

基于细粒度分析的不均衡图像分类算法

刘萍萍^{1,2}, 商文理³, 解小宇¹, 杨晓康³

(1. 吉林大学 计算机科学与技术学院, 长春 130012; 2. 吉林大学 符号计算与知识工程教育部重点实验室, 长春 130012; 3. 吉林大学 软件学院, 长春 130012)

摘要: 针对细粒度属性图像具有复杂性和多样性, 传统的图像分类方法在关注图像细粒度属性方面存在不足, 并在处理不均衡数据集时表现不佳的问题, 提出了一种基于深度度量学习的细粒度图像阈值分类算法。通过引入度量学习方法增强对图像细粒度属性的关注。同时, 通过应用成对损失和代理损失, 提高了模型的分类准确性并加快了模型的收敛速度。为了应对数据不均衡问题, 设计了一个基于阈值分析的分类器。该分类器利用阈值分析技术实现了对细粒度图像的多级分类, 从而改善了在不均衡数据集中少数类别分类准确性较低的问题。实验结果表明, 本文所提出的基于深度度量学习的细粒度图像阈值分类算法在分类准确性方面显著优于其他方法。

关键词: 计算机应用; 深度度量学习; 细粒度分类; 不均衡数据; 阈值分类器

中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5497(2025)06-2122-09

DOI: 10. 13229/j. cnki. jdxbgxb. 20230991

Unbalanced image classification algorithm based on fine-grained analysis

LIU Ping-ping^{1,2}, SHANG Wen-li³, XIE Xiao-yu¹, YANG Xiao-kang³

(1. College of Computer Science and Technology, Jilin University, Changchun 130012, China; 2. Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China; 3. College of Software, Jilin University, Changchun 130012, China)

Abstract: Aiming at the complexity and diversity of fine-grained images, where traditional image classification methods exhibit limitations in focusing on fine-grained attributes and perform poorly when handling imbalanced datasets, a threshold-based fine-grained image classification algorithm utilizing deep metric learning was proposed. The focus on fine-grained attributes of images was enhanced by introducing a metric learning approach. Additionally, the classification accuracy was enhanced and the model convergence was expedited by incorporating pairwise loss and agent loss mechanisms. To address the issue of data imbalance, a classifier was devised grounded in threshold analysis techniques. This innovative classifier harnesses threshold analysis to facilitate multi-level classification of fine-grained images, thereby ameliorating the issue of low classification accuracy for certain categories within an imbalanced dataset. The

收稿日期: 2023-09-15.

基金项目: 吉林省自然科学基金项目(20200201283JC); 吉林省产业关键核心技术攻关项目(20230201085GX); 国家自然科学基金面上项目(62071199).

作者简介: 刘萍萍(1979-), 女, 教授, 博士. 研究方向: 机器学习与图像处理. E-mail: liupp@jlu.edu.cn

results of these experiments unequivocally demonstrate that the proposed threshold classification algorithm for fine-grained images, based on deep metric learning, outperforms alternative methods in terms of classification accuracy.

Key words: computer application; deep metric learning; fine-grained classification; unbalanced data; threshold classifier

0 引言

图像分类是计算机视觉领域中一个经典的研究领域,通常分为语义级图像分类和实例级图像分类。语义级图像分类的目的是确定不同的对象类别,如动植物分类^[1]等;实例级图像分类的目的是确定不同的个体,如人脸识别^[2]等。细粒度图像分类则介于语义级图像分类和实例级图像分类之间,与粗粒度图像分类的任务不同,细粒度图像分类任务通常类与类之间的区别比较细微,一般要靠细微的局部差别才能区别出不同类。这类任务一般称为子类别分类,因为它们的分类界限常位于同一类别的不同子类之间,细粒度图像分类比对象级分类任务所面临的类内差异更大。然而,在处理细粒度分类时,通常需要面对不均衡类别分布的问题,不均衡类别分布会对模型性能产生负面影响,这通常导致模型倾向于学习数量较多的类别,而忽视数量较少的类别,从而降低了模型的准确性。在某些应用场景中,由于数据不可控因素,不均衡类别分布是不可避免的挑战。因此,研究如何在不均衡类别分布条件下实现精确的细粒度图像分类具有重要意义。

学习判别局部区域的特征可以有效地提高细粒度图像分类的性能。因此,此类方法在最近的工作中得到了广泛的研究。通过对细粒度目标部位对应的关键区域进行检测、分割或定位,可以学习到更多具有判别性的局部特征,进一步提高网络的学习能力。Ji等^[3]提出ACNet将注意力机制与卷积二叉树相结合,学习层次细粒度特征,进而提高模型分类性能。Wei等^[4]提出Mask-CNN,通过分割模型获取掩码定位物体,该方法在仅依靠训练时提供的部分注释,不需要边界框,同时测试时不需额外监督信息,在细粒度图像分类任务中取得较好的分类精度。尽管基于判别局部区域定位的方法不断改进,但是在细粒度图像的判别局部特征关系建模方面,以往的工作仍然存在不足。

学习统一的、有区别的特征表示,能有效地模

拟细粒度数据集里不同类别之间的细微差异,此类方法在细粒度图像分类任务中有良好的性能表现。Zheng等^[5]通过对特征通道组应用双线性池化,有效降低了计算成本,同时提高了模型的性能表现。Chang等^[6]构建了通道交互损失,通过简单的通道交互机制,从通道关系中获得特征之间的区别和互补关系进而提高模型性能表现。Bera等^[7]设计了一种基于图神经网络的细粒度分类模型(SR-GNN),通过关系感知的视觉特征变换实现相关局部特征重要性的重加权,并将得到的特征进行自动聚合,得到具有表现力的特征表示。上述方法虽然在图像细粒度分类任务上的性能有所提升,但是忽视了由于部分细粒度图像数据较少而产生的数据不均衡问题。

本研究的贡献在于将深度度量学习方法与阈值分析相结合应用于不均衡的细粒度图像分类领域。实验结果显示,本文所提出的算法在典型的不均衡细粒度图像分类任务中相较于传统方法具有更高的准确性和可靠性。

1 本文方法

近年来,基于深度学习的图像分类研究引入度量学习的概念^[8]。度量学习通过将图像特征映射到一个新的度量空间,拉近同类样本和推远不同类样本的方式,更加关注图像的细粒度属性。以细粒度特征提取为基础的图像分类算法由于关注到图像特征间的细微差异,实际应用时可以有效提高分类结果的准确性。

此外,当数据呈明显的长尾分布,当前常规的分类方法通常很难在少数类中获得优异的分类效果。在本文中,为了进一步提升分类准确性,阈值分类方法被引入。本研究采用视觉转换器(Vision transformer, ViT)作为特征提取网络^[9],并结合度量学习方法,将神经网络提取的特征映射到一个新的度量空间,通过设计阈值分类器,在少数类别上提高了图像分类结果的准确性。

本研究的目标是学习一个嵌入函数 $Z=f(\theta)$ 。 $f(\theta)$ 是一个具有参数 θ 的深度神经网络,将图像 x 在度量空间中映射成特征 Z 。例如, x 和 y 为输入的样本图像,它们在度量空间中的距离为 $d_0(x,y)=\|f_0(x)-f_0(y)\|$ 。一个好的嵌入函数会在度量空间中拉近同类的样本,推远不同类的样本,以便提高分类的准确性。为了实现最佳的分类效果,提出了阈值分类方法应用于不均衡的细粒度图像分类任务。

1.1 阈值分类框架概述

本文所提出的阈值分类框架由两部分组成:特征提取和阈值分类方法。首先,将预处理后的图像输入主干网络进行初始特征提取。其次,通过应用度量损失函数对特征提取网络进行训练。最后,将训练得到的特征提取网络输出的特征向量传递给阈值分类器,以实现对细粒度图像的分类。图 1 展示了阈值分类方法框架的总体训练结构。

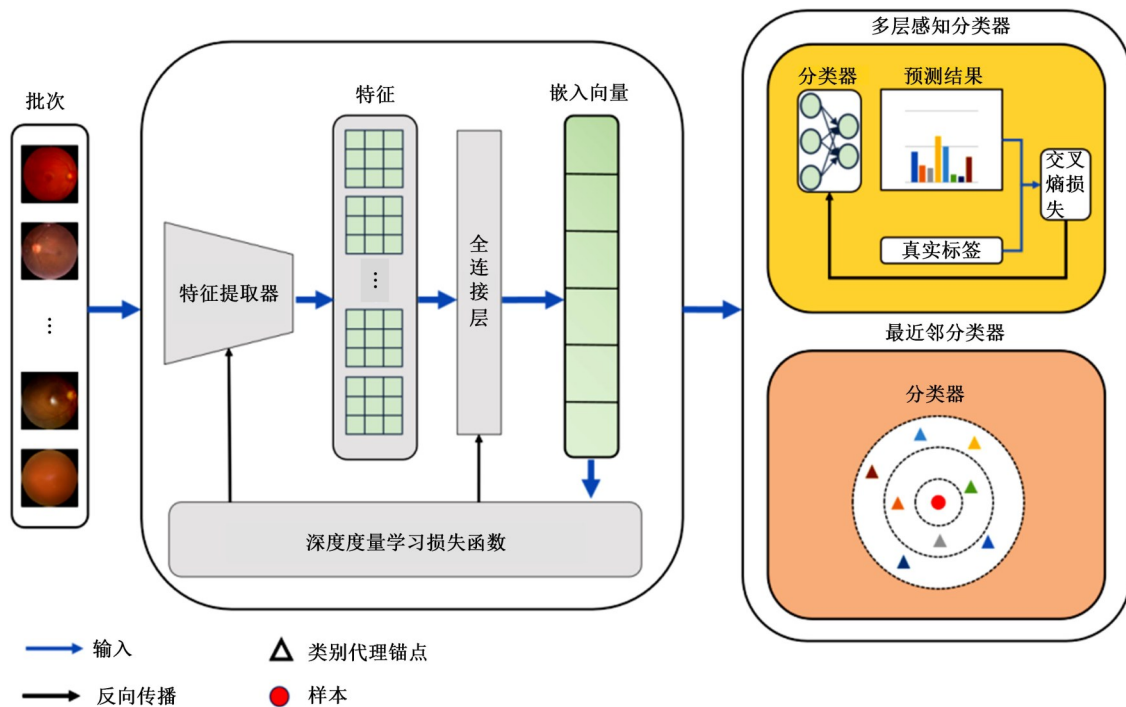


图 1 阈值分类框架的总体训练结构

Fig. 1 Overall training structure of the threshold classification framework

该框架的特征提取阶段旨在从输入图像中提取有用的特征信息。通过主干网络的处理,图像中的局部和全局特征被有效地捕获。为了优化特征提取的性能,更好地关注图像的细粒度属性,度量损失函数被引入,以对特征进行有针对性的训练。鉴于不均衡数据集中少数类样本数量较少,在面对这种严重的数据不均衡问题时,为了减轻数据不均衡对分类结果的影响,借鉴了集成学习^[10]方法,提出了一种基于优势阈值的多级分类方法。在特征提取完成后,所得到的特征向量被传递给阈值分类器,以实现不均衡的细粒度图像分类任务。

1.2 度量学习损失函数

度量学习(Metric learning)可以看作通过度量损失训练特征提取器的过程。在细粒度图像分

类任务中,相似度度量是一个至关重要的问题。传统的图像分类方法通过训练分类器获得分类结果。然而,这种方法忽略了样本之间微小的差异,无法准确度量图像之间的相似性。关注样本之间微小的差异在需要区分相似度较高的不同类别图像时是十分必要的。此外,图像细粒度分类任务通常面临小样本问题。在不均衡图像数据集中,非优势类别中的样本数量通常有限,这导致传统的分类方法可能在面对小样本情况时表现不佳。度量学习方法可以通过学习适当的度量空间,将相似的图像样本聚集在一起,从而有效地利用有限的样本信息,提高分类性能。

目前已有的度量学习损失函数可以分为两类:基于代理的损失和成对损失。基于代理的损失通过引入类别代理,旨在解决训练时模型收敛

慢的问题。代理是训练数据子集的代表,并作为网络参数的一部分进行训练。在基于代理的损失中,样本与样本的关系可以用样本与代理的关系代替,因此可以减少大量的训练时间,更快地达到收敛目标。某个类别中现有的损失将每个数据点视为一个锚,将其与代理而不是其他图像关联,鼓励锚接近同一类的代理,远离不同类别的代理。成对损失则利用数据点之间的细粒度语义关系,同时通过比较样本间的相似性关系建立类别之间的细粒度描述。

在本文的方法中,利用基于代理的损失与成对损失相结合的方法,原因是基于代理的损失可以达到较好的训练开销,可由于在训练的初始阶段无法获得优秀的类别代理,反而无法达到使用代理的价值,因此在训练初始使用了成对损失,而在训练后期将代理损失与成对损失结合,加快模型的收敛。本文选取 Proxy-NCA^[11]和多相似性损失^[12]作为主要损失。

Proxy-NCA 损失为每个类分配一个代理,以便代理数量与类标签数量相同。给定一个输入数据点作为锚点,这个代理和同一类别的输入被认为是正的,其他的代理是负的。Proxy-NCA 损失定义为:

$$L_{NCA} = \sum_{x \in X} -\log \frac{e^{s(x, p^+)}}{\sum_{p^- \in P^-} e^{s(x, p^-)}} \quad (1)$$

式中: x 为输入; p^+ 为正代理; p^- 为负代理; P^- 为负代理集; $s(\cdot, \cdot)$ 为两个向量的余弦相似度。

多相似性损失 (Multi-similarity loss) 通过采样和加权两次迭代,实现对样本高效的训练,并通过定义自相似性和相对相似性,在训练时更加全面地考虑了局部样本分布,从而能更高效精确地对重要样本对进行采用和加权。多相似性损失定义为:

$$L_{MS} = \frac{1}{m} \left\{ \frac{1}{\alpha} \log \left[1 + \sum_{k \in Q_i} e^{-\alpha(S_{ik} - \lambda)} \right] + \frac{1}{\beta} \log \left[1 + \sum_{k \in N_i} e^{\beta(S_{ik} - \lambda)} \right] \right\} \quad (2)$$

式中: S_{ik} 为样本对的余弦相似度; Q_i 为正对样本集; N_i 为负对样本集; m 为训练 batch-size 大小; λ 为 margin 大小; α 和 β 为控制权重尺度的超参数。

为了减少了训练过程中的复杂性并能够实现更快地收敛且兼顾数据点之间的细粒度语义关系,首先使用了多相似性损失函数训练神经网络,

在训练一段时间后加入 Proxy-NCA 损失函数进一步提高框架性能,最终的度量损失函数为:

$$L_{metric} = L_{MS} + 0.01 \times L_{PA} \quad (3)$$

式中: L_{MS} 为多相似性损失; L_{PA} 为 Proxy-NCA 损失。通过采用上述损失函数,可以减少训练过程中的复杂性并能够实现更快的收敛,同时兼顾数据点之间的细粒度语义关系,有效地训练图像分类任务中的特征提取网络,并提高分类框架性能。

最终的总损失为:

$$L = L_{metric} + L_{ce} \quad (4)$$

式中: L_{metric} 为总的度量损失; L_{ce} 为交叉熵损失。

可以看到,特征提取模块的训练目标是训练优秀的度量空间,其中,同类的样本距离更近,不同类的样本距离更远。同时,由于引入基于代理的损失,还会得到一个用训练集生成的类代理集合 Proxy-Set。

鉴于不均衡数据集中少数类样本数量较少,在面对这种严重的数据不均衡问题时,为了减轻数据不均衡对分类结果的影响,借鉴集成学习的思想设计了一种新的阈值分类器。该分类器通过调整分类阈值优化图像的判定标准。

如图 2(a)所示,预先设置一个阈值 τ ,对于待分类的细粒度图像,它首先经过由图 1 训练框架中的 MLP (Multilayer perceptron) 分类器。MLP 分类器是一个多层感知器分类器,经过训练集训练用于对细粒度图像进行分类。分类测试时,当 MLP 分类器输出的预测值大于等于设定的阈值 τ

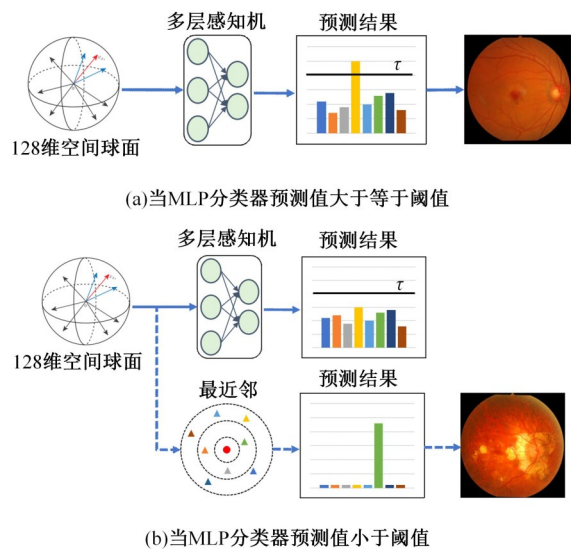


图 2 阈值分类方法示意图

Fig. 2 Schematic diagram of threshold classification method

时,保留MLP分类器预测结果。如果MLP分类器输出的优势标签的概率强度低于阈值 τ ,说明目前的图片无法获得确切的分类结果。这些无法获得确切效果图片以少数类图片居多,将待分类图像送入最近邻(Nearest neighbor, NN)分类器并预测最终结果。其中,使用的近邻中心取为1.2节中的训练的代理集 Proxy-Set。这是为了减小由于数据不均衡造成的影响,并提高分类的准确性。

在图2(b)中展示了当MLP分类器的预测值小于阈值时的处理流程。在这种情况下,图像特征将被送入NN分类器,并由该分类器输出最终的预测结果。NN分类器预测待分类图像与所有类代理的相似度,然后将该图像确定为与之最为相似的类中心点所在的类别。

这种分类方法的设计在面对数据不均衡的情况下都能得到合理的分类结果,从而提高了整体分类的准确性。在实验中,讨论和确定了阈值的设定,以确保分类结果的准确性。

2 实验结果及分析

2.1 数据集及预处理

在医学图像领域,相同器官的不同种类疾病往往有很高的相似度,并且由于罕见病存在,医学图像数据集通常是典型的不均衡的细粒度图像数据集,本文选用了两个典型的医学领域不均衡的细粒度图像数据集验证本文方法。

本文分别在2019年眼科疾病智能识别国际竞赛提供的ODIR5K公开数据集^[13]和COVID Radiography数据集^[14]上验证了所提出的方法。ODIR5K数据集涵盖了8种不同类型的眼底图片。这8种类型分别是正常(Normal)、糖尿病(Diabetes)、青光眼(Glaucoma)、白内障(Cataract)、年龄相关性黄斑变性(AMD)、高血压(Hypertension)、近视(Myopia)和其他疾病(Other)。COVID Radiography数据集包含了4种不同类型的肺部X光图片,分别为正常(Normal)、肺部浑浊(Lung Opacity)、新冠肺炎(COVID)、病毒性肺炎(Viral Pneumonia)。图3展示了这两种数据集中不同类型的图像。

由于数据集上的图像大小不同,为了保证图像的质量,对数据集进行了预处理。根据专家标注信息手动去除无效图像,如图像偏移、图像质量

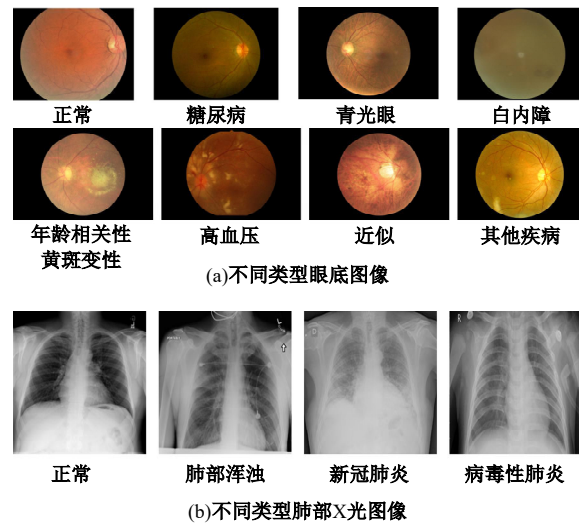


图3 数据集里不同类型的图像

Fig. 3 Different types of images in the dataset

差等。裁剪了黑色边框以减少这种负面影响。该过程包括对图像进行背景前景分割,定位包含病变的前景区域,并根据前景的位置和大小调整图像的大小。所有这些操作均采用自动方式完成。最后,为了方便网络训练,通过归一化将所有图像缩放到[0,1]范围。

从图4(a)中可以看出,ODIR5K数据集中不同疾病的图像数量分布表现出明显的不均衡情况。其中,正常眼底图像约有3 000张,糖尿病视网膜病变图像约有1 500张,其他疾病图像约有

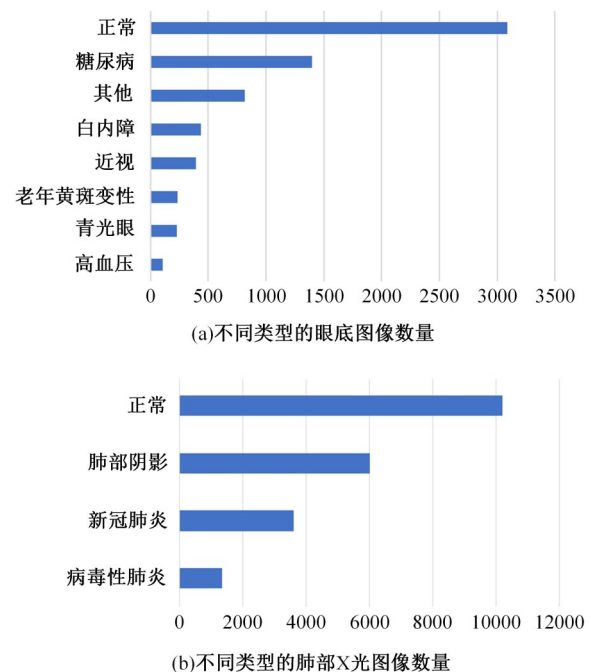


图4 数据集里各类型图像数量

Fig. 4 Number of images of each type in the dataset

1 000 张。剩余 5 个类别中,图像数量都不到 500 张,而其中高血压的图像最少,仅 104 张。从图 4 (b)中可以看出 COVID Radiography 数据集中同样存在数据不均衡的情况,其中正常和肺部浑浊图像共有约 17 000 张,病毒性肺炎图像数量最少约 1 300 张。上述两个数据集所展现出的样本不均衡情况在图像处理领域被认为是一个典型的长尾问题,即少数类别的样本数量明显少于多数类别样本的数量,这种样本不均衡的现象可能会对模型训练和性能产生负面影响。

2.2 实验设置

实验中,使用了 PyTorch 框架和 PyCharm 编译器,并在 NVIDIA GeForce 2080Ti GPU 上进行了训练。其中,70% 的数据用于训练,30% 的数据用于测试。在训练过程中,使用了 ImageNet 预先训练的权重微调模型。模型的输入大小为 224×224 ,优化器为 Adam,设置初始学习率为 0.000 01,每 10 个 epoch 衰减 0.1 倍。Batch-size 大小设置为 32,并设置了 40 个 epoch。在模型经过 1 500 次迭代后加入代理损失。

研究了不同的度量学习损失函数下,当改变嵌入大小从 16 到 1 024 时,网络性能是如何变化的。从表 1 可以明显地观察到模型的性能从 32 开始增加,在 128 处趋于稳定,并在 128 后性能开始下降,因此最终将 Embedding Size 的大小设置为 128。

表 1 不同 Embedding Size 在各度量损失函数下的分类性能

Table 1 Classification performance of different embedding size under each metric loss function

Embedding Size	Multi-similarity		Proxy-NCA	
	ODIR5K	COVID Radiography	ODIR5K	COVID Radiography
16	94.63	93.82	90.87	92.85
32	94.24	93.37	91.03	93.82
64	93.77	94.24	91.10	93.96
128	94.94	95.62	91.13	94.35
256	93.31	94.56	90.28	94.21
512	93.85	94.27	91.59	93.62
1 024	94.01	93.66	90.86	93.24

2.3 消融实验

为了兼顾代理损失和成对损失的优点,降低训练过程的复杂性,实现更快的收敛速度,同时兼顾数据点之间的细粒度语义关系,最终采用了

Proxy-NCA 损失函数和多相似性损失函数训练神经网络。从表 2 可以看出,本文方法不仅在性能上有着良好的表现,而且兼顾了模型的收敛速度。

表 2 不同度量损失在 ODIR5K 数据集上的性能及收敛速度

Table 2 Performance and convergence rate of different metric losses on ODIR5K dataset

损失函数	准确率/%	收敛迭代次数
Triplet+CE	91.05	2.0 k
Contrastive+CE	93.31	1.9 k
Multi-similarity+CE	93.52	2.5 k
Proxy-NCA+CE	92.81	1.3 k
Proxy-Anchor+CE	91.13	1.0 k
本文方法	95.63	1.5 k

探究了不同度量学习损失函数在不同衡量指标下的性能表现,表 3 展示了不同损失函数在 ODIR5K 数据集上的性能表现,表 4 展示了不同损失函数在 COVID Radiography 数据集上的性能表现。

表 3 不同度量损失函数在 ODIR5K 数据集上的表现

Table 3 Performance of different metric loss functions on the ODIR5K dataset

损失函数	准确率	F_1	精准率	召回率
Cross Entropy	92.22	92.50	93.20	92.00
Triplet+CE	91.05	90.50	95.50	87.00
Contrastive+CE	93.31	93.70	95.20	92.30
Multi-Similarity+CE	93.52	94.73	97.32	92.82
Proxy-NCA+CE	92.81	93.60	96.80	91.40
Proxy-Anchor+CE	91.13	91.60	93.40	90.10
本文方法	95.63	95.30	97.60	93.30

表 4 不同度量损失函数在 COVID Radiography 数据集上的表现

Table 4 Performance of different metric loss functions on the COVID Radiography dataset

损失函数	准确率	F_1	精准率	召回率
Cross Entropy	94.19	93.47	96.01	94.10
Triplet+CE	92.94	91.98	96.93	90.12
Contrastive+CE	95.21	94.97	96.24	95.64
Multi-Similarity+CE	95.62	96.52	98.17	95.79
Proxy-NCA+CE	94.35	95.62	97.78	95.92
Proxy-Anchor+CE	92.66	93.61	95.21	93.62
本文方法	96.57	93.40	95.60	91.30

分别验证了 MLP 分类器、NN 分类器和本文所提出的方法在 ODIR5K 数据集和 COVID Radiography 数据集分类任务上的表现,并分别探究了 3 种分类器在不同数据集上不同类别样本上的性能表现。从表 5 中可以看出,本文方法不仅在样本数量较多的类别中表现优异,而且在高血压、青光眼和年龄相关性黄斑变性等样本较少的类别中准确率显著提高。从表 6 中可以看出,在新冠肺炎和病毒性肺炎等样本较少的类别中准确率较高。这表明了本文提出的方法在面对不均衡的细粒度数据时具有出色的适应能力,同时也很好地突出了本文提出的方法在处理样本稀缺的细粒度图像类别时的潜力和价值。

表 5 不同分类器在 ODIR5K 数据集上的表现
Table 5 Performance of different classifiers on the ODIR5K dataset %

类别	准确率		
	本文	MLP	NN
年龄相关性黄斑变性	99.9	95.9	94.8
白内障	100	99.1	96.5
糖尿病	96.8	88.2	91.5
青光眼	99.5	93.0	88.7
高血压	99.8	83.7	92.7
近视	99.9	97.0	99.0
正常	95.2	93.6	93.8
其他疾病	98.2	89.3	92.5

表 6 不同分类器在 COVID Radiography 数据集上的表现
Table 6 Performance of different classifiers on the COVID Radiography dataset %

类别	准确率		
	本文	MLP	NN
正常	95.58	93.3	97.3
肺部浑浊	96.86	96.5	97.8
新冠肺炎	97.50	90.7	94.6
病毒性肺炎	96.60	96.8	89.8

阈值分类器的性能会受到阈值设定的影响。为了深入研究在不同阈值下阈值分类器性能的变化,进行了详尽的研究,并将结果整理在表 7、表 8 中。表 7 展示了在 ODIR5K 数据集中不同阈值下的分类器性能,表 8 展示了在 COVID Radiography 数据集中不同阈值下的分类器性能。从表 7、表 8 中可以看出,随着阈值的增大准确率、 F_1 、精准率、召回率相关衡量指标也会上升,当阈值超过 0.48 时,准确率、 F_1 、精准率、召回率相关衡量指

标会随着阈值的增大而逐步降低。从表 7、表 8 中可以看到当分类器阈值为 0.48 时,模型分类性能达到最佳。因此,最终将阈值分类器的阈值大小设定为 0.48。

表 7 在 ODIR5K 数据集上不同阈值下的分类器性能
Table 7 Classifier performance under different thresholds on the ODIR5K dataset

阈值	准确率	F_1	精准率	召回率
0.46	87.50	89.29	85.00	92.31
0.47	89.36	90.41	87.94	93.65
0.48	95.63	95.30	97.60	93.30
0.49	92.60	94.32	91.87	96.21
0.50	90.96	93.62	91.18	96.43
0.51	80.97	85.64	80.15	92.67

表 8 在 COVID Radiography 数据集上不同阈值下的分类器性能

Table 8 Classifier performance under different thresholds on the COVID Radiography dataset

阈值	准确率	F_1	精准率	召回率
0.46	89.51	90.61	87.20	94.30
0.47	91.38	92.61	89.87	95.54
0.48	96.57	93.40	95.60	91.30
0.49	94.56	95.56	92.85	98.44
0.50	92.91	95.10	92.21	98.19
0.51	82.82	87.91	82.13	94.58

2.4 实验结果和分析

使用了常见的准确率(Accuracy)、 F_1 、精准率(Precision)、召回率(Recall)作为性能评价指标。这些评价指标旨在提供对模型性能的深入了解,以便更全面地分析和比较不同方法所展现出的效果。

2.4.1 评价指标及结果

通过分析表 9 和表 10 中数据可以明显地观察到,在准确率、 F_1 、精准率、召回率这 4 项评价指标结果中,本文方法在各个类别中都有良好的表现;综合来看这 4 项评价指标的结果,本文方法在各个类别中都表现出优秀的性能,这表明该方法并非仅依赖于偏向多数类别提高评价指标的结果,而是能够有效克服不均衡数据所带来的挑战。这表明了本文提出的方法在处理不均衡数据问题上的有效性和可靠性。本文模型不仅能够平衡各个类别之间的性能表现,而且不会受到数据不均衡的影响,这对处理不均衡的细粒度图像数据集的分类任务具有重要意义。这个结果也进一步验证

表 9 在眼底数据集中不同类别下的衡量指标结果

Table 9 Outcomes of metrics under different categories in the fundus dataset %

类别	准确率	F_1	精准率	召回率
年龄相关性黄斑变性	99.9	98.9	100.0	97.9
白内障	100.0	100	100.0	100.0
糖尿病	96.8	92.3	96.9	88.2
青光眼	99.5	91.8	100.0	84.8
高血压	99.8	92.7	95.0	90.5
近视	99.9	99.0	100.0	98.0
正常	95.2	95.2	91.6	99.0
其他疾病	98.2	92.6	97.3	88.3

表 10 在肺炎数据集中不同类别下的衡量指标结果

Table 10 Outcomes of metrics under different categories in the pneumonia dataset %

类别	准确率	F_1	精准率	召回率
正常	95.58	97.83	99.12	96.58
肺部浑浊	96.86	96.36	95.60	97.15
新冠肺炎	97.50	98.94	98.33	99.63
病毒性肺炎	96.60	93.33	92.72	93.96

了本文方法在处理不均衡细粒度图像数据集方面的实用性和潜力。

2.4.2 与其他方法的对比

表 11 和表 12 展示了不同方法在 ODIR5K 数据集和 COVID Radiography 数据集上的表现,现有方法大多数都没有能够在很好地挖掘图像细粒度属性的同时,关注到样本不均衡的问题,可以看到在不同评价指标上本文方法都优于现有方法。

通过对表 11 和表 12 中不同方法在 ODIR5K

表 11 不同方法在 ODIR5K 数据集上的表现

Table 11 Performance of different methods on the ODIR5K dataset %

方法	准确率	F_1	精准率	召回率
EfficientNetB3 ^[15]	92.00	—	71.00	66.00
MCGS-Net ^[16]	—	89.66	65.88	61.60
BFPC-Net ^[17]	94.23	94.16	97.09	93.23
DSRACNN ^[18]	87.90	88.16	88.50	—
本文	95.63	95.30	97.60	93.30

表 12 不同方法在 COVID Radiography 数据集上的表现

Table 12 Performance of different methods on the COVID Radiography dataset

方法	准确率	F_1	精准率	召回率
COVID-CAPS ^[19]	95.70	—	—	—
DRNN ^[20]	92.1	91.1	93.01	—
本文	96.57	93.40	95.60	91.30

和 COVID Radiography 数据集上的表现进行比较分析,可以看到在 ODIR5K 数据集上,本文方法在准确率、 F_1 、精准率和召回率方面都表现出色,分别达到了 95.63%、95.30%、97.60% 和 93.30%。相较之下,EfficientNetB3、MCGS-Net 和 DSRACNN 等方法在其中一些指标上表现较弱,而 BFPC-Net 在准确率和精准率上表现良好,但召回率相对较低。这表明本文方法在 ODIR5K 数据集上具有更全面的性能,不仅在准确性方面表现卓越,而且在处理不同评价指标时能取得优异的成绩;在 COVID Radiography 数据集上,本文方法同样取得了卓越的表现,准确率、 F_1 、精准率和召回率分别达到了 96.57%、93.40%、95.60% 和 91.30%。相较之下,COVID-CAPS 和 DRNN 方法在准确率和 F_1 上表现较好,但在精准率和召回率上可能存在一些不足。本文方法在各项指标上均取得了最佳的性能表现,在 COVID Radiography 数据集上显示出了更强的分类能力。

本文方法不仅在图像细粒度属性的挖掘方面表现出色,而且在处理样本不均衡问题时表现出优异的性能,未来的研究可以进一步验证和探究所提出的方法在更广泛和复杂数据上的适用性和稳定性。

3 结束语

本文提出了一种基于度量学习的阈值分类框架,引入了度量学习方法,提出了一种更有效的阈值分类算法,以提升不均衡细粒度图像分类任务的性能。基于以上改进,在 ODIR5K 数据集和 COVID Radiography 数据集上,阈值分类方法与其他几种方法相比,取得了进步。未来的工作中,将进一步致力于特征提取环节,进一步提升模型的性能。

参考文献:

- [1] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2016: 770-778.
- [2] Huang Z Z, Zhang J P, Shan H M. When age-invariant face recognition meets face age synthesis: a multi-task learning framework[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Kuala Lumpur, Malaysia,

- 2021: 7282-7291.
- [3] Ji R, Wen L, Zhang L, et al. Attention convolutional binary neural tree for fine-grained visual categorization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 10468-10477.
- [4] Wei X S, Xie C W, Wu J, et al. Mask-CNN: localizing parts and selecting descriptors for fine-grained bird species categorization[J]. *Pattern Recognition*, 2018, 76: 704-714.
- [5] Zheng H, Fu J, Zha Z J, et al. Learning deep bilinear transformation for fine-grained image representation[J]. *Advances in Neural Information Processing Systems*, 2019, 32: No. 03621.
- [6] Chang D, Ding Y, Xie J, et al. The devil is in the channels: Mutual-channel loss for fine-grained image classification[J]. *IEEE Transactions on Image Processing*, 2020, 29: 4683-4695.
- [7] Bera A, Wharton Z, Liu Y, et al. SR-GNN: spatial relation-aware graph neural network for fine-grained image categorization[J]. *IEEE Transactions on Image Processing*, 2022, 31: 6017-6031.
- [8] Sundgaard J V, Harte J, Bray P, et al. Deep metric learning for otitis media classification[J]. *Medical Image Analysis*, 2021, 71: No. 102034.
- [9] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[J/OL]. [2023-08-11]. <https://arxiv.org/abs/2010.11929>
- [10] Guo H, Wang S. Long-tailed multi-label visual recognition by collaborative training on uniform and re-balanced samplings[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Kuala Lumpur, Malaysia, 2021: 15089-15098.
- [11] Movshovitz-Attias Y, Toshev A, Leung T K, et al. No fuss distance metric learning using proxies[C]//Proceedings of the IEEE International Conference on Computer Vision, Hawaii, USA, 2017: 360-368.
- [12] Wang X, Han X, Huang W, et al. Multi-similarity loss with general pair weighting for deep metric learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Los Angeles, USA, 2019: 5022-5030.
- [13] International Competition on Ocular Disease Intelligent Recognition[EB/OL]. [2021-11-18]. <https://odir2019.grand-challenge.org/dataset/>
- [14] Rahman T, Khandakar A, Qiblawey Y, et al. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images[J]. *Computers in Biology and Medicine*, 2021, 132: No. 104319.
- [15] Wang J, Yang L, Huo Z, et al. Multi-label classification of fundus images with efficientnet[J]. *IEEE Access*, 2020, 8: 212499-212508.
- [16] Lin J, Cai Q, Lin M. Multi-label classification of fundus images with graph convolutional network and self-supervised learning[J]. *IEEE Signal Processing Letters*, 2021, 28: 454-458.
- [17] Li Z, Xu M, Yang X, et al. Multi-label fundus image classification using attention mechanisms and feature fusion[J]. *Micromachines*, 2022, 13(6): No. 947.
- [18] Yang X, Yi S. Multi-classification of fundus diseases based on DSRA-CNN[J]. *Biomedical Signal Processing and Control*, 2022, 77: No. 103763.
- [19] Afshar P, Heidarian S, Naderkhani F, et al. Covid-caps: a capsule network-based framework for identification of COVID-19 cases from X-ray images[J]. *Pattern Recognition Letters*, 2020, 138: 638-643.
- [20] Panahi A, Askari M R, Akrami M, et al. Deep residual neural network for COVID-19 detection from chest X-ray images[J]. *SN Computer Science*, 2022, 3(2): No. 169.