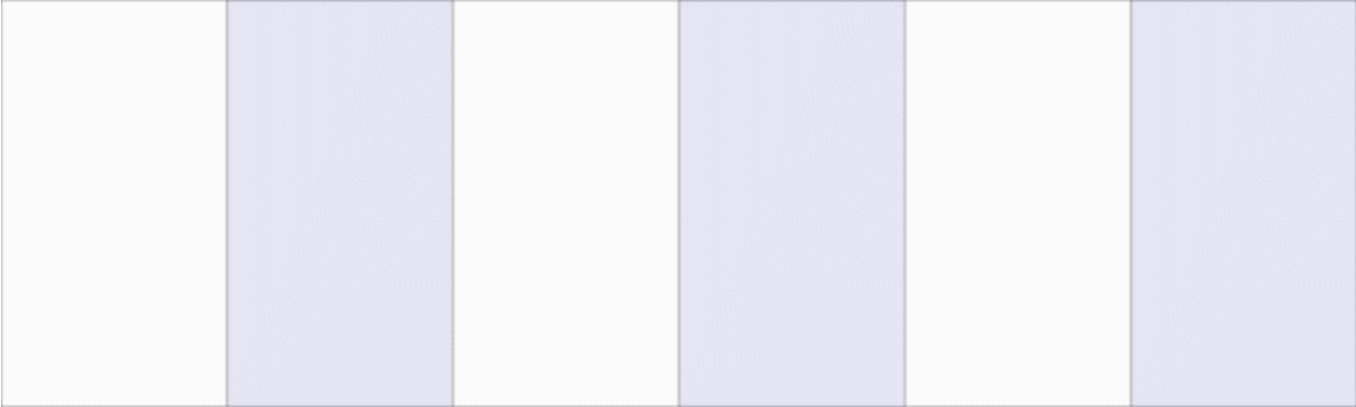


深度对比学习综述

原创 CV开发者都爱看的 极市平台 2023-01-09 22:00:20 发表于广东 手机阅读 𐄞

↑ 点击蓝字 关注极市平台



作者 | 张重生、陈杰、李岐龙、邓斌权、王杰、陈承功
来源 | 河南大学计算机与信息工程学院、河南省大数据分析处理重点实验室
编辑 | 极市平台

极市导读

本文综合考察了对比学习近年的发展和进步，提出一种新的面向对比学习的归类方法，并基于提出的归类方法，对现有对比研究成果进行系统综述，并评述代表性方法的技术特点和区别，系统对比分析现有对比学习方法在不同基准数据集上的性能表现。>>加入极市CV技术交流群，走在计算机视觉的最前沿

深度对比学习综述

张重生¹ 陈杰¹ 李岐龙¹ 邓斌权¹ 王杰¹ 陈承功¹

摘要 在深度学习中,如何利用大量、易获取的无标注数据增强神经网络模型的特征表达能力,是一个具有重要意义的问题,而对比学习是解决该问题的有效方法之一,近年来得到了学术界的广泛关注,涌现出一大批新的研究方法和成果.本文综合考察对比学习近年的发展和进步,提出一种新的面向对比学习的归类方法,该方法将现有对比学习方法归纳为5类,包括:1) 样本对构造;2) 图像增广;3) 网络架构;4) 损失函数;5) 应用.基于提出的归类方法,对现有对比研究成果进行系统综述,并评述代表性方法的技术特点和区别,系统对比分析现有对比学习方法在不同基准数据集上的性能表现.本文还将梳理对比学习的学术发展史,并探讨对比学习与自监督学习、度量学习的区别和联系.最后,本文将讨论对比学习的现存挑战,并展望未来发展方向和趋势.

关键词 对比学习, 深度学习, 特征提取, 自监督学习, 度量学习

引用格式 张重生, 陈杰, 李岐龙, 邓斌权, 王杰, 陈承功. 深度对比学习综述. 自动化学报, 2023, 49(1): 15–39

DOI 10.16383/j.aas.c220421

Deep Contrastive Learning: A Survey

ZHANG Chong-Sheng¹ CHEN Jie¹ LI Qi-Long¹ DENG Bin-Quan¹ WANG Jie¹ CHEN Cheng-Gong¹

论文信息: 张重生, 陈杰, 李岐龙(*)等. 深度对比学习综述. 自动化学报, 2023, 49(1): 15–39

作者单位: 河南大学计算机与信息工程学院 河南省大数据分析与管理重点实验室

论文链接: <http://www.aas.net.cn/cn/article/doi/10.16383/j.aas.c220421>

【注: 原始论文25页, 极市平台出于宣传的需要, 经作者同意, 进行了精简。完整版论文可通过上述链接获取或公众号后台回复“对比学习综述”获取】

摘要:

在深度学习中,如何利用大量、易获取的无标注数据增强神经网络模型的特征表达能力,是一个具有重要意义的问题,而对比学习是解决该问题的有效方法之一,近年来得到了学术界的广泛关注,涌现出一大批新的研究方法和成果.本文综合考察对比学习近年的发展和进步,提出一种新的面向对比学习的归类方法,该方法将现有对比学习方法归纳为5类,包括:1) 样本对构造;2) 图像增广;3) 网络架构;4) 损失函数;5) 应用.基于提出的归类方法,对现有对比研究成果进行系统综述,并评述代表性方法的技术特点和区别,系统对比分析现有对比学习方法在不同基准数据集上的性能表现.本文还将梳理对比学习的学术发展史,并探讨对比学习与自监督学习、度量学习的区别和联系.最后,本文将讨论对比学习的现存挑战,并展望未来发展方向和趋势.

关键词: 对比学习/深度学习/自监督学习/度量学习

1. 引言

近年来, 以深度学习为代表的新一代人工智能技术取得了迅猛发展, 并成功应用于计算机视觉、智能语音等多个领域. 然而, 深度学习通常依赖于海量的标注数据进行模型训练, 才能获得较好的性能表现. 当可用的标注数据较少、而无标注数据较多时, 如何提高深度学习的特征表达能力是亟需解决的重要现实需求. 自监督学习[1]是解决该问题的有效途径之一, 能够利用大量的无标注数据进行自我监督训练, 得到更好的特征提取模型.

然而, 目前, 很少有系统总结对比学习最新进展的英文综述论文[10-12],, 中文综述论文更是极度缺乏. 因此, 学术界迫切需要对深度对比学习的最新文献及进展进行全面系统的总结、归纳和评述, 并分析存在的问题, 预测未来发展趋势. 本文聚焦视觉领域的深度对比学习技术, 系统梳理深度对比学习2018年至今的技术演进, 总结该方向代表性的算法和技术. 如图1所示, 本文首先将深度对比学习的相关技术归纳为样本对构造方法层、图像增广层、网络架构层、损失函数层及应用层5大类型. 然后, 综合归纳现有技术的特点及异同之处, 并分析其性能表现, 指出尚未解决的共性问题及相关挑战, 最后勾勒该领域的未来发展方向与趋势.

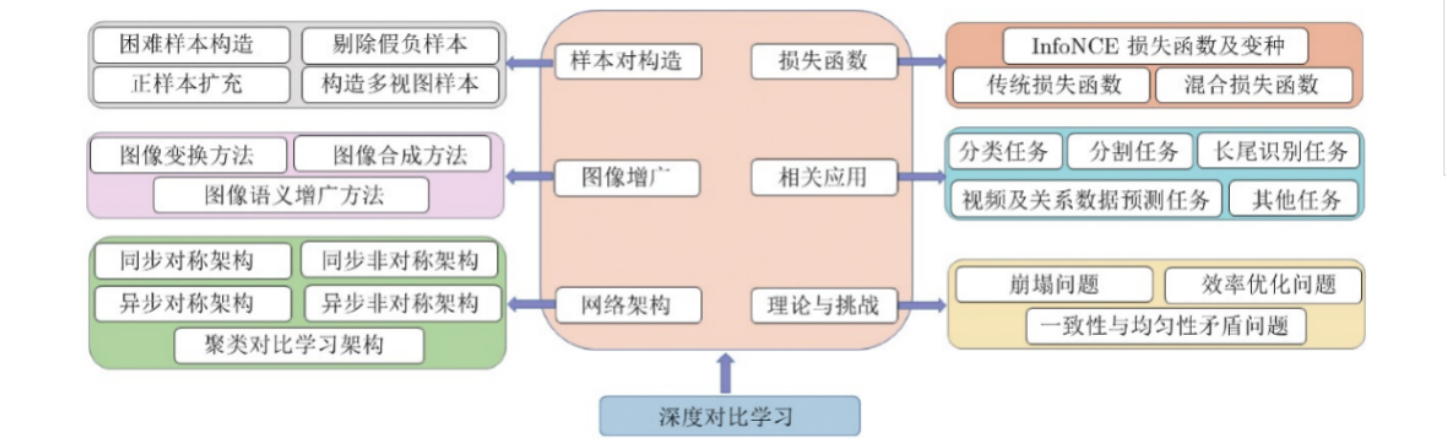


图1 深度对比学习方法归类

2. 对比学习研究现状

在本章内容中, 首先引入所提出的归类方法, 在此基础上归纳总结国内外对比学习研究成果. 值得注意的是, 对比学习亦深受我国学者的关注, 很多优秀的对比学习论文虽然发表在国外学术会议和刊物上, 但作者为国内学者.

2.1 归类方法

本文提出一种新的归类方法, 以对最新的对比学习工作进行系统归纳总结. 如图1所示, 按照对比学习的整体流程, 将现有方法划分为样本对构造层、图像增广层、神经网络架构层、损失函数层以及应用层等类型.

图2将每种类型的对比学习方法进行细分及可视化表示. 其中, 样本对构造层方法可以细分为困难样本构造、剔除假负样本、正样本扩充及构造多视角样本四种方法. 图像增广层可以细分为图像变换、图像合成及图像语义增广三种方法. 网络架构层方法可以分为同步对称、同步非对称、

异步对称、异步非对称对比学习, 及基于聚类的网络架构. 特征提取网络主要使用ResNet[30] (Residual neural network)、Transformer[31]等主流神经网络结构, 损失函数层分为基于互信息的损失函数、传统损失函数和混合损失函数. 下面将分别介绍各类型的方法.

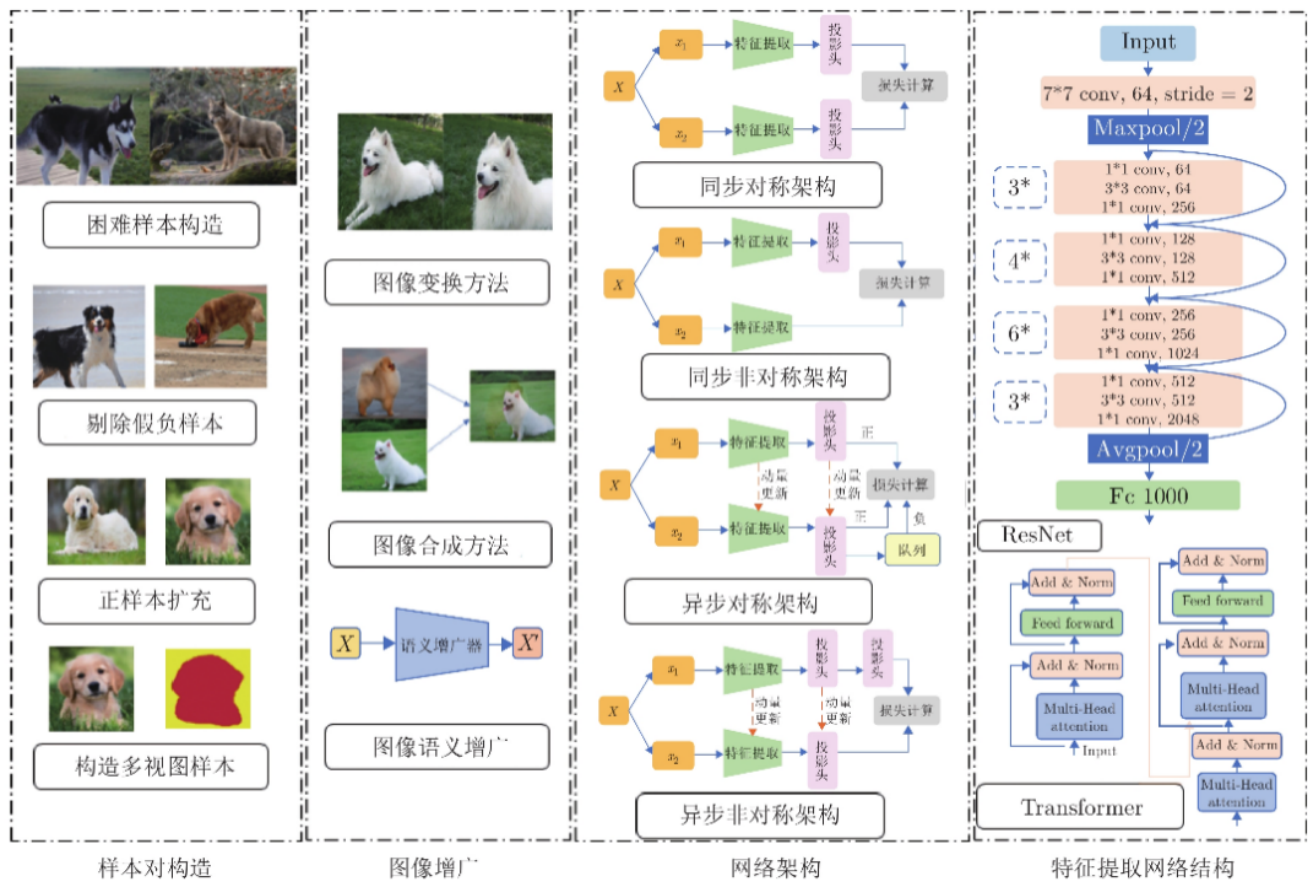


图2 对比学习的整体流程及各模块的细分类方法

2.2 样本处理及样本对构造方法

在对比学习过程中, 样本对的选择指的是对数据集的采样过程. 1) 对无标注数据集, 通常采用随机采样的方法构建一个批次的数据, 因此一个批次的数据可能存在类别分布不均匀的情况, 导致假负样本的出现及困难负样本过少的问题; 2) 对有标注数据集, 通过标签信息采样训练数据, 能够有效提高对比学习效果. 下面将分别介绍各种细分的样本处理及样本对构造方法.

2.2.1 困难样本构造

在选择样本进行模型训练的时候, 有些物体在视觉上相似度很高, 而且在经过神经网络提取特征之后, 他们的特征之间的余弦相似度也比较高, 容易被误认为是相同类的样本, 但是这两张图像中的物体并非同类, 如一张图像中的物体是狗, 另一张是狼. 这种情况称为困难负样本对. 同理, 若两张图片属于同一类, 但特征相似度不高则称为困难正样本对. 多个研究表明, 困难负样本对和困难正样本对在对比学习中具有至关重要的作用.

Zhu等[32]通过对MoCo算法训练过程的可视化分析, 发现增加困难样本在同批次中的比例能够提升网络在下游任务中的表现. 通过这一现象, 作者提出了在特征空间上将负样本对图像对应的

两个特征向量插值, 正样本对图像对应的两个特征向量外推, 构造新的、更加困难的样本对, 最终提高了对比学习的效果. 采用类似的思想, Kalantidis等[33]提出了一种合成困难负样本的方法, 该方法也作用于特征空间中, 首先通过余弦相似性度量方法将输入样本对应的所有负样本进行降序排序并取前 K 个样本, 然后使用与文献[32]相同的方法依次合成 K 个困难负样本. 但是该方法只适用于带队列存储库 (Memory bank) 的对比学习模型. Zhong等[34]提出了一种结合图像混合 (Mixup) 技术的困难样本构造方法, 该方法首先利用Mixup方法合成新的样本, 然后通过余弦相似性度量方法挑选困难的合成样本, 加入到原始数据中进行对比学习训练, 提高了对比学习的训练效果.

2.2.2 剔除假负样本

在同一批次的训练图像中, 如果输入图像和同一个批次的其他某一张图像是同一类, 但是在计算损失的时候把它们误归为负样本对, 这种情况就称为假负样本对. 在对比学习中, 负样本的质量和数量是制约最后训练效果的关键, 如果在训练的批次中出现假负样本, 会降低最后的网络效果.

为了解决该问题, Huynh等[35]提出一种启发式的方法来避免假负样本对训练的影响. 该方法将同一幅图像使用多种增广方法得到的所有图像作为该幅图像对应的支持集, 然后将同一个批次中该图像对应的每个负样本依次与其支持集中的每幅图像进行特征相似性计算, 并将前 K 个最相似的负样本视为假负样本, 认为这些负样本与输入图像具有相同的类标签. 得到假负样本后, 可以直接剔除假负样本, 或将假负样本移入到正样本集合. 通过大量实验, 作者证明了两种处理方式都能提高对比学习的效果. 针对相同的问题, Chuang等[36]提出修正InfoNCE损失函数中所有负样本相似度的指数幂和, 以降低可能采样到的假负样本对损失值带来的影响.

2.2.3 正样本扩充法

扩充正样本有助于提高对比学习的效果. 这里的扩充不包含图像增广方法, 而是指从可用的数据资源中寻找隐藏的、与输入样本类别相同的图像的方法. 在监督对比学习中, 由于同类样本已知, 一般无需扩充正样本. 而在无监督对比学习中, 若图像数据没有标注信息, 一般无法进行正样本扩充. 但在一些特殊场景下, 如行人重识别、遥感图像处理、视频分析, 可以基于某些假设, 对正样本进行扩充.

Kim等[37]将Mixup方法应用到了对比学习领域, 提出MixCo方法, 该方法将经过增广之后的两幅异类图像进行Mixup操作合成新的图像. 然后对所有合成的图像, 计算其InfoNCE损失, 由于每幅合成图像对应两个类别, 因此可以认为是扩充了每个样本对应的正样本个数. 最终损失由原始图像上的对比损失和上述合成图像上的对比损失构成.

在行人重识别领域, 王梦琳[38]提出一种度量不同相机中行人关联性的方法, 将关联性强的样本视作同类样本, 以扩充正样本的数量.

在遥感图像处理中, 基于同一地点的遥感图像中的内容随时间变化较小的特点, Ayush等[39]使用同一地点的不同时刻的遥感图像扩充正样本.

在视频分析中, Qian等[40]基于邻近的视频帧图像可作为同类样本的假设, 扩充正样本, 用于对比学习. Kumar等[41]同样采用这一假设扩充正样本. Han等[42]发现对于两个不同的视频片段, 做的动作相同时, 在RGB视角下, 提取的特征相似度不高, 但在光流 (Optical flow) 视角下, 提取的特征相似度很高, 反之亦然. 基于以上发现, 作者提出将光流视角下相似度高的视频片段作为RGB视角下的正样本, 同样, 可采用相同方法扩充光流视角下的正样本.

在半监督学习领域, Wang等[43]提出通过部分标签训练的分类器, 在当前训练批次中利用余弦相似度指标寻找正样本的方法, 扩充正样本数量. Yang等[44]采用同样的思路设计类感知模块, 扩充正样本数量.

2.2.4 构造多视图样本

多数对比学习方法所使用的数据只有一个来源 (又称为同一个视图). 针对该问题, 研究人员探索使用多视图数据的方法提升对比学习的效果. Tian等[45]提出基于多视图特征的对比学习方法. 该方法将同一幅图像在多个不同视图下的表达分别进行特征提取, 然后进行对比学习, 有利于提升模型的效果. 在视频分析中, Rai等[46]对同一幅图像分别提取光流、语义分割、关键点等多视图特征, 然后进行对比学习, 提升了视频特征表达能力.

2.3 图像增广方法

2.3.1 图像变换方法

图像增广是对比学习的天然组成部分, 以产生网络所需的输入样本对. 传统的图像变换方法有随机裁剪、颜色失真、灰度化等方法. 需要说明的是, 单一的图像变换方法不利于有效的对比学习. SimCLR综合对比了一系列图像变换的效果, 发现由随机裁剪和颜色失真组成的变换组合能够获得更好的对比学习效果.

在众多的图像变换方法之中, 裁剪通常是必须的变换方法之一, 但可能存在经过裁剪后的正样本对信息重叠过多或者完全不重叠的情况. 如果裁剪到的正样本对是两个矩形框中的区域, 那么该图像对并不能促进对比学习, 这种情况称为错误正样本 (False positive, FP). 当裁剪的两个图像区域重叠过多时, 若选择它们作为正样本对, 也无助于对比学习. Peng等[47]针对以上问题, 首先利用热图定位目标区域, 然后采用中心抑制抽样策略, 在目标区域内离物体中心越远的像素点被采样到的概率越高. 得到采样点后, 以采样点为中心, 并使用随机生成的宽高, 进行最终的图像裁剪操作. 在SwAV中, 作者提出了一种多裁剪策略 (Multi-crop), 该策略给编码器提供多个不同尺度的裁剪图像作为正样本, 并验证了该策略有助于提高对比学习算法的性能.

2.3.2 图像合成方法

图像合成不同于图像变换, 前者能够生成新的内容. 在视频分析中, Ding等[48]提出复制粘贴的图像合成方法, 将所选视频帧的前景区域粘贴到其他视频帧的背景图像中, 得到合成的视频帧,

通过这种简单的合成方法,使得对比学习模型能够更加关注前景信息,提取到更加良好的特征表达.

2.3.3 图像语义增广方法

图像语义增广是一种直接对图像中物体的语义进行修改的图像增广方法,如将图像中的物体的颜色或角度进行改变. 现有的一些语义增广算法[49-51]已证明其在不同问题和应用中的有效性. 相较于图像变换方法和图像合成方法,图像语义增广是一种更强的图像增广方法. Tian等[52]认为最小化在不同增广下样本间的互信息有利于提高对比学习的效果,为了最小化训练样本对之间的互信息,作者引入了对抗训练策略寻找可以最小化训练样本对互信息的图像变换编码器 S_g ,通过与InfoNCE联合训练,获得了更适用于下游任务的模型,经过训练得到的图像变换编码器 S_g 可认为是一个语义增广器.

2.4 对比学习网络架构设计

根据对比学习网络架构的更新方式是同步或异步,及该网络架构是对称或非对称,本文将对对比学习所涉及的网络架构划分为同步对称、同步非对称、异步对称和异步非对称4种类型. 同步更新指的是对比学习网络的两个分支(分支1和分支2)同时进行梯度更新,异步更新指的是两个分支网络的权值更新方法不同. 对称指的是分支1和分支2的网络结构完全相同,非对称则指相反的情况. 除此之外,本节还将含有聚类算法的对比学习方法总结为聚类对比学习架构.

2.4.1 同步对称网络架构

在典型的同步对称网络架构中,分支1和分支2采用相同的网络结构,并且同时采用梯度更新.

SimCLR[2]是最早提出采用同步对称网络架构的对比学习工作. 该网络架构采用了结构相同的两个网络分支,每一个分支都包含特征提取网络和投影头. 特征提取网络可以是任意的网络结构,例如ResNet和Transformer,投影头一般使用多层感知机(Multilayer perceptron, MLP). 使用该网络架构进行对比学习的前向过程为: 首先,输入一个样本,经过两种图像增广方法产生成对的训练样本. 然后,通过特征提取网络和投影头将样本输出到投影空间. 最后,在投影空间进行损失计算并回传梯度,更新特征提取网络和投影头的参数. 同步对称网络架构除了可以用于自监督对比学习之外, Khosla等[8]将该类网络架构应用到了监督对比学习之中. 值得注意的是,采用该类网络架构的对比学习算法若想获得良好效果,往往需要在训练时采用很大的批次进行训练,比如1 024或2 048.

2.4.2 同步非对称网络架构

在典型的同步非对称网络结构中,分支1和分支2采用不同的网络结构,但同时采用梯度更新. 同步非对称网络架构与同步对称网络架构的相同之处在于两个分支都进行了梯度更新,不同之处在于前者的两个分支的网络结构不同,而后者相同.

Van等[7]提出一种基于该种网络架构的对比预测编码方法CPC. 在该方法中, 分支1的输入为某一个时间点的数据, 分支2的输入为未来某一个时间点的数据, 训练的目标是利用分支1的输出预测分支2的输出, 然后计算对应的预测损失. 在CPC算法之后, 衍生出很多改进的算法[41, 53], 这些方法均采用典型的同步非对称网络架构, 其主要改进之处在于采样方法或损失函数.

除了上述同步非对称架构之外, 还存在一些其它形式的同步非对称的网络架构, 主要有如下两种形式: 1) 分支1与分支2均含有投影头, 但投影头结构不相同[54-55]. 2) 分支1和分支2的特征提取网络数量不相等[56-57].

Nguyen等[54]将对比学习方法应用于网络结构搜索任务 (Neural architecture search, NAS), 提出CSNAS方法, 该方法构造了一个类似于上述第1种形式的同步非对称网络架构. Misra等[55]在分支2中使用拼图代理任务进行特征学习, 因此在分支2的投影头之前还包含拼接操作以及多层感知机的映射, 构成了符合上述第1种形式的同步非对称网络架构.

Bae等[56]提出自对比学习 (Self contrastive learning, SelfCon) 方法, 该方法构造了一个类似于上述第2种形式的同步非对称网络架构, 在该方法中, 将一个特征提取网络和一个投影头组合在一起, 称为一个主干网络块, 分支1采用多个主干网络块进行特征映射, 分支2从分支1中的某个节点引出, 通过一次主干网络块映射得到另一个视图下的特征, 最后进行损失计算. 类似地, Chaitanya等[57]提出可用于图像分割的对比学习方法, 在该方法中, 分支1和分支2共享编码器, 但分支1直接在编码器后加入投影头, 分支2在编码器后加入一个解码器, 然后再加入投影头.

2.4.3 异步对称网络架构

在异步对称的网络架构中, 分支1和分支2采用相同的网络结构, 但是两个分支的神经网络权值更新方式不同, 因此称作异步更新.

MoCo是经典的采用异步对称网络架构的对比学习方法. 其中, 分支1中的特征提取网络和投影头采用SGD等梯度更新方法更新权值, 而分支2的特征提取网络和投影头采用动量更新的方式, 动量更新的公式如式 (1) 所示:

$$\theta_k^{t+1} \leftarrow m\theta_k^t + (1-m)\theta_q^{t+1} \quad (1)$$

其中, θ_k^t 是分支 2 中动量特征提取网络在时刻 t 的参数值, θ_q^{t+1} 是分支 1 中特征提取网络经过 t 时刻梯度反传训练过后的参数值. m 是 $0-1$ 之间的动量系数. 这种动量更新机制能够有效防止极端样本对参数更新影响过大的问题.

在对比学习中, 负样本的数量对最终的模型效果起决定性的作用. 在原始的对比学习方法中, 负样本的来源是同批次的训练数据, 为了增加负样本的数量, 必须采用大批次的训练数据来支撑训练, 但这样会造成存储资源过大, 计算消耗巨大的问题. MoCo设计了队列 (Queue) 结构, 存储一定数量的之前批次用到的样本的特征, 当队列容量满时, 最老样本的特征出队, 最新样本的特征入

队. 在MoCo方法中, 对比学习所需的负样本将从该队列中抽样产生, 避免了原始对比学习方法为获得较多负样本而构建大的批容量带来的问题. 简言之, 该队列结构减小了对存储资源的要求, 又取消了训练必须采用大批次数据的约束, 提升了计算速度. 后续的基于异步对称网络架构的方法都将该队列结构作为其默认的组成部分.

近年来, 随着Transformer技术的发展, 研究者将其引入到计算机视觉领域, 例如ViT[58]. 在对比学习的研究中, Caron等[59]提出了一种名为DINO (Self-distillation with no labels) 的异步对称对比学习方法, 该方法采用Transformer作为骨干网络, 并将自监督对比学习转换为蒸馏学习的任务, 将分支1视作学生网络, 将分支2视作教师网络. 学生网络采用梯度回传更新参数, 教师网络采用动量更新方式更新参数.

2.4.4 异步非对称网络架构

异步非对称网络架构指的是分支1和分支2采用不同的网络结构, 同时参数更新方式也不同. 异步非对称网络架构与异步对称网络架构的相同之处在于两个分支的更新方式相同, 不同之处在于前者的两个分支的网络结构不同, 而后者相同.

异步非对称网络架构存在两种形式, 一种是采用MoCo作为主干网络, 但是投影头数量不一致, 以BYOL[4]方法为典型代表. 另一种是网络架构不对称, 同时采用梯度交叉更新的方法, 以SimSiam[6]方法为典型代表.

BYOL的主干网络结构是MoCo, 分支2的网络参数采用动量更新机制. BYOL是首个只采用正样本进行对比学习的工作, 但由于训练集中不存在负样本, 如果上下网络分支结构完全相同, 训练就有可能出现“捷径解”的问题 (亦可称为“网络崩塌”问题), “捷径解”指的是对不同的输入, 输出的特征向量完全相同. 为了解决这一问题, BYOL在分支1中额外增加了一个预测头 (MLP), 构成了非对称网络架构, 并将对比学习的实例判别代理任务替换为比对任务, 即分支1最终得到的特征与分支2最终得到的特征要尽可能的相似. BYOL的性能提升归因于以下两点设计: 1) 良好的模型初始化. 2) 在投影头和预测头中加入正则化技术. 如果没有这两种设计, BYOL仍可能出现“捷径解”的问题[60].

Chen等[61]提出了一种基于视觉Transformer的异步非对称网络架构MoCov3. MoCov3采了BYOL型异步非对称网络架构. 在分支1中额外加入了一个预测头, 分支2的编码器采用动量更新, 获得了更好的对比学习效果.

另一种代表性的异步非对称架构是SimSiam. SimSiam方法也不需要利用负样本训练网络. 为了避免网络出现“捷径解”, 该方法的两个分支的网络结构采用交叉梯度更新的方式. 具体而言, 训练分支1的时候, 投影头放在分支1的编码器之后, 计算其损失, 而分支2的梯度停止回传. 在训练分支2的时候, 将分支1的投影头接入到分支2的编码器之后, 并计算损失, 而分支1的梯度停止回传. 这就是交叉梯度更新.

2.4.5 聚类对比学习结构

在上述四种网络架构的基础上, 还可以结合聚类技术进行进一步的优化.

基于实例的对比学习算法在计算损失差别过大, 不利于模型学习到良好的语义特征. 而聚类算法能够在无监督情况下自动学习数据的语义信息. 因此, 一些方法将对比学习与聚类方法结合起来, 嵌入到上述四种网络架构中.

Caron等[62]首次将K-Means聚类算法与无监督深度学习结合, 提出深度聚类算法 (Deep cluster), 该算法通过K-Means聚类产生伪标签, 然后利用伪标签对模型进行自监督训练. 在该方法的基础上, 作者又提出了一种基于聚类的对比学习算法SwAV, 该算法采用同步对称网络架构训练特征提取网络. 在分支1和分支2之间引入聚类中心信息, 并计算相关损失, 帮助网络进行训练. 相关的细节在第2.5.2节中给出.

Li等[63]提出原型对比学习算法 (Prototypical contrastive learning, PCL), 该方法采用异步对称网络架构, 分支1采用梯度更新, 分支2采用动量更新. 在PCL的前向传播过程中, 当分支2对输入样本提取特征之后, 采用K-Means算法依据样本特征进行聚类. 然后, 将分支1得到的每个样本特征与分支2的所有聚类中心计算聚类对比学习损失, 将分支1的样本与其所属的聚类中心的向量视作正样本对, 其余为负样本对. 最终的损失函数由聚类对比学习损失和原有 (实例) 对比学习损失构成. 聚类对比学习损失计算的目标是最大化样本实例与其相对应的聚类中心特征之间的相似度, 因此这种设计也能在一定程度上缓解实例对比学习中的假负样本问题. PCL仅在分支2上进行聚类计算, 后续工作尝试在两个分支上都进行聚类, 如Wang等[64]提出的方法, 将分支1和分支2都进行聚类计算, 然后进行交叉聚类对比学习损失计算.

在现实图像数据集中, 往往存在多层级的语义结构, 例如在“狗”这个分类的数据中, 还存在“哈士奇”、“金毛”、“雪纳瑞”等不同品种的狗. 现有的很多对比学习方法没有考虑到数据集的这个特点. 针对这一问题, Guo等[65]提出了一种分层聚类的对比学习方法 (Hierarchical contrastive selective coding, HCSC), HCSC的网络架构和PCL的神经网络架构相同, 都属于异步对称网络架构. HCSC通过多层聚类来实现对数据集中存在的分层语义现象的模拟. 具体而言, 在分支2的特征提取操作之后, 首先, 通过K-means算法计算第一层聚类中心, 随后针对第一层聚类中心再使用K-means算法寻找下一层的聚类中心, 循环 N 次, 得到 N 个层次的聚类向量, 完成对 N 层语义的模拟. 同时HCSC还提出一种新的负样本采样方法: 在每一层语义结构中寻找与当前输入样本特征的原型相似度较低的样本作为负样本. 这种采样负样本的方法能够有效缓解采样到假负样本的问题.

此外, Li等[66]提出了一种聚类中心可自动学习的聚类对比学习算法. 该网络设计了一个共享特征网络的双分支同步对称网络架构, 两个分支分别构建投影头, 即用于实例对比学习的投影头 I , 和用于聚类对比学习的投影头 C . 投影头 I 采用线性激活, 投影头 C 采用softmax激活. 将投影头 C 输出的矩阵中的列向量作为聚类中心, 通过InfoNCE损失和交叉熵损失联合训练模型, 在训练过程中由于投影头 C 参数在更新, 聚类中心因此也获得自动更新的能力.

Cui等[67]提出了一种新的异步非对称聚类网络架构PaCo, 该方法将聚类思想和MoCo结合在一起, 通过构建一个可随着模型一同学习的参数化类别中心 C , 使数据少的类别在训练中的重要性

提升, 从而在长尾学习中实现对损失的再平衡.

2.5 损失函数设计

为了实现对比学习的优化目标, 研究者使用了多种针对性的损失函数, 可以分为基于互信息的损失函数 (InfoNCE类)、传统损失函数和混合损失函数.

值得注意的是, 对比损失[13]与对比学习损失很容易在字面上产生混淆, 实际上, 对比损失是一种度量学习损失, 只能用于监督学习, 且样本对数据来源与对比学习不同, 因此, 对比损失不一定是对比学习的损失函数.

2.5.1 InfoNCE损失函数及变种

对于深度学习模型来说, 理想的训练目标是获得一个从样本到特征的映射模型 $p(z | x; \theta)$, 其中 x 是样本, z 是模型输出的特征, θ 为需要学习的参数. 很多自监督学习方法通过交叉熵或均方误差训练网络达到上述目标. 然而, 基于交叉熵或均方误差这样的单峰函数的损失往往效果不好. 因此 Van等[7]提出通过最大化互信息的方法训练特征提取网络. 互信息 $I(X; Y)$ 是一个概率论中的概念, 可以衡量两个随机变量 X 和 Y 之间的相关性. 互信息的定义如式(2) 所示:

$$I(X; Y) = E_{x,y \sim p(x,y)} [\ln \frac{p(x|y)}{p(x)}] \quad (2)$$

由于直接最大化互信息是十分困难的, 因此作者提出了通过InfoNCE损失来间接优化互信息的方法. InfoNCE损失的思想是将最大化互信息转换为真实分布占合成分布的概率密度比的预测问题 [68]. 原始InfoNCE公式 (3) 所示:

$$L_{infoNCE} = - \ln \frac{\exp(s(q, h^+))}{\sum_{x_i} \exp(s(q, h_i))} \quad (3)$$

其中, $X = \{x_1, \dots, x_N\}$ 为采样的 N 个样本, f 为特征提取网络, $q = f(x_q)$ 为查询样本的特征向量, $h^+ = f(x_i)$ 为 x_q 为对应的正样本的特征向量. 在训练中, 每个查询样本 x_q 只对应 1 个正样本 x_i , 其余都视为从噪声分布里面采样的负样本. s 为相似度度量函数, 这里采用余弦相似度. Van等[7]和Poole等[69]证明了优化 InfoNCE 损失等价于优化变量之间的互信息的下界, 并且 $I(x_q, x_i) \geq \ln N - L_{infoNCE}$, 可知如果想获得一个较高的互信息值, 需要大批次的样本参与训练.

SimCLR首先将InfoNCE损失引入到对比学习中来, 由于训练所用的样本对来自于同一幅图像的两次不同增广, 因此InfoNCE的优化目标变为最大化同一个样本的两个不同增广图像之间的互信息, SimCLR中采用的InfoNCE如式 (4) 所示:

$$L = -\ln \frac{\exp(\frac{s(z_i, z_i^+)}{\tau})}{\sum_{j=1}^{2N} 1_{j \neq i} \exp(\frac{s(z_i, z_j)}{\tau})}$$

(4)

其中, $X = \{x_1, \dots, x_N\}$ 为采样的 N 个样本, f 为特征提取网络, g 为投影头, T 为图像增广函数, $z_i = g(f(T_1(x_i)))$ 为样本经过第 1 种图像增广之后产生的投影向量, $z_i^+ = g(f(T_2(x_i)))$ 为样本经过第 2 种图像增广之后产生的投影向量, $1_{j \neq i}$ 为一个指示器函数, 当 $j \neq i$ 时为 1, 否则为 0. s 为相似度度量函数, 此处使用余弦相似度函数, 即 $s(u, v) = u^T v / (\|u\| I / \|v\| I)$. 总结而言, 在 SimCLR 的 InfoNCE 公式中, 分子表示同一个样本的两种不同变换得到的正样本对的特征相似度, 分母是不同图像组成的负样本集合中的每个负样本对的相似度的和.

在InfoNCE引入到对比学习之后, 许多方法[3, 9, 32-33, 40, 48, 52-54, 57, 70]对其直接调用并扩展到更多的学习任务中. 表1列出了一些对InfoNCE改动较大的方法, 包括ProtoNCE[63], DCL[71], DirectNCE[72], SCL[8], FNCL[35]. 其中, 第一行为原始的InfoNCE损失函数, 其他部分为对其改进的损失函数, 包括ProtoNCE, DCL, DirectNCE, SCL, FNCL等损失函数.

损失名	文献	年份	会议/刊物	公式	主要改进
InfoNCE	[7]	2019	arXiv	$-\ln \frac{\exp(s(q, h^+))}{\sum_{x_i \in X} \exp(s(q, h_i))}$	初始的 InfoNCE 损失. 文献 [3, 9, 30-31, 38] 等均采用 InfoNCE 损失函数.
ProtoNCE	[63]	2021	ICLR	$-\ln \frac{\exp(s(z_i, c_i)/\tau)}{\sum_{j=0}^r \exp(s(z_i, c_j)/\tau)}$	从两个样本增广间的对比变为样本增广与聚类中心的对比. 注: c_i, c_j 为聚类中心. 总损失包括实例间的对比和实例-原型对比损失, 此处只列出实例-原型对比损失.
DCL	[71]	2021	ECCV	$-\frac{s(z_i, z_i^+)}{\tau} + \ln \sum_{j=1}^{2N} 1_{j \neq i} \exp(s(z_i, z_j)/\tau)$	去除负正耦合系数后通过化简得到该损失.
DirectNCE	[72]	2022	ICLR	$-\ln \frac{\exp(s(\hat{h}_i', \hat{h}_i^+)/\tau)}{\sum_j \exp(s(\hat{h}_i', \hat{h}_j^+)/\tau)}$	$\hat{h}_i' = \hat{h}_i[0:d]$, 即取特征向量的前 d 个维度.
FNCL	[35]	2022	WACV	$-\ln \frac{\exp(s(z_i, z_i^+)/\tau)}{\sum_{j=1}^{2N} 1_{j \neq i, j \neq F_i} \exp(s(z_i, z_j)/\tau)}$	F_i 为第 i 个样本的假负样本集.
SCL	[8]	2020	NIPS	$-\sum_{i \in I} \frac{1}{p(i)} \sum_{p \in p(i)} \ln \frac{\exp(s(z_i, z_p)/\tau)}{\sum_{a \in A(i)} \exp(s(z_i, z_a)/\tau)}$	将标签引入对比学习, $P(i)$ 是与第 i 个样本同类的数据集合.

表1 InfoNCE损失函数及其变种

1. ProtoNCE损失函数. 与原始InfoNCE损失函数相比, ProtoNCE的分子部分将正样本实例之间的相似度修改为正样本与其所在的聚类中心之间的相似度, 分母部分改为计算样本与其它聚类中心之间的相似度. 具体而言, 当前实例特征与其对应的聚类中心互为正样本对, 与其它聚类中心互为负样本对. 在实际操作中, 通过设置不同的聚类中心个数对一个批次的数据进行 M 次聚类, 计算 M 次ProtoNCE损失, 然后求其损失均值, 最后得到的效果会更好.ProtoNCE方法是在同一个语义层级上进行聚类, 但无法获得分层的语义结构. Guo等[65]针对该问题提出分层聚类的方法, 该方法对每个聚类迭代地进行下一级细分聚类, 使特征网络能够学习到分层的语义结构, 获得更好的效果.
2. DCL损失函数. Yeh等[71]通过对InfoNCE的反传梯度进行分析, 发现损失的梯度中存在一个负正耦合系数, 该系数体现了采用InfoNCE训练网络时大批次样本的重要性, 同时揭示当训练

样本中存在简单正样本对和简单负样本时, 会明显降低对比学习训练的效率, 因此作者提出了解耦对比学习损失 (Decoupled contrastive loss, DCL), 解决上述问题. 与InfoNCE损失函数相比, DCL去除InfoNCE损失函数中的负正耦合项, 并将损失展开成两部分, 具体如表1中的DCL损失函数公式所示, 该式展示的是单个样本的DCL损失计算公式, 原InfoNCE包含的负正耦合系数在该式中已去除. 该式的第一项表示正样本对之间的相似度, 第二项是当前样本与同批次其他样本组成的负样本对之间的相似度之和. 从该式可以看出, DCL损失的计算相较于InfoNCE损失更加简单、高效.

3. DirectNCE损失函数. Jing等[72]从解决对比学习中的网络崩塌问题入手, 经过一系列分析, 提出去掉投影头, 然后将特征提取网络输出向量的前 d 个维度单独取出, 计算InfoNCE损失. 由于论文中并没有给改动后的损失起名, 因此为了便于对比, 本论文中将其命名为DirectNCE. DirectNCE与InfoNCE的区别在于所使用的样本特征维度大小, 在DirectNCE中, 样本只取前 d 个维度计算损失.
4. FNCL损失函数. 针对在训练对比学习模型时可能存在的假负样本问题, Huynh等[35]首先提出了一种假负样本的检测策略, 然后对InfoNCE进行改进, 在损失函数中剔除了假负样本的干扰. 本文将该方法暂定名为FNCL. 与InfoNCE相比, FNCL方法首先确定当前正在处理的样本对应的假负样本, 然后在分母中计算负样本对之间的相似度时, 去除假负样本的部分.
5. SCL损失函数. SCL[8]损失函数面向有监督学习, 对InfoNCE损失函数进行改进, 旨在解决深度有监督学习中采用交叉熵损失时神经网络对噪声标签敏感[73]的问题.

令 $P(i)$ 表示当前批中正在处理的样本对应的同类样本集合 (因为数据为有标注数据), $|P(i)|$ 表示该集合中样本的数量, 与 InfoNCE 损失函数相比, SCL 损失函数的分子计算当前样本与其对应的 $P(i)$ 集合中每个正样本之间的相似度的和, 而分母部分并无变化. 该损失在特定情况下可以等价于三元组损失[74]和N-Pair损失[75]. 在SCL的基础上, 研究者还研究了利用对比学习提升长尾学习效果的方法, 此部分工作将在第2.6.4节进行详细介绍.

2.5.2 传统损失函数在对比学习中的应用

如何衡量特征空间中不同特征点之间的距离是对比学习中一个很重要的问题. 欧氏距离是衡量特征点之间距离的一个最直观的方法. 通过最小化均方误差损失 (Mean square error, MSE), 可以直接减小同类特征点之间的欧式距离, 实现让同类特征靠近的目的. BYOL先对两个分支的特征进行正则化, 然后采用MSE损失计算两个特征之间的距离, 对网络进行优化.

除了欧氏距离外, 余弦相似度也能够衡量特征之间的相似性, 因此直接最大化相同样本的不同增广的特征之间的余弦相似度也能达到对比学习的目标. SimSiam[6]对其中一个分支的投影向量 z 和另一个分支的特征向量 h 计算余弦相似度, 直接将负的余弦相似度作为损失函数进行训练, 由于网络结构是一个非对称结构, 为了平衡训练, 作者将投影头依次放在两个分支后面进行损失计算, 从而提出对称损失, 其计算公式如式 (5) 所示:

$$L = -\frac{1}{2} \left(\frac{h_1 z_2}{\|h_1\|_2 \|z_2\|_2} + \frac{h_2 z_1}{\|h_2\|_2 \|z_1\|_2} \right) \quad (5)$$

欧式距离和余弦相似度都是非监督的特征相似度度量方法. 除此之外, 还有一些有监督的相似度计算方法.

在 SwAV 方法中, 首先初始化聚类中心, 然后将聚类中心矩阵分别与分支 1 和分支 2 的特征矩阵相乘, 计算交换预测编码矩阵 Q_1 和 Q_2 , 最后, 采用交叉熵损失训练模型, 训练分支 1 时将 Q_2 作为标签, 训练分支 2 时将 Q_1 作为标签. 聚类中心利用回传梯度进行更新.

此外, Shah等[76]将支持向量机与对比学习结合在了一起, 采用改进后的合页损失 (Hinge loss) 优化对比学习网络.

2.5.3 混合损失函数

在某些情况下, 只采用InfoNCE损失不能获得良好的效果, 而将多种损失函数结合, 有助于提升对比学习的效果.

有监督混合损失. Wang等[77]提出了一种将SCL与交叉熵损失结合起来的损失. 该方法采用一个平滑因子, 在训练早期, SCL占据损失的主导地位, 随着学习过程的进行, 交叉熵损失会逐渐占据主导地位. Li等[78]将ProtoNCE的思想带入到SCL中, 即在SCL的分子中计算每个样本与其同类样本所形成聚类中心之间的相似度. 在该算法中, 聚类中心直接由同类样本的特征求平均值得到. 最后, 作者将上述损失与交叉熵分类损失联合训练, 获得了较好的遥感图像分类效果.

半监督混合损失. Li等[79]设计了一个基于伪标签图结构对比学习方法CoMatch. CoMatch方法采用三部分损失训练网络. 对比学习网络中的一个分支, 对于有标签的数据, 采用交叉熵计算损失, 得到分类模型. 在对比学习网络的另外一个分支, 首先利用分类模型对无标签数据进行预测, 产生软标签, 该分支对每个无标注的数据, 进行特征提取, 并利用预测头产生预测结果, 当样本对应的软标签的置信度较高时, 采用交叉熵损失优化网络, 同时, 每个无标签的样本还将采用InfoNCE计算损失. Yang等[44]采用与CoMatch相似的损失构建方法, 对有标签的数据采用交叉熵损失训练, 对软标签置信度高的无标签数据采用交叉熵训练, 对其余数据采用InfoNCE进行对比学习训练, 在该算法中, 由于包含类感知模块, 可以获得与当前训练样本相同类的样本集合, 因此对比学习损失部分采用实例InfoNCE和SCL相结合的损失函数. 此外, Wang等[43]将交叉熵损失和SCL结合在一起, 获得了良好的效果.

无监督混合损失. Park等[80]将对比学习损失融合到基于GAN的图像风格迁移任务中, 该方法对变换前后相同位置的图像块进行对比学习, 将对比学习损失辅助于GAN损失, 训练图像风格迁移模型, 获得了良好的效果.

其他混合损失. Rai等[46]对于具有多视图特征的数据, 对同一样本在不同视角下的特征, 分别采用InfoNCE、MSE和合页损失计算这些特征之间的相似性, 并将三个相似性度量结果进行混合,

用于网络模型优化. Kim等[37]将Mixup方法用在了对比学习中, 该方法假定合成后的样本同时属于合成前的两个样本的类别, 然后将合成样本分别与合成前的样本进行对比学习, 得到两个损失, 这两个损失的混合系数与合成图像时所产生的混合系数一致. 基于相似的InfoNCE与交叉熵损失结合的思想, Kumar等[41]解决了无监督视频动作分割问题, Yang等[81]解决了文本-图像跨模态特征提取问题, Dong等[82]实现对五种模态数据的跨模态特征提取.

2.6 相关应用

对比学习在分类、分割、预测等下游任务中均有重要应用. 本文针对每种下游任务, 按照数据的类型, 介绍相关的应用. 本文将数据的类型概括为静态数据和序列数据, 其中静态数据主要有图像、关系型数据、点云和图结构等类型, 序列数据主要有视频、音频、信号等类型.

2.6.1 分类任务

分类任务是对比学习最常见的下游应用. 在静态数据中, 针对图像分类任务, Hou等[83]提出基于对比学习的半监督高光谱图像分类算法, 解决有标注数据不足时高光谱图像分类问题. 该算法分为2个阶段对模型进行训练, 第1阶段, 对于无标签样本, 利用对比学习方法对模型进行预训练. 第2阶段, 利用有标注的样本对模型进行监督学习. 针对小样本遥感场景分类问题, Li等[78]将无监督对比学习方法融合到小样本学习的框架中, 提高了模型的特征提取能力. 郭东恩等[84]将监督对比学习方法引入到遥感图像场景分类任务中, 通过监督对比学习预训练, 提高了遥感图像分类精度. Aberdam等[85]将对比学习应用到文本图像识别任务. 由于对文本图像采用随机增广的方法可能会导致文本内容的丢失等问题, 因此, 作者首先设计可对齐的文本图像增广技术, 然后, 基于同步对称网络架构进行对比学习训练, 最终提高了文本图像识别的准确率. 在细粒度分类问题中, Zhang等[86]直接采用数据集中包含的分层语义标签, 利用SCL损失构建了一个细粒度分类对比学习算法. 对遥感图像数据而言, 同一个地理位置的图像语义信息几乎不随时间的变化而变化. 基于该特点, Ayush等[39]设计了一个针对遥感图像的对比学习方法, 该方法将同一地理位置不同时间的两幅图像作为对比学习中的正样本对. 基于MoCo架构, 该方法将图像定位的代理任务添加到其中一个分支的特征提取网络之后, 辅助模型训练, 从而提高了下游任务的预测性能.

卢绍帅等[87]将监督对比学习应用到了文本数据的情感分类研究中. 在弱监督预训练阶段, 采用三元组损失预训练模型; 随后, 在下游的分类器训练阶段, 采用SCL损失和交叉熵损失联合优化网络, 获得更好的分类结果.

在序列数据的分类中, 也涌现出了一些基于对比学习的算法. 李巍华等[88]将MoCo方法迁移到故障信号诊断研究领域中, 首先, 通过对信号进行无监督对比学习预训练, 获得良好的特征提取网络. 然后, 再进行分类网络训练, 解决了信号故障诊断问题.

自监督对比学习的训练通常分为两个独立的阶段, 即特征提取网络训练和分类器训练. 在分类器训练阶段, 有是否冻结特征提取网络参数的两种选择. Wang等[77]认为这种两阶段的学习方式会损害特征提取网络和分类器的兼容性, 因此提出一个混合框架进行特征提取和分类器的联合学习. 该方法的对比学习部分采用同步对称网络架构, 并在特征提取网络后面加入一个分类器. 在

训练过程中, 通过一个平滑因子来调整两个损失的权重, 使得对比学习在训练开始时起主导作用, 随着训练时间的推移, 分类器学习过程逐渐主导训练.

2.6.2 分割任务

分割任务指的是对图像的语义分割、实例分割, 视频中的动作分割等任务. 图像分割任务关注像素级的分类, 因此在此类任务中, 特征提取网络能否学习到良好的局部特征至关重要. Wang等[89]为了更好地学习到图像的局部空间特征, 提出密集对比学习算法 (Dense contrastive learning, DenseCL). 该方法提出全局对比学习框架和局部对比学习框架, 每个框架均采用同步对称网络架构, 两个框架共享同一个特征提取网络. 其中, 局部对比学习框架对卷积得到的特征取消拉平操作, 从而保留特征的空间信息, 使得学习到的特征提取网络更适合于分割任务. 在医学图像分割领域, 由于数据的标注过程非常依赖专家知识, 获取大量的有标注数据代价十分高昂, 因此, 如何利用大量的无标签医学数据训练图像分割模型是一个很重要的研究问题. Chaitanya等[57]将对比学习的思想应用到该领域, 提出基于自编码器框架的全局-局部对比学习网络, 在该方法中, 全局对比学习目标是学习图像的全局语义信息, 局部对比学习目标是学习局部特征信息. 全局网络和局部网络共享同一个编码器. 医学图像中有一个“卷” (Volume) 的概念, 对于全局网络, 正样本对来自于同一幅图像的不同卷. 对于局部网络, 正样本对来自同一幅图像编码后特征的同一个空间位置. 两个分支均采用InfoNCE进行损失计算. 康健等[90]采用监督对比学习方法解决高分辨率SAR图像的分割问题, 通过改进的SCL损失提高同类建筑像素特征之间的相关性, 最终提高模型对建筑物的分割精度. Wang等[91]在Mask R-CNN[92]框架中加入对比学习模块, 提高了像素级特征的可分辨能力, 获得更好的图像分割结果.

在视频动作分割问题中, Kumar等[41]提出基于对比学习的无监督视频动作分割方法. 该方法对SwAV算法进行改进, 且不需要图像增广, 利用“视频数据的相邻帧为同类样本”这一假设, 将相邻帧的图像作为正样本对进行对比学习, 完成视频动作分割任务.

2.6.3 视频及关系数据预测任务

在视频预测问题中, 研究者使用密集预测编码 (Dense predictive coding, DPC)[53], 预测视频未来帧的信息. Han等[42]将DPC与存储库思想结合起来, 提出存储增强密集预测编码方法, 该方法将视频的特征保存到存储库模块中, 并设计了存储库寻址机制. 通过该存储库模块的设计, 网络在训练过程中能够考虑更长时间段的特征, 使预测结果更好. 此外, 为了更好地捕捉到视频中的重要信息, Zhang等[93]提出对编码视频同时进行帧间以及帧内的对比学习, Han等[53]对不同视角的特征进行对比学习.

Bahri等[94]将对比学习方法应用到了关系型数据预测任务中. 为了构建训练所需的正样本对, 作者提出了一种面向关系型数据的增广方法. 该方法受启发于关系型数据中同一维度 (属性) 下的信息语义相同的特点, 先在输入样本的对应维度上随机抹除一部分数据, 然后, 从其他样本的相同维度的数据中随机抽取信息填充到当前输入样本被抹除的位置中, 最后, 基于同步对称架构的网络进行对比学习. 训练得到的模型可用于预测关系型数据中丢失区域的信息.

2.6.4 长尾识别任务

粗略地说, 长尾数据是指尾部类众多的不均衡数据, 而长尾学习的主要研究目标是提升尾部类的识别正确率. 最近几年, 研究者们开始尝试将对比学习的思想和技术应用到长尾学习任务中.

文献[95]提出K-正样本对比损失 (K-positive contrastive loss, KCL), 将对比学习与长尾识别任务结合起来. 具体而言, 在长尾学习的特征学习阶段, KCL使用对比学习方法, 但每个训练样本仅随机选取个同类样本. 而在后续的分类器训练阶段, 仍采用传统的交叉熵损失, 但使用类均衡采样, 以平衡不同类的样本量、提升少数类的分类准确度. 文献[96]提出目标监督对比学习 (Targeted supervised contrastive learning, TSC) 方法, 将监督对比学习用于长尾识别的任务中. 该方法首先在特征空间中设定均匀分布的聚类中心, 然后, 在KCL损失的基础上, 增加样本到其聚类中心之间的距离计算的损失项, 以将样本逼近其聚类中心, 使得不同类别之间的分类界限更加清晰.

文献[97]提出平衡对比学习损失 (Balanced contrastive learning, BCL), 用于长尾识别. 该方法将类别中心加入到对比学习计算中, 并求批中每个类别的样本的梯度的平均值, 以减少多数类样本在梯度方面的影响.

2.6.5 其他任务

对比学习除了在分类、分割、预测这些任务中有广泛的应用以外, 在多模态学习任务中, 也有重要作用. 多模态学习任务通常包含两大类子问题, 即模态内的特征学习问题和模态间特征对齐问题.

Yang等[81]提出了一种视觉-语言跨模态对比学习方法. 该方法提出将对比学习方法应用于模态内特征提取网络的训练, 使各模态的特征提取网络能力更强, 解决了跨模态学习过程中模态内的特征学习问题. Dong等[82]将对比学习运用在包含5个模态数据的模型训练中. 针对模态内特征学习问题, 该方法采用掩码恢复任务作为模态内模型的代理任务. 然后, 针对不同模态特征之间的对齐问题, 设计模态间对比学习模块, 利用模态间的对齐得分矩阵衡量不同模态间信息的相似度, 进行更好的对比学习. Afham等[98]将对比学习算法引入到3D点云表示学习中. 该模型采用两个并行的对比学习方法, 其中一个对比学习方法基于SimCLR网络结构, 在点云数据内部进行对比学习, 另一个对比学习方法引入二维图像数据, 将点云数据与对应的二维图像数据进行跨模态的对比学习.

Laskin等[99]将对比学习结合到强化学习中, 用于提高特征提取网络的能力. 该方法基于MoCo架构, 分支1的输出, 送入强化学习中; 分支1和分支2的输出, 组对送入对比学习中.

3. 综合对比分析

本节根据前文提出的对比学习归类方法, 对现有的方法进行归纳汇总和整体分析. 表2汇总了代表性的对比学习方法及其所属归类 (包括具体样本对构造、图像增广方法、网络架构类型和损失

函数).

文献	年份	会议/刊物	样本对构造	图像增广	网络架构	损失函数	数据标注
[7]	2019	arXiv	随机采样	图像变换	同步非对称	InfoNCE 类	无
[2]	2020	ICML	随机采样	图像变换	同步对称	InfoNCE 类	无
[3]	2020	CVPR	随机采样	图像变换	异步对称	InfoNCE 类	无
[4]	2020	NIPS	随机采样	图像变换	异步非对称	传统损失	无
[5]	2020	NIPS	随机采样	图像变换	聚类/同步对称	传统损失	无
[8]	2020	NIPS	随机采样	图像变换	同步对称	InfoNCE 类	有
[9]	2020	NIPS	随机采样	图像变换	同步对称	混合损失	部分
[33]	2020	NIPS	困难样本构造	图像变换	异步对称	InfoNCE 类	无
[36]	2020	NIPS	剔除假负样本	图像变换	同步对称	InfoNCE 类	无
[37]	2020	arXiv	正样本扩充	图像变换	异步对称	InfoNCE 类	无
[42]	2020	NIPS	正样本扩充	图像变换	同步对称	InfoNCE 类	无
[45]	2020	ECCV	构造多视图	图像变换	同步对称	InfoNCE 类	无
[52]	2020	NIPS	随机采样	语义增广	同步对称	InfoNCE 类	无
[55]	2020	CVPR	随机采样	图像变换	同步非对称	InfoNCE 类	无
[57]	2020	NIPS	随机采样	图像变换	同步非对称	InfoNCE 类	无
[70]	2020	arXiv	随机采样	图像变换	异步对称	InfoNCE 类	无
[109]	2020	ECCV	随机采样	图像变换	同步非对称	InfoNCE 类	无
[6]	2021	CVPR	随机采样	图像变换	异步非对称	传统损失	无
[32]	2021	ICCV	困难样本构造	图像变换	异步对称	InfoNCE 类	无
[34]	2021	CVPR	困难样本构造	图像变换	同步对称	混合损失	部分
[39]	2021	ICCV	正样本扩充	图像变换	异步对称	混合损失	无
[40]	2021	CVPR	正样本扩充	图像变换	同步对称	InfoNCE 类	无
[46]	2021	CVPRW	构造多视图	图像变换	同步对称	混合损失	无
[61]	2021	ICCV	随机采样	图像变换	异步非对称	InfoNCE 类	无
[63]	2021	ICLR	随机采样	图像变换	聚类/异步对称	InfoNCE 类	无
[64]	2021	CVPR	随机采样	图像变换	聚类架构	InfoNCE 类	无
[66]	2021	AAAI	随机采样	图像变换	聚类/同步对称	InfoNCE 类	无
[77]	2021	CVPR	随机采样	图像变换	同步对称	混合损失	有
[79]	2021	ICCV	随机采样	图像变换	同步对称	混合损失	部分
[83]	2021	TGRS	随机采样	图像变换	同步对称	混合损失	部分
[78]	2021	TGRS	随机采样	图像变换	同步对称	InfoNCE 类	无
[85]	2021	CVPR	随机采样	图像变换	同步对称	InfoNCE 类	无
[35]	2022	WACV	剔除假负样本	图像变换	同步对称	InfoNCE 类	无
[41]	2022	CVPR	正样本扩充	图像变换	同步非对称	混合损失	无
[43]	2022	ICLR	正样本扩充	图像变换	异步对称	混合损失	部分
[44]	2022	CVPR	正样本扩充	图像变换	同步对称	混合损失	部分
[47]	2022	CVPR	随机采样	图像变换	任意架构	InfoNCE 类	无
[48]	2022	CVPR	随机采样	图像合成	异步对称	InfoNCE 类	无
[54]	2022	TAI	随机采样	图像变换	同步非对称	InfoNCE 类	无
[65]	2022	CVPR	剔除假负样本	图像变换	聚类/异步对称	InfoNCE 类	无
[72]	2022	ICLR	随机采样	图像变换	同步对称	InfoNCE 类	无
[76]	2022	AAAI	随机采样	图像变换	同步对称	传统损失	无
[94]	2022	ICLR	随机采样	图像变换	同步对称	InfoNCE 类	无
[98]	2022	CVPR	随机采样	图像变换	同步对称	InfoNCE 类	无
[103]	2022	ICLR	随机采样	图像变换	同步对称	混合损失	无

表2 对比学习方法整体归类分析

4. 未来发展方向

本文在对目前的对比学习论文进行归纳和总结后,认为该研究领域还存在许多可以探索的问题,同时存在一些可以与其他领域互相借鉴和发展的方向,具有广泛的研究前景,以下是对该领域发展的展望:

1. 对比学习中样本对的选择方法仍存在发展空间,在训练过程中剔除假负样本以及选择合适的正样本对能够有效地提高特征学习网络的学习效果.因此如何更加合理地剔除假负样本和选择正样本对是一个值得研究的关键问题.
2. 解决对比学习训练过程中的一致性与均匀性矛盾是一个十分重要的问题,如果该问题得到解决,能在很大程度上提高特征提取网络在下游任务上的泛化能力.
3. 主动学习是一种通过最少的标注样本获得最好的训练效果的学习技术[106].在深度主动学习领域,网络模型需要首先在一个含有标签的数据集 L_0 上进行预训练,然后通过查询策略从无标签数据集 U 中筛选最有用的样本给专家进行标注,最后更新当前训练的有标签数据集 L ,采用 L 的数据继续训练网络.重复以上过程直到标注预算耗尽或触发停止策略[106].对比学习是一种良好的模型预训练方法,可以自发的通过无标签数据或少量标签数据训练出特征提取模型,因此可以将对比学习算法引入到主动学习的网络模型预训练过程中,或作为辅助主动学习挑选待标注样本的方法.
4. 对比学习和无监督域自适应[107-108]的结合.在无监督域自适应问题中,源域数据存在标签,目标域数据不存在标签,源域数据和目标域数据分布相近或相同,且拥有相同的任务[107],如何将源域数据和目标域数据一同训练,使得模型能够在目标域上获得良好的效果是无监督域自适应的核心问题.在无监督域自适应研究中,源域数据和目标域数据可以通过自监督训练方法联合训练模型,对比学习就是一种先进的自监督训练算法,因此如何将对比学习方法与无监督域自适应方法进行有效结合是一个值得研究的问题.
5. 目前对比学习主要的下游应用是分类任务,如何设计更多的对比学习方法应用到检测、追踪等下游任务中,也将是未来的发展方向之一.

5. 结束语

对比学习是近年的研究热点.本文系统梳理了对比学习的研究现状,提出一种将现有方法划分为样本对构造层、图像增广层、网络架构层、损失函数层和应用层的归类方法,并从自监督对比学习算法入手,分析和归纳近四年主要的对比学习方法.而且,本文还全面对比了不同方法在各种下游任务中的性能表现,指出了对比学习现存的挑战,勾勒了其未来发展方向.对比学习研究作为一个快速发展的研究领域,在理论依据、模型设计、损失函数设计及与下游任务结合等方面还有较大的研究空间.

公众号后台回复“[对比学习综述](#)”获取最新对比学术综述PDF



极市平台

为计算机视觉开发者提供全流程算法开发训练平台，以及大咖技术分享、社区交流、竞...
848篇原创内容

公众号

极市干货

技术干货：损失函数技术总结及Pytorch使用示例 | 深度学习有哪些trick？ | 目标检测正负样本区分策略和平衡策略总结

实操教程：GPU多卡并行训练总结（以pytorch为例） | CUDA WarpReduce 学习笔记 | 卷积神经网络压缩方法总结

极市原创作者激励计划

极市平台深耕CV开发者领域近5年，拥有一大批优质CV开发者受众，覆盖微信、知乎、B站、微博等多个渠道。通过极市平台，您的文章的观点和看法能分享至更多CV开发者，既能体现文章的价值，又能让文章在视觉圈内得到更大程度上的推广，并且极市还将给予优质的作者可观的稿酬！

我们欢迎领域内的各位来进行投稿或者是宣传自己/团队的工作，让知识成为最为流通的干货！

对于优质内容开发者，极市可推荐至国内优秀出版社合作出书，同时为开发者引荐行业大牛，组织个人分享交流会，推荐名企就业机会等。

- 投稿须知：
- 1.作者保证投稿作品为自己的原创作品。
 - 2.极市平台尊重原作者署名权，并支付相应稿费。文章发布后，版权仍属于原作者。
 - 3.原作者可以将文章发在其他平台的个人账号，但需要在文章顶部标明首发于极市平台

投稿方式：

添加小编微信Fengcall（微信号：fengcall19），备注：姓名-投稿

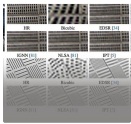
// 点击阅读原文进入CV社区
收获更多技术干货

阅读原文

喜欢此内容的人还喜欢

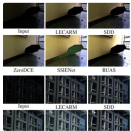
ICCV 2023 | 南开程明明团队提出适用于SR任务的新颖注意力机制（已开源）

极市平台



ICCV23 | 将隐式神经表征用于低光增强，北大张健团队提出NeRCo

极市平台



YOLOv5帮助母猪产仔？南京农业大学研发母猪产仔检测模型并部署到Jetson Nano开发板

极市平台

