软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, 2022,33(4):1156−1169 [doi: 10.13328/j.cnki.jos.006468]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

考虑多粒度类相关性的对比式开放集识别方法*

朱鹏飞, 张琬迎, 王 煜, 胡清华

(天津大学 智能与计算学部, 天津 300350)

通信作者: 王煜, E-mail: armstrong_wangyu@tju.edu.cn

摘 要: 深度神经网络在分类任务上不断取得性能突破, 但在测试中面对未知类样本时, 会错误地给出一个已知类预测结果. 开放集识别任务旨在解决该问题, 要求模型不仅精确地分类已知类, 同时对未知类样本进行准确判别. 现有方法虽然取得了不错的效果, 但由于未对开放集识别任务的影响因素进行分析, 因而大多基于某种假设启发式地设计模型, 难以保证对于实际场景的适应性. 分析了现有方法的共性, 通过设计一个新的决策变量实验, 发现模型对已知类的表示学习能力是其中的一个关键影响因素. 基于该结论, 提出了一种基于模型表示学习能力增强的开放集识别方法. 首先, 由于对比式学习已展示出的强大表示学习能力以及开放集识别任务所包含的标签信息, 引入了监督对比式学习方法, 提高模型对已知类的建模能力; 其次, 考虑到类别间的相关性是在类别层次上的表示, 且类别之间往往呈现分层结构关系, 设计了一种多粒度类相关性的损失函数, 通过在标签语义空间构建分层结构并度量多粒度类相关性的方式, 约束模型学习不同已知类间的相关关系, 进一步提高其表示学习能力; 最后, 在多个标准数据集上进行了实验验证, 证明了所提出方法在开放集识别任务上的有效性.

关键词: 开放集识别; 表示学习; 对比式学习; 多粒度类相关性; 分类

中图法分类号: TP18

中文引用格式: 朱鹏飞, 张琬迎, 王煜, 胡清华. 考虑多粒度类相关性的对比式开放集识别方法. 软件学报, 2022, 33(4): 1156−1169. http://www.jos.org.cn/1000-9825/6468.htm

英文引用格式: Zhu PF, Zhang WY, Wang Y, Hu QH. Multi-granularity Inter-class Correlation Based Contrastive Learning for Open Set Recognition. Ruan Jian Xue Bao/Journal of Software, 2022, 33(4): 1156−1169 (in Chinese). http://www.jos.org.cn/1000-9825/6468.htm

Multi-granularity Inter-class Correlation Based Contrastive Learning for Open Set Recognition

ZHU Peng-Fei, ZHANG Wan-Ying, WANG Yu, HU Qing-Hua

(College of Intelligence and Computing, Tianjin University, Tianjin 300350, China)

Abstract: In recent years, deep neural networks have continuously achieved breakthroughs in the classification task, but they will mistakenly give a wrong known class prediction when faced with unknown samples in the testing phase. The open set recognition is a possible way to solve the problem, which requires the model not only to classify the known classes, but also to distinguish the unknown samples accurately. Most of the existing methods are designed heuristically based on certain assumptions. Despite keeping the performance increasing, they have not analyzed the key factors that affect the task. This study analyzes the commonalities of existing methods by designing a new decision variable experiment and find that the ability of model to learn representations of known classes is an important factor. Then an open set recognition method is proposed based on model representation learning ability enhancement. Firstly, due to the powerful representation learning capabilities demonstrated by the contrastive learning and the label information contained in

* 基金项目: 国家重点研发计划(2019YFB2101904); 国家自然科学基金(62106174, 61732011, 61876127); 天津市自然科学基金(17JCZDJC30800); 青海省应用基础研究项目(2019-ZJ-7017); 中国博士后科学基金资助项目(2021TQ0242, 2021M690118)
本文由“面向开放场景的鲁棒机器学习”专刊特约编辑陈恩红教授、李宇峰副教授、邹权教授推荐.
收稿时间: 2021-05-08; 修改时间: 2021-07-16; 采用时间: 2021-08-27; jos 在线出版时间: 2021-10-26

the open set recognition task, supervised contrastive learning is introduced to improve the modeling ability of the model to known classes. Secondly, considering that the correlation among the categories is the representation learning at the category level, and the hierarchical structure relationship among the categories is often presented, a multi-granularity inter-class correlation loss is designed by building the hierarchical structure in the label semantic space and measuring the multi-granularity inter-class correlation. The multi-granularity inter-class correlation loss constrains the model to learn the correlation among different known classes to further improve the representation learning ability of model. Finally, experimental results on multiple standard datasets verify the effectiveness of the proposed method on open set recognition tasks.

Key words: open set recognition; representation learning; contrastive learning; multi-granularity inter-class correlation; classification

近年来，随着机器学习领域的发展，深度学习方法的性能不断取得突破[1,2]，在计算机视觉的多个任务中得到了广泛应用[3,4]. 但传统深度神经网络(deep neural network, DNN)只能处理封闭场景任务，即训练集和测试集中包含的类别相同. 若某样本为训练集中未出现的新类样本，DNN 会将样本预测为训练集中某一已知类别，从而给出一个错误结果. 开放集识别任务(open set recognition, OSR)旨在对已知类别样本进行正确分类的同时，对出现的未知类别样本进行准确地判别. 在开放集识别任务中，未知类样本的不可预知性是开放集识别任务的主要挑战之一，因为训练过程中，模型无法获得开放集样本的相关信息，这为模型判别未知类带来了极大的挑战.

解决开放集识别任务最直观的方法是基于DNN的SoftMax输出概率结合阈值进行判别，如果最高概率低于阈值，则认为是未知类样本. 然而，由于其学习目标是区分已知类样本，因此这种方法只能学习到用于分类的判别性特征，无法对类别的所有特性进行全面描述，不利于未知类判别[5,6]. 针对该问题，学者们提出了许多方法进行解决，大致可概括为两类：第一类方法是 OpenMax 方法，在建模中考虑未知类信息，尽管基于 SoftMax 层的输出也可进行未知类判别，但由于该输出一般对应各自已知类的概率，未能体现模型对未知类的判别情况. 因此，Bendale 等学者[7]在此基础上提出了 OpenMax 函数代替 SoftMax 函数，将已知各个类别的概率进行重新调整，以显式地给出未知类别的概率. 此后，许多学者基于 OpenMax 对于未知类进行显式度量的思想，提出了基于已知类概率调整和未知类样本生成的模型[7–9]. 这一类方法相比传统 SoftMax 方法在建模中显式地考虑了未知类信息，因此对类别的描述更加丰富. 但由于对未知类知识的获取较为困难，因此性能有所局限. 针对该问题，学者们提出了基于自编码器(auto-encoder, AE)的方法，从已知类和未知类差异的角度进行开放集识别. 这类方法假设已知类样本通过训练可较好地进行重构，而未知类样本由于缺乏训练样本导致重构误差较大，从而能够被识别出来. 学者们通过分类-重构的多任务学习[10]、多阶段学习[5]以及基于变分自编码器的隐层表示建模[11]等方式，对已知类样本进行了全面地描述和表示，从而使得模型更好地对未知类样本进行判别.

尽管上述方法已经取得了不错的性能，但仍然存在一些开放性问题：首先，影响开放集识别的关键因素尚未进行深入探讨，大量工作针对开放集识别中的未知类判别提出了一些有效的算法，基于模型对未知样本概率低或者重构误差高的假设进行判别，但并未分析未知样本的判别特性和影响因素；其次，现有方法大多基于某些假设，如未知样本重构误差高、数据表示服从高斯分布等，但这些假设主要通过启发式设计，并未证明出其在真实应用的合理性.

针对上述问题，本文通过分析网络对未知样本判别的影响因素，发现模型的表示学习能力是决定开放集识别性能的一个关键问题. 具体地，通过实验分析发现，模型学习到的样本特征表示的一范数是未知类判别的一个有效决策变量，由于一范数可反映模型对样本表示的描述，因此发现模型对已知类别的表示学习能力是影响其未知类判别性能的一个关键因素.

基于以上发现，本文旨在提高模型对已知类特征的刻画能力，提出了考虑多粒度类相关性的对比式学习模型(multi-granularity inter-class correlation based contrastive learning method, MICC)，以更加有效地解决开放集识别任务. 首先，引入了对比式学习方法. 对比式学习已经被广泛证明其在特征提取方面的强大能力，考虑到开放集识别本质上属于监督式学习任务，本文利用监督对比式学习方法，在不增加模型推理复杂度的基础上，通过增强实例级别的特征表示能力，实现更加丰富的表示学习. 其次，考虑了类别间的多粒度相关性. 对

比式学习从实例级别增强了模型的表示学习能力，但未进行类级别的表示增强. 类间关系是刻画类别表示的一种重要方式，通过约束模型学习，可反映该关系的特征表示，提高模型对已知类的描述能力. 由于在实际场景中类别间往往呈现分层结构[12,13]，因此提出利用类别的语义信息构建类别层次结构，通过学习类别间的多粒度相关关系，进一步提升模型对已知类的刻画和表示学习能力.

在开放集识别任务的多个公开数据集上，本文所提出方法均获得了当前最优的性能，展示出其有效性；此外，所提出方法相比于现有基于自编码器的方法大大减少了参数量，简单易用.

本文工作的主要贡献总结如下.

(1) 探究了开放集识别任务的影响因素，发现模型的表示学习能力是关键影响因素之一；通过实验证明，样本表示向量的一范数可以作为未知类判别的一种有效决策变量；

(2) 引入了对比式学习方法，利用其强大的表示学习能力增强模型在样本级别上的特征表示，提高模型对已知类的刻画能力，提升模型的开放集识别能力；

(3) 设计了一个新颖的多粒度类相关性损失函数，通过构建并度量不同类别在分层结构上相关性，约束模型学习可反映这种关系的特征，实现类别层次上的表示学习增强，从而进一步提升模型的开放集识别能力.

本文相关工作部分总结了目前存在的大部分开放集识别方法及一些对比式学习方法. 方法部分详细介绍了本文所提出的 MICC 模型以及其中各模块的作用. 实验部分主要介绍了 MICC 在一些公开数据集上开放集识别任务的实验结果，并对模型性能进行了分析. 最后总结本文的主要内容，并对未来工作进行展望.

1 相关工作

1.1 开放集识别

开放集识别方法可大致分为两类：基于 OpenMax 的方法和基于自编码器的方法. 如前文所述，解决开放集识别问题的一种简单直接的方法是在经典监督学习分类网络中使用 SoftMax 层的输出作为判断依据，若其预测某样本的概率低于所设定的阈值，则识别该样本属于未知类. 但 SoftMax 层的输出一般为对应各自已知类的概率，未能体现未知类判别的置信度. Bendale 等学者[7]在此基础上提出了 OpenMax 函数代替 SoftMax 函数，将已知各个类别的概率进行重新调整，以显式地给出未知类别的置信度. 此后，许多学者基于 OpenMax 方法对于未知类进行显式度量的思想，提出不同的未知类判别建模方式. Shu 等学者[8]深度开放分类模型，设计了一个基于 Sigmoid 函数的一对多分类层代替 SoftMax 层建模未知类. 除了对分类器进行针对性设计外，一些学者尝试通过生成未知类别的代表性样本，将无监督的未知类判别问题转换为传统监督学习问题进行解决. Ge 等学者[9]提出了生成式 OpenMax 模型，使用生成模型，基于已知类样本合成未知类样本.

自编码器(auto-encoder, AE)的开放集识别模型展现出了优异的性能，并受到了广泛的关注. 其假设是已知类样本通过训练可被模型较好地重构，未知类由于缺乏训练样本导致重构误差较大，从而能够被判别出来. Yoshihashi 等学者[10]将卷积神经网络与自编码器重构结合起来，将每一个隐层的输出均进行重构，以获取更丰富的样本描述性表示. Oza 和 Patel[5]提出了一种两步训练的类条件引导的自编码器模型，首先进行传统的监督分类训练；第 2 步，在固定特征提取网络的基础上，通过样本采样引入相应的类别信息来约束自编码器的重构过程；最后，通过极值理论(extreme value theory, EVT)进行决策分数拟合. Sun 等人[11]引入了变分自编码器(variational auto-encoder)，将隐层表示映射为不同的高斯分布，并通过闭集分类器进行已知类样本分类和未知检测器判别新类样本.

除上述两大类方法外，还有一些传统识别方法，如: Scheirer 等人[14]提出了一种基于 SVM 的方法，该方法给传统 SVM 学习一个额外的超平面来判别未知样本; Jain 等人[15]提出了 P_I-SVM 算法, P_I-SVM 算法采用 EVT 模块建模，训练正样本作为决策边界，从而能够拒绝未知样本; Cevikalp 等人[16,17]用一组拟线性多元二次曲线函数为已知类样本定义了接受区域; Zhang 等学者[18]指出，具有鉴别性的信息大多隐藏在重构误差分布中，故而提出了基于稀疏表示的开放集识别模型(SROSR); Junior 等人[19]提出了最近邻距离率(NNDR)方法，利用两

个最近似类别之间的相似度分数实现开放集识别.

1.2 对比式学习

近年来, 自监督学习模式, 尤其是对比式学习被广泛应用在计算机视觉相关任务中. Hadsell 等学者[20]提出一种对比正例与负例进行表示学习的方法, 根据这一思路, Dosovitskiy 等学者[21]提出以参数形式将每个实例都视为由特征向量表示的类. Wu 等学者[22]提出使用记忆库来存储实例类表示向量, 这是最近几篇论文采用和扩展的一种方法(Zhuang 等人[23], 2019; Tian 等人[24], 2019; He 等人[25], 2019; Misra 和 Vander Maaten[26], 2019 年). 另外一些工作探索了使用批内样本对负例进行采样的方法, 取代了记忆库负采样方法(Doersch & Zisserman[27], 2017; Ye 等人[28], 2019; Ji 等人[29], 2019). 最近的文献将其方法的良好表现与特征表示之间的相互信息最大化联系起来(Oord 等人[30], 2018; Hénaff 等人[31], 2019; Hjelm 等人[32], 2018; Bachman 等人[33], 2019). 众多文献和实验已经证明, 对比式学习方法在特征表示学习方面具有优势. 其中, 监督对比式学习方法能够有效地利用标签信息, 同时考虑负例样本和与每个锚点相同类别的多个正例样本, 使对比式学习方法适用于监督学习任务. 在嵌入空间中, 监督对比式学习方法将属于同一类的样本在特征空间中的表示学习地更接近, 并将来自不同类的样本在特征空间内远离[34], 从而可以获得更加准确的特征表示, 提高模型对已知类的刻画能力, 以提高模型在下游任务中的表现. 受此启发, 本文模型针对开放集识别问题引入监督对比式学习方法.

2 MICC 方法

2.1 任务描述与MICC方法概述

对于开放集识别任务, 给定一组有标签的已知类样本 $X\in\mathbb{R}^{m\times n}$ 作为训练集(即闭集), 训练集样本对应标签 $y\in\{1,\ldots,k\}^n$, 其中, k 是已知类的数量, n 是训练集样本的总数, m 是每个实例的维度. 通过训练模型网络, 期望学习到模型 $f:X\rightarrow\{1,\ldots,k+1\}^n$, 以使模型将测试集中已知类实例进行准确地分类为第 c 类, 其中, $c=1,\ldots,k$, 而将未知类实例(即开放集实例)判别为第 $k+1$ 类.

2.2 决策变量

深度神经网络的 SoftMax 层输出概率和自编码器的重构误差是进行未知类判别常用的两种决策变量.

(1) SoftMax 输出. 在大多数网络中[35,36], 最后一层全连接层输出的 k 维表示向量经过归一化指数函数, 即 SoftMax 函数, 得到在 k 个已知的类标签上产生的一个概率分布. 若模型未学习过某未知类样本, 则其对该样本属于各个已知类的预测结果较为不确信, 即最高概率值较低; 若低于某一设定的阈值, 可认为该测试样本为未知类样本;

(2) 自编码器重构误差. 自编码器结构中, 编码器从输入样本中获得压缩编码, 解码器从编码中重构数据. 编码本质上是学习样本的低维特征表示[37,38]. 在训练过程中, 通过最小化重构误差对自编码器模型进行训练, 使其将已知类特征表示映射到低维流形上, 保留样本相关的重要信息; 在测试过程中, 模型根据学习到的特征表示对测试样本进行重构. 使用自编码器进行未知样本判别的假设在于: 相对于未知类测试样本, 已知类测试样本会获得更小的重构误差[39,40]. 因此, 将重构误差作为未知类判别的决策变量, 重构误差小, 测试样本为已知类样本; 反之, 测试样本为未知类样本.

上述两种方式中, 深度神经网络的 SoftMax 层输出是由其学习到的样本特征表示进行变换得到, 自编码器方法本质上学习了样本的低维特征表示. 由此可以推测, 模型对已知类别的表示学习能力对于未知类判别任务有着重要的作用. 为了分析此假设的合理性, 本文设计了一个验证实验, 通过定义与样本特征表示相关的新决策变量判别未知样本, 探究表示学习与未知类判别的关系. 具体地, 定义了一个决策变量 r, 形式化表示为

$$r=\frac{1}{\|\boldsymbol{z}\|_1} \tag{1}$$

其中, $\boldsymbol{z}=Enc(x)$为编码器网络学习到的表示向量, x 为训练集实例, $\|\boldsymbol{z}\|_1$ 为表示向量 $\boldsymbol{z}$ 的范数. 本文假设未知类的

判别和表示向量 $\boldsymbol{z}$ 相关，利用 r 作为决策变量设计实验．表 1 展示了深度神经网络方法在 CIFAR10 数据集上未知类判别任务的实验结果，以接收机工作特性曲线(AUROC)[41]下面积作为评估度量，AUROC 值越大，代表模型判别未知样本的能力越强．其中，Plain 为根据本文设置实现的传统深度网络基准模型，以 SoftMax 分数作为决策变量判别未知样本．可以看出，决策变量 r 在未知类判别任务上的实验结果优于或持平于 SoftMax 分数作为决策变量的实验结果．

表 1　CIFAR10 数据集上不同决策变量在未知类判别任务中的 AUROC 值

方法	SoftMax score	r
Plain	0.850	0.860
MICC	0.936	0.935

图 1 展示了利用 r 作为决策变量时，MICC 的训练集样本、已知类测试集样本和未知类测试集样本的分布情况．其中，绿色直方图表示已知类训练集样本，紫色直方图表示已知类测试样本，粉色直方图表示未知类测试样本．可以发现：r 作为决策变量时，已知类样本和未知类样本得到了很好的分离．实验结果表明：模型学习到的特征表示可作为未知类判别任务的决策变量，很好地区分未知类样本和已知类样本，从而证明了未知类的判别和模型对已知类的表示学习能力相关．因此，在开放集识别任务的建模中，应增强已知类别的表示学习能力，未知类判别任务实际需要模型对已知类进行更加全面和准确的刻画．因此，只需设计模型学习到更加丰富、准确、具有判别性的特征表示即可提升模型的任务表现(如图 2 所示)．

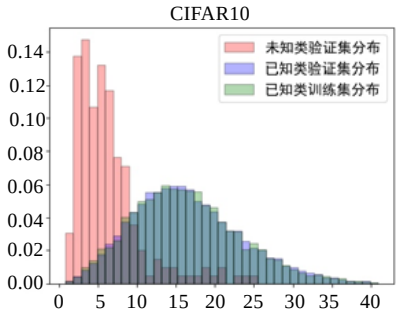


图 1　r 作为决策变量的样本分布直方图

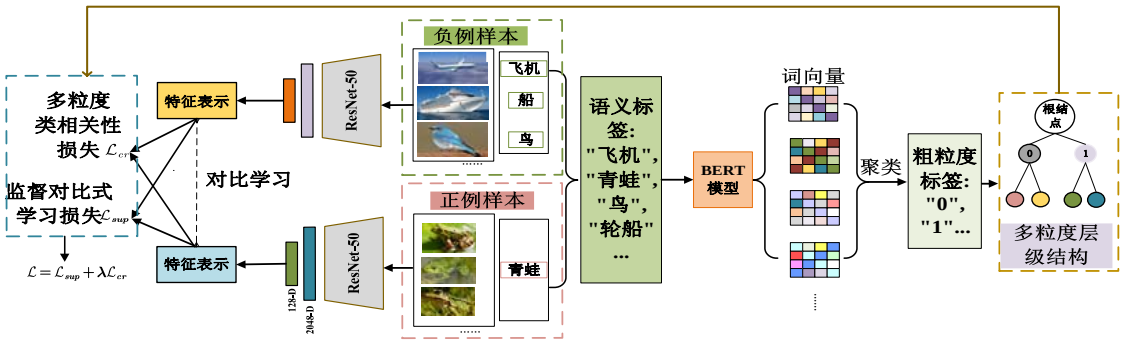


图 2　本文模型方法框架图

2.3　对比式学习

对比式学习方法是近年来的研究热点之一，通过对比式学习在特征空间中将相似样本拉近，将不相似样本拉远，相比于传统监督式训练，实现了更加丰富的语义特征表示学习[34,42]．由第 2.2 节的分析可知，模型对已知类的表示学习能力是影响未知类判别的一个关键因素．因此，在开放集识别任务中引入对比式学习方法有利于未知样本的判别．对比式学习主要面向无监督表示学习提出[42]，在无需标签信息的情况下学习通用的

语义特征表示. 在监督学习任务中, 标签信息需进行充分利用, 进一步增强网络的表示学习能力. 下面分别对自监督对比式学习和监督对比式学习进行介绍.

- 自监督对比式学习

Chen 等人[42]在 2020 年提出了一个简单有效的对比式学习方法, 该方法对于一组由 N 个实例-标签对随机组成的样本$\{x_k,y_k\}_{k=1\ldots N}$ 进行两种不同形式的数据增广, 获得包含 $2N$ 对实例-标签数据的训练集 $\{\tilde{x}_\ell,\tilde{y}_\ell\}_{\ell=1\ldots 2N}$, $\tilde{x}_{2k}$ 与 $\tilde{x}_{2k-1}$ 是由实例 x_k 进行两种不同的数据增广获得, 并且 $\tilde{y}_{2k-1}=\tilde{y}_{2k}=y_k$, 即 $\tilde{x}_{2k}$, $\tilde{x}_{2k-1}$ 与 x_k 拥有同一标签. 模型网络主要包含 3 个部分.

(1) 编码器网络 $Enc(\cdot)$: 编码器网络将实例 x 映射到表示向量: $\boldsymbol{z}=Enc(x)\in\mathcal{R}^{D_E}$, 模型对于实例 x 进行两次不同的数据增广, 将两个增广后的样本分别输入到同一编码器网络中, 针对实例 x 得到一对表示向量, 并将表示向量 $\boldsymbol{z}$ 归一化至单位超球面 $\mathcal{R}^{D_E}$ (D_E=2048, 表示向量的维度)上, 从而得到有效的特征表示;

(2) 映射网络 $proj(\cdot)$: 将编码器网络学习到的表示向量 $\boldsymbol{z}$ 映射到向量 $\hat{\boldsymbol{z}}=Proj(\boldsymbol{z})\in\mathcal{R}^{D_P}$, $\mathcal{R}^{D_P}$ 为向量的维度, 本文将 $proj(\cdot)$设置为含有一个隐藏层、输入向量大小为 2 048 维、输出向量大小为 D_P=128 的多层感知机或输出向量大小为 D_P=128 的线性层. 对于映射网络的输出, 模型再次进行归一化操作, 使其位于单位超球面上, 从而可以使用内积来测量映射空间中的距离;

(3) 映射网络 $Classifier(\cdot)$: 本文将分类器设置为一层线性网络的线性分类器, 进行后续闭集分类任务和开放集识别任务.

在训练过程中, 假设 $i\in I\equiv\{1\ldots 2N\}$为训练集中实例的索引, 第 i 个增广样本作为锚点, 定义 $j(i)$为起源于同一实例除锚点外的增广实例的索引, $j(i)$即为正例, 其他 $2(N-1)$个实例($\{k\in A(i)\backslash\{j(i)\}\}$)则为对比式学习的负例, 其损失函数定义为

$$\mathcal{L}_{self}=\sum_{i\in I}\mathcal{L}_i^{self}=-\sum_{i\in I}\log\frac{\exp(\hat{\boldsymbol{z}}_i\cdot\hat{\boldsymbol{z}}_{j(i)}/\mathcal{T})}{\sum_{a\in A(i)}\exp(\hat{\boldsymbol{z}}_i\cdot\hat{\boldsymbol{z}}_a/\mathcal{T})}\tag{2}$$

其中, $\hat{\boldsymbol{z}}_\ell=Proj(Enc(x_\ell))\in\mathcal{R}^{D_P}$, $\cdot$代表内积, $\mathcal{T}\in\mathcal{R}^+$为温标参数, 设置为$\mathcal{T}$=0.1, $A(i)\equiv I\backslash\{i\}$.

由于开放集识别任务需要同时准确地分类已知类和判别未知类, 因此是一种监督学习任务. 自监督对比式学习方法由于将同一类别的不同样本看作负例, 导致类内差异大, 故而在监督学习任务中其性能难以达到监督训练方法.

- 监督对比式学习

Khosla 等人[34]面向监督学习任务提出了监督对比式学习, 即利用训练样本的标签信息, 针对每一个锚点 i, 不仅拉近其自身的增广样本, 同时拉近其同类的正例及正例的增广样本, 通过利用标签信息进一步提高模型在监督学习任务中的表示学习能力, 其损失函数定义为

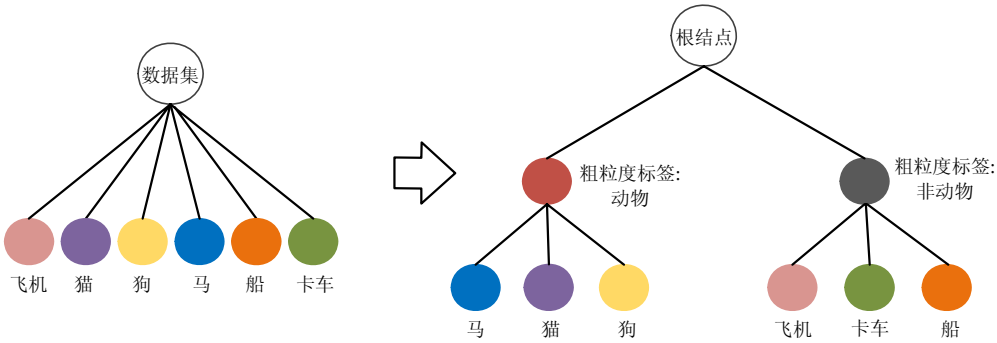
$$\mathcal{L}_{\sup}=\sum_{i\in I}\mathcal{L}_i^{\sup}=\sum_{i\in I}\frac{-1}{|P(i)|}\sum_{p\in P(i)}\log\frac{\exp(\hat{\boldsymbol{z}}_i\cdot\hat{\boldsymbol{z}}_p/\mathcal{T})}{\sum_{a\in A(i)}\exp(\hat{\boldsymbol{z}}_i\cdot\hat{\boldsymbol{z}}_a/\mathcal{T})}\tag{3}$$

其中, $P(i)\equiv\{p\in A(i):\tilde{y}_p=\tilde{y}_i\}$, 即训练集样本中除锚点 i 外, 与锚点 i 拥有相同标签的全部实例均作为对比式学习的正例. 温标参数设置为$\mathcal{T}$=0.1.

2.4 多粒度类相关性损失

开放集识别模型检测未知类样本的过程实际上是比较该样本类别与各个已知类别相似或相关程度, 因此, 学习到已知类之间的相关关系可以有利于开放集识别任务. 对比式学习模块主要在样本级别上进行表示学习, 但并未利用类别间的相关关系. 因此, 在利用监督对比式学习思想的基础上, 考虑在训练过程中评估实例的监督信息, 对实例类别间的相关性进行建模. 受人脑认知机制的启发, 考虑通过分层抽象的方式, 将不同类别的概念监督信息在不同粒度上依据类别间相关性映射到新的粗粒度标签, 即同一类别在不同的粒度层次

上对应不同的抽象概念标签：相似类别在同一粒度层次上对应相同的抽象概念标签，不相似类别在同一粒度层级上对应不同的抽象概念标签．这样，分层结构中，每一层结点中存储相较于子结点更为抽象的粗粒度标签信息，利用每一层结点信息间的相关性，可以实现对不同类别间关系的判断能力．以 CIFAR10 数据集中部分类别数据为例，如示意图 3 为两层粒度层级结构：图 3(a)为现有方法中的原始数据集结构中的类别关系，仅将不同类别区分开，未考虑类别间的相关关系；图 3(b)为本文所提出的多粒度类相关性层级结构，在区分不同类别的同时，依据类别标签的语义相关性，通过构建多粒度分层结构度量类别间的相关关系，从而进一步提升模型对已知类的表示学习能力．由此，本文定义一个新的多粒度类相关性损失，以达到更好的判别开放集实例的目的．



(a) 现有方法中的类别关系　　(b) 多粒度类相关性

图 3　多粒度结构示意图

对于训练集样本$\{x_k,y_k\}_{k=1\ldots N}$，利用标签的语义信息构建 $X_{\alpha,\alpha=0,\ldots,\alpha}$层的层级结构，其中，$\alpha$为层级数，$X_{\alpha=0}$ 为数据原始标签 y_k 作为叶子结点所在的层级．利用 $Word2Vec(\cdot)$模块对原始标签进行语义特征的提取，获得词向量 $\boldsymbol{vec}$．利用 K-means 方法对 $\boldsymbol{vec}$ 进行粗粒度层级(即第 $X_{\alpha=1}$ 层级)上的聚类，得到粗粒度标签 $\hat{y}_{\alpha-s}\in\{1,\ldots,t\}^N$，$t<k$，进而得到实例-粗粒度标签对样本 $\{\hat{x}_s,\hat{y}_{\alpha-s}\}_{\alpha=0,\ldots,\alpha;s=1,\ldots,k}$．以此类推，通过提取 $X_{\alpha-1}$ 层的标签信息,可获得 X_α 层的粗粒度标签信息．其中，t 为粗粒度类别数，$t<k$．本文设置$\alpha=1$，即仅讨论两层分层结构的情况,将$X_{\alpha=1}$层级粗粒度标签简化为 $\hat{y}_s\in\{1,\ldots,t\}^N$，$t<k$．对于一批训练样本，在模型计算监督对比损失函数的同时,对实例的多粒度类相关性损失函数进行计算，本文定义多粒度类相关性损失为

$$\mathcal{L}_{cr}=-\sum_{k}^{N}\hat{y}_s\log p_s(\hat{y}'_s\mid x_k) \tag{4}$$

其中，$p_s(\hat{y}'_s\mid x_k)$ 为实例 x_k 在粗粒度标签 $\hat{y}_s$ 上的期望分布．

MICC 的最终损失函数定义为

$$\mathcal{L}=\mathcal{L}_{sup}+\lambda\mathcal{L}_{cr} \tag{5}$$

其中，$\lambda>0$ 为平衡权重，本文在所有实验中设置$\lambda=1$．

算法 1 是 MICC 训练过程和测试过程的详细步骤．

算法 1．

1. Train process:
2. Input: 训练集样本$\{x_k,y_k\}_{k=1\ldots N}$;
3. Function
4. $\boldsymbol{vec}\leftarrow Word2Vec(y_k)$
5. $\hat{y}_s\leftarrow Cluster(\boldsymbol{vec})$
6. **For** $\{x_k,y_k\}_{k=1\ldots N}$ in batch:
7. $\hat{\boldsymbol{z}}\leftarrow Proj(Enc(x_k))$

Calculation: $\mathcal{L}_{\text{sup}} = \sum_{i\in I}\mathcal{L}_i^{\text{sup}} = \sum_{i\in I}\frac{-1}{|P(i)|}\sum_{p\in P(i)}\log\frac{\exp(\hat{\boldsymbol{z}}_i\cdot\hat{\boldsymbol{z}}_p/\mathcal{T})}{\sum_{a\in A(i)}\exp(\hat{\boldsymbol{z}}_i\cdot\hat{\boldsymbol{z}}_a/\mathcal{T})}$

$\mathcal{L}_{cr} = -\sum_{k}^{N}\hat{y}_s\log p_s(\hat{y}'_s\mid x_k)$

$\mathcal{L}=\mathcal{L}_{sup}+\lambda\mathcal{L}_{cr}$

8. **End** train
9. Test process:
10. Input: 测试样本 x.
11. $\hat{\boldsymbol{z}}\leftarrow Proj(Enc(x))$
12. $score\leftarrow Classifier(\boldsymbol{z})$
13. 根据输出分数 $score$ 预测样本 x 为某一已知类别或为未知类.
14. **End** test

MICC 模型在训练阶段, 首先将类别的语义标签利用 BERT 转换为词向量, 随后利用 K-means 方法聚类构建多粒度层次结构; 随后, 随机初始化网络, 同时更新监督对比式学习损失(见公式(3))、多粒度类相关性损失(见公式(4))和总损失函数(见公式(5)). 测试阶段, 先对测试样本进行训练, 生成表示向量, 再输入进分类器进行分类.

3 实验结果与分析

本文在多个标准数据集上进行了实验, 并评估了模型在开放集识别任务中的性能表现. 为证明 MICC 的有效性, 将 MICC 在两种主要评价指标下的实验结果与一些现有方法进行了比较. 另外, 本文设计消融实验对模型中的部分对性能的影响进行了探讨, 并通过实验分析了超参数对于模型性能的影响.

3.1 实验细节

- 评价指标

本文使用宏平均 $F1$ 分数[43]和接收机工作特性曲线(AUROC)[41]下面积作为评估度量. 宏平均 $F1$ 分数是分类任务中综合考虑查准率和查全率的性能度量, 表 2 中的实验结果综合评价了模型在开放集识别任务中的表现. AUROC 值可以解释为正例被赋予比负例更高检测分数的概率, 是一个与阈值无关的度量. AUROC 值越大,代表模型判别未知样本的能力越强. 后文表 4 中的实验结果展示了模型判别未知样本的性能.

表 2 不同方法在开放集识别任务中的宏平均 $F1$ 分数对比

方法	ImageNet-crop	ImageNet-resize	LSUN-crop	LSUN-resize
SoftMax	0.639	0.653	0.642	0.647
OpenMax[6]	0.660	0.684	0.657	0.668
CROSR[12]	0.721	0.735	0.720	0.749
C2AE[7]	0.837	0.826	0.783	0.801
CGDL[41]	0.840	0.832	0.806	0.812
MICC	**0.917**	**0.916**	**0.924**	**0.892**

- 实验设置

在实验过程中, MICC 模型分为训练阶段和测试阶段两部分. 在实验训练阶段, 使用 ResNet-50 作为编码器网络. 一层线性网络作为映射网络, 一层全连接线性网络作为分类器. 实验过程中, 数据批大小始终设置为 512. 训练过程中, 学习率设置为 0.5, 训练轮数 1 000 次; 测试过程中, 学习率设置为 5. 模型池化层(D_E=2048)的归一化激活用作表示向量. 在监督对比式学习的数据增广模块 $Aug(\cdot)$, 实验中使用了自动增广方法(AutoAugment)[34]和堆叠式随机增广方法(stacked RandAugment)[34]. 对于 $Word2Vec(\cdot)$模块, 实验中利用 BERT[44]模型生成训练集样本原始标签的词向量. 公式(2)、公式(3)中, $\mathcal{T}\in\mathcal{R}^+$为温标参数, 是可调的超参数, 实验中设置为$\mathcal{T}$=0.1. 对于开放集识别任务, 本文利用宏平均 $F1$ 分数综合体现模型在已知类分类和未知类判别

任务中的表现, 利用接收机工作特性曲线(AUROC)下面积体现模型在判别未知类任务中表现的优劣.

- 数据集划分

本文使用的数据集包括 CIFAR10, CIFAR+10, CIFAR+50, ImageNet-crop, ImageNet-resize, LSUN-crop, LSUN-resize, TinyImageNet.

对于开放集识别任务, 需要在实验数据集上将已知类划分为训练集和测试集, 其余类别作为未知类样本. 本文实验采取随机方式对标准数据集进行 5 次划分, 获得多组已知类样本和未知类样本. CIFAR10 数据集、CIFAR100 数据集和 TinyImageNet 数据集均有固定的训练集和测试集划分, 因此使用原有的设置进行实验. 对于 CIFAR10 数据集, 随机选取其训练集中 6 类作为已知类, 其余 4 类作为未知类. 对于 CIFAR+10[20]和 CIFAR+50[20]数据集, 实验中随机选取 CIFAR10 数据集的训练集中的 4 类作为已知类; 在 CIFAR100 数据集的训练集中, 分别随机选取 10 类和 50 类作为未知类. TinyImageNet 数据集共包含了 200 个不同的对象类, 具有相对复杂的决策空间. 在实验中, 本文随机选取其训练集中 20 类作为已知类, 其余 180 类作为未知类. 实验中, 已知类训练集和已知类测试集的样本数量按照已知类样本的 4:1 比例随机划分, 所有未知类样本作为未知类测试集.

ImageNet-crop, ImageNet-resize, LSUN-crop, LSUN-resize 数据集分别是通过对 ImageNet[45]和 LSUN[46]进行裁剪或调整原始图像的大小(以适应 CIFAR10 数据集中图像大小)获得的, 每个数据集包含 10 000 个样本. 实验中, 以 CIFAR10 数据集的训练集作为已知类训练集, 测试集作为已知类测试集, 上述 4 种数据集分别作为未知类测试集. 本文实验结果均为 5 次随机划分数据实验结果的平均值.

3.2 开放集识别

为了评估 MICC 的有效性, 本文将实验结果与 5 种最新的方法进行了比较, 包括 SoftMax, OpenMax[7], CROSR[10], C2AE[5]和 CGDL[41]. 表 2 中展示了在 ImageNet-crop, ImageNet-resize, LSUN-crop, LSUN-resize 数据集上, MICC 与上述 5 种对比方法的宏平均 $F1$ 分数, 其中, 对比方法的实验结果引用自 C2AE[5]方法和 CGDL[41]方法的文献. 实验结果综合体现了本文模型在已知类分类和未知类判别任务上的性能, 可以发现: SoftMax 方法聚焦学习相对简单的已知类判别性特征, 难以全面刻画已知类别的特点, 导致在未知类判别任务中表现不佳. 基于 AE 的方法(例如 CROSR[10], C2AE[5]和 CGDL[41])通过低维特征重构学习到了类别的多样化表示, 但存在两方面问题: (1) 像素点对像素点的重构会保留与当前类别无关的背景信息, 影响模型对类别关键信息刻画的准确性; (2) 无关信息不利于判别已知类, 影响已知类分类性能. MICC 利用对比式学习方法和多粒度类间相似性信息, 提高了模型对已知类的刻画和表示能力, 因此可提高开放集识别性能. 表 3 将本文模型的参数量与 CGDL[41], CROSR[10]和 MLOSR[47]方法的参数量进行了对比, 可以看出: MICC 模型所需参数量更少, 简单易用, 且可扩展性更强.

表 3 不同方法模型参数量对比

方法	MICC	CGDL	CROSR	MLOSR
参数量 (k)	**174**	170 367	423	7 165

3.3 未知类判别

本节中的实验遵循文献[41]中的评估体系, 通过 AUROC 评价模型的未知类判别效果, 同时比较模型的闭集精度以评价已知类分类性能. 本节实验补充了另外 3 种方法进行比较, 包括 G-OpenMax[9], OSRCI[48]和 MLOSR[47]. 表 4 展示了不同方法在 TinyImageNet, CIFAR10, CIFAR+10 和 CIFAR+50 数据集上的未知类判别实验结果, 其中, 对比方法的实验结果引用自 CGDL[41]方法和 MLOSR[47]方法的文献. 可以看出: 相对于现有方法, MICC 实验结果存在明显优势, 证实了其有效性. 图 4 展示了 MICC 在 4 种数据集下样本分布的直方图, 横轴为样本的输出分数, 纵轴为可视化直方图. 其中, 粉色区域代表测试中未知类测试样本的分布, 紫色区域代表测试中已知类测试样本的分布, 绿色区域代表已知类训练样本的分布. 可以看出: 已知类测试样本与训练集样本分布一致, 未知类测试样本则被显式分离. 此外, 本文比较了不同模型的已知类分类精度(CIFAR10

数据集). 由于其他方法未公布闭集性能或实验设置不同, 本实验只比较了 SoftMax 方法和 MICC 方法. 其中, SoftMax 方法的闭集精度为 96.43%, MICC 为 98.74%, 表明所提出方法能够显著提高已知类分类精度. 相比之下, 现有方法除了 SoftMax 方法和 CGDL[41]方法可实现闭集精度不下降之外, 其他方法难以保持已知类分类性能. 综上所述, MICC 可同时提高已知类分类和未知类判别的性能, 从而能更好地解决开放集识别问题.

表 4 不同方法在未知检测任务中的 AUROC 值

方法	CIFAR10	CIFAR+10	CIFAR+50	TinyImageNet
SoftMax	0.677	0.816	0.805	0.577
OpenMax[6]	0.695	0.817	0.796	0.576
G-OpenMax[17]	0.675	0.827	0.819	0.580
OSRCI[45]	0.699	0.838	0.827	0.586
MLOSR[46]	0.845	0.895	0.877	0.718
CGDL[41]	0.890	0.928	0.915	0.762
MICC	**0.936**	**0.963**	**0.964**	**0.855**

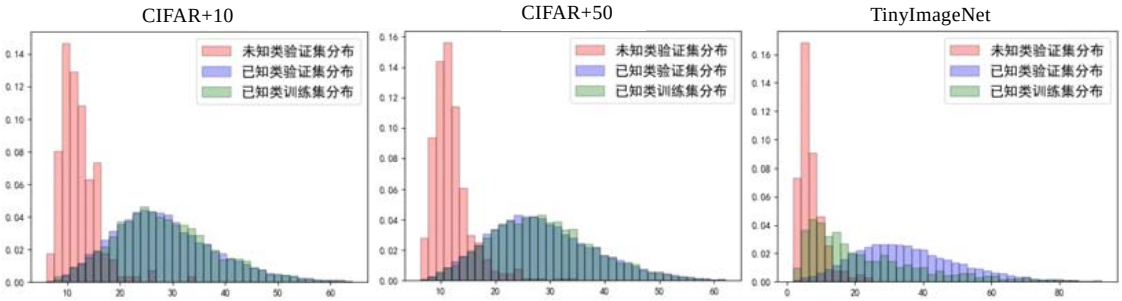


图 4 样本分布直方图

3.4 消融实验

本节实验主要讨论和分析 MICC 模型各部分对于模型性能的影响.

- Plain: 根据本文设置实现的传统深度网络基准模型, 以 ResNet-50 作为主体网络架构, 一层全连接层作为分类器 ,交叉熵函数作为作为损失函数, SoftMax 分数作为未知类判别的决策变量. 实验中, 遍历所有数据 500 次, 数据批大小设置为 512, 学习率设置为 0.8;
- Supcon: 监督对比式学习模型, 其实现参考文献[34]. 训练过程和测试过程分两步进行, 同样使用 ResNet-50 作为编码器网络, 一层线性网络作为映射网络, 一层全连接线性网络作为分类器, 以监督对比式学习损失作为损失函数, 训练轮数 1 000 次, 数据批次大小设置为 512, 学习率设置为 0.5. 测试时先对分类器进行训练, 训练轮数 100 次, 数据批次(batch size)大小设置为 512, 学习率设置为 5.

表 5 中展示了标准深度网络模型、监督对比式学习模型和本文模型在 CIFAR10, CIFAR+10, CIFAR+50, TinyImageNet 数据集上的实验结果.

表 5 消融模块在不同数据集上的 AUROC 值

方法	CIFAR10	CIFAR+10	CIFAR+50	TinyImageNet
Plain	0.850	0.904	0.901	0.838
Supcon	0.907	0.963	0.961	0.841
MICC	**0.936**	**0.964**	**0.964**	**0.855**

根据实验结果, 结论如下.

(1) 监督对比式学习方法可以学习到丰富的已知类语义表示, 通过全面刻画已知类特点提高模型的未知类判别能力;

(2) 多粒度类相关性损失函数利用多粒度层级上实例的相关性信息训练网络, 提升了模型的任务表现. 这表明: 在开放集识别任务中, 有效建模类间相关关系能提升模型对已知类的表示能力, 有利于未知类判别;

(3) 多粒度类相关性损失对 CIFAR+10 和 CIFAR+50 提升有限，对 CIFAR10 和 TinyImageNet 提升较为明显. 其原因在于: CIFAR+M 实验已知类仅有 4 类，类别数过少，难以形成明显的多粒度关系；随着已知类数目的增多，其效果愈发明显.

对于第 2.2 节本文针对决策变量的计算方式设计消融实验进行了探究:

$$r_{2\text{-}norm}=\frac{1}{\|\boldsymbol{z}\|_2} \tag{6}$$

其中，$\|\boldsymbol{z}\|_2$ 为表示向量 $\boldsymbol{z}$ 的二范数. 表 6 中展示了本文所提出的决策变量 r 和 $r_{2\text{-}norm}$ 的实验结果对比(对于不同的模型，决策变量的计算方式可能会影响模型的任务表现). 可以发现，表示向量的不同范数影响模型的任务表现. 两种范数都可以描述样本的特征表示，但由于整体上一范数更为稳定，因此本文第 2.2 节实验选用表示向量的一范数作为决策变量.

表 6 不同计算方式的决策变量下的实验结果对比

方法	r	$r_{2\text{-}norm}$
Plain	0.860	0.851
MICC	0.935	0.938

3.5 参数分析

本节对实验中出现的参数进行分析探讨. 图 5 展示了不同粗粒度类别数下模型的 AUROC 值的变化折线图，图中横轴为粗粒度类别数，纵轴为 AUROC 值. 红色直线为 CIFAR10 数据集上实验结果，蓝色点线为 CIFAR+10 数据集上实验结果，黄色短线、点相间虚线为 CIFAR+50 数据集上实验结果，绿色虚线为 TinyImageNet 数据集上实验结果. 可以发现: 对于 CIFAR+10 和 CIFAR+50 两组实验，使用相同的训练集，改变粗粒度类别数对模型性能影响不明显，模型对参数变化较为鲁棒. 其原因可能与已知类别数量较少有关，对于 CIFAR10 和 TinyImageNet，改变粗粒度类别数，AUROC 值存在波动，粗粒度类别数为 2 时(t=2)模型效果最好. 因此，本文实验中均使用该设置进行性能比较.

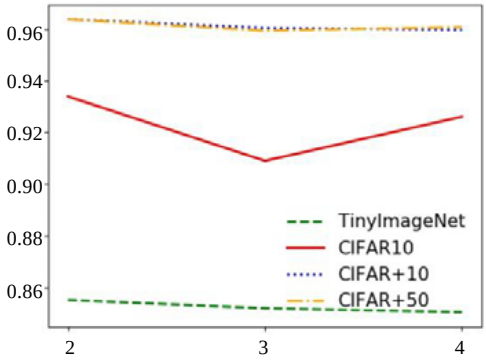
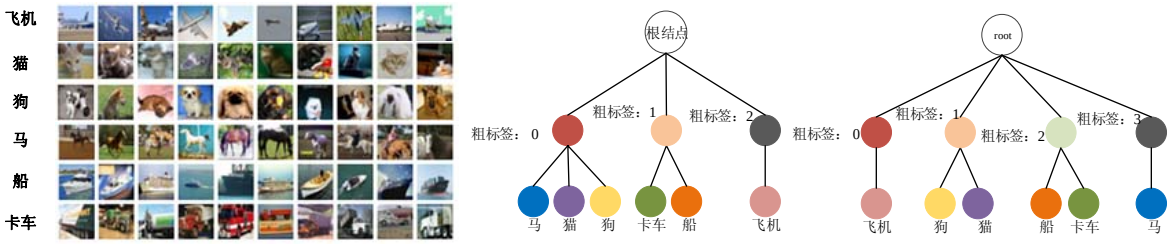


图 5 不同粗粒度类别数下 MICC 的 AUROC 值折线图

对构建的层级结构，通过可视化对其进行了进一步分析. 图 6 所示是以 CIFAR10 数据集的一组实验为例，不同粗粒度类别数的层级结构示意图(随机划分已知类样本 6 类，其余 4 类为未知类样本)，其中，6 类已知类分别为飞机、猫、狗、马、船、卡车. 可以发现: 将 6 类已知类样本划分粗粒度类别为动物(粗粒度标签为 0)、非动物(粗粒度标签为 1)，即两层 2 类粗粒度类别的层级结构时(如图 3(b)所示)，已知类训练集样本间差异最大，能够为模型引入合理的粗粒度类别下细粒度类别间的相关性信息和粗粒度类别间的差异性信息. 图 5 中的折线整体呈下降趋势，其原因可能在于较多的粗粒度类别数为模型引入了过多细粒度层次信息，无法准确度量其粗粒度层次信息. 另外，实验中数据集均按照随机方式划分，样本原始标签的语义信息可能会影响粗粒度聚类结果，从而造成粗粒度类别中样本不平衡等情况，影响模型效果.



(a) 实验中的训练集数据样例 (b) 构建两层 3 类粗粒度类别的多粒度层级结构(α=1, t=3) (c) 构建两层 4 类粗粒度类别的多粒度层级结构(α=1, t=4)

图 6 不同粗粒度类别数的层级结构示意图

4 总结与展望

本文面向开放集识别任务，提出了一种考虑多粒度类相关性的对比式学习方法. 首先，探索并分析了影响任务性能的关键影响因素，发现模型学习的样本特征表示可作为决策变量进行未知类判别. 基于该结论，引入了对比式学习模型，借助其强大的表示学习能力，通过在样本级别上的语义特征学习，增强模型的开放集识别性能. 此外，进一步考虑增强模型在类别层次上的表示学习，提出了一个新的多粒度类相关性损失函数，利用训练样本标签语义信息的相关性进行建模，通过度量不同类别间的多粒度相关性，使得模型学习到可反映类间关系的特征表示. 本文进行了大量实验，证明所提出方法达到了当前最优的开放集识别性能；同时，相比于流行的自编码器类方法减少了参数量，模型简单易用. 由于实际场景中模型需面对的类别数量更多，任务更加复杂，未来的工作中可深入探索更加复杂的多粒度层级结构对于开放集识别任务的影响.

References:

[1] Li ZX, Wei HY, Huang FC, *et al*. Combine visual features and scene semantics for image captioning. Chinese Journal of Computers, 2020, 9(3): 1624−1640 (in Chinese with English abstract).

[2] Tian X, Wang L, Ding Q. Review of image semantic segmentation based on deep learning. Ruan Jian Xue Bao/Journal of Software, 2019, 30(2): 440−468 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5659.htm [doi: 10.13328/j.cnki.jos.005659]

[3] Du PF, Li XY, Gao YL. Survey on multimodal visual language representation learning. Ruan Jian Xue Bao/Journal of Software, 2021, 32(2): 327−348 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6125.htm [doi: 10.13328/j.cnki.jos.006125]

[4] Feng YG, Yu J, Sang JT, *et al*. Survey on knowledge-based zero-shot visual recognition. Ruan Jian Xue Bao/Journal of Software, 2021, 32(2): 370−405 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6146.htm [doi: 10.13328/j.cnki.jos.006146]

[5] Oza P, Patel VM. C2AE: Class conditioned auto-encoder for open-set recognition. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2019. 2307−2316.

[6] Chen GY, Qiao LM, Shi YM, *et al*. Learning open set network with discriminative reciprocal points. In: Proc. of the European Conf. on Computer Vision (ECCV). 2020. 507−22. [doi: 10.1007/978-3-030-58580-8_30]

[7] Bendale A, Boult TE. Towards open set deep networks. In: Proc. of the 2016 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2016. 1563−1572.

[8] Shu L, Xu H, Liu B. Doc: Deep open classification of text documents. In: Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing. 2017.

[9] Ge ZY, Demyanov S, Chen ZT, *et al*. Generative openmax for multi-class open set classification. In: Proc. of the British Machine Vision Conf. 2017. 2017. [doi: 10.5244/C.31.42]

[10] Yoshihashi R, Shao W, Kawakami R, *et al*. Classification reconstruction learning for open-set recognition. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2019. 4016−4025.

[11] Sun X, Yang ZN, Zhang C, *et al*. Conditional gaussian distribution learning for open set recognition. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2020. 13480−13489.

[12] Wang Y, Hu QH, Zhu PF, *et al*. Deep fuzzy tree for large-scale hierarchical visual classification. IEEE Trans. on Fuzzy Systems, 2020, 28(7): 1395−1406. [doi: 10.1109/TFUZZ.2019.2936801]

[13] Wang Y, Wang Z, Hu QH, *et al*. Hierarchical semantic risk minimization for large-scale image classification. IEEE Trans. on Cybernetics, 2021. [doi: 10.1109/TCYB.2021.3059631]

[14] Scheirer WJ, de Rezende Rocha A, Sapkota A, *et al*. Toward open set recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012, 35(7): 1757−1772. [doi: 10.1109/TPAMI.2012.256]

[15] Jain LP, Scheirer WJ, Boult TE. Multi-class open set recognition using probability of inclusion. In: Proc. of the European Conf. on Computer Vision (ECCV). Springer, 2014. 393−409.

[16] Hakan C, Yavuz HS. Fast and accurate face recognition with image sets. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision Workshops (ICCVW). 2017. 1564−1572. [doi: 10.1109/ICCVW.2017.184]

[17] Cevikalp H, Triggs B. Polyhedral conic classifiers for visual object detection and classification. In: Proc. of the 2017 IEEE/ CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2017. 4114−4122.

[18] Zhang H, Patel VM. Sparse representation-based open set recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2016, 39(8): 1690−1696. [doi: 10.1109/TPAMI.2016.2613924]

[19] Junior PRM, de Souza RM, de O Werneck R, *et al*. Nearest neighbors distance ratio open-set classifier. Machine Learning, 2017, 106(3): 359−386. [doi: 10.1007/s10994-016-5610-8]

[20] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. In: Proc. of the 2006 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Vol.2. 2006. 1735−1742.

[21] Alexey D, Springenberg JT, Riedmiller M, *et al*. Discriminative unsupervised feature learning with convolutional neural networks. Advances in Neural Information Processing Systems, 2014, 27: 766−774.

[22] Wu ZR, Xiong YJ, Yu SX, *et al*. Unsupervised feature learning via non-parametric instance discrimination. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. 3733−3742.

[23] Zhuang CX, Zhai A, Yamins D. Local aggregation for unsupervised learning of visual embeddings. In: Proc. of the 2019 IEEE/CVF Int'l Conf. on Computer Vision (ICCV). 2019. 6002−6012.

[24] Tian YL, Krishnan D, Isola P. Contrastive multiview coding. In: Proc. of the European Conf. on Computer Vision (ECCV). 2019. 776−794.

[25] He KM, Fan HQ, Wu YX, *et al*. Momentum contrast for unsupervised visual representation learning. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2020. 9729−9738.

[26] Ishan M, van der Maaten L. Self-supervised learning of pretext-invariant representations. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2020. 6707−6717.

[27] Carl D, Zisserman A. Multi-task self-supervised visual learning. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision (ICCV). 2017. 2070−2079.

[28] Mang Y, Zhang X, Yuen PC, *et al*. Unsupervised embedding learning via invariant and spreading instance feature. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2019. 6210−6219.

[29] Xu J, Vedaldi A, Henriques J. Invariant information clustering for unsupervised image classification and segmentation. In: Proc. of the 2019 IEEE/CVF Int'l Conf. on Computer Vision (ICCV). 2019. 9865−9874.

[30] van den Oord A, Li YZ, Vinyals O. Representation learning with contrastive predictive coding. arXiv: 1807.03748, 2018.

[31] Olivier JH, Srinivas A, De Fauw J, *et al*. Data-efficient image recognition with contrastive predictive coding. arXiv: 1905.09272, 2019.

[32] Devon HR, Fedorov A, Lavoie-marchildon S, *et al*. Learning deep representations by mutual information estimation and maximization. In: Proc. of the Int'l Conf. on Learning Representations. 2018.

[33] Philip B, Devon HR, Buchwalter W. Learning representations by maximizing mutual information across views. Advances in Neural Information Processing Systems, 2019, 32: 15535−15545.

[34] Prannay K, Teterwak P, Wang C, *et al*. Supervised contrastive learning. Advances in Neural Information Processing Systems, 2020, 33.

[35] Ken C, Simonyan K, Vedaldi A, *et al*. Return of the devil in the details: Delving deep into convolutional nets. In: Proc. of the British Machine Vision Conf. 2014.

[36] Christian S, Liu W, Jia YQ, *et al*. Going deeper with convolutions. In: Proc. of the 2015 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2015. 1−9.

[37] Yoshua B, Lamblin P, Popovici D, *et al*. Greedy layer-wise training of deep networks. Advances in Neural Information Processing Systems, 2006, 19:153−160.

[38] Kingma DP, Welling M. Auto-encoding variational bayes. In: Proc. of the Int'l Conf. on Learning Representations (ICLR 2014). 2014.

[39] Mahmudul H, Choi JH, Neumann J, *et al*. Learning temporal regularity in video sequences. In: Proc. of the 2016 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2016. 733−742.

[40] Zhao YR, Deng B, Shen C, *et al*. Spatio-temporal autoencoder for video anomaly detection. In: Proc. of the 25th ACM Int'l Conf. on Multimedia. 2017. 1933−1941.

[41] Sun X, Yang ZN, Zhang C, *et al*. Conditional Gaussian distribution learning for open set recognition. arXiv: 2003.08823, 2020.
[42] Chen T, Kornblith S, Norouzi M, *et al*. A simple framework for contrastive learning of visual representations. In: Proc. of the 37th Int'l Conf. on Machine Learning (ICML 2020), Vol.1. 2020. 1597–1607.
[43] Powers, Martin D. Evaluation: From precision, recall and *F*-measure to ROC, informedness, markedness and correlation. Int'l Journal of Machine Learning Technology, 2011, 2(1): 37–63. [doi: 10.9735/2229-3981]
[44] Jacob D, Chang MW, Lee K, *et al*. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Vol.1: Long and Short Papers. 2018. 4171–4186.
[45] Olga R, Deng J, Su H, *et al*. ImageNet large scale visual recognition challenge. Int'l Journal of Computer Vision, 2015, 115(3): 211–252.
[46] Xiao JX, Hays J, Ehinger KA, *et al*. SUN database: Large-scale scene recognition from abbey to zoo. In: Proc. of the 2010 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2010. 3485–3492.
[47] Poojan O, Patel VM. Deep CNN-based multi-task learning for open-set recognition. Int'l Journal of Computer Vision. arXiv: 1903.03161, 2019.
[48] Lawrence N, Olson ML, Fern XZ, *et al*. Open set learning with counterfactual images. In: Proc. of the European Conf. on Computer Vision (ECCV). 2018. 620–635.

附中文参考文献:

[1] 李志欣, 魏海洋, 黄飞成, 等. 结合视觉特征和场景语义的图像描述生成. 计算机学报, 2020, 9(3): 1624–1640.
[2] 田萱, 王亮, 丁琪. 基于深度学习的图像语义分割方法综述. 软件学报, 2019, 30(2): 440–468. http://www.jos.org.cn/1000-9825/5659.htm [doi: 10.13328/j.cnki.jos.005659]
[3] 杜鹏飞, 李小勇, 高雅丽. 多模态视觉语言表征学习研究综述. 软件学报, 2021, 32(2): 327–348. http://www.jos.org.cn/1000-9825/6125.htm [doi: 10.13328/j.cnki.jos.006125]
[4] 冯耀功, 于剑, 桑基韬, 等. 基于知识的零样本视觉识别综述. 软件学报, 2021, 32(2): 370–405. http://www.jos.org.cn/1000-9825/6146.htm [doi: 10.13328/j.cnki.jos.006146]

朱鹏飞(1983－), 男, 博士, 副教授, 博士生导师, CCF 专业会员, 主要研究领域为机器学习, 计算机视觉.

王煜(1991－), 男, 博士, 助理研究员, CCF 学生会员, 主要研究领域为复杂场景多粒度建模, 动态开放环境机器学习.

张琬迎(1995－), 女, 硕士生, 主要研究领域为机器学习, 开放集识别, 计算机视觉.

胡清华(1976－), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为不确定性人工智能, 机器学习, 模式识别.