torque sensors for minimallyinvasive surgery robotic tools: An overview. IEEE Transactions on Medical Robotics and Bionics, 2023, 5(3):

458-471

[2] Teng S, Hu X, Deng P, et al. Motion planning for autonomous driving: The state of the art and future perspectives. IEEE Transactions on Intelligent Vehicles, 2023, 8(6): 3692-3711

[3] Mohsan S A H, Othman N Q H, Li Y, et al. Unmanned aerial vehicles (UAVs): Practical aspects, applications, open challenges, security issues, and future trends. Intelligent Service Robotics, 2023, 16(1): 109-137

[4] Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 427-436

[5] Scheirer W J, de Rezende R A, Sapkota A, et al. Toward open set recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(7): 1757-1772

[6] Padmanabhan R, Padmanabhan D R T, Thanigaivelu K. Convolutional neural networks for vehicle damage detection. AIP Conference Proceedings, 2023, 2790(1): 100332.

[7] Krentzel D, Shorte S L, Zimmer C. Deep learning in imagebased phenotypic drug discovery. Trends in Cell Biology, 2023, 33(7): 538-554

[8] Liu M Q, Zhang Z J, Chen Y F, et al. Adversarial attack and defense on deep learning for air transportation communication jamming//Proceedings of the IEEE Transactions on Intelligent Transportation Systems. Piscataway, USA, 2023: 973-986

[9] Cheng G, Yuan X, Yao X, et al. Towards large-scale small object detection: Survey and benchmarks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 13467-13488

[10] Thisanke H, Deshan C, Chamith K, et al. Semantic segmentation using vision transformers: A survey. Engineering Applications of Artificial Intelligence, 2023, 126: 106-669

[11] Berroukham A, Housni K, Lahraichi M, et al. Deep learningbased methods for anomaly detection in video surveillance: A review. Bulletin of Electrical Engineering and Informatics, 2023, 12(1): 314-327

[12] Zhang J, Lai Z P, Li X, et al. Cross-domain Chinese word segmentation based on new word discovery. Journal of Electronics & Information Technology, 2022, 44(9): 3241-3248 (in Chinese)
(张军, 赖志鹏, 李学等. 基于新词发现的跨领域中文分词方法. 电子与信息学报, 2022, 44(9): 3241-3248)

[13] Zhu J, Guo Y, Wan Y, Tian K. New word detection based on branch entropy-segmentation probability model. Computer Science, 2023, 50(7): 221-228 (in Chinese)
(祝钰莹, 郭燕, 万亿兆, 田凯. 基于信息熵-切分概率模型的新词发现方法. 计算机科学, 2023, 50(7): 221-228)

[14] Shu S, He S, Wang H, et al. A generalized unbiased risk estimator for learning with augmented classes//Proceedings of the AAAI Conference on Artificial Intelligence. Washington, USA, 2023: 9829-9836

[15] Wu M, Pan S, Zhu X. Openwgl: Open-world graph learning//Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM). Sorrento, Italy, 2020: 681-690

[16] Geng C, Huang S, Chen S. Recent advances in open set recognition: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(10): 3614-3631

[17] Salehi M, Mirzaei H, Hendrycks D, et al. A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges. arXiv preprint arXiv: 2110.14051, 2021

[18] Mahdavi A, Carvalho M. A survey on open set recognition//Proceedings of the 2021 IEEE 4th International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). Laguna Hills, USA, 2021: 37-44

[19] Gao F, Yang L, Li H. A survey on open set recognition. Journal of Nanjing University, 2022, 58(1): 115-134 (in Chineses)
(高菲, 杨柳, 李晖. 开放集识别研究综述. 南京大学学报(自然科学), 2022, 58(1): 115-134)

[20] Sun J, Dong Q. A Survey on open-set image recognition. arXiv preprint arXiv: 2312.15571, 2023

[21] Bendale A, Boult T. Towards open world recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1893-1902

[22] Hendrycks D, Gimpel K. A baseline for detecting misclassified and out-of-distribution examples in neural networks. arXiv preprint arXiv: 1610.02136, 2016

[23] Liang S Y, Li Y X, Srikant R. Enhancing the reliability of out-of-distribution image detection in neural networks. arXiv preprint arXiv: 1706.02690, 2017

[24] Dong X, Guo J F, Li A, et al. Neural mean discrepancy for efficient out-of-distribution detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022: 19217-19227

[25] Liu W T, Wang X Y, Owens D. O, et al. Energy-based outof-distribution detection//Proceedings of the International Conference on Neural Information Processing Systems. Oriando,USA 2020: 21464-21475

[26] Liu X X, Lochman Y, Zach C. Gen: Pushing the limits of softmax-based out-of-distribution detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada, 2023: 23946-23955

[27] Zhang Q, Li Q C, Zhang P, et al. A dynamic variational framework for open-world node classification in structured sequences//Proceedings of the 2022 IEEE International Conference on Data Mining (ICDM). Oriando, USA, 2022: 703-712

[28] Zhang Q, Liu Z Q, Li Q C, et al. Open-world structured sequence learning via dense target encoding. Information Sciences, 2024, 680:121147

[29] Bendale A, Boult T E. Towards open set deep networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 1563-1572

[30] Yoshihashi R, Shao W, Kawakami R, et al. Classification-reconstruction learning for open-set recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

Recognition. Piscataway，USA，2019：4016-4025

[31] Lyu Z，Gutierrez N B，Beksi W J. MetaMax：Improved open-set deep neural networks via weibull calibration//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway，USA，2023：439-443

[32] Miller D，Sunderhauf N，Milford M，et al. Class anchor clustering：A loss for distance-based open set recognition//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway，USA，2021：3570-3578

[33] Yang H M，Zhang X Y，Fei Y，et al. Convolutional prototype network for open set recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence，2022，44(5)：2358-2370

[34] Chen G Y，Qiao L M，Shi Y M，et al. Learning open set network with discriminative reciprocal points//Proceedings of the 16th European Conf on Computer Vision. Berlin，Germany，2020：507-522

[35] Chen G Y，Peng P X，Wang X Q，et al. Adversarial reciprocal points learning for open set recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence，2021，44(11)：8065-8081

[36] Lu J，Xu Y L，Li H，et al. Pmal：Open set recognition via robust prototype mining//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto，USA，2022：1872-1880

[37] Liu Z，Fu Y，Pan Q，et al. Orientational distribution learning with hierarchical spatial attention for open set recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence，2022，45(7)：8757-8772

[38] Park J，Park H，Jeong E，et al. Understanding open-set recognition by jacobian norm and inter-class separation. Pattern Recognition，2024，145：11

[39] Guo Y R，Camporese G，Yang W J，et al. Conditional variational capsule network for open set recognition//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway，USA，2021：103-111

[40] Liu J M，Tian J，Han W，et al. Learning multiple gaussian prototypes for open-set recognition. Information Sciences，2023，626：738-753

[41] Zhang Q，Li X W，Lu J X，et al. ROG_PL：Robust open-set graph learning via region-based prototype learning//Proceedings of the the AAAI Conference on Artificial Intelligence. Vancouver，Canada，2024：9350-9358

[42] Zhang H，Ding H H. Prototypical matching and open set rejection for zero-shot semantic segmentation//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal，Canada，2021：6954-6963

[43] Huang S Y，Ma J W，Han G X，et al. Task-adaptive negative envision for few-shot open-set recognition//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans，USA，2022：7161-7170

[44] Oza P，Patel V M. C2ae：Class conditioned auto-encoder for open-set recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach，USA，2019：2307-2316

[45] Perera P，Morariu V I，Jain R，et al. Generative discriminative feature representations for open-set recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway，USA，2020：11814-11823

[46] Sun X，Yang Z N，Zhang C，et al. Conditional gaussian distribution learning for open set recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA，2020：13480-13489

[47] Zhang H，Li A，Guo J，et al. Hybrid models for open set recognition//Proceeding of the 16th European Conference on Computer Vision. Glasgow，UK，2020：102-117

[48] Alexander C，Yuan L，Diego K. Open-set recognition with gaussian mixture variational autoencoders//Proceedings of the AAAI Conference on Artificial Intelligence. New York，USA，2020：6877-6884

[49] Yang Y Y，Gao R Y，Xu Q. Out-of-distribution detection with semantic mismatch under masking//Proceedings of the 17th European Conf on Computer Vision. Berlin，Germany，2022：373-390

[50] Huang H，Wang Y，Hu Q，et al. Class-specific semantic reconstruction for open set recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence，2022，45：4214-4228

[51] Liu C，Chun Y，Yin X C. Open-set text recognition via shape-awareness visual reconstruction//Proceedings of the International Conference on Document Analysis and Recognition. San Jose，USA 2023：89-105

[52] Zhang X L，Cheng X L，Zhang D H，et al. Learning network architecture for open-set recognition//Proceedings of the AAAI Conference on Artificial Intelligence. USA 2022：36

[53] Liao N，Liu Y，Xiaobo L，et al. Cohoz：Contrastive multimodal prompt tuning for hierarchical open-set zero-shot recognition//Proceedings of the 30th ACM International Conference son Multimedia. St. Petersburg，Russia，2022：3262-3271

[54] Liao N，Zhang X，Cao M，et al. R-tuning：Regularized prompt tuning in open-set scenarios. arXiv preprint arXiv：2303.05122，2023

[55] Ren H，Tang F，Pan X，et al. A2Pt：Anti-associative prompt tuning for open set visual recognition. IEEE Transactions on Multimedia，2023，26：8419-8431

[56] Dhamija A R，Günther M，Boult T. Reducing network agnostophobia//Proceedings of the Advances in Neural Information Processing System. Montréal，Canada，2018：31

[57] Hendrycks D，Mazeika M，Dietterich T. Deep anomaly detection with outlier exposure. arXiv preprint arXiv：1812.04606，2018

[58] Papadopoulos A A，Rajati M R，Shaikh N，et al. Outlier exposure with confidence control for out-of-distribution detection. Neurocomputing，2021，441：138-150

[59] Zhang J Y，Inkawhich N，Linderman R，et al. Mixture outlier exposure：Towards out-of-distribution detection in fine-grained environments//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa，

USA，2023：5531-5540

[60] Cho W，Choo J. Towards accurate open-set recognition via background-class regularization//Proceedings of the European Conference on Computer Vision (ECCV). Berlin，Germany，2022：658-674

[61] Kong S，Ramanan D. Opengan：Open-set recognition via open data generation//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal，Canada，2021：813-822

[62] Ge Z Y，Demyanov S，Chen Z T，et al. Generative openmax for multi-class open set classification. arxiv preprint arXiv：1707.07418，2017

[63] Neal L，Olson M，Fern X，et al. Open set learning with counterfactual images//Proceedings of the European Conference on Computer Vision (ECCV). Berlin，Germany，2018：613-628

[64] Moon W J，Park J，Seong H S，et al. Difficulty-aware simulator for open set recognition//Proceedings of the 17th European Conf on Computer Vision. Berlin，Germany，2022：365-381

[65] Zhou D W，Ye H J，Zhan D C. Learning placeholders for open-set recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville，USA，2021：4401-4410

[66] Yue Z，Wang T，Sun Q，et al. Counterfactual zero-shot and open-set visual recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville，TN，USA，2021：15404-15414

[67] Xu B，Shen F，Zhao J. Contrastive open set recognition//Proceedings of the AAAI Conference on Artificial Intelligence. Washington，D.C. USA，2023：10546-10556

[68] Jiang G，Zhu P，Wang Y，et al. OpenMix＋：Revisiting data augmentation for open set recognition. IEEE Transactions on Circuits and Systems for Video Technology，2023，33(11)：6777-6787

[69] Sun J，Dong Q. Conditional feature generation for transductive open-set recognition via dual-space consistent sampling. Pattern Recognition，2024，146：110046

[70] Zhang Q，Shi Z L，Zhang X L，et al. G2Pxy：Generative open-set node classification on graphs with proxy unknowns//Proceedings of the 32nd International Joint Conference on Artificial Intelligence. Macao，China，2023：4576-4583

[71] Ding C B，Pang G S and Shen C H. Catching both gray and black swans：Open-set supervised anomaly detection//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans，USA，2022：7378-7388

[72] Müller A. Integral probability metrics and their generating classes of functions. Advances in Applied Probability，1997，29(2)：429－443

[73] Ioffe S，Szegedy C. Batch normalization：Accelerating deep network training by reducing internal covariate shift//Proceedings of the International Conference on Machine Learning. Lille，France，2015：448-456

[74] LeCun Y，Chopra S，Hadsell R，et al. Predicting structured dara：A tutorial on energy-based learning. Predicting Structured Data，Cambridge，USA：MIT Press，2006

[75] Dawid A P，Musio M. Theory and applications of proper scoring rules. Metron，2014，72：169-183

[76] Kotz S，Nadarajah S. Extreme value distributions：Theory and applications. River Edge，NJ：World Scientific，2000

[77] Scheirer W J，Rocha A，Micheals R J，et al. Metarecognition：The theory and practice of recognition score analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence，2011，33(8)：1689-1695

[78] Sohn K. Improved deep metric learning with multi-class n-pair loss objective//Proceedings of the International Conference on Neural Information Processing Systems. Barcelona，Spain 2016：29

[79] Abouzaid S，Jaeschke T，Kueppers S，et al. Deep learning based material characterization using FMCW radar with open-set recognition technique. IEEE Transactions on Microwave Theory and Techniques，2023，71(11)：4628-4638

[80] Yang H M，Zhang X Y，Yin F，et al. Robust classification with convolutional prototype learning//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City，UT，USA，2018：3474-3482

[81] Snell J，Swersky K，Zemel R. Prototypical networks for few-shot learning//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach，USA，2017：4080-4090

[82] Vaswani A，Shazeer N，Parmar N，et al. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach，USA，2017：6000-6010

[83] Hinton G E，Krizhevsky A，Wang S D. Transforming autoencoders//Proceedings of the 21st International Conference on Artificial Neural Networks. Berlin，Germany，2011：44-51

[84] Zhang Q，Lu J X，Li X W，et al. CONC：Complex-noiseresistant open-set node classifcation with adaptive noise detection//Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI-24)，Jeju Island，Republic of Korea，2024

[85] Denouden T，Salay R，Czarnecki K，et al. Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance. arXiv preprint arXiv：1812.02765，2018

[86] Perez E，Strub F，De Vries H，et al. FiLM：Visual reasoning with a general conditioning layer//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto，USA，2018：32

[87] Gidaris S，Singh P，Komodakis N. Unsupervised representation learning by predicting image rotations. arXiv preprint arXiv：1803.07728，2018

[88] Zhang R，Isola P，Efros A A，et al. The unreasonable effectiveness of deep features as a perceptual metric//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City，USA，2018：586-595

[89] Liu P，Yuan W，Fu J，et al. Pre-train，prompt，and predict：A systematic survey of prompting methods in natural language

processing. ACM Computing Surveys，2023，55(9)，1-35

[90] Lake B M，Salakhutdinov R，Tenenbaum J B. Human-level concept learning through probabilistic program induction. Science，2015，350(6266)：1332-1338

[91] Pan S J，Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering，2010，22(10)：1345-1359

[92] Zhang H，Cisse M，Dauphin Y N，et al. Mixup：Beyond empirical risk minimization. arXiv preprint arXiv：1710.09412，2017

[93] Yun S，Han D，Oh S J，et al. Cutmix：Regularization strategy to train strong classifiers with localizable features//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul，Republic of Korea，2019：6023-6032

[94] Verma V，Lamb A，Beckham C，et al. Manifold mixup：Better representations by interpolating hidden states//Proceedings of the International Conference on Machine Learning. California，USA，2019：6438-6447

[95] Khosla P，Teterwak P，Wang C，et al. Supervised contrastive learning//Proceedings of the International Conference on Neural Information Processing Systems. Vancouver，Canada，2020：18661-18673

[96] Cubuk E D，Zoph B，Shlens J，et al. Randaugment：Practical automated data augmentation with a reduced search space//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA，2020：702-703

[97] Krizhevsky A，Sutskever I，Hinton G E. Imagenet classification with deep convolutional neural networks//Proceedings of the International Conference on Neural Information Processing Systems. Lake Tahoe，USA，2012：25

[98] Simonyan K，Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv：1409.1556，2014

[99] He K M，Zhang X Y，Ren S Q，et al. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas，USA，2016：770-778

[100] Zagoruyko S，Komodakis N. Wide residual networks. arXiv preprint，arXiv：1605.07146，2016

[101] Huang G，Liu Z，van Der Maaten L，et al. Densely connected convolutional networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu，HI，USA，2017：4700-4708

[102] Salimans T，Goodfellow I，Zaremba W，et al. Improved techniques for training gans//Proceedings of the International Conference on Neural Information Processing Systems. Barcelona，Spain，2016：29

[103] Vaze S，Han K，Vedaldi A，et al. Open-set recognition：A good closed-set classifier is all you need? //Proceedings of the International Conference on Learning Representations. Virtual，2022：1-26

[104] Yang T，Wang D，Tang F，et al. Progressive open space expansion for open-set model attribution//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver，Canada，2023：15856-15865

[105] Xu J，Grohnfeldt C，Kao O. OpenIncrement：A unified framework for open set recognition and deep class-incremental learning//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris，France，2023：3303-3311

[106] Pal D，More D，Bhargav S，et al. Domain adaptive fewshot open-set learning//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris，France，2023：18831-18840

[107] Grcić M，Bevandić P，Šegvić S. Densehybrid：Hybrid anomaly detection for dense open-set recognition//Proceedings of the European Conference on Computer Vision. Tel-Aviv，Israel，2022：500-517

[108] Cai J，Wang Y，Hsu H M，et al. Luna：Localizing unfamiliarity near acquaintance for open-set long-tailed recognition//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver，Canada，2022：131-139

[109] LeCun Y，Bottou L，Bengio Y，et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998，86(11)：2278-2324

[110] Krizhevsky A，Hinton G. Learning multiple layers of features from tiny images. Handbook of Systemic Autoimmune Diseases，2009，1(4)：3-54

[111] Netzer Y，Wang T，Coates A，et al. Reading digits in natural images with unsupervised feature learning//Proceedings of the Conference on Neural Information Processing Systems Workshop on Deep Learning & Unsupervised Feature Learning. Granada，Spain，2011：1-27

[112] Deng J，Dong W，Socher R，et al. ImageNet：A large-scale hierarchical image database//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami，USA，2009：248-255

[113] Russakovsky O，Deng J，Su H，et al. ImageNet large scale visual recognition challenge. International Journal of Computer Vision，2015，115：211-252

[114] Wang Z，Xu Q，Yang Z，et al. OpenAUC：Towards AUC-oriented open-set recognition//Proceedings of the International Conference on Neural Information Processing Systems. New Orleans，USA，2022：25033-25045

[115] Larochelle H，Erhan D，Bengio Y. Zero-data learning of new tasks//Proceedings of the AAAI Conference on Artificial Intelligence. Chicago，USA，2008：3

[116] Huang S，Yang W L，Zhang Y，et al. Feature generation approach with indirect domain adaptation for transductive zero-shot learning. Journal of Software，2022，33(11)：4268-4284 (in Chinese)
(黄晟，杨万里，张译等. 基于间接域适应特征生成的直推式零样本学习方法. 软件学报，2022，33(11)：4268-4284)

[117] Ren W，Tang Y，Sun Q，et al. Visual semantic segmentation based on few/zero-shot learning：An overview. IEEE/CAA Journal of Automatica Sinica，2023，(1)：1-21

[118] Wang W，Zheng V W，Yu H，et al. A survey of zeroshot learning：settings，methods，and applications. IEEE Transactions on Intelligent Systems and Technology (TIST)，2019，10(2)：1-37

[119] Huang Q, Zhang H, Xue M, et al. A survey of deep learning for low-shot object detection. ACM Computing Surveys, 2023, 56(5): 1-37

[120] Li X X, Liu Z Y, Wu J J, et al. Total relation network with attention for few-shot image classification. Chinese Journal of Computers, 2023, 46(2): 371-384 (in Chinese)
(李晓旭, 刘忠源, 武继杰等. 小样本图像分类的注意力全关系网络. 计算机学报, 2023, 46(2): 371-384)

[121] Song Y, Wang T, Cai P, et al. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. ACM Computing Surveys, 2023, 55(271): 1-40

[122] Chai Y M, Yuan W L, Wang L M, et al. A cross-domain recommendation model based on dual attention mechanism and transfer learning. Chinese Journal of Computers, 2020, 43(10): 1924-1942 (in Chinese)
(柴玉梅, 员武莲, 王黎明等. 基于双注意力机制和迁移学习的跨领域推荐模型. 计算机学报, 2020, 43(10): 1924-1942)

[123] Sisodia P S, Ameta G K, Kumar Y, et al. A review of deep transfer learning approaches for class-wise prediction of Alzheimer's disease using MRI images. Archives of Computational Methods in Engineering, 2023, 30(4): 2409-2429

[124] Zhu Z, Lin K, Jain A K, et al. Transfer learning in deep reinforcement learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 13344-13362

[125] Yu Z, Li J, Du Z, et al. A comprehensive survey on Sourcefree domain adaptation. arXiv preprint arXiv:2302.11803, 2023

[126] Pazho A D, Noghre G A, Purkayastha A A, et al. A survey of graph-based deep learning for anomaly detection in distributed systems. IEEE Transactions on Knowledge and Data Engineering, 2023, 36(1): 1-20

[127] Sikder M N K, Batarseh F A. Outlier detection using AI: A survey. AI Assurance, 2023, 231-291

[128] Ren P Z, Xiao Y, Chang X J, et al. A survey of deep active learning. ACM computing surveys (CSUR), 2021, 54: 1-40

[129] Cui P, Wang J. Out-of-distribution (OOD) detection based on deep learning: A review. Electronics, 2022, 11(21): 3500

[130] Zhou Z Y, Dou W S, Li Shuo, et al. DiTing: Semisupervised Adversarial Training Approach for Robust Outof-distribution Detection. Journal of Software, 2024, 35(6): 2936-2950 (in Chinese)
(周志阳, 窦文生, 李硕等. 谛听: 面向鲁棒分布外样本检测的半监督对抗训练方法. 软件学报, 2024, 35(6): 2936-2950)

[131] Wu J, Li X, Xu S, et al. Towards open vocabulary learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(7): 1-20

[132] Zareian A, Rosa K D, Hu D H, et al. Open-vocabulary object detection using captions//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 14393-14402

[133] Bao Z, Hossain S, Lang H, et al. A review of high-definition map creation methods for autonomous driving. Engineering Applications of Artificial Intelligence, 2023, 122: 106125

[134] Hu Y, Xie Q, Jain V, et al. Toward general-purpose robots via foundation models: A survey and meta-analysis. arXiv preprint arXiv:2312.08782, 2023

[135] Hadipour-Rokni R, Asli-Ardeh E A, Jahanbakhshi A, et al. Intelligent detection of citrus fruit pests using machine vision system and convolutional neural network through transfer learning technique. Computers in Biology and Medicine, 2023, 155: 106611

[136] Parmar J, Chouhan S, Raychoudhury V, et al. Openworld machine learning: Applications, challenges, and opportunities. ACM Computing Surveys, 2023, 55(10): 1-37

ZHANG Qin, Ph. D., associate professor. Her current research interests include open set learning, natural language processing and graph learning.

LIU Zi-Qi, master candidate. Her current research interests are open set learning and graph learning.

ZHANG Xiao-Lin, Ph. D. He current research interests are computer vision and digital man technology.

ZHANG Peng, Ph. D., associate professor. He current research interests are open set learning, computer vision and effective computing.

LIU Han, Ph. D., assistant professor. He current research interests are open set learning and natural language processing.

CHEN Xiao-Jun, Ph. D., professor. He current research interests are data mining and machine learning.

Background

Deep learning has become the main method for many computer vision problems, and has achieved amazing results in many visual recognition tasks, which has contributed to the maturity and implementation of a large number of intelligent systems. However, there is a common assumption in most current research methods for recognition systems that the training process takes place in closed environmental scenarios and all test categories can be seen during the training phase. This assumption causes existing machine learning models to fail to perceive emerging classes during the testing phase and instead over-confidently misclassify them into one of the existing classes during the training phase. In practical applications, it may cause irreparable losses in some key scenarios, such as new word detection, medical diagnosis, automatic driving and other fields.

Considering the requirements of real-world applications, open set recognition problem has attracted more and more researchers' attention in recent years, and has become a hot research direction in the field of machine learning. The open set recognition task was proposed to solve the problem that most existing machine learning models cannot be directly applied to open dynamic environments, which is crucial to ensure the reliability and safety of machine learning systems. It aims to enable machine learning models to deal with samples of unknown classes, rather than only known classes. Such tasks require models with good generalization performance and adaptability to make accurate predictions in unseen situations. The research on open set recognition task is of great significance to promote the application of machine learning systems in the real world. It provides us with more reliable and safe machine learning solutions, so that the models can better adapt to the changing environment and unknown situations.

In this survey, we conduct a systematic survey of the research on open set recognition in recent years, focusing on the open set recognition methods based on deep learning, combing and introducing the classical models, deeply comparing the advantages and disadvantages between different types of methods, and comparing and evaluating more than thirty open set recognition algorithms according to categories on six commonly used datasets of open set recognition task. Finally, we discuss the similarities and differences between the open set recognition problem and other similar research problems, as well as the main challenges and future research directions of the open set recognition task based on deep learning.

This work was supported by the National Natural Science Foundation of China (Grant No. 62206179, 62202280, 62106147), the Guangdong Provincial Natural Science Foundation (Grant No. 2022A1515010129), the University Stability Support Program of Shenzhen (Grant No. 20220811121315001), the Shandong Province Natural Science Foundation (Grant No. ZR2024QF034, ZR2021QF017), the Youth Expert Program of the Taishan Scholars Project in Shandong Province (Grant No. TSQN202312196), and the Shenzhen Science and Technology Program (Grant No. ZDSYS20220527171400002).