

DOI:10.13232/j.cnki.jnju.2023.04.005

基于特征类内紧凑性的不平衡医学图像分类方法

孟 元, 张轶哲*, 张功萱, 宋 辉

(南京理工大学计算机科学与工程学院, 南京, 210094)

摘 要: 近些年, 基于深度学习的算法和模型在各种图像分析任务中都取得了显著的成功, 与常见的自然图像相比, 医学图像数据集依然面临高度不平衡的问题, 不平衡数据会导致特征空间里的决策边缘倾向样本多的类别, 导致分类效果的下降. 为了解决该问题, 提出一种基于卷积神经网络考虑特征类内紧凑性的不平衡医学图像分类方法 (Z -Score Compactness-based Convolutional Neural Network, ZC3NC). 首先, 从一个卷积神经网络的最后一层卷积层提取训练集样本与测试集样本的特征图, 随后引入一个新的 Z 分数来度量测试集数据的特征图相对训练集每个类在特征空间上的偏离度, 偏离度的度量基于类内的紧凑度, 其主要关注样本的分布特性, 对各类样本数量的不平衡性不敏感. 最终, 根据计算的偏离度, 对测试集的数据进行分类. 在 DermaMNIST 数据集上的实验表明, 在不对数据和神经网络模型做任何额外增强的情况下, 该方法的平衡准确率比原卷积神经网络模型平均提高 11.15%, 最多提高 14.08%, 证明提出的分类方法能有效地提高多种卷积神经网络对不平衡医学图像数据的分类性能. 此外, 和最先进的不平衡分类方法 Under-Bagging KNN 相比, 该方法的性能平均提升了 2.36%.

关键词: 卷积神经网络, 类别不平衡, 医学图像, 特征, 分类

中图分类号: TP391.4

文献标志码: A

Imbalanced medical image classification based on intra-class compactness of features

Meng Yuan, Zhang Yizhe*, Zhang Gongxuan, Song Hui

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China)

Abstract: In recent years, algorithms and models based on deep learning have achieved significant success in various image analysis tasks. However, compared to common natural images, medical image datasets often face highly imbalanced problems, which lead to decreased classification performance. Imbalanced data causes decision boundaries in the feature space to tend towards the class with more samples. To solve this problem, this paper proposes an imbalanced medical image classification method based on convolutional neural networks considering intra-class compactness of features (Z -Score Compactness-based Convolutional Neural Network, ZC3NC). First, feature maps of training and testing set samples are extracted from the last convolutional layer of a convolutional neural network. Then, we introduce a new Z -score based measure to test the deviation of the testing set data feature maps relative to each class of the training set in the feature space. The measure of deviation is based on intra-class compactness, which focuses on the distribution characteristics of the samples and is insensitive to the imbalance of the number of samples in each class. Finally, based on the calculated deviation, we classify the testing set data. Experiments on the DermaMNIST dataset show that without any additional data or neural network model enhancements, the balanced accuracy of the proposed method increases by an average of 11.15% compared to the original convolutional neural network model, with a maximum increase of 14.08%. This verifies that the proposed

基金项目: 国家自然科学基金(62272232, 62201263), 江苏省自然科学基金(BK20220949)

收稿日期: 2023-06-19

* 通讯联系人, E-mail: zhangyizhe@njjust.edu.cn

classification method effectively improves the classification performance of various convolutional neural networks for imbalanced medical image data. Furthermore, compared to the state-of-the-art imbalanced classification method, Underbagging KNN, the average improvement of ZC3NC is 2.36%

Key words: convolutional neural network, class imbalance, medical image, feature, classification

目前的医疗实践中,临床医生对疾病监测和诊断的主观判断可能存在误差^[1],为了克服这一问题,研究者提出了许多计算机辅助检测和诊断方法来帮助临床医生更有效地阅读医学图像并作出诊断决策.医学图像分类是其中的热点研究之一,通过该方法可以将从CT、X射线、皮肤镜等医学检测中得到的不同图像正确地划分到各自的类别中,协助医生得到更可靠的诊断结果^[2].

近年来,深度学习已经成为医学图像分类中使用最广泛的技术之一^[3],其优势是在处理非常复杂的数据时,可以在训练过程中自动提取特征,并利用多层次的神经网络对数据进行抽象表示,提高分类器的性能.然而,由于在医学环境中收集罕见疾病或其他不寻常的实例的固有困难以及获取医学图像数据和标记的高昂成本或隐私风险,许多医学图像的数据集具有高度不平衡的长尾分布(Long-Tailed Distributions)的特点^[4],即常见病拥有最多的样本数量,而罕见疾病的样本数量极少.如图1所示,在色素性皮肤病变的多源皮肤镜图像数据集 DermaMNIST^[5]的训练集中,占比最高的一个疾病类别的样本量比其他的类别样本的总和更多.由于深度学习技术普遍需要足够的图像来进行训练,这种高度不平衡会导致模型在部分疾病类上的表现明显下降^[6],这对于临床诊断是不可接受的.

为了提高深度学习对不平衡医学图像数据集的分类性能,研究者主要从数据和模型两个层面

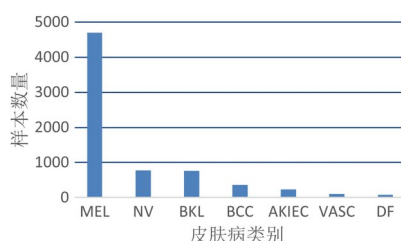


图1 DermaMNIST数据集中各类别的数量

Fig. 1 Class distribution of the DermaMNIST dataset

进行改进.从数据的角度出发,主要是结合欠采样(Under-Sampling)和过采样(Over-Sampling)两种方法^[7].Hassan and Haque^[8]采用可调Q因子小波变换信号处理技术和RUSBoost分类模型来设计自动阻塞性睡眠呼吸暂停的筛选方案,结合随机欠采样方法训练分类网络,实现睡眠呼吸暂停的自动检测.Xu et al^[9]设计了一种混合采样算法,将面向错误分类的合成少数过采样技术(M-SMOTE)和基于随机的编辑最近邻欠采样(Edited Nearest Neighbours, ENN)相结合,并动态调整M-SMOTE和ENN的几率,可以利用甲状腺功能体检图像来筛查甲状腺功能的异常.在模型层面,研究者使用修改模型来强调少数类的重要性.Ghorbani et al^[10]基于图卷积网络,提出一种重新加权的对抗图卷积网络,在训练时对类样本加权并改变每个样本对分类器的重要性.此外,还有一些研究者使用集成学习的方法将多个模型结合到一个任务中,在数据不平衡场景下也取得了很好的表现.然而,以上大部分解决方案都缺乏通用性,对其他模型不兼容,或者和原模型相比使用难度更高^[11].

针对上述挑战,本文提出一种适用于多种卷积神经网络的基于特征类内紧凑性的不平衡医学图像分类方法ZC3NC,无须重新训练即可进一步优化已有卷积神经网络模型在不平衡数据集上的分类结果.该方法的大致流程:首先,得到训练集训练完成的模型最后一层卷积层的特征图并按类集群,计算得到类内紧凑性信息;然后得到测试集特征图,引入Z分数来衡量未知类特征图偏离每个类的程度,实现对网络的分类优化.在DermaMNIST和EyePacs数据集上,分别将该方法应用于ResNet18, ResNet50, ResNeXt50和GoogLeNet四类卷积神经网络进行分类,结果其平衡准确率分别比原模型平均提升11.15%和7.02%,显示出良好的性能和对不同网络的适用性.

1 背景知识

1.1 卷积神经网络 卷积神经网络(Convolutional Neural Network, CNN)是一种深度学习模型,由一系列卷积层、池化层、全连接层组成,现已被广泛应用于图像分析和语音信号处理等任务^[12],如图2所示。

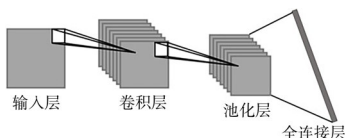


图2 卷积神经网络的结构

Fig. 2 The architecture of CNN

其中,卷积层是CNN的核心,每个卷积层包括多个可学习的卷积核,负责对输入图像进行扫描,生成一组高维特征图。每个卷积核可以捕捉输入图像的某种特定模式,如边缘、纹理、形状等,其计算如式(1)所示:

$$y = f(w_i \otimes x_i + b_i) \quad (1)$$

其中, x_i 是第 i 层卷积的输入数据, \otimes 是卷积计算, w_i 是卷积核权重矩阵, b_i 是偏置项, f 表示激活函数, y 是输出特征图。

池化层是CNN的降采样层,可以降低特征图的维度,减少模型参数和计算量,在保留重要信息的同时避免过拟合。

全连接层是CNN的输出层,负责将池化层输出的特征图展开成一维向量,并通过权重矩阵和偏置向量与输出节点相结合,最后由输出节点对应模型的分类或回归任务。

1.2 Z分数(Z-Score) 也称标准分数(Standard Score),是统计学中常用的一个概念,表示某一数据点与其所在数据集的平均值之间的差异,以标准差的形式来衡量,其计算如式(2)所示:

$$Z = \frac{(x - \mu)}{\sigma} \quad (2)$$

其中, x 是需要测量的值, μ 代表总体的平均值, σ 代表总体标准差。

Z分数是一种测量某分数在分布中相对位置的方法^[13],可以描述需要测量的值高于或低于某特定总体平均值的标准偏差。Z分数的绝对值越大,说明其偏离该平均值的程度越大,差异越显

著。如今,Z分数已被广泛应用于数据归一化、异常检测、财务状况预测等领域^[14]。

1.3 类内紧凑性 类内紧凑性(Intra-Class Compactness)被用于衡量同一类别内部样本的近似程度和聚集性,机器学习任务中通常希望同一类别的样本彼此之间更加接近,形成更紧凑的簇,这有助于提高分类准确性,降低泛化误差,增强模型的可解释性^[15]。类内紧凑性可以通过计算样本之间的距离或相似度来度量,例如方差、类内平均距离和DB指数等。较小的类内距离表示同一类别内的样本更相似,而较大的类内距离则意味着类内分散性较高,样本之间的差异性更大。

本文不关注如何修改模型或样本来提高某一类别特征的类内紧凑性,而是考虑利用不同类之间类内紧凑性的差异,比较测试集特征图和不同类训练集特征图的实际偏离程度,进而提升不平衡数据集的分类效果。

2 ZC3NC方法

本文提出的ZC3NC方法的基本结构如图3所示。ZC3NC主要包含两个部分:基础网络模块和度量模块。其中,基础网络模块即需要提升的卷积神经网络,用于对训练集和测试集数据进行特征提取;度量模块在基础网络模块的最后一个卷积层末端引入,用于度量未知类特征图偏离测试集各类特征图集合的程度。

2.1 特征提取 特征(feature)指图片在经过CNN堆叠的卷积层和池化层后产生的结果,Ali Amirshahi et al^[16]证明,通过CNN提取图像的特征在大多数质量衡量标准中优于其他方法制作的特征。

本文用 $Q: \mathbb{R}^n \rightarrow [0, 1]^c$ 表示分类神经网络,其中, $x \in \mathbb{R}^n$ 是输入数据, c 是类别数。设网络 Q 有 L 层, Q^l 代表网络的第 l 层, q^l 代表第 l 层的输出(其中 $q^0 = x$),即输入 x 通过第一层到第 l 层的非线性变换后获得的特征图。网络 Q 的每一层都允许提取和输入 x 的独特特征,而卷积神经网络的最后一层卷积层,即网络的倒数第二层 Q^{l-1} 通常会提取对于分辨输入类很重要的特征^[17]。因此,本文方法将从卷积神经网络中提取训练集和测试集

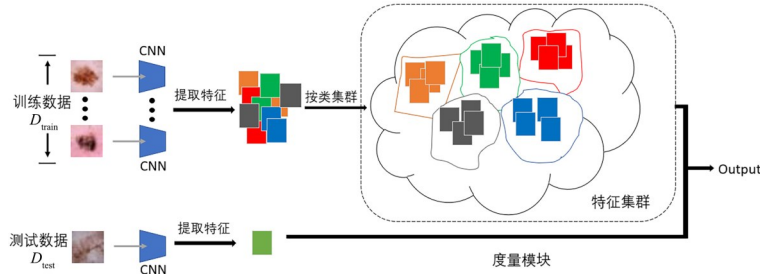


图3 基于特征类内紧凑性的不平衡医学图像分类方法框架

Fig. 3 The framework for imbalanced medical image classification based on intra-class compactness of features

图像的 q^{L-1} 用于度量模块。

图4展示了在DermaMNIST数据集上训练的ResNet18模型倒数第二层提取的特征的二维表示,通过DensMAP算法^[18-19]进行了可视化,可以直观地看出数据集中不同类的特征已被较好地分隔开。

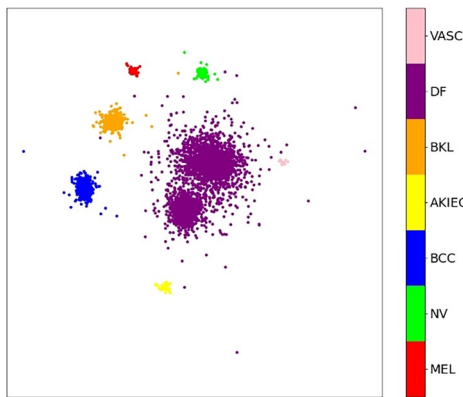


图4 在DermaMNIST上训练的ResNet18模型的倒数第二层中提取的特征的二维表示

Fig. 4 Two-dimensional representations of features extracted from a ResNet18 model trained on the DermaMNIST dataset

2.2 度量模块原理 通过提取训练集和测试集的特征图来分析图像是很常见的方法,然而其中大多数都是直接比较特征之间的距离或重新训练新的分类器模型。这类方法仅能体现类内或类间的相对关系^[20],却没有考虑同类特征图之间的距离、方差等类内要素,忽略了测试集特征图和训练集特征图的实际偏离程度。在特征空间中分类新样本时可能出现如图5所示的情况,无论是将新样本特征图与最近邻居 A^* , B^* 比较,或与两类中心点值 A , B 比较,都可能错误地将该样本分类为

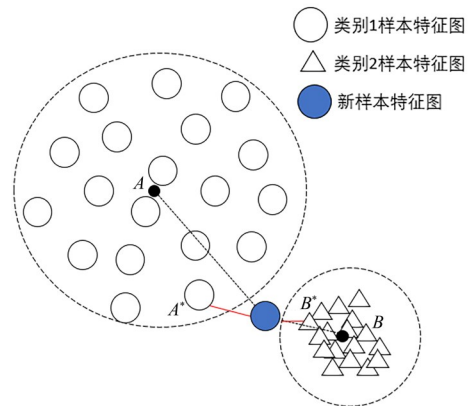


图5 在特征空间比较新样本的示意图

Fig. 5 Illustration of comparing new samples in the feature space

类别2,因此无法提升分类效果。

针对上述问题,本文认为应当在样本分类时重点考虑某些类别样本特征图内部的紧凑性,以此为依据,重新确定拥有更远绝对距离但更近相对距离的特征图的分类情况。因此,本文提出利用Z分数对测试集特征与不同类训练集特征集群之间的偏差程度进行度量并分类。

具体地,度量模块首先将训练集样本的特征图按类集群,分别计算每类集群内特征图两两之间的距离,并求平均距离及其标准差,以此表示不同类训练集特征图的类内紧凑性。然后,将新样本特征图分别与每个类的样本特征计算平均距离,最后代入Z分数得到新样本相对于每个类的偏差程度,以此判断新样本的类别。

2.3 ZC3NC算法 本方法的算法流程图如图6所示,算法流程如下。

输入: 不平衡数据集的训练集 D_{train} , 测试集未知类图像 D_{test} 。

输出: D_{test} 分类结果。

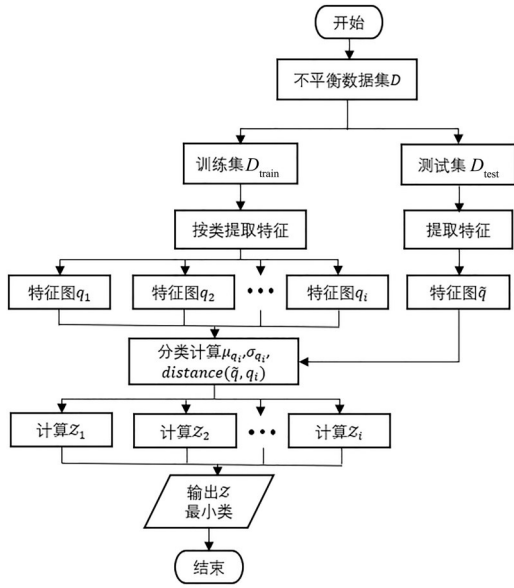


图 6 ZC3NC 算法的流程图

Fig. 6 The algorithm flowchart of ZC3NC

1. 使用训练好的卷积神经网络模型,从最后一层卷积层 Q^{l-1} 分别提取 D_{train} 中每类的特征图 q_i 和 D_{test} 特征图 \tilde{q} .

2. 计算训练集每类特征集合 q_i 内部特征两两之间的距离矩阵 $A_{(N \times N)}^i$, 如式(3)所示:

$$A_{(N \times N)}^i = \begin{bmatrix} \|q_i^1, q_i^1\| & \cdots & \|q_i^1, q_i^N\| \\ \vdots & \ddots & \vdots \\ \|q_i^N, q_i^1\| & \cdots & \|q_i^N, q_i^N\| \end{bmatrix} \quad (3)$$

其中, $\|x, y\|$ 表示特征 x 和 y 之间的欧式距离, N 是集合 q_i 中特征的数量.

3. 计算集合 q_i 的矩阵 $A_{(N \times N)}^i$ 的平均距离 μ_{q_i} 和标准差 σ_{q_i} , 如式(4)和式(5)所示:

$$\mu_{q_i} = \frac{1}{N^2} \sum_{a=1}^N \sum_{b=1}^N A_{a,b} \quad (4)$$

$$\sigma_{q_i} = \sqrt{\frac{1}{N^2} \sum_{a=1}^N \sum_{b=1}^N (A_{a,b} - \mu_{q_i})^2} \quad (5)$$

4. 使用测试集特征 \tilde{q} , 分别与训练集每类特征图集合 q_i 的计算平均距离 $\text{distance}(\tilde{q}, q_i)$, 如式(6)所示:

$$\text{distance}(\tilde{q}, q_i) = \frac{1}{N} \sum_{n=1}^N \|\tilde{q}, q_i^n\| \quad (6)$$

5. 将训练集特征集 q_i 的平均距离 μ_{q_i} 、标准差 σ_{q_i} 和特征 \tilde{q} 与集合 q_i 的平均距离 $\text{distance}(\tilde{q}, q_i)$ 代入式(1), 计算新样本相对于类 i 的偏差程度 z_i 并求绝对值, 若 $\text{distance}(\tilde{q}, q_i) - \mu_{q_i}$ 取负, 绝对值越大同样说明特征过于紧凑, 如式(7)所示:

$$z_i = \frac{|\text{distance}(\tilde{q}, q_i) - \mu_{q_i}|}{\sigma_{q_i}} \quad (7)$$

其中, 训练集特征集 q_i 相对于 m 类集合的偏差程度 z_i , 形成数列 $Z = [z_1, z_2, \dots, z_m]$.

6. 取数列 Z 的最小值所在索引作为分类结果, 如式(8)所示:

$$\hat{y} = \text{argmin}[z_1, z_2, \dots, z_m] \quad (8)$$

3 实验与结果分析

本文对方法的验证主要包括三部分:

(1) 构建 ResNet18, ResNet50, ResNeXt50 和 GoogLeNet 四类卷积神经网络并设计网络参数, 在数据集上进行训练和测试.

(2) 从上述四类卷积神经网络中提取特征图, 应用 ZC3NC 方法与原模型进行比较, 验证方法的有效性.

(3) 应用本文方法与多种不平衡数据集分类方法进行比较.

3.1 实验数据集 为了验证提出的 ZC3NC 分类方法的有效性, 在色素性皮肤病变的多源皮肤镜图像数据集 DermaMNIST 和糖尿病视网膜病变图像数据集 EyePacs^[21] 上进行实验. 数据集的大小、类别数和不平衡比等详细信息如表 1 所示.

表 1 实验使用的数据集信息

Table 1 Information of datasets used in experiments

Datasets	Instances	Classes	Majority	Minority	IR
DermaMNIST	10015	7	66.95%	1.15%	58.30
EyePacs	88702	5	73.67%	2.16%	34.14

其中, 不平衡比率 (Imbalance Ratio, IR)^[22] 是最多数类的样本量与最少类类的样本量之比, IR 越大, 代表数据集的不平衡程度越大.

DermaMNIST 是基于色素性皮肤病变的多源皮肤镜图像数据集 HAM10000^[23] 处理得到的, 作为 MedMNIST 的一个子集用于多类分类任务. 该数据集由 10015 张皮肤镜图像组成, 被标记成七个不同类别, 分别是黑色素瘤 (Melanoma, MEL)、黑色素细胞痣 (Melanocytic nevus, NV)、基底细胞癌 (Basal Cell Carcinoma, BCC)、光化性角化病 (Actinic Keratosis, AKIEC)、良性角化病 (Benign Keratosis, BKL)、皮肤纤维瘤 (Dermatofib-

broma, DF) 和 血管 病 变 (Vascular Lesion, VASC). 以 7:1:2 的比例将数据集图像分为训练集、验证集和测试集,并将图像由 $3 \times 600 \times 450$ 调整为 $3 \times 28 \times 28$,示例图片如图 7 所示.

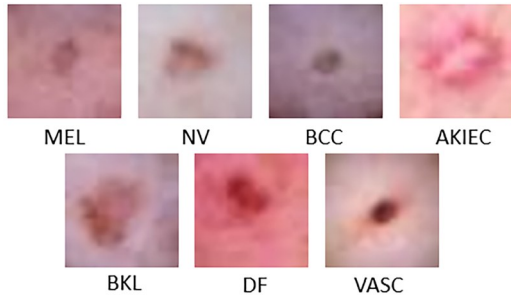


图 7 DermaMNIST 数据集的部分示例图片

Fig 7 Partial examples of images of the DermaMNIST dataset

EyePacs 是糖尿病视网膜病变 (Diabetic Retinopathy, DR) 分级任务中最大的公开数据集,被标记了属于五种可能的 DR 分级,范围从 DR0 (无 DR) 到 DR4 (增生性 DR). 该数据库包括约 35000 个训练样本和 55000 个测试集样本,并使用 10% 的训练集图像作为验证集,如图 8 所示.

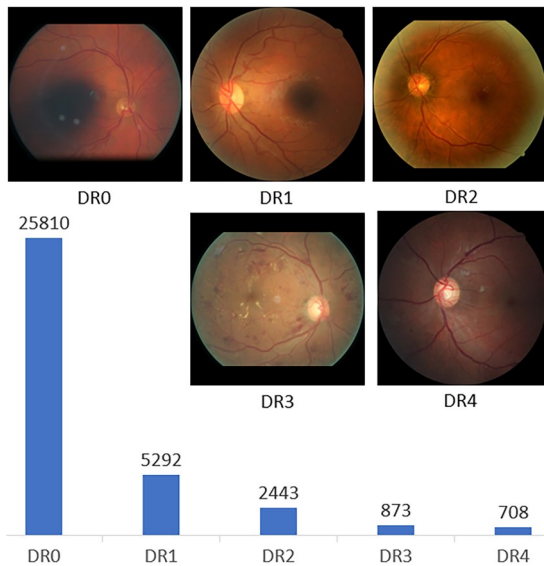


图 8 EyePacs 数据集的 DR 分级

Fig. 8 DR grading of the EyePACS dataset

3.2 实验细节 采用 Pytorch 框架,在 NVIDIA GeForce2080 Ti GPU 48 GB RBM 的 Ubuntu18 服务器上训练. 选用 Torchvision^[24] 提供的结构用于对比和提取特征的卷积神经网络模型. 依照数

据集推荐的超参数设置,训练网络时 Batch 大小为 256,最初学习率为 0.01,使用 Multistep 策略调整,衰减学习率为 0.1,epoch 的最大训练值为 100. 选取交叉熵损失函数 (Cross-Entropy Loss) 和自适应矩估计 (Adaptive Moment Estimation, Adam) 优化器,并加入样本类权重进行训练.

3.3 评价指标 对于不平衡数据集的分类,准确率不能很好地反映一个分类算法的性能. 本文选取了更适应不平衡数据集的评价指标——平衡准确度 (Balanced Accuracy, Balanced ACC) 和加权精确度 (Weighted-Precision, Weighted-P).

Balanced ACC 本质上是召回率的算术平均值,它给予每个类以同样的重要性,可以更好地预测整个不平衡数据集^[25-26],如式 (9) 所示:

$$\text{Balanced ACC} = \frac{1}{C} \sum_{i=0}^{C-1} \frac{TP_i}{TP_i + FN_i} \quad (9)$$

Weighted-P 是一种综合考虑各个类别样本数量和预测准确性的指标,其原理是将每个类别的精确度 (Precision) 和该类别在总样本中的比例相乘,然后求和^[27],如式 (10) 所示:

$$\text{Weighted-P} = \sum_{i=1}^C \frac{TP_i}{TP_i + FP_i} \times \frac{TP_i + FN_i}{\sum_{j=1}^C (TP_j + FN_j)} \quad (10)$$

式 (9) 和式 (10) 中, C 代表数据集的种类数; TP 代表真正类,即实际为正类且预测也为正的样本数; FN 代表假负类,即实际为正类且预测为负类的样本数; FP 为假正类,即实际为负类且预测为正类的样本数.

3.4 实验结果与分析 为了验证 ZC3NC 的通用性和有效性,分别在 DermaMNIST 和 EyePacs 数据集上使用四类卷积神经网络 ResNet18, ResNet50^[28], ResNeXt50^[29] 和 GoogLeNet^[30] 进行测试,将本文方法与原模型结果进行比较.

在 DermaMNIST 数据集上的对比实验结果如表 2 所示,表中黑体字表示最优的性能. 由表可见, ZC3NC 的两个指标比未应用本方法的原型模型都有提升,其中 Balanced ACC 平均提升 11.15%, Weighted-P 平均提升 2.33%.

图 9 展示了在 DermaMNIST 数据集上训练的 ResNet18 模型倒数第二层提取的训练集和部

表 2 在 DermaMNIST 数据集上的实验结果

Table 2 Experimental results of different algorithms on the DermaMNIST dataset

模型	原模型		原模型+ZC3NC	
	Balanced ACC	Weighted-P	Balanced ACC	Weighted-P
ResNet18	56.51%	74.64%	64.46%	76.46%
ResNet50	45.62%	70.46%	53.10%	72.39%
ResNeXt50	53.18%	72.84%	57.30%	75.77%
GoogLeNet	61.98%	77.71%	67.36%	77.90%
Avg	54.32%	73.91%	60.56%	75.63%

分测试集特征的二维表示,其中圆点代表训练集特征图,三角代表测试集中原模型未成功分类而 ZC3NC 成功分类的特征图. 由图可见,部分三角距离其他类的绝对距离更近但仍被 ZC3NC 方法成功分类.

在 EyePacs 数据集上的对比实验结果如表 3 所示,表中黑体字表示最优的性能. 应用本文方法的 Balanced ACC 均大于原模型,只在

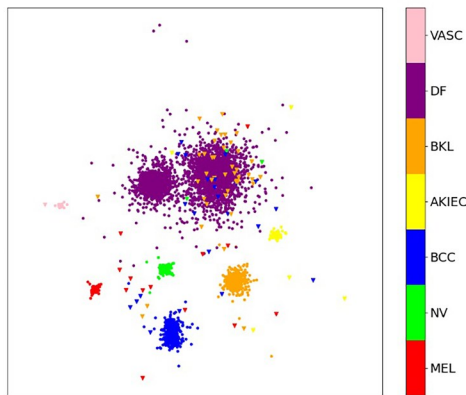


图 9 训练集和部分测试集特征的二维表示

Fig 9 Two-dimensional representation of training set and partial test set features

GoogLeNet 模型上 Weighted-P 稍低,但差距很小. 平均来说,和原模型相比,应用 ZC3NC 的 Balanced ACC 提升 7.02%, Weighted-P 提升 0.16%.

表 3 在 EyePacs 数据集上的实验结果

Table 3 Experimental results of different algorithms on the EyePacs dataset

模型	原模型		原模型+ZC3NC	
	Balanced ACC	Weighted-P	Balanced ACC	Weighted-P
ResNet18	47.19%	72.48%	51.11%	72.54%
ResNet50	47.70%	73.65%	51.92%	73.94%
ResNeXt50	48.19%	73.93%	52.34%	74.39%
GoogLeNet	48.46%	72.57%	49.63%	72.22%
Avg	47.89%	73.16%	51.25%	73.27%

在 DermaMNIST 和 EyePacs 数据集上,分别与经典处理高度不平衡数据集的方法 SMOTE (Synthetic Minority Oversampling Technique)^[31], K-Means SMOTE (KMSMOTE)^[32]和目前最先进的处理高度不平衡数据集的集成学习方法 Under-Bagging KNN (UbKNN)^[33]进行对比实验. UbKNN 利用 Bagging 结合欠采样技术生成平衡数据集,并使用 KNN 算法进行分类,从理论和实验两方面验证了 ZC3NC 的 Balanced ACC 的优势. 对比实验结果如表 4 和表 5 所示,表中黑体字表示最优的性能. 由表可见,ZC3NC 的 Balanced ACC 和 Weighted-P 两个指标在所有卷积神经网络模型上都稍优于上述三种方法.

图 10 和图 11 展示了 DermaMNIST 和 EyePacs 数据集上,上述各种方法在四类卷积神经网络中的平均表现,图中黑体字表示最优的性能.

表 4 各算法在 DermaMNIST 数据集上的对比实验结果

Table 4 Experimental results of different algorithms on the DermaMNIST dataset

模型	SMOTE		KMSMOTE		UbKNN		ZC3NC	
	Balanced ACC	Weighted-P	Balanced ACC	Weighted-P	Balanced ACC	Weighted-P	Balanced ACC	Weighted-P
ResNet18	60.65%	73.25%	59.00%	74.92%	64.07%	76.12%	64.46%	76.46%
ResNet50	49.87%	71.89%	47.15%	70.71%	52.94%	72.16%	53.10%	72.39%
ResNeXt50	55.34%	73.35%	52.60%	72.35%	57.05%	73.84%	57.30%	75.77%
GoogLeNet	64.75%	76.92%	66.57%	77.59%	66.91%	76.62%	67.36%	77.90%
Avg	57.66%	74.35%	56.33%	73.89%	60.24%	74.69%	60.56%	75.63%

表5 各算法在EyePacs数据集上的对比实验结果

Table 5 Experimental results of different algorithms on the EyePacs dataset

模型	SMOTE		KMSMOTE		UbKNN		ZC3NC	
	Balanced ACC	Weighted-P	Balanced ACC	Weighted-P	Balanced ACC	Weighted-P	Balanced ACC	Weighted-P
ResNet18	46.54%	70.77%	48.09%	72.26%	48.77%	71.05%	51.11%	72.54%
ResNet50	46.02%	73.91%	46.66%	73.01%	48.83%	72.52%	51.92%	73.94%
ResNeXt50	47.46%	74.28%	49.72%	74.35%	48.19%	72.78%	52.34%	74.39%
GoogLeNet	47.38%	70.00%	46.89%	70.34%	49.05%	71.41%	49.63%	72.22%
Avg	46.85%	72.24%	47.89%	72.49%	48.71%	71.94%	51.25%	73.27%

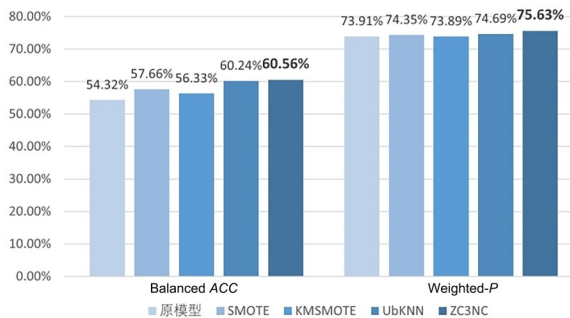


图10 各算法在DermaMNIST数据集上对比实验的平均结果

Fig. 10 Average experimental results of different algorithms on the DermaMNIST dataset

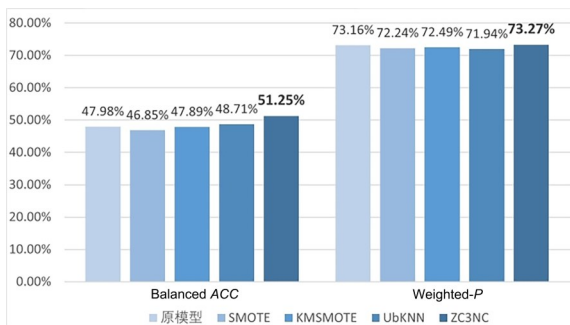


图11 各算法在EyePacs数据集上对比实验的平均结果

Fig. 11 Average experimental results of different algorithms on the EyePacs dataset

综上,ZC3NC方法有效提升了各类卷积神经网络在不平衡医学图像数据集上的分类性能,与经典不平衡分类方法以及最先进的方法相比,均表现相近或有一定优势,证明本文方法是一种有效的改进算法,适用于多种卷积神经网络。

4 结论

本文针对卷积神经网络面对不平衡医学图像数据集的分类困难,提出一种基于特征类内紧凑

性的不平衡医学图像分类方法ZC3NC,通过比较测试集特征图和不同类训练集特征图的实际偏离程度进行分类,达到了较好的分类效果,并在多个卷积神经网络和数据集中进行了验证。

未来将在比较特征类内紧凑性的同时,结合类间松散度进一步提升方法性能,并进一步探索特征空间信息,使其能适应平衡样本分类任务。

参考文献

- [1] Chen X X, Wang X M, Zhang K, et al. Recent advances and clinical applications of deep learning in medical image analysis. Medical Image Analysis, 2022(79):102444.
- [2] 郑光远,刘峡壁,韩光辉. 医学影像计算机辅助检测与诊断系统综述. 软件学报, 2018, 29(5): 1471—1514. (Zheng G Y, Liu X B, Han G H. Survey on medical image computer aided detection and diagnosis systems. Journal of Software, 2018, 29(5): 1471—1514.)
- [3] Cai L, Gao J Y, Zhao D. A review of the application of deep learning in medical image classification and segmentation. Annals of Translational Medicine, 2020, 8(11):713.
- [4] Singh R, Bharti V, Purohit V, et al. MetaMed: Few-shot medical image classification using gradient-based meta-learning. Pattern Recognition, 2021(120): 108111.
- [5] Yang J C, Shi R, Wei D L, et al. MedMNIST v2: A large-scale lightweight benchmark for 2D and 3D biomedical image classification. Scientific Data, 2023, 10(1):41.
- [6] Queller G, Lamard M, Conze P H, et al. Automatic detection of rare pathologies in fundus photographs

- using few-shot learning. *Medical Image Analysis*, 2020(1):101660.
- [7] Batista G E A P A, Prati R C, Monard M C. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, 2004, 6(1):20—29.
- [8] Hassan A R, Haque M A. An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting. *Neurocomputing*, 2017(235):122—130.
- [9] Xu Z Z, Shen D R, Nie T Z, et al. A hybrid sampling algorithm combining M-SMOTE and ENN based on random forest for medical imbalanced data. *Journal of Biomedical Informatics*, 2020(107):103465.
- [10] Ghorbani M, Kazi A, Baghshah M S, et al. RA-GCN: Graph convolutional network for disease prediction problems with imbalanced data. *Medical Image Analysis*, 2022(75):102272.
- [11] Frid-Adar M, Diamant I, Klang E, et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, 2018(321):321—331.
- [12] Li Z W, Liu F, Yang W J, et al. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(12):6999—7019.
- [13] Curtis A E, Smith T A, Ziganshin B A, et al. The mystery of the Z-score. *Aorta*, 2016, 4(4):124—130.
- [14] Kirkwood B R, Sterne J A C. *Essential medical statistics*. New York: John Wiley & Sons, 2010:126—128.
- [15] Wen Y D, Zhang K P, Li Z F, et al. A discriminative feature learning approach for deep face recognition//The 14th European Conference on Computer Vision. Springer Berlin Heidelberg, 2016:499—515.
- [16] Ali Amirshahi S, Pedersen M, Yu S X. Image quality assessment by comparing CNN features between images. *Journal of Imaging Science and Technology*, 2016, 60(1):060410.
- [17] Abdelzad V, Czarnecki K, Salay R, et al. Detecting out-of-distribution inputs in deep neural networks using an early-layer output. 2019, arXiv:1910.10307.
- [18] Narayan A, Berger B, Cho H. Assessing single-cell transcriptomic variability through density-preserving data visualization. *Nature Biotechnology*, 2021, 39(6):765—774.
- [19] McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. 2018, arXiv:1802.03426.
- [20] Pang T Y, Xu K, Dong Y P, et al. Rethinking softmax cross-entropy loss for adversarial robustness//The 8th International Conference on Learning Representations. Addis Ababa, Ethiopia: OpenReview.net, 2020, arXiv:1905.10626.
- [21] Galdran A, Carneiro G, Ballester M A G. Balanced-mixup for highly imbalanced medical image classification//The 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Springer Berlin Heidelberg, 2021:323—333.
- [22] Zhu R, Guo Y W, Xue J H. Adjusting the imbalance ratio by the dimensionality of imbalanced data. *Pattern Recognition Letters*, 2020(133):217—223.
- [23] Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 2018, 5(1):180161.
- [24] Marcel S, Rodriguez Y. Torchvision the machine-vision package of torch//Proceedings of the 18th ACM International Conference on Multimedia. Firenze, Italy: ACM, 2010:1485—1488.
- [25] Grandini M, Bagli E, Visani G. Metrics for multi-class classification: An overview. 2020, arXiv:2008.05756.
- [26] Flach P A. Performance evaluation in machine learning: The good, the bad, the ugly, and the way forward//Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu, HI, USA: AAAI Press, 2019:9808—9814.
- [27] Xu H Y, Zhang H, Han K, et al. Learning alignment for multimodal emotion recognition from speech//Interspeech 2019, the 20th Annual Conference of the International Speech Communication Association. Graz, Austria: ISCA, 2019:3569—3573.
- [28] 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, NV, USA: IEEE, 2016:770—778.

- [29] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 5987—5995.
- [30] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions// Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015: 1—9.
- [31] 刘娜. 面向非平衡数据的医疗智能诊断与决策支持研究. 博士学位论文. 天津: 天津大学, 2021. (Liu N. Research on medical intelligent diagnosis and decision support based on imbalanced data. Ph. D. Dissertation. Tianjin: Tianjin University, 2021.)
- [32] Douzas G, Bacao F, Last F. Improving imbalanced learning through a heuristic oversampling method based on k - means and SMOTE. Information Sciences, 2018(465): 1—20.
- [33] Hang H Y, Cai Y C, Yang H F, et al. Under-bagging nearest neighbors for imbalanced classification. The Journal of Machine Learning Research, 2022, 23 (1): 118.

(责任编辑 杨可盛)