

基于对比学习的小样本金属表面损伤分类

吴冠荣, 李元祥, 王艺霖, 陆雨寒, 陈秀华

(上海交通大学航空航天学院, 上海 200240)

摘要: 现有小样本分类方法局限于从每轮支持信息中归纳出类内共性, 忽略了在迭代过程中类间关联性以及样本本身携带的类别信息。由于金属损伤纹理细微、多变, 因此所形成的特征分布类间距离小、类内距离大。因特征分布聚合性差导致小样本分类性能降低且新类泛化性变差, 提出一种基于内外双层训练模型架构的小样本金属表面损伤分类方法。内模型在利用度量手段完成元分类任务的同时, 引入双模态特征作为外模型特征空间的信号, 即在新映射空间下利用类别标签信息有监督地对比不同类别的图像特征、优化特征分布, 使类间区分度更大、类内聚合度更高。在训练阶段中外模型反传对比损失, 间接加强原有特征空间的表征能力, 从而提高内模型的度量水平, 提升分类精度。同时, 利用类别嵌入作为动态类别中心, 可以有效减少小样本问题中的噪声干扰, 加强模型泛化性能。在 GC10、NEU 及 APSD 3 个常用的金属损伤数据集上的实验结果表明, 相比 ProtoNet、MatchingNet、RelationNet 等主流方法, 该方法具有较优的分类精度, 特别是新类别的泛化能力得到大幅提升, 5-way 5-shot 设定下分类精度至少提高了 5.24、1.39 和 6.37 个百分点, 分类错误下降率分别为 36.00%、17.94% 和 66.15%; 此外, 新类分类精度分别从 36.53%、82.43%、31.89% 提升至 69.12%、91.57%、48.23%。5-way 1-shot 设定下分类精度分别至少提高 8.34、3.01 和 4.61 个百分点, 分类错误下降率分别为 28.32%、23.37% 和 46.57%。

关键词: 金属表面损伤; 对比学习; 度量学习; 元学习; 小样本分类

源代码链接: <https://github.com/wgr573echo/mentalDefect.git>

中图分类号: TP391.1

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0067599

Few-Shot Metal Surface Defect Classification Based on Contrastive Learning

WU Guanrong, LI Yuanxiang, WANG Yilin, LU Yuhua, CHEN Xiuhua

(School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China)

[Abstract] The existing few-shot classification methods are limited to inducing intra class commonalities from each round of support information, ignoring inter class correlations and category information carried by the samples themselves during the iteration process. Due to the fine and varied texture of metal damage, the resulting feature distribution has small inter class distance and large intra class distance. A few-shot metal surface damage classification method based on an inner and outer two-layer training model architecture is proposed, as the poor aggregation of feature distribution leads to a decrease in the performance of few-shot classification and a decrease in the generalization of new classes. The inner model uses metric methods to complete the metal classification task, while the outer model incorporates bimodal features as signals in the feature space. In the new mapping space, category label information is used to supervise the comparison of image features from different categories and optimize the feature distribution, resulting in improved inter-class discrimination and intra-class aggregation. During the training phase, the external model enhances the representation ability of the original space through backpropagation contrastive loss, thereby enhancing the measurement level of the internal model and improving classification accuracy. Additionally, the use of category embedding as a dynamic category center effectively reduces noise interference in small sample problems and enhances model generalization performance. Experimental results on three commonly used metal damage datasets, GC10, NEU, and APSD, demonstrate that the proposed method achieves superior classification accuracy compared to mainstream methods such as ProtoNet, MatchingNet, and RelationNet. In particular, the generalization ability of new categories is significantly improved. Under the 5-way 5-shot setting, the classification accuracy is by at least 5.24, 1.39, and 6.37 percentage points, with classification error reduction rates of 36.00%, 17.94%, and 66.15%, respectively. Specifically, the accuracy of new class classification increases from 36.53%, 82.43%, and 31.89% to 69.12%, 91.57%, and 48.23%, respectively. Under the 5-way 1-shot setting, the classification accuracy is improved by at least 8.34, 3.01, and 4.61 percentage points, with classification error reduction rates of 28.32%, 23.37%, and 46.57%, respectively.

收稿日期: 2023-05-10 修回日期: 2023-06-25

通信作者 E-mail: wgr_echo@163.com

[Key words] metal surface defect; contrastive learning; metric learning; meta-learning; few-shot classification

0 引言

对金属表面存在的损伤进行分类是金属类工业产品质量检验和反馈工作中不可或缺的一环,通常需要有专业知识的人借助电镜等器材进行判断^[1]。但是损伤类型多、视觉疲劳、工业流水线的压力等因素容易影响主观判断,人工分类的准确率忽高忽低。随着机器视觉技术的发展,神经网络对视觉特征的强大学习能力可以弥补人工分类的不足。

基于机器学习的金属损伤分类按照是否使用初始化参数,分为迁移学习方法和深度学习方法。迁移学习方法希望从 ImageNet 等大型训练集中学习到最通用的特征,在目标数据集上进行微调^[2]。而深度学习方法通过调整模型参数来拟合当前数据的特征分布^[3],或者通过加深模型层数,提高模型对多维特征的表达能^[4]。但金属损伤的纹理特征和常见的大型数据集差异很大,公开的预训练模型迁移效果有限,并且金属损伤种类繁多,一些偶发性、新产生的损伤不满足神经网络对训练数据量的需求,若直接使用神经网络进行有监督训练,模型易于过拟合、泛化性能差。因此,在很多应用场景下需要考虑小样本问题。

较早的小样本问题解决方案通常是学习合成伪数据^[5]。这类方法依赖于扩充样本的质量,如果合成器效果一般,则容易引入噪声数据,效果有时不升反降^[6]。目前,小样本分类研究聚焦于如何学习1个更好的特征表示,使下游任务中即使只有少量标注信息,分类模型也能表现良好^[7]。常见方法包括元学习框架^[8-10]、度量学习^[11-12]、知识蒸馏^[13-14]等,都是为了找到更好构造特征空间的方式。ProtoNet^[8]使用度量手段进行原型学习,由于其具有一定的灵活性和高效性,因此仍广泛应用在目前算法中。文献[9]在通用的两阶段微调训练机制基础上,增加了一步预训练,利用 DINO 机制在额外的无标签数据上训练主干网络的特征提取能力,再在 ProtoNet 上训练模型,实现了小样本分类任务的 SOTA(State-Of-The-Art)效果。

文献[15]首先将小样本解决方案引入金属损伤分类,利用生成对抗网络(GAN)生成大量无标注的钢材表面损伤样本,在扩充样本量后,再用残差网络进行分类。文献[16]利用 Wasserstein 距离的生成对抗网络(WGAN)有效扩充钢板气泡损伤,提高检测精度。然而,生成伪样本的方法容易模式固化,尤其对于金属表面损伤问题,金属材料种类多,损伤种类更多,这些方法仅扩充某种材料或某类损伤的数据量,因此算法的适用性不高。文献[17]针对铝材常见损伤类别各提取一种特征图 CCMs,并利用注意力模块生成小样本类别的建议特征图(RCMs),对 CCMs 和 RCMs 进行映射和转化,分类结果由 CCMs 和 RCMs 共同决定。该方法基于原型学习思路,将

类内共性作为区分标准,预测时衡量当前样本与各个类别原型特征的相似度,最相似的类别作为分类结果,构造原型的范式在一定程度上缓解了模型泛化性差的问题。但是,金属损伤纹理细微、多变决定了特征分布的类间距离小、类内距离大,现有方法局限于在弱区分度的特征分布基础上根据类内共性完成分类,导致少样本、新类别分类精度低。

为此,本文提出一种基于对比学习的小样本金属表面损伤分类方法 CLFS。该方法在元学习机制下,以外层模型对比、内层模型度量作为元任务,引入类别标签嵌入作为外模型对比不同类别图像特征的监督信号,使得在新的特征空间下,类内特征距离近,类间特征距离远,视觉特征相似的类别中心距离近。在金属损伤分类任务中,外模型对比学习代理任务可以间接使得原有特征空间中不同损伤类别之间的差异更加明显,从而增强了内模型的度量学习能力,有效提高分类精度;引入类别信息作为监督信号可以减少对比学习过程中样本噪声的影响,提高小样本原型向量估计的鲁棒性,从而提升对新类别的泛化能力。

1 小样本学习描述

关于小样本问题的任务设定,本文采用最常见的目标因子分析(TFA)数据划分方法(TFA-split)和全类别测试方案^[18]。TFA-split将数据集分为训练集和测试集:训练集选取具有较高资源的基础类别(C_{base}),共 M_b 类;测试集包含训练集中未出现的新类别(C_{novel}),新类别是低资源的小样本问题,新类别数共 M_n 类。TFA-split在迭代过程中保持训练集和测试集固定不变。不同于仅测试 C_{novel} 的狭义小样本问题,本文同时测试模型对 C_{base} 和 C_{novel} 的分类性能。

本文模型的训练采用元学习模式。元学习模式在训练过程中将传统的单个循环多批次的训练任务拆分成数个元任务。每个元任务目标一致、构成一致,从所有类别中选取 K 个类别,每个类别各选 N 个样本作为元任务输入。1个元任务又称 K -way N -shot 任务,1个元任务的工作周期为1个 episode。在训练过程中多次迭代元任务使得模型满足最终的小样本测试任务要求。元学习模式实现了从以数据为中心向以任务为中心的转变,从而减少对样本量的依赖,并提高小样本学习的性能。

本文根据资源高低将1个金属损伤样本集按类别区分为 C_{base} 和 C_{novel} ,并将 C_{base} 样本划分出训练集 D_{train} 和测试子集 D_{test}^1 , C_{novel} 样本作为测试子集 D_{test}^2 ,则测试集样本为 $D_{test} = D_{test}^1 \cup D_{test}^2$ 。由于所有样本需要被统一为元任务的输入格式,因此按照 K -way N -shot 格式重构 D_{train} 和 D_{test} 。

在训练阶段,每个元任务先从 C_{base} 中抽取 K ($K \leq M_b$)个类别,针对抽出来的每个类别对应从 D_{train} 中抽取 N 张图片作为训练样本,称这 $K \times N$ 张样本为1个

元任务的支持集 S 。对同样的 K 个类别再从 D_{train} 取 N 张不同的图片作为验证样本,称为 1 个元任务的查询集 Q 。 S 和 Q 的类别相同、样本不同,合在一起作为 1 个元任务的输入样本。元学习过程是指在每次元任务中利用 S 的信息对 Q 进行分类,从而更新模型参数。 D_{train} 被重构成 m 个元任务,即 $D_{\text{train}} = \sum_{i=1}^m (S_i, Q_i)$ 。本文采用 2 种常用的 K, N 设定,即 5-way 5-shot 和 5-way 1-shot。

在测试阶段,将 D_{test} 按照 K -way N -shot ($K \leq M_b + M_n$) 格式重构成多个元任务,模型在多个元任务上的平均分类性能即最后性能。

2 网络架构

虽然元学习模式能够让模型快速适应新的任务,但是因金属损伤的类内方差大、类间相似度大等特性,使得很多小样本学习方法的效果明显下降。为提高

模型的泛化能力,本文提出引入类别标签信息作为对比学习的监督信号。利用对比损失函数惩罚不同类别特征之间的距离,能有效地划分外模型特征空间的类间区域,间接优化内模型的特征分布,使得内模型在分类时可以不受到无规律特征分布的影响。

2.1 训练模型

CLFS 分类模型在训练阶段分为内模型和外模型,内模型负责图像分类,外模型负责优化特征分布的代理任务,内外模型联合训练。图 1 所示为在 K -way N -shot 设定下 CLFS 的网络架构,原始样本经过预处理分为图像输入和类别标签嵌入作为整个训练模型的输入信号。图 1 中 Query set 即所需训练的样本编码得到的向量集合。内模型主要包括利用 ResNet 18 的特征提取模块 F 和原型向量(包括支持向量和查询向量)生成模块。外模型除了与内模型共用的特征提取模块 F 之外,主要由类别标签嵌入、特征转译(特征编码和类别标签编码)和对比分支 3 个部分组成。

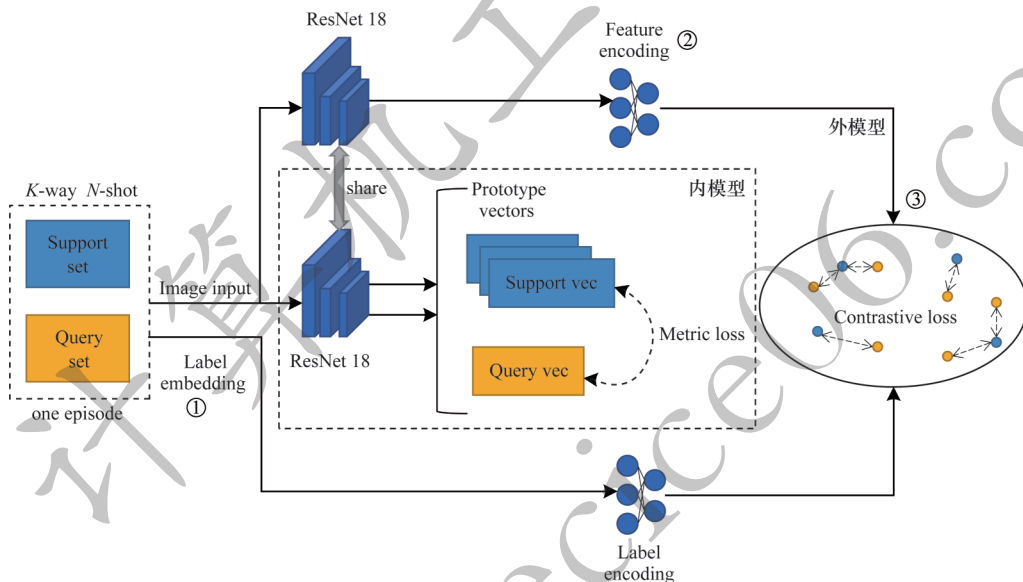


图 1 在训练阶段 K -way N -shot 设定下 CLFS 的网络结构

Fig.1 Structure of CLFS network with the K -way N -shot setting in the training phase

2.1.1 内模型

本文使用 ResNet 18 作为特征提取模块,ResNet 18 由 4 层残差模块构成,选择特征层 8 倍下采样结果作为 F 的输出向量。支持集样本和训练集样本分别经过 F 映射得到支持向量和查询向量,以计算原型向量集,并采取度量学习的方式完成分类任务^[8]。

元任务的内模型工作流程的输入为 $X: K \times N$ 张支持集图片; $X': K \times N$ 张查询集图片,输出为 K^* 查询集分类结果。工作流程如下:

1) 第 k 类的 N 张支持集图片 X_k 经过特征提取模块 F 的投影后,形成支持向量集 $S_k = F(X_k)$;

2) 元任务的所有类别支持向量集为 $S = \bigcup_{k=1}^K S_k$;

3) 查询集图片 X' 经过特征提取模块 F 的投影后,形成查询向量集 $Q = \bigcup_{i=1}^{K \times N} q_i$;

4) 根据 S 计算每个类别的原型向量 $c_k = \frac{1}{N} \times S_k$;

5) 计算查询向量和每个原型向量之间的欧氏距离 $D_k = \|Q - c_k\|_2$;

6) 第 i 个查询向量 q_i 距离最近的原型向量所代表的类别为 q_i 对应查询图片分类结果 $k_i^* = \operatorname{argmin}_k D_k$;

7) 输出 $K^* = \bigcup_{i=1}^{K \times N} k_i^*$ 。

2.1.2 外模型

在训练阶段增加的外模型中,利用类别标签嵌入信息作为确定的类别中心,减少对对比学习过程中样本噪声的影响,从而间接提高内模型原型向量估计的鲁棒性。

1) 类别标签嵌入。CLFS 所用的 3 个数据集标

签如表 1 所示。本文的类别标签嵌入流程:当前数据集的所有类别标签经过独热编码后,按照类别顺

序存储为 Hash 表,在每次元任务中对用到的类别进行 Hash 映射,得到初始类别嵌入向量。

表 1 数据集标签信息

Table 1 Label information of datasets

数据集	类别
GC10 ^[19]	crease,rolled pit,crescent gap,silk spot,inclusion,waist folding,oil spot,water spot,punching hole,welding line,共 10 类
NEU ^[20]	rolled_in scale,patches,crazing,pitted surface,inclusion,scratches,共 6 类
APSD ^[21]	non-conductive,bump,rubbing,convex powder,concave,crack,coating,orange peel,dirty spots,leaky bottom,共 10 类

2)特征转译。该部分包括 2 个模块:(1)特征编码器,通过特征提取模块得到的结果向量在进入对比空间前也须经过 1 层多层感知机(MLP)映射;(2)类别标签编码器,在预处理阶段嵌入的初始类别标签向量先经过 1 层可学习的 MLP 映射,再进行有监督对比学习。本文设计 2 个特征转译模块主要是为了解决在对比学习中不同模态之间的差异性,提升对比特征空间的一致性,同时,构建辅助特征空间使得对比学习过程不影响内模型的度量空间。此外,在特征转译模块提供的新特征空间下,对

比学习的动态性^[22]让类别标签嵌入和图像特征互相影响。类别标签嵌入的结果作为类别中心监督特征分布,通过反传损失给特征编码器、特征提取模块逐步优化分类边界。图像特征信息通过反传给可学习的类别标签编码器,类别中心的分布位置可以继承视觉特征相似性。类别标签监督不同类别之间的特征分布情况如图 2 所示。从图像纹理层面钢材油斑类别(oil_spot)和钢材夹杂物(inclusion)损伤类别相较于油斑和折痕(crease)更相似,则训练后油斑类别的特征分布更接近夹杂物而不是折痕。

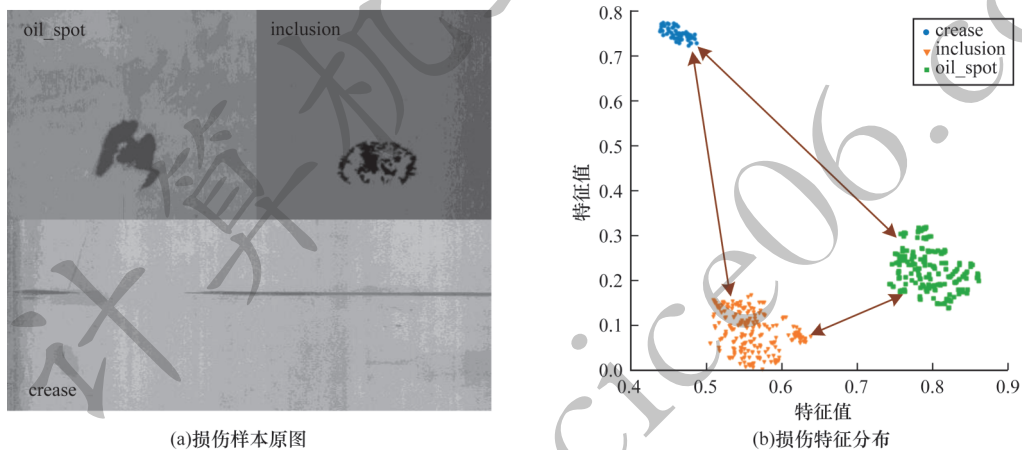


图 2 类别标签监督不同类别之间的特征分布情况

Fig.2 Features distribution among different categories in category labels supervise

3)对比分支。对比学习过程结合 2 个可学习的特征转译模块,在训练过程中拉近 2 个新映射的特征空间距离,可以有效减小不同模态特征之间的差异。经过相同结构的特征转译模块,类别标签和图像 2 种模态的特征被映射到同一空间,

主要目标是在新的特征空间下,利用离散类别标签先验信息对图像特征分布进行有监督对比学习,使得同类别的特征分布聚合性更好,不同类别的特征区分度更高。对比分支示意图如图 3 所示。

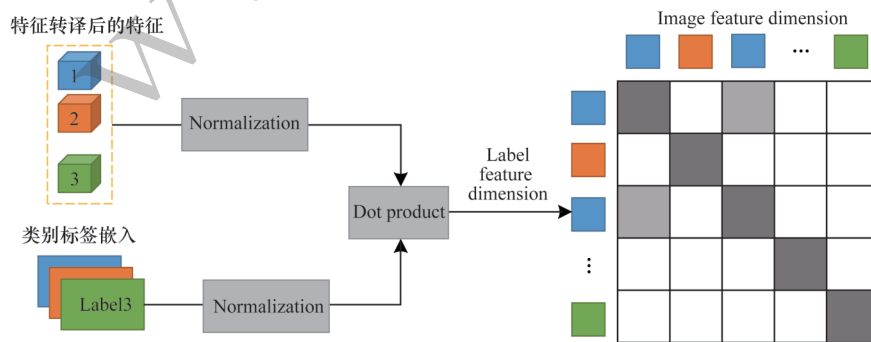


图 3 对比分支示意图

Fig.3 Schematic diagram of contrasting branches

在特征转译后的图像特征和类别标签嵌入经过归一化(Normalization)和点积(dot product)操作,得到余弦相似度矩阵。

2.1.3 损失函数

损失函数分为度量损失和对比损失,分别是内模型对小样本元任务分类结果的监督和外模型对特征分布情况的监督。

度量损失 L_{distance} 是指在1次元任务的训练过程中,携带真值的 $K \times N$ 张查询图片,经过特征提取形成查询向量集 $\mathbf{Q} = \bigcup_{i=1}^{K \times N} \mathbf{q}_i$ 。对于查询向量 \mathbf{q}_i ,计算其与对应真值类别原型 $\tilde{\mathbf{c}}_i$ 之间的 Softmax 负概率分布,即惩罚当前所得特征与真值类别的原型向量之间的距离^[8]:

$$L_{\text{distance}} = -\frac{1}{K \times N} \times \sum_{i=1}^{K \times N} \log_a \frac{\exp(-\|\tilde{\mathbf{c}}_i - \mathbf{q}_i\|_2)}{\sum_{k \neq i} \exp(-\|\mathbf{c}_k - \mathbf{q}_i\|_2)} \quad (1)$$

受 SimCLR^[22] 的启发,本文选取余弦相似度作为转译后的图片特征和类别标签特征之间的对比损失 L_{contrast} 。图3中最终构成的余弦相似度矩阵分别从标签特征维度和图像特征维度惩罚同类别之间的距离,从而间接增大类间距离。假设1个 episode 由 $2N$ 张样本组成,包括 N 张支持样本和 N 张查询样本,标准化后标签特征向量表示为 \mathbf{v}_j , 对应真值为 \hat{y}_j 。查询图片和支持图片所得到的标准化向量均表示为 \mathbf{w}_i , 对应真值为 \hat{y}_i , 1 表示当 {} 中条件为真时,则取值为 1, 反之为 0。则 1 个 episode 的对比损失表达式如下:

$$L_{\text{contrast}} = -\frac{1}{4N} \times \sum_i \sum_j \mathbf{1}_{\{\hat{y}_i = \hat{y}_j\}} \left(\log_a \frac{\exp(\mathbf{v}_j \cdot \mathbf{w}_i / \tau)}{\sum_{m \neq i} \exp(\mathbf{v}_j \cdot \mathbf{w}_m / \tau)} \right) \quad (2)$$

其中:温度系数 τ 用来调节对比学习中的分类细粒度, τ 越大,对不同样本间差异性的重视程度越高。

总的损失函数表达式如下:

$$L = \alpha \times L_{\text{distance}} + L_{\text{contrast}} \quad (3)$$

通过一定的权重系数 α 将内外模型的损失函数进行统一,外模型的转译层参数由外模型对比损失单独更新,其他模块的可学习参数由整体损失函数共同优化。

2.2 预测模型

经过训练后,多模态对比学习模块加强了内模型对不同类别特征分布的区分能力,元学习模式也使得内模型提前适应小样本任务。因此测试阶段只需要内模型进行前向推理,不需要外模型进行监督和辅助。整个测试推理过程与训练阶段内模型的1次前向传播步骤一致,1次 K -way N -shot 的全类别测试任务包括训练阶段未涉及的新类别 C_{novel} , 体现了 CLFS 小样本分类的泛化性能。预测模型架构如图4所示。预测模型为训练模型的内模型,最终的

分类结果 k^* 由距离查询向量最近的原型向量的所属类别决定。

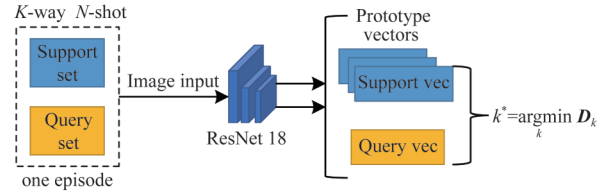


图4 预测模型架构

Fig.4 Framework of prediction model

3 实验

3.1 数据集与实验设置

本文实验选取金属表面损伤常用的3个数据集:GC10钢板表面损伤数据集^[19]、东北大学钢材损伤分类数据集(NEU)^[20]、天池铝型材料表面损伤数据集(APSD)^[21]。

GC10数据集包括10种类型的钢材表面缺陷,每个类别的样本数量不等,共3570张灰度图。由于从生产线统一采集,因此GC10样本尺寸均为2048×1000像素。NEU数据集由东北大学机器视觉与机器人团队发布,包括6类常见钢材表面损伤,每类有300张灰度图,尺寸固定为200×200像素。APSD数据集来自天池2018铝材表面瑕疵识别比赛,本文选取其中9个类别用于训练,每个类别的样本量从50到200不等,尺寸统一为256×256像素。本文实验将原始RGB图像统一转为灰度图。

为统一输入图像尺寸对骨干网络的影响,在预处理阶段将所有样本保持初始的长宽比压缩至150×150像素左右,再用背景像素填充成方阵。CLFS采用ResNet 18作为图像特征提取层,1层MLP作为转译层,所有网络使用Adam优化器,设定的初始学习率为0.001,每20轮减小1次学习率,设置对比损失中温度系数 $\tau=2$,度量损失占整体损失的权重 $\alpha=0.2$ 。

实验软件配置包括Python 3.8.10、PyTorch 1.7.0、OpenCV 4.6、Numpy 1.21,硬件平台为GPU-NVIDIA GeForce RTX 3090和CPU-Intel® Xeon® Platinum 8255C。

3.2 实验结果与分析

本文重复10组对比实验,取所有分类精度结果的标准差作为偏差范围。所选取的比较对象不仅包括传统的两阶段微调(TFA)^[23],而且包括MatchingNet^[24]、RelationNet^[25]、ProtoNet^[8]等主流的小样本分类方法。同时,在全类别测试的基础上,通过混淆矩阵比较新类的分类性能,并进行特征分布可视化分析。

3.2.1 分类精度对比

CLFS在5-way 5-shot和5-way 1-shot设定下,全类别测试的分类精度都得到了较大幅度的提升,如表2所示,加粗表示最优数据。

表 2 不同方法的分类精度比较(全类别测试)

Table 2 Classification accuracy comparison among different methods (full category test)

方法	骨干网络	GC10 数据集		NEU 数据集		APSD 数据集	
		5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
TFA ^[23]	Conv64	20.18±0.22	33.83±1.33	29.33±1.28	74.61±0.78	27.70±0.53	63.52±0.11
	ResNet 18	20.73±0.41	37.18±0.92	52.33±1.01	80.71±0.45	36.03±0.10	62.87±1.29
	WRN	32.33±0.57	58.33±2.52	33.46±0.57	76.10±0.22	40.47±1.08	61.04±2.16
ProtoNet ^[8]	Conv64	70.10±0.18	85.30±1.33	82.88±0.98	87.03±0.36	79.88±0.56	86.49±1.40
	ResNet 18	70.55±0.41	85.44±0.05	87.12±1.10	90.61±0.91	77.42±1.28	82.74±1.08
MatchingNet ^[24]	LSTM	66.13±0.51	83.04±0.13	86.50±0.10	92.25±0.63	74.68±1.56	89.23±0.12
RelationNet ^[25]	Conv64	65.16±0.76	81.98±0.18	81.49±0.11	85.50±0.53	90.10±0.08	90.37±0.37
CLFS	ResNet 18	78.89±0.22	90.68±0.56	90.13±0.35	93.64±0.61	94.71±0.29	96.74±0.53

在 5-way 5-shot 设定下,CLFS 的分类精度明显优于 TFA,而且 CLFS 比 ProtoNet、MatchingNet、RelationNet 3 个主流小样本方法的分类精度至少提高了 5.24、1.39 和 6.37 个百分点(与次优结果比),分类错误下降率分别为 36.00%、17.94% 和 66.15%。对于 5-way 1-shot 设定,CLFS 相较于 3 个主流小样本方法,分类精度提升效果更显著,分别至少提高 8.34、3.01 和 4.61 个百分点(与次优结果比),分类错误下降率分别为 28.32%、23.37% 和 46.57% 以上。

3.2.2 混淆矩阵分析

在 GC10 数据集 5-way 5-shot 下 ProtoNet 与 CLFS 分类混淆矩阵如图 5 所示。前 7 行(列)类别为

此次实验中设置的 C_{base} ,最后 3 行(列)为小样本的 C_{novel} ,混淆矩阵以行为单位包括真正例(TP)和假反例(FN)类结果,以列为单位包括 TP 和假正例(FP)类结果。从全类别测试结果中可以看出,对于 GC10 数据集,ProtoNet 不仅对 C_{novel} 的分类效果较差,而且对 C_{base} 漏报率和误报率也较高。CLFS 在同样情况下 C_{base} 的 FN 类、FP 类结果基本清零,即真值为 C_{base} 的类别基本分类正确。特别地,CLFS 有效提高模型对小样本 C_{novel} 的泛化性,相较于 ProtoNet, C_{novel} 分类精度由 36.53% 大幅提升至 69.12%。针对 NEU 数据集和 APSD 数据集, C_{novel} 分类精度分别由 82.43% 提升至 91.57%、31.89% 提升至 48.23%。

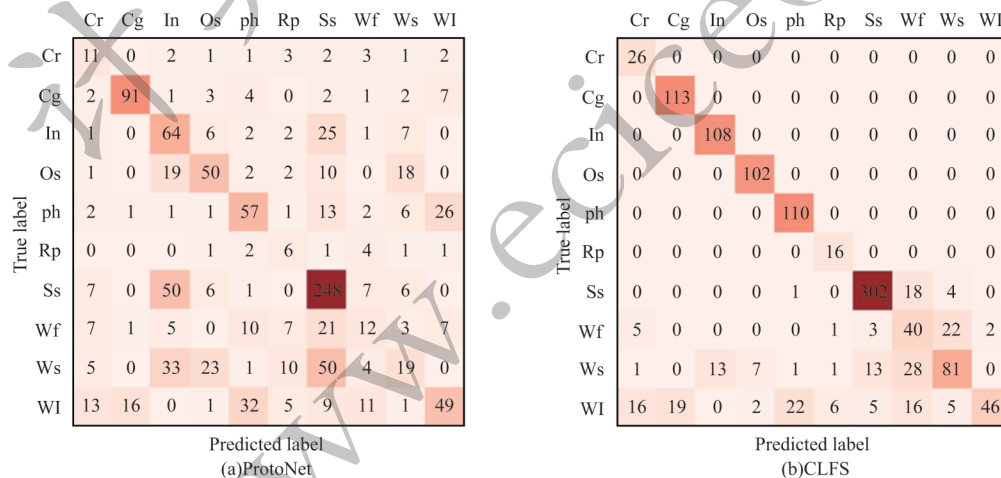


图 5 在 GC10 数据集上 5-way 5-shot 下 ProtoNet 与 CLFS 分类混淆矩阵

Fig.5 Classification confusion matrix for ProtoNet and CLFS in 5-way 5-shot on GC10 dataset

3.2.3 特征分布可视化

为动态监督特征分布,本文采取 t -SNE 对统一 Backbone 后的 ProtoNet 和 CLFS 的同一特征层进行降维映射。在 GC10 数据集 5-way 5-shot 下 ProtoNet 和 CLFS 的特征分布对比如图 6 所示(彩色效果见《计算机工程》官网 HTML 版)。从图 6(a)~图 6(c) 可以看出,ProtoNet 的特征分布中大多数类别的特征点互相穿插, C_{base} 和 C_{novel} 都没有区分度。图 6(d)~图 6(f)所示为同样条件下 CLFS 在测试过程中前向

预测得到的特征分布,类别之间的间距明显、类内聚合度也显著提高。特别地,将 C_{base} 和 C_{novel} 分开分析, C_{base} 的特征边界清晰,类别分布的相邻关系保留了视觉特征的相似性。由于测试阶段才引入 C_{novel} ,因此 CLFS 模型对于 C_{base} 和 C_{novel} 之间的区分度下降,但是与图 6(c)的 C_{novel} 分布对比可以看出,CLFS 有效提高了类内聚合性。实验结果表明,CLFS 利用外模型双模态对比学习,有效地解决了类间相似性较高的金属损伤分类问题。

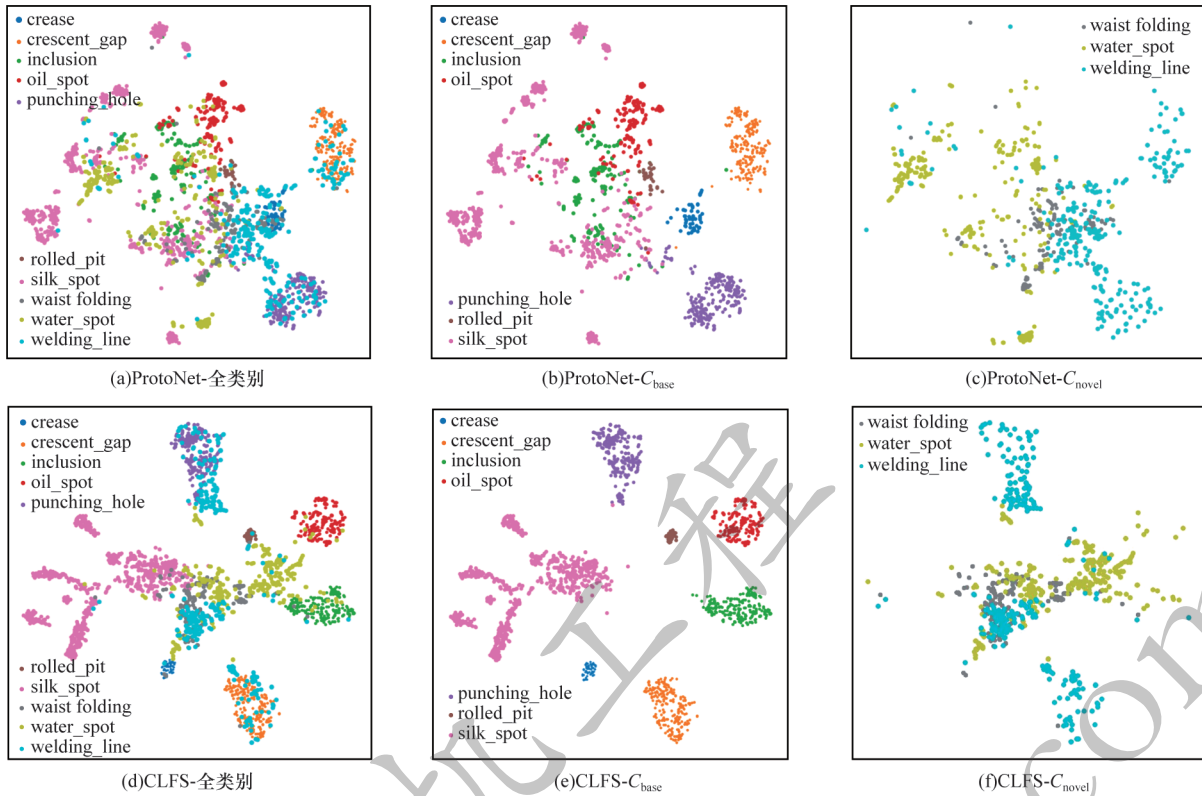


图 6 在 GC10 数据集 5-way 5-shot 下 ProtoNet 和 CLFS 的特征分布对比

Fig.6 Comparison of feature distribution between ProtoNet and CLFS under 5-way 5-shot for GC10 dataset

3.3 消融实验

为了分别验证对比学习和引入类别标签信息能否提升模型对小样本的分类能力,本文使用 ProtoNet 作为 baseline,拆分 CLFS 的外模型为 baseline+对比损失、baseline+标签嵌入,在 3 个数据集上完成 5-way 5-shot 的消融实验。

消融实验结果如表 3 所示。加入对比损失和引入类别标签嵌入的方法相较于 baseline 分别提升了 1%~2% 的分类精度,验证了对比损失项和类别标签嵌入的有效性。CLFS 所设计的外模型通过构造辅助特征空间,将对比损失和标签嵌入 2 种方法组合,对分类效果有更显著的提升,单个数据集的分类精度相较于 baseline 分别提升 5.24、3.03 和 14.00 个百分点,分类错误下降率分别为 36.00%、32.27% 和 81.11%。CLFS 的内外模型机制具有一定的有效性。

表 3 消融实验结果

Table 3 Results of ablation experiment %

方法	分类精度		
	GC10	NEU	APSD
baseline	85.44	90.61	82.74
baseline+对比损失	86.61	92.27	86.74
baseline+标签嵌入	85.79	91.03	84.43
CLFS	90.68	93.64	96.74

3.4 超参数分析

CLFS 涉及 2 个超参数 τ 和 α , GC10 数据集在 5-way 5-shot 设定下进行超参数分析,根据经验^[22]确

定超参数 τ 和 α 的取值范围分别为 $[0.1, 1, 2, 5]$ 和 $[0.1, 0.2, 0.5, 2]$ 。本文使用网格法对每组参数重复 3 次实验。超参数对模型性能的影响如表 4 所示,在合适的超参数范围内,分类性能对超参数的敏感度较低,超参数优化空间较为平缓,峰值处的误差范围更小,因此鲁棒性更好。从表 4 可以看出,当 $\tau=2$ 、 $\alpha=0.2$ 时,模型性能最佳。

表 4 超参数对模型性能的影响

Table 4 The impact of hyperparameters on model performance %

参数	$\alpha=0.1$	$\alpha=0.2$	$\alpha=0.5$	$\alpha=2$
$\tau=0.1$	86.20±1.02	86.81±0.53	87.12±0.66	87.43±0.59
$\tau=1$	88.17±0.40	88.68±0.21	88.88±0.35	88.70±0.44
$\tau=2$	87.35±0.31	90.68±0.11	89.44±0.18	89.16±0.23
$\tau=5$	88.31±0.43	88.42±0.14	89.17±0.49	89.08±0.42

4 结束语

针对小样本金属损伤特征分布聚合性低、新损伤类别泛化性差的问题,本文通过设计内外模型机制,提出一种基于对比学习的小样本金属表面损伤分类方法。内模型以传统度量学习为目标,完成小样本分类的任务;外模型构造辅助特征空间,在该空间下利用类别标签嵌入信息,动态对比不同类别的特征分布,从而实现聚合同类别特征,得到明确清晰的分类边界,提高模型对 base 类别的分类精度,而且在较大程度上提升对 novel 类别的泛化能力。在

3个常见的金属损伤数据集上验证本文方法的有效性,特征分布的可视化进一步展现了本文方法在增加类间距离、聚合类内特征的任务上具有显著的提升效果。目前,小样本金属损伤分类研究尚处于探索阶段,可参考的数据集和实验结果有限。本文仅初步探索了对比学习结合多模态优化特征分布在小样本分类算法方面的可行性,下一步将完善标签嵌入模型,便于从语义相似性角度补充小样本金属损伤的先验信息,进一步优化分类性能,同时使得结果具有可解释性。

参考文献

- [1] DUAN G F, SONG Y G, LIU Z Y, et al. Cross-domain few-shot defect recognition for metal surfaces[J]. *Measurement Science and Technology*, 2022, 34(1): 015202.
- [2] HE Y, SONG K C, MENG Q G, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features[J]. *IEEE Transactions on Instrumentation and Measurement*, 2019, 69(4): 1493-1504.
- [3] SAMSUDIN S S, AROF H, HARSUN S W, et al. Steel surface defect classification using multi-resolution empirical mode decomposition and LBP[J]. *Measurement Science and Technology*, 2020, 32(1): 015601.
- [4] 刘欢, 刘晓佳, 王宇斐, 等. 基于复合卷积层神经网络结构的焊缝缺陷分类技术[J]. *航空学报*, 2022, 43(S1): 165-172.
LIU H, LIU X J, WANG Y F, et al. Weld defect classification technology based on compound convolution neural network structure [J]. *Acta Aeronautica et Astronautica Sinica*, 2022, 43(S1): 165-172. (in Chinese)
- [5] 唐泽宇, 邹小虎, 李鹏飞, 等. 基于迁移学习的小样本OFDM目标增强识别方法[J]. *上海交通大学学报*, 2022, 56(12): 1666-1674.
TANG Z Y, ZOU X H, LI P F, et al. A few-shots OFDM target augmented identification method based on transfer learning[J]. *Journal of Shanghai Jiao Tong University*, 2022, 56(12): 1666-1674. (in Chinese)
- [6] 葛轶洲, 刘恒, 王言, 等. 小样本困境下的深度学习图像识别综述[J]. *软件学报*, 2022, 33(1): 193-210.
GE Y Z, LIU H, WANG Y, et al. Survey on deep learning image recognition in dilemma of small samples[J]. *Journal of Software*, 2022, 33(1): 193-210. (in Chinese)
- [7] LUO X, XU J, XU Z. Channel importance matters in few-shot image classification[C]//*Proceedings of International Conference on Machine Learning*. Stroudsburg, USA: Association for Computational Linguistics, 2022: 14542-14559.
- [8] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. New York, USA: ACM Press, 2017: 4080-4090.
- [9] 张玲玲, 陈一苇, 吴文俊, 等. 基于对比约束的可解释小样本学习[J]. *计算机研究与发展*, 2021, 58(12): 2573-2584.
ZHANG L L, CHEN Y W, WU W J, et al. Interpretable few-shot learning with contrastive constraint[J]. *Journal of Computer Research and Development*, 2021, 58(12): 2573-2584. (in Chinese)
- [10] HU S X, LI D, STUHMER J, et al. Pushing the limits of simple pipelines for few-shot learning; external data and fine-tuning make a difference [C]//*Proceedings of Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2022: 9068-9077.
- [11] YAO H X, ZHANG C X, WEI Y, et al. Graph few-shot learning via knowledge transfer[EB/OL]. [2023-04-08]. <https://arxiv.org/abs/1910.03053v2>.
- [12] 冯磊, 蒋磊, 许华, 等. 基于深度级联孪生网络的小样本调制识别算法[J]. *计算机工程*, 2021, 47(4): 108-114.
FENG L, JIANG L, XU H, et al. Small sample modulation recognition algorithm based on depth cascade siamese network[J]. *Computer Engineering*, 2021, 47(4): 108-114. (in Chinese)
- [13] JIAN Y R, TORRESANI L. Label hallucination for few-shot classification[EB/OL]. [2023-04-08]. <https://arxiv.org/abs/2112.03340>.
- [14] 张睿, 杨义鑫, 李阳, 等. 自监督学习下小样本遥感图像场景分类[J]. *中国图象图形学报*, 2022, 27(11): 3371-3381.
ZHANG R, YANG Y X, LI Y, et al. Self-supervised learning based few-shot remote sensing scene image classification[J]. *Journal of Image and Graphics*, 2022, 27(11): 3371-3381. (in Chinese)
- [15] HE Y, SONG K C, DONG H W, et al. Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network[J]. *Optics and Lasers in Engineering*, 2019, 122: 294-302.
- [16] 李钧正, 殷子玉, 乐心怡. 基于小样本学习的钢板表面缺陷检测技术[J]. *航空科学技术*, 2021, 32(6): 65-70.
LI J Z, YIN Z Y, LE X Y. Surface defect detection for steel plate with small dataset [J]. *Aeronautical Science and Technology*, 2021, 32(6): 65-70. (in Chinese)
- [17] ZHANG D F, SONG K C, XU J, et al. Unified detection method of aluminum profile surface defects; common and rare defect categories[J]. *Optics and Lasers in Engineering*, 2020, 126: 105936.
- [18] PEREZ E, KIELA D, CHO K. True few-shot learning with language models[EB/OL]. [2023-04-08]. <https://arxiv.org/abs/2105.11447>.
- [19] LV X M, DUAN F J, JIANG J J, et al. Deep metallic surface defect detection; the new benchmark and detection network[J]. *Sensors*, 2020, 20(6): 1562.
- [20] SONG K C, YAN Y H. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects[J]. *Applied Surface Science*, 2013, 285: 858-864.
- [21] Guangdong Industrial Intelligent Manufacturing Innovation Competition. Recognition of surface defects of Aluminum profiles [EB/OL]. [2023-04-08]. <https://tianchi.aliyun.com/dataset/140666>.
- [22] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations [EB/OL]. [2023-04-08]. <https://arxiv.org/abs/2002.05709v2>.
- [23] SUN B, LI B H, CAI S C, et al. FSCE; few-shot object detection via contrastive proposal encoding [C]//*Proceedings of International Conference on Computer Vision*. Washington D. C., USA: IEEE Press, 2021: 7352-7362.
- [24] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[EB/OL]. [2023-04-08]. <https://arxiv.org/pdf/1606.04080.pdf>.
- [25] SUNG F, YANG Y X, ZHANG L, et al. Learning to compare; relation network for few-shot learning[EB/OL]. [2023-04-08]. <https://arxiv.org/pdf/1711.06025.pdf>.