

DOI:10.16136/j.joel.2022.05.0619

# 基于 PiT 的皮肤镜图像分类方法研究

韦春苗<sup>1</sup>, 徐 岩<sup>1\*</sup>, 蒋新辉<sup>2</sup>, 魏一铭<sup>1</sup>

(1. 兰州交通大学 电子与信息工程学院, 甘肃 兰州 730070; 2. 柳州铁道职业技术学院 人工智能学院, 广西 柳州 545616)

**摘要:**随着计算机技术的进步, 现有的 Transformer 被扩展成处理计算机视觉任务的网络结构。为提高黑色素瘤的早期确诊率以提高皮肤病患者的治愈率, 本文提出一种改进的基于 PiT (pyramid pooling transformer) 的网络模型来实现对 7 种皮肤病变的皮肤镜图像进行自动分类。本文模型主要由 PiT 模块和抗干扰模块等 2 个部分组成, Pit 继承了 ViT 的优点, 并通过池化进行空间尺寸转换来提高模型的鲁棒性, 经过预训练的 PiT 网络拥有大量的自然图像特征, 且 PiT 部分网络可为下游的分类任务提供所需的图像特征, 本文设计出抗干扰模块, 用来抵抗皮肤镜图像中的干扰因素 (如毛发、异物遮挡) 的影响, 从而提高模型性能、提高分类精度。实验结果表明, 本文模型在 ISIC2018 验证集上的分类准确率、精确率、召回率、 $F1$ -score 值分别高达 91.58%、83.59%、89.92%、86.34%, 每秒传输帧数 (frames per second, FPS) 达到 85 Hz 与现有的几种先进的分类网络相比, 分类性能和模型效率都有所提高, 具有相对优势, 证明本文模型具有一定的实用价值。

**关键词:**图像处理; 图像分类; PiT; 抗干扰模块

**中图分类号:** TP391.41 **文献标识码:** A **文章编号:** 1005-0086(2022)05-0505-08

## Study on dermoscope image classification method based on PiT

WEI Chunmiao<sup>1</sup>, XU Yan<sup>1\*</sup>, JIANG Xinhui<sup>2</sup>, WEI Yiming<sup>1</sup>

(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China; 2. College of Artificial Intelligence, Liuzhou Railway Vocational and Technical College, Liuzhou, Guangxi 545616, China)

**Abstract:** With the advancement of computer technology, the existing Transformer has been expanded into a network structure for processing computer vision tasks. In order to improve the early diagnosis rate of melanoma and the cure rate of skin disease patients, this paper proposes an improved network model based on PiT (pyramid pooling transformer) to realize automatic classification of dermoscopic images of seven skin lesions. The model of this paper is mainly composed of the PiT module and the anti-interference module. Pit inherits the advantages of ViT and uses pooling to perform spatial size conversion to improve the robustness of the model. The pre-trained PiT network has a large number of natural image features, and the PiT part of the network can provide the required image features for downstream classification tasks. In this paper, an anti-interference module is designed to resist the influence of interference factors (such as hair and foreign object occlusion) in the dermoscopic image, thereby improving the performance of the model. Improve classification accuracy. Experimental results show that the classification accuracy, precision, recall, and  $F1$ -score values of this model on the ISIC 2018 verification set are as high as 91.58%, 83.59%, 89.92%, 86.34%, and the number of frames per second (FPS) reaches 85 Hz. Compared with several existing advanced classification networks, the classification performance and model efficiency have been improved, and it has relative advantages, which proves that the model in this paper has certain practical value.

**Key words:** image processing; image classification; PiT; anti-interference module

\* E-mail: xuyan@mail.lzjtu.cn

收稿日期: 2021-09-01 修订日期: 2021-10-03

## 1 引 言

皮肤病是全球最重视的健康问题之一,黑色素瘤是一种致死率极高的皮肤癌,据美国癌症协会估计,2020 年黑色素瘤的新增病历超过 10 万,死亡率估计高达 7 000 例<sup>[1]</sup>,若在其病发早期能被及时发现并接受治疗,则黑色素瘤可被治愈的概率极高<sup>[2]</sup>,而黑色素瘤的识别与分类是早期确诊和治疗的关键。传统皮肤病变的确诊方式主要依靠专业皮肤科医生的视觉经验,将病变区域皮肤与健康区域皮肤做比较,后逐个对局部病变区域的皮肤进行评估<sup>[3]</sup>,皮肤镜是一种可放大病变区域皮肤、消除皮肤表面反射的非侵入式成像技术<sup>[4]</sup>,借助皮肤镜技术可获取清晰度较高的皮肤表面斑点图,可为准确诊断黑色素瘤提供可靠的依据。基于人工的视觉检查皮肤镜图像手动分类不仅效率低,而且主观性极强(即使是专业性极强的专业皮肤科医生也会产生差距很大的诊断结果),因此借助计算机辅助诊断技术实现皮肤镜图像自动分类具有很重要的现实意义。

然而,基于皮肤镜图像的黑色素瘤自动识别分类方向的研究仍然面临严峻的挑战,如皮肤镜图像病灶区域存在光照不均、毛发或异物遮挡、皮肤纹理不清晰、病变皮肤与良性痣相似度高等一系列的问题会影响算法的精度,此外,可供模型训练的可靠数据集较少,且现有的各种分类模型仍具有一定的局限性,算法性能有待提高,世界各地学者都在为致力于开发出性能更高的黑色素瘤自动分类系统而努力。

早期的识别方法主要是通过利用颜色、形状等手工特征来识别黑色素瘤,GANSTER 等<sup>[5]</sup>提出利用特征选择算法来选择一些合适的特征,并结合这些特征来提高分类精度,但这些低级制作的手工特征无法处理黑色素瘤和非黑色素瘤的病变之间高度相似性及皮肤病变的伪影。近年来,随着计算机技术的飞速发展,卷积神经网络(convolutional neural network, CNN)在医学图像分类任务中取得了划时代的成功,许多学者积极投入到基于深度学习的黑色素瘤分类系统的研发中,并取得了许多优秀的成果。KAWAHARA 等<sup>[6]</sup>使用基于 AlexNet 的完全预训练 CNN,该网络在自然图像数据集上进行预训练以获取自然图像特征,进而在转向分类任务的学习时可大大减少训练时间,得到的分类精度高达 81.8%。ESTEVA 等<sup>[7]</sup>使用 3 种经过预训练的网络: CNN、GoogleNet、inception-V3 结构,对 12 945 0 张皮肤病数据集进行分类,并荣获佳绩。文献[8]使用基

于 VGG16 的 CNN 网络,使用 3 种迁移学习的方法对黑色素瘤进行二分类,并取得不错的成绩。LI 等<sup>[9]</sup>提出一种由 2 个全卷积神经网络(full convolutional neural network, FCN)组成的深度学习框架,同时完成皮肤镜图像的特征提取及病变分类。现有的基于 CNN 的分类网络模型复杂,对特征的提取有待加强,且目前的皮肤镜图像检测主要实现的二分类任务,分类算法在分类精度和多分类任务的研究中都有很大的进步空间。

在过去的十几年, CNN 成为计算机视觉领域的主导,并在各种视觉任务中的处理水平不断提高,而在自然语言处理(natural language processing, NLP)领域的关键技术是 Transformer<sup>[10]</sup>, Transformer 是基于自注意力机制的特性使其可轻松捕获全局依赖关系,而全局信息对视觉任务必不可少, Transformer 能很好地克服 CNN 的缺点——CNN 需堆叠更多层网络来扩大感受野,所以,许多基于 Transformer 的视觉任务也取得了巨大的成功。DOSOVITSKIY 等<sup>[11]</sup>将 Transformer (ViT)运用于图像分类并取得成功,首先将输入图像分为小块(Patches),后在 NLP 应用程序中将每条路径做 1 个标记(token),后输入 Transformer 中进行分类。然而,若序列数据过长,则会提高 Transformer 的空间计算复杂度提高,且对于高分辨的输入数据,会导致更多细节信息的丢失。为解决此问题, Swin Transformer<sup>[12]</sup>提出在输入数据的小窗口范围内计算多头自注意力(multi-head self-attention, MHSA),并移动窗口来扩大感受野,然而却不能解决序列顺序问题。PiT<sup>[13]</sup>继承了 ViT 的优势,并利用 CNN 的设计原理,通过提高通道维度并减小空间维度来进一步提高模型鲁棒性和分类准确度。为得到一个更高性能的分类网络,本文在基于 PiT 网络的基础上做改进,设计 1 个抗干扰模块进行网络优化,得到结构化的皮肤镜图像自动识别分类模型,实现皮肤镜图像(黑色素瘤)的 7 分类任务,本研究在黑色素瘤分类任务上的准确率、网络运行速率等有很大的进步,具有一定的应用价值。

## 2 网络架构设计

模型主要由 2 个部分组成: PiT 和抗干扰模块。输入图像首先进入 PiT 进行处理, PiT 在基于 Transformer 架构有效性的前提下,还充分发挥空间维度转换的作用,通过减少空间维度来增强模型的分类能力,经过特征映射后输入到抗干扰模块处理,条形池注意力机制和动态激活函数(dynamicrelu)



用来防止梯度消失或爆炸问题,为防止过拟合,在训练过程中对部分网络按照一定的概率进行随机丢弃(Dropout)进行网络优化,抗干扰模块中跳跃连接的使用有效提高了图像分类的准确性,抗干扰模块整体部分如图1所示。下面对 SPAM 和动态激活函数 2 个模块进行详细描述。

### 1) 条形池注意力机制

条形池继承了全局平均池化的优点,不仅能很好地捕获远程依赖关系,且对于形状不规则的病灶,也能很好地关注局部细节信息,同时它克服了全局平均池化的缺点,条形池不需对整个特征场景进行大量计算,能很好地避免较远位置间没必要的连接。条形池是一个轻量级的模块,主要利用水平池化和垂直池化来捕获不同维度的空间上下文信息,以达到扩大主干网络感受野的目的,它可被看做一种注意力机制,无需从头进行训练,可轻松嵌入到经过预训练的 Transformer 之后,实现捕获远程空间依赖和通道间依赖的功能。条形池结构图如图3所示。

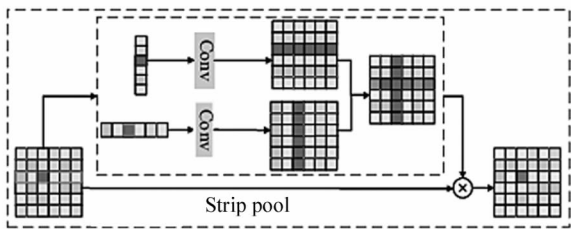


图3 条形池结构图

Fig. 3 Bar pool structure diagram

### 2) 动态激活函数(dynamic reLU, DY-ReLU)

ReLU 是一种简单却强大的被广泛用于提高神经网络性能的激活函数,ReLU 是静态的,可实现神经网络的非线性化,对于输入数据,其不能根据数据变化而随之变化,而是有自己特定的变化规律,而 DY-ReLU 是动态的,相对于 ReLU 只是增加了少量的计算量,却拥有更强大的表征能力,特别是对轻量型网络而言。DY-ReLU 是一个分段函数,如图4所示。

DY-ReLU 主要包含辅助激活函数  $Q(x)$  和激活函数  $f_{\theta(x)}(x)$ ,辅助函数的作用是计算激活函数的参数,激活函数  $f_{\theta(x)}(x)$  主要用来计算输入的激活输出。 $f_{\theta(x)}(x)$  可表示为:

$$y_c = f_{\theta(x)}(x_c) = \max_{1 \leq k \leq K} \{a_c^k(x) x_c + b_c^k(x)\}, \quad (1)$$

式中,  $x_c$  为向量  $x$  在第  $C$  通道的输入,  $a_c^k, b_c^k$  为  $Q(x)$  的 2 个计算输出。具体形式如下:

$$[a_1^1, \dots, a_C^1, \dots, a_1^K, \dots, a_C^K, b_1^1, \dots, b_1^K, \dots, b_C^K]^T = \theta(x), \quad (2)$$

式中,  $K$  为分段函数个数,  $C$  表示通道数。

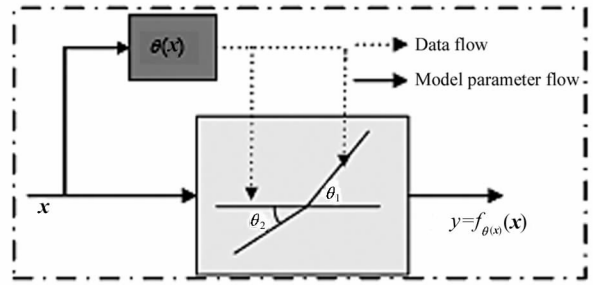


图4 DY-ReLU 激活函数示意图

Fig. 4 Dynamic ReLU activation function diagram

$Q(x)$  由 SE 模块实现,对于本文算法来说, DY-ReLU 的输入图片大小为  $8 \times 8 \times 20$ , 先被传送到全局池化层实现空间信息压缩,后通过 1 个全连接层(该层夹着个 ReLU),最后采用 sigmoid 进行标准化。输出结果为初始值和残差的加权和,其计算式为:

$$a_c^k(x) = \alpha^k + \lambda_a \Delta a_c^k(x), \quad (3)$$

$$b_c^k(x) = \beta^k + \lambda_b \Delta b_c^k(x), \quad (4)$$

式中,  $\alpha^k, \beta^k$  分别表示  $a_c^k$  和  $b_c^k$  的初始值,  $\lambda_a, \lambda_b$  表示残差范围的控制标量,  $\alpha^k, \beta^k, \lambda_a, \lambda_b$  都为超参数。

DY-ReLU 有 3 种类型: DY-ReLU-A、DY-ReLU-B、DY-ReLU-C。本文算法所用到的类型是 DY-ReLU-A: 所有空间位置与通道共享相同的分段线性激活函数, DY-ReLU-A 相对于 DY-ReLU-B 来说计算量较小,且在本算法的应用中能取得更好的结果。其原理图如图5所示。

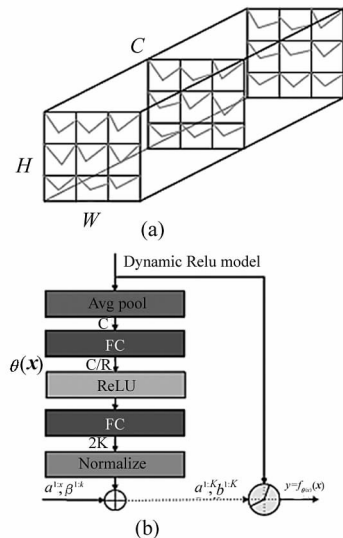


图5 DY-ReLU-A 原理图: (a) DY-ReLU-A 空间与通道共享示意图; (b) DY-ReLU-A 流程图

Fig. 5 DY-ReLU-A schematic diagram: (a) The Dy-reLU-A space shares the schematic diagram with the channel; (b) DY-ReLU-A flow chart

### 2.3 交叉熵损失函数

为解决多分类任务中类别不均的问题,本文采用交叉熵损失函数来提高模型的分类准确率,交叉熵损失函数的表达式为:

$$L_{CE} = - \sum_{i=1}^m p(x) \log q(x), \quad (5)$$

式中,  $m$  表示每批次的样本数,  $p(x)$  为 7 种皮肤病变类型的真实分布概率,  $q(x)$  为 7 种皮肤病变类型的预测分布概率。

$q(x)$  的表达式为:

$$q(x) = \frac{e^{w_i^T x_i + b_i}}{\sum_{j=1}^n e^{w_j^T x_i + b_j}}, \quad (6)$$

式中,  $n$  为皮肤病变类型数,  $y_i$  表示第  $i$  种皮肤病变类型的编号,  $x_i (x_i \in R_d)$  为模型提取的第  $i$  个深度特征,  $d$  为模型输出特征维度,  $w_j$  表示最后一个连接层中权重  $w$  的第  $j$  列,  $b$  为偏置项。

模型通过前向传播方式衡量预测概率,再利用交叉熵损失  $L_{CE}$  来减小  $p(x)$  和  $q(x)$  之间的分布差距,从而提高分类准确率。

## 3 实验结果及分析

### 3.1 实验环境及数据集

本实验采用 Pytorch1.5.0 深度学习框架,编程语言为 Python3.7.7, Windows 操作系统,显卡为 NVIDIA GTX 1660TI 6G,处理器为 I5 9400F。采用国际皮肤成像协会(The International Skin Imaging Collaboration, ISIC)2018 年皮肤镜图像分类公开挑战赛的公开数据集 ISIC 2018。该数据集由国际上较为权威的临床中心提供,被专业的皮肤科医生进行注释和标记。ISIC 2018 分为 8 012 张训练集、500 张验证集和 1 503 张测试集,每个数据集中每个包含 7 种病变类型的皮肤镜图像:光化性角化病(Akiec)、皮肤纤维瘤(Df)、基底细胞癌(Bcc)、黑色素瘤(Mel)、良性角化病(Bkl)、血管病变(Vasc)、色素细胞痣(Nv)。7 种类型的皮肤病图像在 3 个数据集的分布情况如表 1 所示。

表 1 7 种类别的皮肤病在 3 个数据集上的分布情况

Tab. 1 Distribution of seven categories of skin diseases in three data sets

| Sample           | Akiec | Df | Bcc | Mel | Bkl | Vasc | Nv    |
|------------------|-------|----|-----|-----|-----|------|-------|
| The training set | 262   | 92 | 411 | 890 | 879 | 113  | 5 364 |
| Validation set   | 16    | 6  | 26  | 56  | 55  | 7    | 335   |
| The test set     | 49    | 17 | 77  | 167 | 165 | 21   | 1 006 |

从表中可知, ISIC 2018 数据集存在严重的不平衡问题,数据不平衡会导致分类器对数量多的皮肤病变类型识别准确率高,但对数量少的皮肤病类别表现不好,这导致分类器的总体识别精准度并不客观,为减少数据不平衡对分类结果的影响,对 3 个数据集进行随机缩放、翻转、水平或竖直平移等操作来实现数据增强,使 7 种数据集达到平衡,平衡后的数据如图 6 所示。

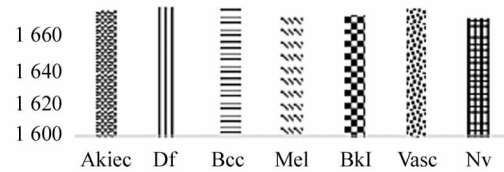


图 6 平衡后的数据集

Fig. 6 The balanced data set

同时数据集尺寸不一,为了减小网络计算量,对所有图像沿中心进行裁剪,统一归一化成  $224 \times 224 \times 3$  的标准尺寸。

### 3.2 评价标准

本文采用公认的多分类任务评价指标进行评估,包括准确率(Accuracy,  $Acc$ )、精确率(Precision,  $P$ )、召回率(Recall,  $R$ )、 $F1-score$  值,其中,  $Acc$  反映了模型对整个样本的判定能力,精确率反映了实际正例在模型判定的所有与正例中的占比,召回率反映了被正确检测到的正样本在总的正样本中的比重;  $F1-score$  综合了精确率和召回率的结果,一般来说,  $F1-score$  越高表示模型对单个类别的分类效果越好。它们的表达式分别为:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}, \quad (7)$$

$$P = \frac{TP}{TP + FP}, \quad (8)$$

$$R = \frac{TP}{TP + FN}, \quad (9)$$

$$F1-score = \frac{2 \frac{TP}{TP + FP} \times \frac{TP}{TP + FN}}{\frac{TP}{TP + FP} + \frac{TP}{TP + FN}}, \quad (10)$$

式中,  $TP$  为真阳性(true positive),代表实例在分类器中的预测结果与实际结果都为正,  $FP$  为假阳性(false positive),代表实例在分类器中的预测结果为正,实际结果为负,  $TN$  为真阴性(true negative),代表实例在分类器中的预测结果与实际结果都为负,  $FN$  为假阴性(false negative),代表实例在分类器中的预测结果为负,而实际结果为正。  $Acc$ 、 $P$ 、召回率

和  $F1-score$  值越高,说明模型分类效果越好。

### 3.3 实验结果分析

为证明本文算法的有效性,使用本文模型与 PiT 分类模型(未加入抗干扰模块)在数据集 ISIC 2018 测试集上进行分类测试对比,2 个模型在 7 种皮肤病分类训练所得结果对比数据如表 2 所示。

从表中可知,本模型在 7 种皮肤病类型上的精确率、召回率和  $F1-score$  的表现相对于 PiT 模型来说都有所进步,证明了本文模型相对于 PiT 分类模型而言总体的分类性能得到很大改善。

为全面衡量本算法的分类能力,将本模型与  $PiT^{[13]}$ 、DenseNet 121<sup>[17]</sup>、ResNet 50<sup>[18]</sup>、Shuffle-

表 2 7 类皮肤病在本模型和 PiT 模型的精确率、召回率和  $F1-score$  对比数据

Tab. 2 The precision, recall and  $F1-score$  compared between the P2T model and the PiT model

| The evaluation index | Ours  |      |      |      |      |      |      | PiT   |      |      |      |      |      |      |
|----------------------|-------|------|------|------|------|------|------|-------|------|------|------|------|------|------|
|                      | Akiec | Bcc  | Bkl  | Df   | Mel  | Nv   | Vasc | Akiec | Bcc  | Bkl  | Df   | Mel  | Nv   | Vasc |
| Precision            | 0.81  | 0.89 | 0.86 | 1.00 | 0.79 | 0.95 | 1.00 | 0.65  | 0.86 | 0.85 | 0.88 | 0.76 | 0.94 | 1.00 |
| Recall               | 0.81  | 0.92 | 0.82 | 0.73 | 0.75 | 0.97 | 0.86 | 0.88  | 0.84 | 0.73 | 0.64 | 0.73 | 0.96 | 0.79 |
| $F1-score$           | 0.81  | 0.90 | 0.84 | 0.84 | 0.77 | 0.96 | 0.92 | 0.75  | 0.85 | 0.79 | 0.74 | 0.75 | 0.95 | 0.88 |

Netv2<sup>[19]</sup>这几种分类模型在 ISIC 2018 验证集上进行测试,采用准确率、精确率、召回率、 $F1-score$  以及每秒传输帧数(frames per second, FPS)这 5 种指标

进行分类评估结果,如表 3 所示。

从表中可知,本文算法在准确率、精确率、召回率、 $F1-score$  表现最佳,表明本文模型相对于其他网

表 3 ISIC 2018 验证集上各算法分类评估结果,粗体标记最优值

Tab. 3 The classification evaluation results of each algorithm on ISIC 2018 validation set are marked with the optimal value in bold

| Methods      | Accuracy/%   | Precision/%  | Recall/%     | $F1\_score$ /% | FPS/Hz     |
|--------------|--------------|--------------|--------------|----------------|------------|
| Ours         | <b>91.58</b> | <b>83.59</b> | <b>89.92</b> | <b>86.34</b>   | 85         |
| PiT          | 89.47        | 79.48        | 84.83        | 81.40          | 94         |
| DenseNet121  | 90.45        | 82.52        | 87.32        | 84.34          | 45         |
| ResNet50     | 89.68        | 80.74        | 82.34        | 81.45          | 61         |
| ShuffleNetv2 | 84.79        | 68.65        | 76.32        | 71.53          | <b>129</b> |

络模型,具有更高的分类精度。FPS 代表算法的检测速度,从表中可知,本文算法检测速度与其他算法对比,虽不能获得最优值,但在牺牲很少的速度的同时获得了更高的精度,且本文算法符合一个高效分类网络的速度要求。所以,总体而言,本文模型是个高效率、高性能的分类网络,具有较高的实际应用价值。

图 7 为本模型与其他 4 种模型在 ISIC 2018 验证集和测试集上的描述性统计。其中,图 7(a)、(c)分别为本模型与其他 4 种模型在 ISIC 2018 验证集上训练过程中的正确率和训练 Loss 对比图,图 7(b)、(d)为本模型与其他 4 种模型在 ISIC 2018 测试集上训练过程中的正确率和训练 Loss 对比图。

从图 7(a)和(b)可得出,本文模型在 ISIC 2018 训练集和测试集上的训练准确率经过 10 个 epoch 就

趋于平稳,且训练获得的  $Acc$  值较高,说明相对于其他 4 种模型,本文算法更加稳定且分类准确率更高。从图 7(c)和(d)可得出,训练过程中,本文模型的损失函数的数值不断变小,在第 10 个 epoch 之后降到最低且逐渐趋于平稳,说明本文模型的泛化能力强、稳定性好。

本文模型与 PiT 模型的可视化混淆矩阵如图 8 所示,混淆矩阵的对角线数据表示模型预测标签为真实标签的点数,非对角线的元素为被模型错误标记的点,对角线值越高表示模型的分类能力越强。

从图 8 可知,本文模型相对于 PiT 模型来说,具有更强的皮肤病变识别能力,综上,说明本文在 PiT 上改进得到的模型对皮肤病图像分类具有较高的准确率和效率,且本文模型具有较强的泛化性。

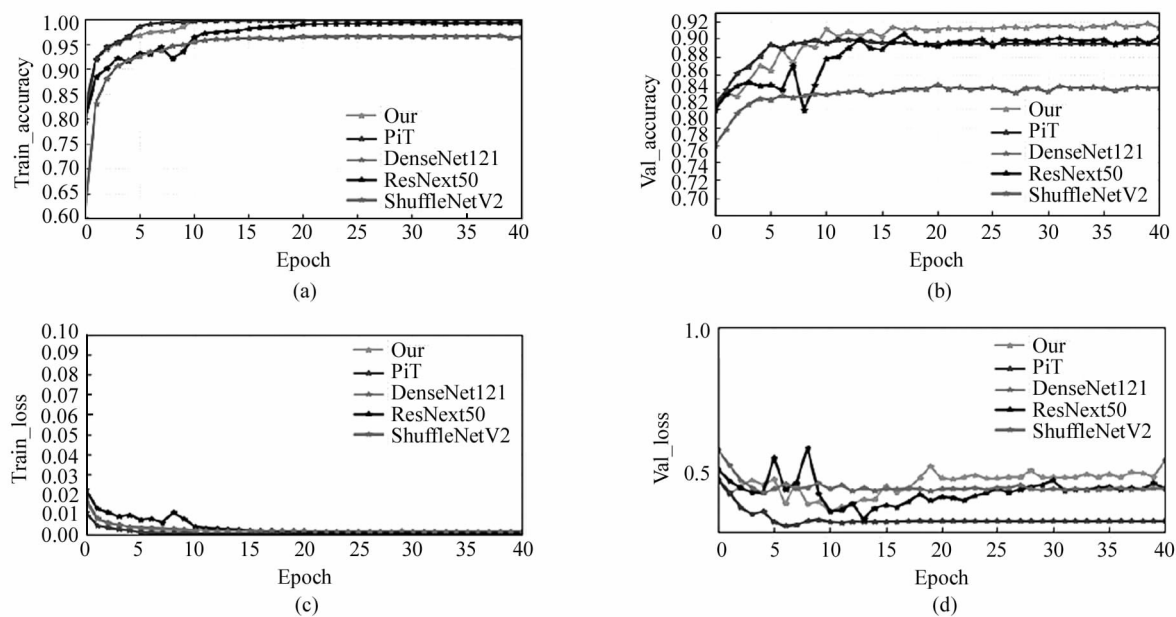


图 7 各算法在 ISIC 2018 训练集和测试集的性能测试图:(a)(b) 各算法在 ISIC 2018 训练集、测试集上的准确率迭代图;  
(c)(d) 各算法在 ISIC 2018 训练集、测试集上的损失迭代图

Fig. 7 Performance test graph of each algorithm in ISIC 2018 training set and test set;  
(a)(b) The accuracy iteration graph of each algorithm on ISIC 2018 training set and test set;  
(c)(d) Loss iteration graph of each algorithm on ISIC 2018 training set and test set

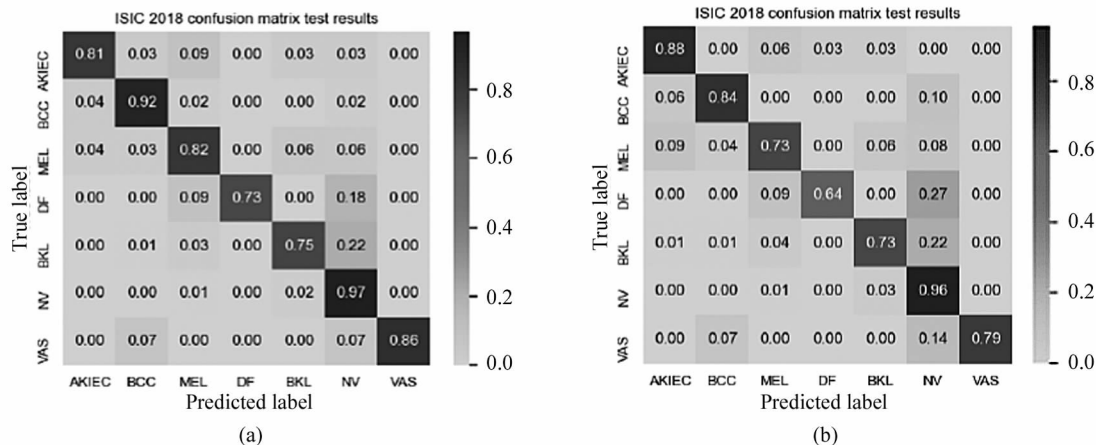


图 8 可视化混淆矩阵:(a) 本模型分类结果混淆矩阵;(b) PiT 模型分类结果混淆矩阵

Fig. 8 Visual confusion matrix;(a) The confusion matrix of the classification results of this model;  
(b) Confusion matrix of PiT model classification results

4 结 论

皮肤镜图像分类在辅助医生的临床诊断中具有重要意义,为实现 7 种皮肤病类型的皮肤镜图像的自动分类,本文先利用 PiT 网络对输入图像数据进行全局建模,经过预训练的 PiT 作为分类模型的重要组成,能减少网络计算量的同时还能解决结构信息的丢失,PiT 处理后得到的图像通过抗干扰模块,

该模块通过滤除皮肤镜图像中各种其他因子干扰来提高模型分类精度,利用交叉熵损失函数缩小样本预测概率与真实概率的差距从而提高模型性能,本文模型是在 PiT 的基础上做出改进,实验表明,本模型对 7 类皮肤镜图像进行分类的精确率、召回率和  $F1-score$  值整体优于 PiT 模型,且与 PiT、DenseNet121、ResNet50、ShuffleNetv2 这几种优秀的分类网络相比,虽然牺牲了较少的网络运算速度,

但在准确率、精确率、召回率、 $F1$ -score 上获得了最优值,分别为 91.58%、83.59%、89.92%、86.34%,证明本文算法具有一定的优越性和实用价值。

### 参考文献:

- [1] SEER. 2020. Melanoma of the skin-cancer stat facts[N/QL]. <https://seer.cancer.gov/statfacts/html/melan.html>, 2020.
- [2] TRINIDAD C M, TORRES-CABALA C A, CURRY J L, et al. Update on eighth edition american joint committee on cancer classification for cutaneous melanoma and overview of potential pitfalls in histological examination of staging parameters[J]. Journal of Clinical Pathology, 2019, 72(4): 265-270.
- [3] GACHON J, BEAULIEU P, SEI J F, et al. First prospective study of the recognition process of melanoma in dermatological practice[J]. Archives of dermatology, 2005, 141(4): 434-438.
- [4] YU L, CHEN H, DOU Q, et al. Automated melanoma recognition in dermoscopy images via very deep residual networks[J]. IEEE Transactions on Medical Imaging, 2016, 36(4): 994-1004.
- [5] GANSTER H, PINZ P, ROHRER R, et al. Automated melanoma recognition[J]. IEEE Transactions on Medical Imaging, 2001, 20(3): 233-239.
- [6] AWAHARA J, BENTAIEB A, HAMARNEH G. Deep features to classify skin lesions[C]//13th International Symposium on Biomedical Imaging (ISBI), April 13-16, 2016, Prague, Czech Republic. New York: IEEE, 2016: 1397-1400.
- [7] ESTEVA A, KUPREL B, NOVOA R A, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. Nature, 2017, 542(7639): 115-118.
- [8] LOPEZ A R, GIRO-I-NIETO X, BURDICK J, et al. Skin lesion classification from dermoscopic images using deep learning techniques[C]//13th IASTED International Conference on Biomedical Engineering (BioMed), February 20-21, 2017, Innsbruck, Austria. New York: IEEE, 2017: 49-54.
- [9] LI Y, SHEN L. Skin lesion analysis towards melanoma detection using deep learning network[J]. Sensors, 2018, 18(2): 556.
- [10] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. (2017-12-06) [2021-09-01]. <https://arxiv.org/abs/1706.03762>.
- [11] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth  $16 \times 16$  words; Transformers for image recognition at scale[EB/OL]. (2021-06-03) [2021-09-01]. <https://arxiv.org/abs/2010.11929>.
- [12] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[EB/OL]. (2021-08-17) [2021-09-01]. <https://arxiv.org/abs/2103.14030>.
- [13] Heo B, Yun S, Han D, et al. Rethinking spatial dimensions of vision transformers[EB/OL]. (2021-08-18) [2021-09-01]. <https://arxiv.org/abs/2103.16302v2>.
- [14] Hou Q, Zhang L, Cheng M M, et al. Strip pooling: Rethinking spatial pooling for scene parsing[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 4003-4012.
- [15] CHEN Y, DAI X, LIU M, et al. Dynamic relu[C]//European Conference on Computer Vision. August 23-28, 2020, Glasgow, UK. Berlin: Springer, 2020: 351-367.
- [16] BA J L, KIROUS J R, HINTON G E. Layer normalization[EB/OL]. (2016-07-21) [2021-09-01]. <https://arxiv.org/abs/1607.06450>.
- [17] HUANG G, LIU Z, LAURENS V D M, et al. Densely connected convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 4700-4708.
- [18] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 1492-1500.
- [19] Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 6848-6856.

### 作者简介:

徐 岩 (1963—),男,硕士生导师,主要从事图像处理、语音处理和铁路通信等方面的研究。