



上下文空间与实例信息的皮肤镜图像自监督分类

刘嘉轩, 胡非易, 张辉, 张金洲, 李玲

引用本文:

刘嘉轩,胡非易,张辉,张金洲,李玲. 上下文空间与实例信息的皮肤镜图像自监督分类[J]. 智能系统学报, 2023, 18(4): 783–792.
LIU Jiaxuan,HU Feiyi,ZHANG Hui,ZHANG Jinzhou,LI Ling. Dermoscopic images classification based on context and instance-level feature of self-supervised learning[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(4): 783–792.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202211010>

您可能感兴趣的其他文章

融合迁移学习和神经网络的皮肤病诊断方法

A skin diseases diagnosis method combining transfer learning and neural networks
智能系统学报. 2020, 15(3): 452–459 <https://dx.doi.org/10.11992/tis.201811015>

海底声呐图像智能底质分类技术研究综述

Survey of the intelligent seabed sediment classification technology based on sonar images
智能系统学报. 2020, 15(3): 587–600 <https://dx.doi.org/10.11992/tis.202004026>

融合整体与局部信息的武夷岩茶叶片分类方法

Classification of Wuyi rock tealeaves by integrating global and local information
智能系统学报. 2020, 15(5): 919–924 <https://dx.doi.org/10.11992/tis.202003018>

基于注意力融合的图像描述生成方法

An image caption generation method based on attention fusion
智能系统学报. 2020, 15(4): 740–749 <https://dx.doi.org/10.11992/tis.201910039>

基于深度学习的椎间孔狭窄自动多分级研究

Deep learning based automatic multi-classification algorithm for intervertebral foraminal stenosis
智能系统学报. 2019, 14(4): 708–715 <https://dx.doi.org/10.11992/tis.201806015>

SUCE:基于聚类集成的半监督二分类方法

SUCE: semi-supervised binary classification based on clustering ensemble
智能系统学报. 2018, 13(6): 974–980 <https://dx.doi.org/10.11992/tis.201711027>

DOI: 10.11992/tis.202211010

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20230321.1924.029.html>

上下文空间与实例信息的皮肤镜图像自监督分类

刘嘉轩¹, 胡非易², 张辉³, 张金洲¹, 李玲¹

(1. 长沙理工大学 电气与信息工程学院, 湖南 长沙 410114; 2. 湖南大学 电气与信息工程学院, 湖南 长沙 410000; 3. 湖南大学 机器人学院, 湖南 长沙 410012)

摘要: 针对基于深度学习的皮肤疾病检测方法受限于皮肤镜图像标注难度大、成本高、耗时耗力等问题, 本文提出了一种针对皮肤镜图像的自监督学习方法, 利用无标注的皮肤镜图像来提高皮肤疾病分类准确率。根据皮肤镜图像空间结构相似度高、信息单一的特点, 使用基于旋转预测自监督代理任务, 通过模型预测图像旋转的角度, 约束模型对图像病变区域的关注, 学习上下文空间特征; 同时使用个体判别任务, 利用正负样本对比学习, 获得皮肤镜图像实例信息, 为皮肤镜图像的分类提供全局分类信息指导; 融合上下文空间特征与实例语义信息得到初始的皮肤疾病检测模型。实验结果表明: 本文所提的自监督学习方法, 从无标记的皮肤镜样本挖掘出重要信息, 更有效地关注皮肤镜图像病变区域, 通过融合皮肤镜图像特征信息, 使分类结果更加准确。

关键词: 皮肤病诊断; 自监督学习; 特征提取; 图像分类; 迁移学习; 卷积神经网络; 医学图像处理; 辅助诊断
中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2023)04-0783-10

中文引用格式: 刘嘉轩, 胡非易, 张辉, 等. 上下文空间与实例信息的皮肤镜图像自监督分类 [J]. 智能系统学报, 2023, 18(4): 783-792.

英文引用格式: LIU Jiaxuan, HU Feiyi, ZHANG Hui, et al. Dermoscopic images classification based on context and instance-level feature of self-supervised learning[J]. CAAI transactions on intelligent systems, 2023, 18(4): 783-792.

Dermoscopic images classification based on context and instance-level feature of self-supervised learning

LIU Jiaxuan¹, HU Feiyi², ZHANG Hui³, ZHANG Jinzhou¹, LI Ling¹

(1. School of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410114, China; 2. School of Electrical and Information Engineering, Hunan University, Changsha 410000, China; 3. School of Robotics, Hunan University, Changsha 410012, China)

Abstract: The skin disease detection method based on deep learning is limited by the difficulty of annotations, high cost, time consuming and labor-intensive labeling of dermoscopy images. This paper proposes a self-supervised learning method for dermoscopy images, which uses unmarked dermoscopy images to improve the classification accuracy of skin diseases. According to the characteristics of high similarity and single information of the spatial structure of dermoscopic images, a rotation prediction pretext task is designed to predict the rotation angle of the image through the model, so as to constrain the model's attention to pathological areas of the image, and learn the context feature. At the same time, the paper presents an instance discrimination task and comparative learning of positive and negative instances to obtain the instance information of dermatoscope images, which provides guidance for global classification information in the classification of dermatoscope images. The initial skin disease detection model is obtained by fusing context features and instance semantic information. The experimental results show that the self-supervised learning method proposed in this paper can mine important information from unmarked dermoscopy samples, pay more attention to the lesion area of dermoscopy images, and make the classification results more accurate by fusing the characteristic information of dermoscopy images.

Keywords: skin disease diagnosis; self-supervised learning; feature extraction; image classification; transfer learning; convolutional neural network; medical image processing; auxiliary diagnosis

收稿日期: 2022-11-10. 网络出版日期: 2023-03-22.

基金项目: 国家重大研究计划重点支持项目(92148204); 科技创新 2030“新一代人工智能”重大项目(2021ZD0114503); 国家自然科学基金项目(61971071, 62027810, 62133005); 湖南省杰出青年科学基金项目(2021JJ10025); 湖南省重点研发计划(2021GK4011, 2022GK2011); 长沙市科技重大专项(kh2003026); 机器人学国家重点实验室联合开放基金项目(2021-KF-22-17); 中国高校产学研创新基金项目(2020HYA06006).

通信作者: 张辉. E-mail: zhanghuihy@126.com.

皮肤病变通常是指黑色素、棕色色素导致皮肤呈现棕色、黑色甚至蓝色, 被称为色素沉着皮肤病。虽然大多数情况下是良性的, 但在某些情况下, 色素沉着皮肤病可能是恶性的皮肤癌症。据美国癌症协会统计, 有将近1/5的美国人会在一生中患上皮肤癌^[1], 并且过去 30 年中皮肤癌的数量

比其他癌症数量之和还多^[2]。恶性皮肤癌具有恶性程度高、转移速度快、难以发现等特点。因此,研究一种有效的皮肤疾病检测技术,精准并且快速地早期诊断皮肤癌症有着重要的现实意义。

随着人工智能深度学习分析方法在计算机视觉领域兴起^[3],基于深度学习的医学图像分析技术在X、计算机断层扫描(computed tomography, CT)、磁共振成像(magnetic resonance imaging, MRI)、核医学和皮肤镜图像等多个领域^[4]得到了进一步的研究推广,甚至一定程度上超过了人类医生的水平。通过计算机辅助诊断系统(computer aided design, CAD)^[5]分析海量的皮肤镜图像,不但可以节约医生时间和人力成本,而且提升患者的治愈几率^[6]。但期望将基于深度学习的皮肤疾病检测方法部署于实际的应用下,仍然存在着诸多的难题有待解决。首先,皮肤镜样本数据少,并且无法人工合成^[7]。并且受限于伦理道德,医院无法直接与科研机构交换数据,因此整体皮肤镜样本数据较少。其次,不同于自然图像,医学图像标注难度大、成本高、耗时耗力,医学图像的标注往往需要相关从业人员进行标注^[8]。因此,研究一种无监督方法,从无标注皮肤镜图像挖掘先验信息,协助皮肤病变检测是非常有必要的。

近年来,随着自监督学习技术在自然图像上的应用与发展,自监督学习的分类成为图像分析的主流^[9-10]。自监督学习是一种具有特殊监督形式的无监督学习方法,自监督学习使用数据集本身的信息来构造伪标签从而进行表征学习(representation learning, RL)^[11]。自监督学习属于无监督学习的分支,因此不需要耗时耗力、并且有噪声干扰的手工图像标注。因此,在目前缺乏大规模医学影像数据集的情况下,运用自监督学习方法尝试从无标记的数据中挖掘出具有鉴别性的信息,是切合当前医学图像领域资源特点的有效途径。但是,针对自监督学习分类模型的研究大多数集中于自然图像的分类和目标检测上^[12],相比于自然图像,皮肤疾病检测显然更加棘手,其难点主要如下:1)皮肤镜图像的比例与自然图像差异大,图像包含大量相似的空间结构(病变区域总是在皮肤镜图像中央),病变区域关键信息占比少,皮肤疾病分类可能取决于少量像素;2)与随手可得的自然图像相比,皮肤疾病图片获取相对困难,数量较少,并且标注相对更困难;3)皮肤镜图像与自然图片存在色彩、光照等差异。

针对上述难点,本文提出一种针对色素沉着皮肤病分类任务的自监督学习方法,基于个体判别与旋转代理任务从海量无标注皮肤镜图像挖掘

数据先验知识。该方法主要贡献包括:1)将色素沉着皮肤病变数据中的上下文空间特征与实例级别语义特征深度融合,自监督学习皮肤镜图像的特征。2)对方法中的特征学习代理任务进行解释,进一步阐明学习机理。3)在基准皮肤镜图像数据集与小样本数据集上进行实验,进一步证明本文提出的自监督学习技术增强了皮肤疾病识别模型的检测能力。

1 融合特征的自监督分类方法

本文提出了一种基于旋转预测任务和个体判别任务自监督学习分类方法,利用大量的无标注图片,解决皮肤疾病分类中存在的难题。自监督学习框架如图1所示。根据自监督学习框架,本文针对皮肤镜图像中空间结构相似这一特性,利用旋转预测任务学习上下文空间信息,约束模型集中关注于图像前景(皮肤病灶区域);并融合个体判别任务(instance discrimination, InstDisc),对比正样本和负样本来学习图像表征,为皮肤镜图像的分类提供全局分类信息指导,整体框架如图2所示。

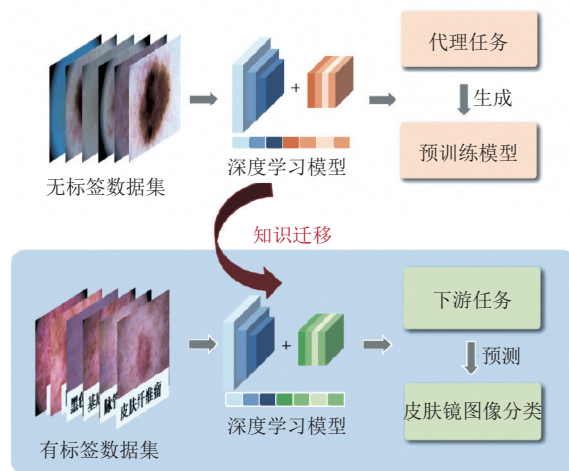


图1 自监督学习框架

Fig. 1 Self-supervised learning framework

1.1 旋转预测任务

在皮肤镜图像数据中,图像的空间结构较为相似,图像前景为皮肤病变区域,图像后景为正常皮肤组织。图像前景相对复杂并且对于皮肤疾病分类较为重要,图像后景相对单一、简单。为此,根据数据空间结构相似的特点,使用旋转任务产生旋转伪标签 $y \in \{0, 1, 2, 3\}$ ($0^\circ, 90^\circ, 180^\circ, 270^\circ$)。通过模型预测图像旋转的角度,以促使模型关注图像病变区域(前景)来判断图像旋转角度,从无标注皮肤镜图像学习空间语义信息(context information, CI)^[13]。具体来说,图片 x_i 首先经过随机

旋转的图像变化得到 $\mathbf{x}_{i,y}$, 其中 $y \in \{0,1,2,3\}$ 作为模型^[14] (ResNet18) 的输入随后输出特征 $\mathbf{f}_{i,y}$ 。此后, 利用全连接层再次将特征 $\mathbf{f}_{i,y} \in \mathbf{R}^{H \times C \times 512}$ 降维为 $\mathbf{f}_{i,y} \in \mathbf{R}^4$ 。用这样的方式, 本文将原始的三维特征转化为用于判断旋转预测的一维特征, 得到用于旋转预测的特征图。特征图输入到Sigmoid激活函数

将特征图数值压缩到 0~1, 并使用交叉熵损失计算梯度为

$$L_R = \frac{1}{N} \sum_{i=1}^N \sum_{y=0}^3 l(\text{Sigmoid}(\mathbf{f}_{i,y}), y) \quad (1)$$

式中: l 为交叉熵损失, $y \in \{0,1,2,3\}$ 为旋转伪标签。

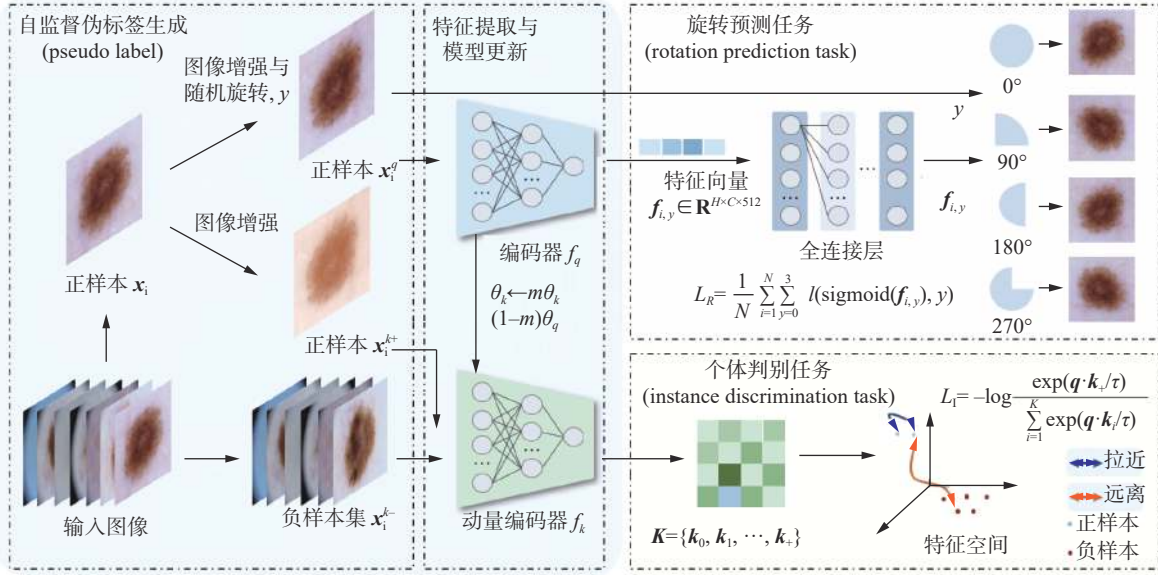


图2 基于旋转任务和个体判别任务的自监督学习分类方法整体结构

Fig. 2 Diagram of self-supervised learning classification method based on rotation task and instance discrimination task

1.2 个体判别任务

个体判别任务 (instance discrimination) 启发于有监督学习, 旨在将每一张图片都视为一个类别, 约束模型学习每一张图片的特征, 从而把图片都区分开来。个体判别任务的核心目标是让模型学习“如何去对比”, 即如何抓住样本的本质特征从而区分图片, 促使模型学习到实例级别语义信息^[15]。目前的个体判别任务^[15]一般使用负对比估计 (negative contrastive estimation, NCE) 损失计算模型梯度, 具体为

$$L_{\text{NCE}} = E_{\mathbf{x}^+, \mathbf{x}^-} \left[-\log \left(\frac{e^{f(\mathbf{x}^+)^\top f(\mathbf{x}^-)}}{e^{f(\mathbf{x}^+)^\top f(\mathbf{x}^+)} + e^{f(\mathbf{x}^-)^\top f(\mathbf{x}^-)}} \right) \right] \quad (2)$$

式中: \mathbf{x}^+ 为与输入 \mathbf{x} 相似的样本, \mathbf{x}^- 为与 \mathbf{x} 不同的样本。

首先, 本文设计了符合皮肤镜数据的图像增强策略, 对于图片 \mathbf{x}_i , 使用图像调整统一大小、随机翻转、亮度、对比度、饱和度、色调、灰度变化的图像增强策略得到正样本 \mathbf{x}_i^q 和 \mathbf{x}_i^{k+} , 而与图片 \mathbf{x}_i 不相似的图片则为负样本 \mathbf{x}_i^{k-} 。本文使用编码器 f_q (ResNet18) 编码 \mathbf{x}_i^q 得到特征编码 $\mathbf{q} = f_q(\mathbf{x}_i^q)$ 。并且为了保证个体判别任务学习的效率, 使用动量编码器 f_k 字典特征编码 $\mathbf{k} = f_k(\mathbf{x}_i^k)$, 从而得到一个大并

且动态的字典 $\mathbf{K} = \{\mathbf{k}_0, \mathbf{k}_1, \dots, \mathbf{k}_K\}$ 。其次, 使用对比损失来计算模型梯度, 当来自同一输入图像的图像表征相似时, 对比损失将降低。本文使用 InfoNCE^[16] 损失计算模型对比损失:

$$L_I = -\log \frac{\exp(\mathbf{q} \cdot \mathbf{k}_+/ \tau)}{\sum_{i=0}^K \exp(\mathbf{q} \cdot \mathbf{k}_i / \tau)} \quad (3)$$

式中: τ 为损失温度超参数, K 为负样本数量。

最后, 个体判别任务需要得到大量并且一致的负样本, 从而避免模型通过不同批次的负样本学习到捷径解, 本文应用了动量编码器来编码负样本, 利用动量更新来更新动量编码器为

$$\theta_k \leftarrow m\theta_k + (1-m)\theta_q \quad (4)$$

式中: θ_q 为正样本编码器的参数, θ_k 为动量编码器的参数, $m \in (0,1)$ 为动量更新系数。

1.3 模型介绍

1.3.1 损失函数

模型的总损失函数是旋转预测任务与个体判别任务的线性组合, 模型的损失函数表示为

$$L = \lambda \cdot L_R + L_I \quad (5)$$

式中 λ 为旋转预测的加权因子。在实验中, 本文

设定超参数 $\lambda = 0.1$, 同时本文在消融实验中分析了 λ 的影响。

1.3.2 模型架构

模型的编码器网络基于 ResNet18^[14], 并且本文将 ResNet18 最后一层残差块的池化层换为自适应池化层。当图片 x_i 通过主干网络之后, 特征图将会转为一个向量, 分别通过全连接层、批量归一化、Relu 激活函数, 并使用全连接层降维至 128 和 4, 从而得到特征 f_i 和 f_R 。特征 f_i 的维度与之前工作^[15-16] 保持一致, 特征 f_R 将通过 Softmax 激活函数产生 4 个概率分布来预测图像旋转角度。

1.4 算法流程

在自监督学习阶段, 无监督学习上下文空间特征与实例语义特征, 挖掘出具有鉴别性的信息。随后, 使用 K 最近邻 (K-nearest neighbors, KNN) 分类器与迁移学习来验证方法学习到的特征。

输入 训练数据集 $D = \{x_n\}_{n=1}^N$ 、ResNet18 模型结构、训练次数、学习率的初始值和衰减系数。

1) 初始化编码器 θ_q 和动量编码器 θ_k , 产生旋转角度标签 $y \in \{0, 1, 2, 3\}$;

2) 取数据集中一张图片经旋转变换以及图像增强得到的图片 x_i^q , 同时进行图像增强得到的第 2 张图片 x_i^{k+} , 其他图片作为负样本 x_i^{k-} ;

3) 将旋转变换以及图像增强得到的图片 x_i^q 输入网络, 进行网络正向传播, 预测旋转角度, 计算旋转后交叉熵损失 L_R ;

4) 将 x_i^{k+} 以及 x_i^{k-} 输入动量编码器得到动态字典 $K = \{k_0, k_1, \dots, k_n\}$, 根据编码器提取的特征与动态字典包含的特征进行正负样本的判别, 计算对比损失 L_i ;

5) 梯度下降, 编码器和动量编码器参数更新;

6) 判断是否达到预设的训练次数, 若达到训练

次数, 则停止训练, 若没有达到训练次数, 则返回 2);

输出 参数 (θ_q) 已知的 ResNet18 网络结构。

2 实验与结果分析

本节进行大量实验验证所提出方法的有效性。首先使用所提出的自监督方法与其他对比方法在预训练数据集上训练模型, 再采用 KNN 分类器与迁移学习验证本文方法在下游分类数据集上的有效性; 并使用不同比例的训练数据微调训练模型, 验证模型在小样本数据情况下的有效性。

2.1 数据集及运行环境配置

2.1.1 自监督训练数据集

本文所采用自监督训练数据集为无标注的皮肤镜数据国际皮肤成像协作组织 2020 年挑战 (ISIC challenge 2020) 数据集作为训练数据^[17]。实验选取 ISIC2020 中公开的 Training JPEG, 共包含不同种类的皮肤镜图像 33 126 张, 其中包含了恶性与良性病例皮肤镜数据。由于图像的分辨率不同, 因此将自监督训练数据大小调整至 224 像素 \times 224 像素。

2.1.2 下游分类数据集

本文所采用的皮肤疾病分类任务的数据集为色素沉着皮肤疾病 HAM10000 公共数据集^[18], 由 10 015 张皮肤镜图像组成。数据集包含日光化性角化病 (actinic keratoses)、基底细胞癌 (basal cell carcinoma)、良性角化样病变 (benign keratosis-like lesion)、皮肤纤维瘤 (dermatofibroma)、黑色素瘤 (dermatofibroma)、黑色素细胞痣 (melanocytic nevi) 和血管病变 (vascular lesions) 共 7 个类别的数据, 每幅图片包含其中对应的标签, 7 个类别具体如图 3 所示。由于每幅图像大小不一, 本文统一调整至 224 像素 \times 224 像素大小。

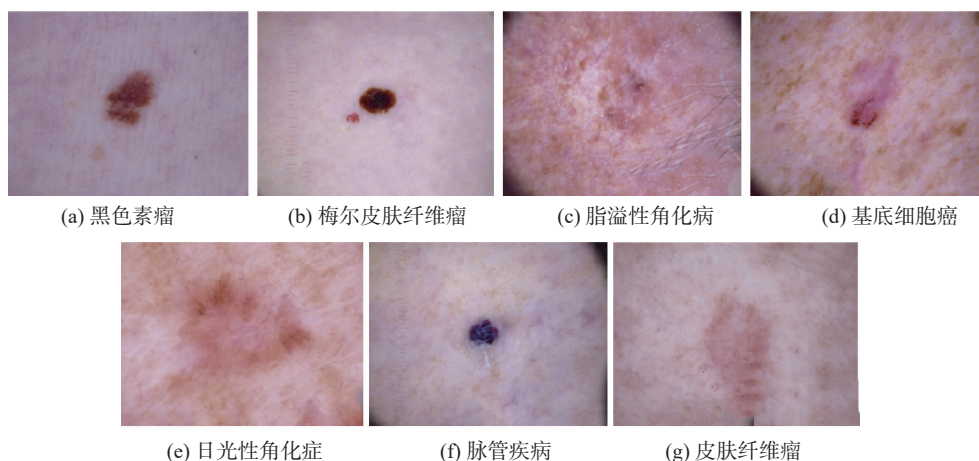


图 3 7 类典型的皮肤疾病图片

Fig. 3 Seven typical skin disease images

2.1.3 运行环境配置

实验环境在单机系统搭建,处理器为 Intel(R) Xeon W-2150B@3.00 GHz,显卡为 NVIDIA RTX 3090, 128 GB 内存,操作系统为 Ubuntu20.04,通过 Pytorch1.8.2 深度学习框架进行相关实验的运行。在自监督学习和皮肤疾病识别实验中,采用 ResNet18 作为自监督学习模型和皮肤疾病识别的基础模型。

2.2 评价指标

模型的评价指标采用准确率 (accuracy, ACC)、召回率 (Recall)、精确率 (Precision) 和 F_1 得分 (F_1 -score) 进行评估,准确率是指预测正确的次数占总次数的比例:

$$Acc = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (6)$$

召回率是指正确预测为正的占全部实际为正的的比例:

$$Rec = \frac{T_P}{T_P + F_N} \quad (7)$$

精确率是指正确预测为正的占全部预测为正的的比例:

$$P_{re} = \frac{T_P}{T_P + F_P} \quad (8)$$

F_1 得分是召回率与精确率的平均:

$$F_1 = \frac{2 \cdot P_{re} \cdot Rec}{P_{re} + Rec} \quad (9)$$

式中: T_P 为混淆矩阵中被模型预测为正类的正样本的数量, T_N 为被模型预测为负类的负样本数量, F_P 为被模型预测为正类的负样本数量, F_N 为被模型预测为负类的正样本数量。

2.3 对比试验

为验证本文提出的自监督方法在皮肤疾病数据集上的优越性,本文进行了多组对比实验并对不同的实验结果进行分析。对比实验包含3个部分:第1部分在下游分类数据集 HAM10000 上进行了实验,将所提出的方法与当前最先进的自然图像自监督方法进行对比,对比方法包括动量对比 (momentum contrast, MoCo)^[15]、简单对比学习的视觉表示 (simple contrastive learning of visual representations, SimCLR)^[19]、利用旋转预测学习视觉表示 (unsupervised representation learning by predicting image rotations, Rotation)^[20]、引导你自己的潜在表示 (bootstrap your own latent, BYOL)^[21] 和自监督学习不变视觉表示 (pretext-invariant representation learning, PIRL)^[22]。此外,基于有监督的方法,如在 ImageNet^[23] 和随机初始化的权重 Ran-

dom Weights 也被用以进行了对比实验;第2部分为小样本数据集上的对比实验,对比的方法与第1部分保持一致;第3部分为消融实验,对所提2种模块进行验证并分析结果,同时分析了部分重要超参数的设定。为了公平,所有方法都基于 ResNet18 的基础模型。

2.3.1 自监督方法对比实验

为了验证所提出的自监督方法的有效性,本文设计与基于自然图像自监督方法之间的对比实验。在实验的过程中,本文首先在自监督训练数据集上使用自监督学习训练模型,完成自监督训练后,再使用下游分类数据集分类皮肤疾病,验证模型学习得到的特征。在自监督学习阶段,训练选择的优化器为 (stochastic gradient descent, SGD)^[24],批大小设为 512,使用 0.06 的学习率,并使用余弦退火策略自适应调节学习率大小。完成自监督训练后,模型学习到许多用于皮肤疾病识别的重要特征。同时与之前的工作^[11,15,19]保持一致,本文使用 KNN 分类器与迁移学习来验证评估自监督方法学习到的特征。其中,KNN 分类器中的 k 设为 100,迁移学习的优化器为 SGD,批大小设为 64,使用 0.001 的学习率。KNN 分类的具体结果如表 1 所示,不难看出在 KNN 分类器上,本文提出的方法准确率为 74.55%,与当前最先进的自监督方法 Rotation、BYOL、MoCo、SimCLR 和 Invariant 对比,模型精度分别提升了 6.39%、2.09%、1.80%、1.50% 和 2.49%。同时,与使用有监督分类的 ImageNet 数据集训练模型实验结果表明,使用本文提出的自监督学习方法相比于有监督分类的 ImageNet 数据集训练模型提高了 5.79%,这是因为自然图像 (ImageNet 数据集) 与色素沉着皮肤病变数据集 (HAM10000 下游分类数据集) 的数据分布存在鸿沟,使用自监督学习的模型能够更好地使皮肤镜数据在高维空间中线性可分,而 ImageNet 训练的模型虽然采用了海量的自然图像,但由于数据分布与待测数据差异较大,限制了识别皮肤疾病的性能。

为了验证本文提出的自监督学习方法在迁移学习中的有效性,在自监督训练数据集上使用前文提到的基于自然图像的自监督方法进行预训练,完成自监督预训练后,再迁移至下游分类任务使用下游分类有监督数据集进行微调。实验使用 K 折交叉验证^[25],并在训练中选择 K 数值为 5,即 5 折交叉验证。具体的实验结果如表 2 所示。

表 1 KNN 分类结果
Table 1 Results of KNN

%

检测模型	Acc	Precision	Recall	F_1 -score
Random	68.26	52.49	15.08	56.13
ImageNet	68.76	56.50	15.64	57.48
Rotation	68.16	68.44	14.68	55.56
BYOL	72.46	69.63	23.05	65.03
MoCo	72.75	72.97	23.37	65.29
SimCLR	73.05	73.76	24.08	65.93
Invariant	72.06	66.74	24.41	65.34
本文模型	74.55	71.93	29.41	70.51

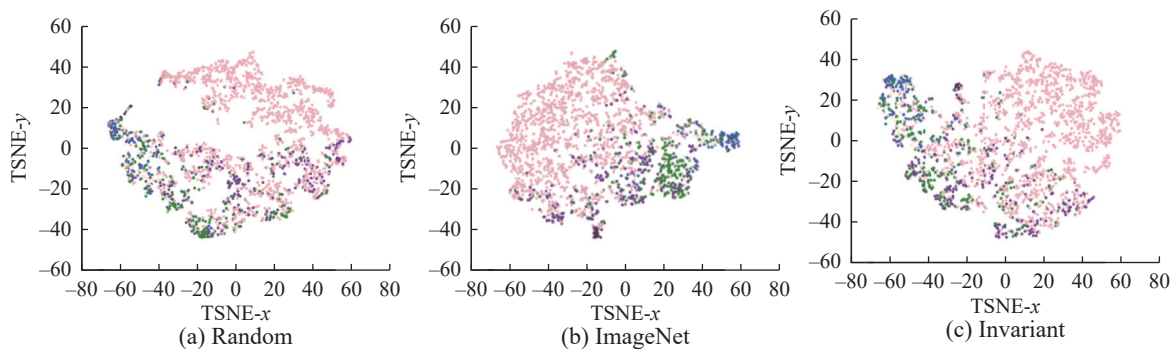
表 2 迁移学习对比结果
Table 2 Results of transfer learning

%

检测模型	下游任务微调后结果				冻住主干网络下游任务微调结果			
	Acc	Precision	Recall	F_1 -score	Acc	Precision	Recall	F_1 -score
Random	71.16±0.51	73.84±0.01	47.33±0.10	70.64±0.22	66.48±0.01	49.37±0.07	16.46±0.01	55.25±0.01
ImageNet	79.26±0.24	81.84±0.04	59.15±0.11	78.65±0.03	74.63±0.01	71.80±0.02	39.42±0.17	71.19±0.03
Rotation	80.47±0.02	81.47±0.03	63.81±0.19	79.81±0.04	67.87±0.01	54.85±0.09	18.81±0.01	58.64±0.08
BYOL	76.97±0.09	78.30±0.02	50.98±0.10	74.81±0.07	74.66±0.01	70.32±0.01	35.48±0.01	71.17±0.01
MoCo	82.19±0.01	82.03±0.01	62.59±0.16	80.93±0.01	74.50±0.01	69.64±0.01	32.05±0.01	69.89±0.01
SimCLR	79.23±0.05	81.60±0.02	63.01±0.26	78.96±0.02	76.16±0.01	72.13±0.01	38.94±0.04	73.02±0.01
Invariant	75.05±0.01	73.20±0.09	44.09±0.69	71.48±0.12	70.40±0.01	63.12±0.03	24.48±0.03	63.57±0.04
本文方法	84.28±0.01	84.58±0.01	68.81±0.10	83.63±0.01	76.43±0.01	74.16±0.02	43.35±0.06	74.12±0.02

从表 2 中可以看出,在不冻结主干网络 (Res-Net) 的情况下,使用本文提出的自监督学习的模型准确率为 84.28%,其他使用基于自然图像的自监督学习模型的准确率为 80.47%、76.97%、82.19%、79.23% 和 75.05%,分别提升了 3.81%、7.31%、2.09%、5.05% 和 9.23%。同时,与在 ImageNet 数据集预训练的模型相比,使用本文提出的自监督学习方法准确率提升了 5.02%。在冻结主干网络的情况下,使

用本文提出的自监督学习的模型准确率为 76.43%。实验结果表明本文提出的将个体判别任务提取的高维度特征与旋转任务提取的高维度特征相融合的自监督学习方法在皮肤疾病识别上效果显著提升。最后,可视化自监督方法在迁移学习分类中的检测结果可视化如图 4 所示。相比于其他对比实验方法,本文提出的方法聚类情况更为集中,错误样本更少,进而反映出高维空间中样本线性可分。



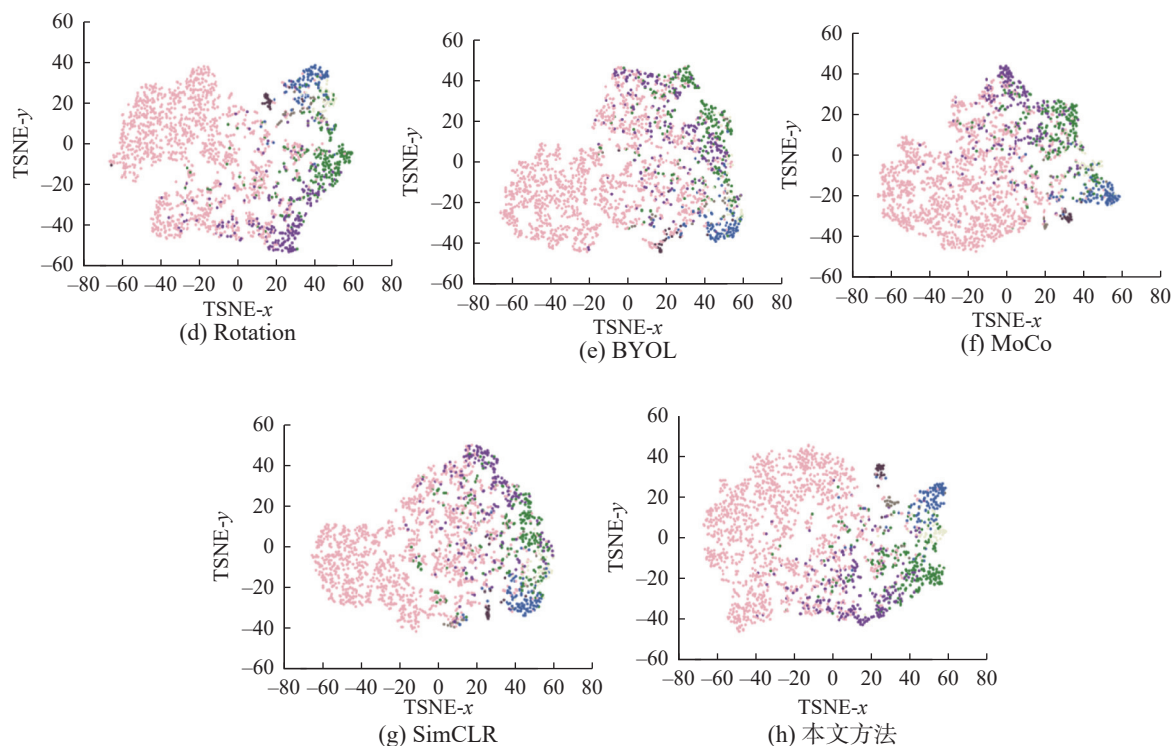


图4 自监督方法在迁移学习分类中的检测结果可视化

Fig. 4 Visualization of detection results of self-supervised method in transfer learning classification

2.3.2 小样本数据集对比实验

为了进一步模拟真实场景下皮肤镜标注数据稀少的特点,并验证本文提出的自监督方法在小样本数据集上的有效性,本文在下游数据集 HAM-10000 上进行了不同训练集比例的实验,并将所提出的方法与 MoCo、Rotation、BYOL、SimCLR 和 Invariant 相比,此外,还与 ImageNet 数据集预训练和随机初始化 Random 的模型进行对比。与 2.3.1 节相同,在实验过程中,本文首先在自监督训练数据集上使用自监督学习训练模型,完成自监督训练后,使用迁移学习来验证评估自监督方法学习到的特征。首先,实验使用 5 折交叉验证,并选取训练集和测试集样本比例为 4:1;然后,从训练集中随机选择不同比例的训练数据来微调模型,同时保持测试数据不变。本文选取包含训练样本比例分别为 1%、2%、4%、6%、8% 和 10% 的子训练集,具体的实验结果如图 5 所示。可以看出,当微调的训练数据减少时,不使用自监督学习的模型识别性能明显下降,而使用了自监督学习的模型虽然识别精度有所下降,但相比于不使用自监督学习的模型识别精度下降较少。同时可以发现,使用本文提出的自监督学习的模型可以更加稳定的识别不同种类的皮肤病变。通过实验,进一步验证了在小样本情况下,本文提出的自监督方法训练具有显著优势。

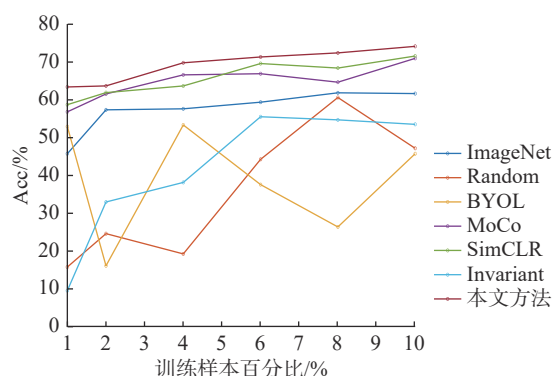


图5 自监督方法在小样本数据集上迁移学习分类结果

Fig. 5 Result of self-supervised method transfer learning on small sample datasets

2.3.3 消融实验

为了进一步定性和定量地分析旋转任务提供的上下文空间信息,并更进一步的分析部分超参数设定。本文首先设计了可视化消融实验,并进一步通过实验分析了超参数设定。具体实验设置与对比实验 2.3.1 所保持一致。可视化实验中,本文采用可视化特征激活图^[26](class activation map, CAM)。CAM 中,网络越感兴趣的区域温度越高,越不感兴趣的区域温度越低。实验分别可视化了随机初始化(random initialize)、单独使用旋转预测、个体判别任务和本文方法的特征提取网络输出的特征图激活映射,如图 6 所示,证实了本文提出的自监督方法的有效性。

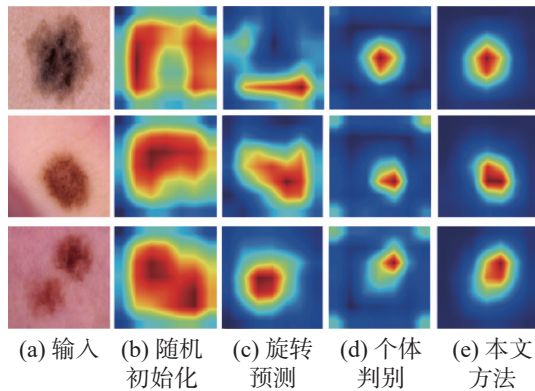


图6 可视化不同任务特征激活图

Fig. 6 Visualize activation maps of different tasks

通过观察图6可以发现,相比于单独使用旋转预测和随机初始化,本文提出的方法的感兴趣区域集中于皮肤病变区域(图像前景),基本不会出现错误识别感兴趣区域的情况。而对于皮肤病变识别而言,错误识别感兴趣区域是非常致命的。相比于单独使用个体判别任务,本文方法的识别质量更高,稳定性更好,感兴趣区域基本集中于目标区域,不会错误关注背景。具体来说,如图6中(b)、(c)测试图中,基于个体判别任务训练的模型错误地将图像背景识别为感兴趣区域,而本文提出的方法融合了个体判别任务和旋转辅助任务,检测效果相比单独使用一种任务都有提升,可视化结果均将皮肤病变区域识别为感兴趣区域,因而有效地提高了模型对皮肤病变识别的检测能力。

此外,将本文方法与单独采用旋转任务、个体判别任务进行实验对比,使用KNN分类器作为模型分类器。实验内容以及结果如表3所示。

表3 不同任务消融实验结果

Table 3 Results of ablation experiments on different tasks

方法	Acc	Precision	Recall	F_1 -score
旋转任务	68.16	68.44	14.68	55.56
个体判别	73.95	40.52	26.50	26.79
本文方法	74.55	71.93	29.41	70.51

从表3可以看出,与单独基于旋转的自监督学习方法相比,在HAM10000数据集上,本文提出自监督学习方法精度提升6.39%。同时,相比于单独使用基于个体判别任务,使用本文提出自监督学习方法精度提升了0.6%。实验结果表明,在旋转辅助任务所得到上下文空间特征和个体判别任务的实例级别的特征融合所带来性能提升的

同时,旋转辅助任务以病变图像旋转角度为指导,约束模型集中关注皮肤病变区域,提升了皮肤疾病识别精度。

本文对超参数 λ 与负样本数量 K 进行实验分析,使用KNN分类器作为模型分类器。其中,超参数 λ 为0表示为不使用旋转辅助任务(即只使用个体判别自监督学习),超参数 λ 增加表示旋转辅助任务的影响提升,实验结果如表4所示。实验结果表明,当超参数 λ 为0.1时,皮肤疾病识别的准确率为74.55%。同时,本文分析了负样本数量对皮肤疾病识别精度的影响,实验的结果如表5所示。可以看出,当负样本数量 K 为8192时,皮肤疾病识别的准确率最高,为74.95%。

表4 超参数 λ 对分类结果影响Table 4 Effect of hyperparameters λ on classification results

λ	0	0.1	0.5	1	5	10
Acc	73.45	74.55	73.85	73.95	72.15	71.75

表5 负样本数量对分类结果影响

Table 5 Effect of negative sample numbers on classification results

K	128	512	1024	2048	4096	8192	16384
Acc	73.65	74.55	73.85	73.95	74.85	74.95	74.55

3 结束语

针对皮肤镜图像的皮肤病检测识别,本文提出一种医学自监督分类,利用旋转预测与个体判别代理任务,融合空间上下文信息的特征与实例级语义特征,实现从海量无标注皮肤镜图像挖掘数据先验知识,提高皮肤疾病检测识别精度。旋转预测代理任务通过预测皮肤镜图像旋转角度以约束模型关注图像病变区域,从而获取空间上下文语义信息;个体判别任务利用正负样本对比学习,获得皮肤镜图像实例级语义特征。将学习到的先验知识通过导入训练权重和迁移学习的方式,解决传统有监督学习皮肤镜图像标注成本高的问题。实验结果证明,本文提出的自监督学习方法对HAM10000数据集的KNN检测精度为74.55%,迁移学习的检测精度为84.28%。相较于自然图像的自监督学习学习技术,可以在皮肤疾病检测识别任务中达到较高检测精度,提升皮肤疾病检测能力。

虽然本文的医学自监督学习技术缓解了皮肤镜图像获取难度大、成本高、耗时耗力的问题,但

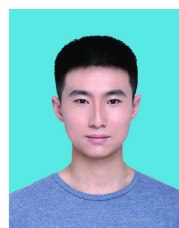
是在自监督学习的下游任务(皮肤疾病检测识别任务)中,仍然需要少量的有标注样本配合完成皮肤疾病检测识别。因此,如何进一步结合无监督学习和自监督学习,完全利用无标注样本的方式完成皮肤疾病检测识别也将是我们下一步的研究工作。

参考文献:

- [1] MIRBEIK-SABZEVARI A, TAVASSOLIAN N. Ultrawideband, stable normal and cancer skin tissue phantoms for millimeter-wave skin cancer imaging[J]. *IEEE transactions on bio-medical engineering*, 2019, 66(1): 176–186.
- [2] HOSNY K M, KASSEM M A, FOAUD M M. Skin cancer classification using deep learning and transfer learning[C]//2018 9th Cairo International Biomedical Engineering Conference. Piscataway: IEEE, 2018: 90–93.
- [3] 李晨, 张辉, 张邹铨, 等. 融合多尺度特征与全局上下文信息的X光违禁物品检测[J]. *中国图象图形学报*, 2022, 27(10): 3043–3057.
LI Chen, ZHANG Hui, ZHANG Zouquan, et al. Integrated multi-scale features and global context in X-ray detection for prohibited items[J]. *Chinese journal of image and graphics*, 2022, 27(10): 3043–3057.
- [4] SHEN Dinggang, WU Guorong, SUK H I. Deep learning in medical image analysis[J]. *Annual review of biomedical engineering*, 2017, 19: 221–248.
- [5] SHIN H C, ROTH H R, GAO Mingchen, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning[J]. *IEEE transactions on medical imaging*, 2016, 35(5): 1285–1298.
- [6] SADEGHI M, LEE T K, MCLEAN D, et al. Detection and analysis of irregular streaks in dermoscopic images of skin lesions[J]. *IEEE transactions on medical imaging*, 2013, 32(5): 849–861.
- [7] MA Yuhui, LIU Jiang, LIU Yonghuai, et al. Structure and illumination constrained GAN for medical image enhancement[J]. *IEEE transactions on medical imaging*, 2021, 40(12): 3955–3967.
- [8] BAILO O, HAM D, SHIN Y M. Red blood cell image generation for data augmentation using conditional generative adversarial networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2019: 1039–1048.
- [9] JING Longlong, TIAN Yingli. Self-supervised visual feature learning with deep neural networks: a survey[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2021, 43(11): 4037–4058.
- [10] 莫宏伟, 傅智杰. 基于迁移学习的无监督跨域人脸表情识别[J]. *智能系统学报*, 2021, 16(3): 397–406.
MO Hongwei, FU Zhijie. Unsupervised cross-domain expression recognition based on transfer learning[J]. *CAAI transactions on intelligent systems*, 2021, 16(3): 397–406.
- [11] 杜航原, 张晶, 王文剑. 一种深度自监督聚类集成算法[J]. *智能系统学报*, 2020, 15(6): 1113–1120.
DU Hangyuan, ZHANG Jing, WANG Wenjian. A deep self-supervised clustering ensemble algorithm[J]. *CAAI transactions on intelligent systems*, 2020, 15(6): 1113–1120.
- [12] WU Zhirong, XIONG Yuanjun, YU S X, et al. Unsupervised feature learning via non-parametric instance discrimination[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3733–3742.
- [13] FENG Zeyu, XU Chang, TAO Dacheng. Self-supervised representation learning by rotation feature decoupling[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 10356–10366.
- [14] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770–778.
- [15] HE Kaiming, FAN Haoqi, WU Yuxin, et al. Momentum contrast for unsupervised visual representation learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 9726–9735.
- [16] ZIMMERMANN R S, SHARMA Y, SCHNEIDER S, et al. Contrastive learning inverts the data generating process[EB/OL]. (2021-02-17)[2022-11-10]. <https://arxiv.org/abs/2102.08850>.
- [17] ROTEMBERG V, KURTANSKY N, BETZ-STABLEIN B, et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context[J]. *Scientific data*, 2021, 8(1): 34.
- [18] TSCHANDL P, ROSENDAHL C, KITTLER H. The HAM10000 dataset, a large collection of multi-source

- dermatoscopic images of common pigmented skin lesions[J]. *Scientific data*, 2018, 5: 180161.
- [19] CHEN Ting, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 1597–1607.
- [20] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised representation learning by predicting image rotations [EB/OL]. (2018–03–21)[2022–11–10].<https://arxiv.org/abs/1803.07728>.
- [21] GRILL J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent a new approach to self-supervised learning[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 21271–21284.
- [22] MISRA I, VAN DER MAATEN L. Self-supervised learning of pretext-invariant representations[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 6706–6716.
- [23] DENG Jia, DONG Wei, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 248–255.
- [24] ROBBINS H, MONRO S. A stochastic approximation method[J]. *The annals of mathematical statistics*, 1951, 22(3): 400–407.
- [25] RODRIGUEZ J D, PEREZ A, LOZANO J A. Sensitivity analysis of k-fold cross validation in prediction error estimation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2010, 32(3): 569–575.
- [26] JIANG Pengtao, ZHANG Changbin, HOU Qibin, et al. Layercam: exploring hierarchical class activation maps for localization[J]. *IEEE transactions on image processing*, 2021, 30: 5875–5888.

作者简介:



刘嘉轩, 硕士研究生, 主要研究方向为医学图像分析和计算机视觉。



胡非易, 硕士研究生, 主要研究方向为图像识别、医学图像分析。



张辉, 教授, 博士生导师, 主要研究方向为计算机视觉。获省部级科学技术奖励一等奖 8 项, 获 2022 年湖南省第十三届教学成果特等奖等, 主持科技创新 2030——“新一代人工智能”重大项目、国家自然科学基金共融机器人重大研究计划重点项目、国家重点研发计划子课题、国家科技支撑计划项目子课题等 20 余项, 授权发明专利 38 项。发表学术论文 50 余篇。