

基于双线性特征融合的皮肤病分类研究

黄志伦^{1,2}, 刘俊^{1,2}, 郑萌³

- (1. 武汉科技大学 计算机科学与技术学院, 湖北 武汉 430065;
2. 智能信息处理与实时工业系统湖北省重点实验室(武汉科技大学), 湖北 武汉 430065;
3. 武汉理工大学 计算机科学与技术学院, 湖北 武汉 430070)

摘要: 皮肤覆盖肌肉、骨骼和身体的每个部分, 是人体中最大的器官。由于其暴露于外界, 所以感染更容易发生在皮肤上。皮肤病作为一种常见疾病, 利用计算机技术对其进行辅助诊断, 有助于减轻医生负担。针对常规卷积神经网络应用于皮肤病图像分类时由于不同种皮肤病图像之间的类间相似性以及同种皮肤病图像之间具有类内差异性导致分类困难的问题, 提出一种改进双线性特征融合模型。使用经过剪枝的 Inception-ResNet-v1 和 v2 版本作为特征提取器并行提取图像特征, 对特征进行双线性融合, 获取更多阶数的特征信息可以提高模型对图像细节的敏感度。然后添加额外的软注意力模块, 通过加权的方式进行过滤或者加强, 给图像每个位置给予不同的权重以达到对模型的加强效果。在 skin-cancer-classesic 数据集上的 7 种皮肤病图像上进行训练, 与 S-CNN、MobileNet 和 Incremental CNN 的对比证明了该模型的有效性, 在 Precision、Recall 和 F1-Score 指标上该模型均为最优。

关键词: 皮肤病; 深度学习; 双线性特征融合; 注意力机制; 图像分类

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2023)02-0161-06

doi: 10.3969/j.issn.1673-629X.2023.02.024

Classification of Skin Diseases Based on Bilinear Feature Fusion

HUANG Zhi-lun^{1,2}, LIU Jun^{1,2}, ZHENG Meng³

- (1. School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China;
2. Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System
(Wuhan University of Science and Technology), Wuhan 430065, China;
3. School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China)

Abstract: Skin covers muscles, bones and every part of the body, which is the largest organ in the human body. Because it is exposed to the outside world, the infection is more likely to occur on the skin. Skin disease is a common disease. Using computer technology to diagnose it will help to reduce the burden of doctors. In view of the difficulty of classification caused by the inter class similarity between different skin disease images and the intra class difference between the same skin disease images when the conventional convolutional neural network is applied to the classification of skin disease images, an improved bilinear feature fusion model is proposed. Using the pruned version of Inception-ResNet-v1 and v2 as the feature extractor to extract image features in parallel, bilinear fusion of features and obtain more order feature information can improve the sensitivity of the model to image details. Then add an additional soft attention module to filter or strengthen the image by weighted sum, and give different weights to each position of the image to achieve the effect of strengthening the model. Training on seven kinds of skin disease images on skin cancer classic dataset, the comparison with S-CNN, MobileNet and Incremental CNN proves the effectiveness of this model, which is the best in Precision, Recall and F1-Score.

Key words: skin disease; deep learning; bilinear feature fusion; attention mechanism; image classification

0 引言

皮肤容易受到感染并引发炎症, 因为皮肤作为人体最大的器官, 是直接和外界细菌或者病毒等感染

源接触, 所以皮肤病在所有年龄段的人群都比较常见。它主要由 3 个部分组成, 表皮、真皮和皮下组织^[1]。位于隔绝外界的第一层防护是表皮, 表皮处于整个皮肤

收稿日期: 2022-03-14

修回日期: 2022-07-19

基金项目: 教育部科技基础工作专项项目(2014FY110900)

作者简介: 黄志伦(1998-), 男, 硕士研究生, CCF 会员(G9777G), 通信作者, 研究方向为医学图像处理; 刘俊(1977-), 男, 博士, 教授, CCF 会员(14985G), 研究方向为计算机视觉、医学图像处理。

结构的最外层,可以直接观察并触摸到。位于隔绝外界的第二层防护是真皮,这是一种结缔组织,用于连接第一层防护和第二层防护。位于隔绝外界的第三层防护是皮下组织,皮下组织的作用主要是使表皮和真皮可以贴合地附着在骨骼和肌肉之上。实际上,导致感染皮肤病的原因一般并不是单一的某个外界因素,而是内部因素和外部因素结合在一起共同作用的结果。常见的内部因素有细胞核分层遗传、激素分泌过度、免疫系统异常等。

但不管导致皮肤感染并发生炎症的原因是什么,有相关研究人员发现,只要在早期阶段及时发现这些皮肤病,及时开展治疗行动,就一定可以很快得到治愈。皮肤病的早期发现一般是通过肉眼观察和直接触摸,然后通过皮肤镜分析进一步诊断。对于可能感染皮肤病的患者前往医院的常见检测方案一般有贴片测试、表皮培养^[2]。为了确定感染源,患者可能需要花费一笔不菲的费用用于多个步骤的检测,收入较低的患者难以承受。

Celebi 等人^[3]提出了一种顺序流水线方法,他要求实验者必须遵循一系列明确定义的步骤,每一个步骤的固定输出作为另一个步骤的输入。这种方法能有效地将皮肤镜图像分为良性或者黑色素瘤,分为以下 4 个步骤:(1)分割皮肤病灶;(2)从病变边界和病变内的区域提取基于颜色和形状的特征;(3)筛选出判别特征的子集;(4)使用机器学习分类器进行区分。在这之后,陆续出现了很多这种顺序流水线的变体,改进或者省略其中的步骤,但是这种机器学习的方法步骤过于繁琐。Codella 等人^[4]发现可以从数据中学习特征,而不是针对特定皮肤病学标准设计的特征。他应用了一种无监督稀疏编码方法来学习稀疏数量的模式,从而最大限度地减少图像重构误差。Miglan 等人^[5]尝试使用预训练的 EfficientNet 架构,通过微调进行皮肤病图像的分类训练,但是性能还存在不少提升空间。李亚娥等人^[6]提出一种自适应的参数设置方法,通过确定 delta 的最佳取值范围来提高分类的精度。

针对常规分类方法存在的不足,该文提出一种改进双线性特征融合模型。首先,使用 Inception-ResNet-v1 和 v2 用来并行提取图像特征映射,这两个子网络需要进行适当的剪枝以适用于文中分类任务,两个不同的特征提取网络可以提取到不同阶数的特征图,然后考虑到皮肤镜病理图像中不同区域的可参考性不尽相同,在经过双线性特征融合后设计额外的软注意力模块,对每个位置给予不同的权重以加强模型的效果,与多个网络训练结果的对比验证了该方法的有效性。

1 相关工作

双线性网络(Bilinear CNN)^[7]由 2 个 CNN 特征提取器组成,其输出在图像的每个位置使用外部积乘以图像表示。由于 BCNN 可以以平移不变的方式对局部成对特征交互进行建模,所以该模型对于细粒度分类尤其有效。当用于分类任务时,有定义一个四元组 $M = (f_A, f_B, P, C)$ 。其中 f_A 和 f_B 是 2 个 CNN 特征函数, P 是池化函数, C 是分类函数。通过 BCNN 提取的深度视觉为:

$$\varphi(I) = \sum_{l \in L} \text{bilinear}(l, I, f_A, f_B) \quad (1)$$

式中, $\text{bilinear}(l, I, f_A, f_B)$ 是每个位置 $l \in L$ 处的 f_A 和 f_B 的双线性特征组合。映射函数 f 为每个输出位置的特征向量,分类时使用图像特征 φ 训练函数 C 。

2 文中方法

2.1 双线性特征融合结构

该文使用基于 Inception 模块的更高级的 Inception-ResNet-v1 和 v2 作为双线性特征融合提取器。Inception-ResNet-v1 是 Inception 根据残差思想设计出来的变体,具体的做法是在卷积层输出的地方加上一个基于残差思想设计的跳跃连接,通过这种跳跃连接做到直接把 Inception 的输出作为输入添加到残差计算中,此时需要在原始卷积层中添加一个 1×1 的卷积用于保证在进行卷积运算之后输入的维度和输出的维度是一致的,这样原先的 Inception 模块中的一个池化操作要被更新为残差连接。

对于另外一个特征提取器,使用的是 Inception-ResNet-v2,这是一个结合了 Inception v4^[8] 和 ResNet^[9] 的混合模型。此前的 Inception-ResNet-v1 设计之初是为了与 Inception-v3^[10] 进行横向对比,因为在大多数的情况下所能使用的硬件资源是有限的,所以 Inception-ResNet-v1 并没有对模型加深很多和添加更多的参数,在保持和 Inception-v3 的计算开销和资源消耗几乎持平的情况下探究这种融合设计在同样使用环境下的极限性能。而这里的 Inception-ResNet-v2 没有这种限制性的目的。

完整的 Inception-v4 总网络图从对输入图像的预处理模块 Stem 开始,然后连接多个 Inception 子模块,最后进行平均池化和正则化并进行分类,图 1 包含 Inception-ResNet-A 的模块图。这里使用的子模块同样使用了多层卷积并行处理的思想,同时需要注意的是,与以往的人工神经网络相比,Inception 模块并不用人指定使用哪一个过滤器,而是把所有这些参数的值全部添加进该人工神经网络模型中。Inception 模块可以做到自行学习,并可以确定理想使用的参数。在

并行卷积然后汇总在一起进行拼接特征图映射的时候,这里增加的 1×1 Conv 是为了匹配输入进来的特征维度。因为可以看到,在之前通过卷积层的特征提取过程中有使用不带激活函数的卷积层来通过过滤器进行降维,这时输入的图像的通道数会被压缩。而由于这里使用了基于跳跃连接的残差思想,所以会把输

入的特征映射与并行卷积后得到的特征映射联合,这里为了保证两者的通道数相同,需要添加额外的卷积层来扩增并行拼接的特征映射的通道数。剩余的 Inception-ResNet-B 和 Inception-ResNet-C 同样也在图 1 给出。v1 和 v2 版本的 Inception-ResNet-A、B、C 结构上是相同的,区别在于参数设置,用表 1 罗列出。

表 1 两个版本的 A 模块不同卷积的通道数

	v1-A	v2-A	v1-B	v2-B	v1-C	v2-C
1×1 Conv	32	32	128	192	192	192
1×1 Conv	32	32	128	128	192	192
1×1 Conv	32	32	-	-	-	-
1×3 Conv	-	-	-	-	192	224
1×7 Conv	-	-	128	160	-	-
3×1 Conv	-	-	-	-	192	256
3×3 Conv	32	32	-	-	-	-
3×3 Conv	32	48	-	-	-	-
3×3 Conv	32	48	-	-	-	-
7×1 Conv	-	-	128	192	-	-

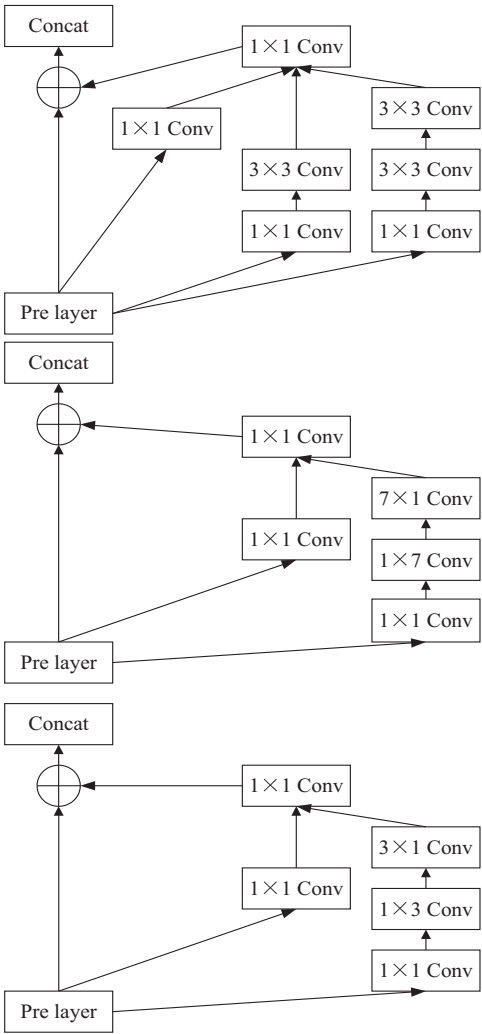


图 1 Inception-ResNet-A(上)、B(中)、C(下)结构

Inception-ResNet-A、B、C 三个子模块只是整个 Inception-ResNet 的一部分,完整的剪枝网络结构通过

嵌入 Reduction 结构以耦合。这是一个融合了稀疏思想和残差思想的网络结构,这两种思想都可以在网络深度增加的时候缓解模型的退化问题。在这里的子模块使用残差思想的跳跃连接会把输入层的特征映射和卷积之后的特征映射拼接在一起,单独使用稀疏结构的 Inception 模块只会并行使用不同大小的卷积层提取特征信息,结合了跳跃连接的残差结构可以做到相比较于单独的 Inception 模块性能得到一定程度的提升。

2.2 软注意力模块

为了加强训练,该文考虑使用软注意力提高模型的性能。注意力机制是目前深度学习人工卷积神经网络领域非常流行的一个概念,这种思想借鉴于人类的视觉注意力。使用注意力可以做到更加关注与当前目标和期望相关的特征信息,在一定程度上忽视掉图像中与训练目标无关的源域信息^[11]。一般来说,注意力被分为硬注意力和软注意力。硬注意力指只选择到某一个位置或者几个位置的信息,软注意力则是会选择全部的信息,计算所有输入信息的加权聚合。一般来说,硬注意力会更加强调点的变化,强调一个随机的预测过程。但是,如果是选取概率最高的那一个点,这个步骤是不可微分的,这使得硬注意力相对于软注意力难以被训练^[12]。

软注意力的做法是让所有的信息都去加权,通过权重比分离出最需要的信息,避免受到过多无用的噪声干扰,最终提高准确率。软注意力是可以微分的,这意味着可以通过神经网络计算梯度,并通过反馈直接学习到软注意力的权重,这是软注意力相比较于硬注

意力更加被广泛使用的主要原因。但是软注意力也有不太完美的地方,部分输入信息对结果是完全没有影响的,但由于软注意力的全局信息再加权的机制,还是必须要让这些完全不会对结果产生影响的信息加入运算,这会造成一定程度的资源消耗^[13]。相关的计算公式如下:

$$x = \sum_{i=1}^{K \times K} l_i X_i \quad (2)$$

其中, l_i 是通过预测 softmax 函数作为原始图像中与特定区域对应的第 i 个位置上的 $K \times K$ 图像的概率而应用的注意力图, X_i 表示从特征提取器模块中提取的图像特征, i 的取值范围为 1 到 $K \times K$ 。实际上,软注意力就是两个特征映射矩阵对应位置元素进行乘积的过程,即注意力特征图和输入特征图。

完整的皮肤病图像分类模型见图 2。分别使用 Inception-ResNet-v1 和 Inception-ResNet-v2 的剪枝模型来提取皮肤镜下病理图像的特征信息,然后进行双线性融合。由于结合了稀疏思想和残差思想,人工卷积神经网络可以做到在网络变的更深的情况下保证性能加强。考虑到皮肤镜下病理图像的不同区域对于进行文中目标工作图像分类产生影响的权重不同,所以考虑使用软注意力来加强对无用特征的过滤。注意力一般分为硬注意力和软注意力,由于硬注意力难以计算,所以一般使用软注意力加强对重要信息的聚合。最后的步骤是整合提取的特征信息并使用 softmax 分类。

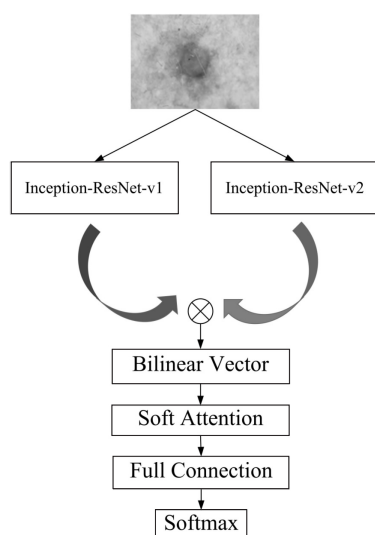


图 2 双线性特征融合网络整体结构

3 实验与分析

3.1 数据集介绍和实验环境

实验使用的数据集是 skin-cancer-classesisc, 包括 2 923 张皮肤病图像, 其中黑色素瘤图像、皮肤痣和良性角化病的图像占据多数, 分别有 454 张、873 张

和 478 张, 实例图像见图 3。实验环境为 64 位 Windows 10 操作系统, GPU 为 GTX 1080Ti, 显存是 11G, 使用 Python 3.6 作为编程语言, 使用 Tensorflow 作为深度学习框架。

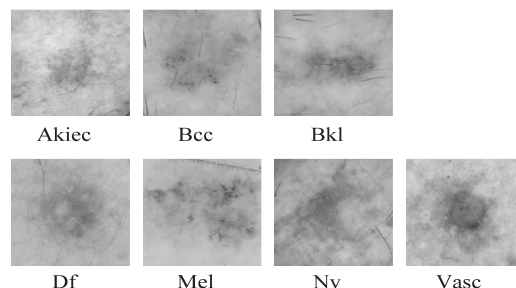


图 3 共 7 种皮肤病图像展示

3.2 预处理

由于该数据集包含的图像数量过少, 该文进行镜像、翻转等方式的数据增强。扩充后设置训练集、测试集、验证集的比例为 80%、15%、5%, 见表 2, 然后把图像大小全部设置为 128×128 并归一化到 0~1。

表 2 扩充后的训练集、验证集、测试集

类别	训练集	验证集	测试集	合计
Akiec	3 139	196	588	3 923
Bcc	3 368	211	633	4 212
Bkl	3 359	210	630	4 199
Df	3 098	194	582	3 874
Mel	3 612	225	675	4 512
Nv	3 532	220	660	4 412
Vasc	3 243	203	609	4 055
合计	23 351	1 459	4 377	29 187

3.3 评价指标

选择合适的评价指标是至关重要的, 因为它给出了一个尺度和标准来衡量深度神经网络模型的优劣, 然后在接下来的完善过程中又会遵循这个评价指标来优化。一般综合使用多种参考指标, 以减少偶然性。为了便于表示, 这里使用 TP 代表真阳性, FP 代表假阳性, TN 代表真阴性, FN 则代表假阴性。

精度的含义是成功预测阳性样本的能力, 即占有所有预测为阳性样本(无论是否预测正确) 的比例, 其计算公式如下:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

召回率是从另一个角度衡量预测阳性样本的能力, 表示为成功预测为阳性样本的个数占有所有阳性样本个数(其中包含误测) 的比例, 其计算公式如下:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

F1 Score 综合考虑精度和召回率, 兼顾了阳性样

本的误报和漏报,是一种相对更均衡的评价标准,其计算公式如下:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5}$$

3.4 实验结果与分析

本节实验设置使用 Inception-ResNet-v1 和 Inception-ResNet-v2 作为双线性特征提取器的网络为 BCNN (IRv1, IRv2), 在训练时初始学习率设置为 0.005, 批量大小设置为 64。每经过 10 个 epoch 将学习率衰减为原先的 80%。在第 100 个 epoch 之后把批量大小设置为 16。

表 3 显示了文中使用的 BCNN (IRv1, IRv2) 的分类结果。Precision 主要表示预测的阳性样本中有多少是被正确预测的,即衡量模型是否预测得准确。Recall 主要表示有多少阳性样本是被正确预测的,即衡量模型是否预测得全面。在实际预测中通常还需要 F1-Score,这是一个同时兼顾分类模型精确率和召回率的评价指标。可以看出,文中算法尽管在黑色素瘤上的表现不是特别优秀,但是在其他 6 种皮肤疾病上都维持一个不错的水平。在皮肤痣上,文中算法可以在 Precision 上达到 0.96,在 Recall 上达到 0.98, F1-Score 同样高达 0.97。分类结果的混淆矩阵见图 4。

表 3 文中模型分类结果

类别	Precision	Recall	F1-Score
Akiec	0.88	0.86	0.87
Bcc	0.84	0.86	0.85
Bkl	0.90	0.84	0.87
Df	0.95	0.78	0.86
Mel	0.66	0.53	0.59
Nv	0.96	0.98	0.97
Vasc	0.91	0.89	0.90

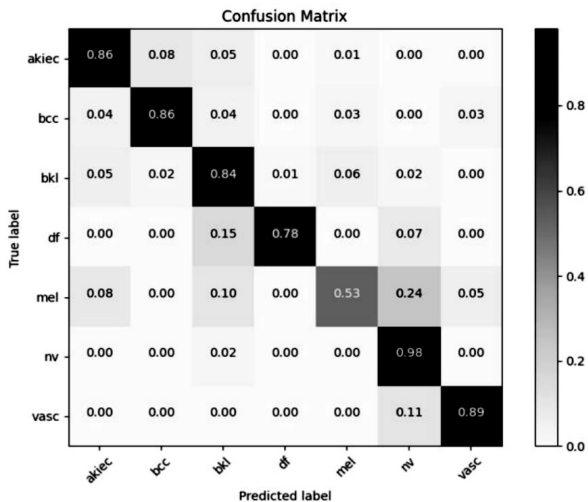


图 4 分类结果的混淆矩阵

表 4 中则是列出文中分类模型和其他 3 种模型的

实验比较。文献[14]设计了一个基于 CNN 的皮肤病分类的交互 Web 平台,便于用户的使用。文献[15]设计了轻量级的卷积网络,同时也可以取得不错的效果。文献[16]设计带有增量结构的分类模型,同样可以提高分类的性能。表中的三种指标都是加权平均值,可以看出在这 3 个指标上文中模型均优于另外三种,由此可见,使用剪枝后的 Inception-ResNet-v1 和 v2 并行提取图像特征,然后对提取到的映射图进行双线性融合,辅以软注意模块加强训练,能够取得较好的效果。

表 4 使用共 4 种分类模型的表现

模型	Precision	Recall	F1-Score
S-CNN ^[14]	0.78	0.76	0.77
MobileNet ^[15]	0.79	0.77	0.78
Incremental CNN ^[16]	0.82	0.80	0.81
BCNN (IRv1, IRv2)	0.87	0.81	0.84

4 结束语

为了解决皮肤病图像在分类时由于皮肤病图像之间存在类间相似性和类内差异性而分类困难的问题,采用双线性特征融合的思想,使用剪枝后的 Inception-ResNet-v1 和 v2 作为双线性特征提取器。这种特殊的网络结构采用稀疏思想和残差思想,能够在一定程度上保证网络模型变深的同时性能不会因此退化。然后考虑到皮肤镜下病理图像中不同区域的特征信息对分类目标结果的影响大小不同,在进行双线性特征汇合后添加软注意力模块,使网络模型可以屏蔽掉无用的特征,专注于更重要的纹理部分。最后在实验部分验证双线性特征提取器的优越性和使用软注意力的加强效果。该工作还可以进一步做分割训练,会更加便于为医师提供计算机辅助诊断,未来的研究将会着眼于此。

参考文献:

[1] MANERKAR S, SHASHWATA H, JUHI S, et al. Classification of skin disease using multi SVM classifier[C]//3rd international conference on electrical, electronics, engineering trends, communication, optimization and sciences (EEE-COS). Tadepalligudem: Institution of Engineering and Technology, 2016: 363-368.

[2] 党宁宁. 不同皮肤病生活质量的比较研究[J]. 中国皮肤性病学杂志, 2004, 18(4): 221-222.

[3] CELEBI M E, KINGRAVI H A, UDDIN B, et al. A methodological approach to the classification of dermoscopy images[J]. Computerized Medical Imaging and Graphics, 2007, 31(6): 362-373.

[4] CODELLA N F C, GUTMAN D, CELEBI M E, et al. Skin

- lesion analysis toward melanoma detection; a challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC) [C]//International symposium on biomedical imaging (ISBI). Washington D. C; IEEE, 2018; 168–172.
- [5] MIGLANI V, BHATIA M. Skin lesion classification: a transfer learning approach using efficientnets [C]//International conference on advanced machine learning technologies and applications. Jaipur; Springer, 2020; 315–324.
- [6] 李亚娥, 汪西莉. 一种自适应的半监督图像分类算法[J]. 计算机技术与发展, 2013, 23(2): 112–114.
- [7] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear CNN models for fine-grained visual recognition [C]//Proceedings of the IEEE international conference on computer vision (ICCV). Santiago; IEEE, 2015; 1449–1457.
- [8] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]//Thirty-first AAAI conference on artificial intelligence. San Francisco; AAAI Press, 2017; 4278–4284.
- [9] 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 (CVPR). Las Vegas; IEEE, 2016; 770–778.
- [10] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). Las Vegas; IEEE, 2016; 2818–2826.
- [11] PAN Y, ZHANG L. Dual attention deep learning network for automatic steel surface defect segmentation [J]. Computer - Aided Civil and Infrastructure Engineering, 2021, 15(6): 1–20.
- [12] MNIH V, HEES N, GRAVES A. Recurrent models of visual attention [C]//Proceedings of the 27th international conference on neural information processing systems. Cambridge; MIT Press, 2014; 2204–2212.
- [13] XIAO T, XU Y, YANG K, et al. The application of two-level attention models in deep convolutional neural network for fine-grained image classification [C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). Boston; IEEE, 2015; 842–850.
- [14] SUBRAMANIAN R R, ACHUTH D, KUMAR P S, et al. Skin cancer classification using convolution neural networks [C]//11th international conference on cloud computing, data science & engineering (Confluence). Noida; IEEE, 2021; 433–442.
- [15] CHATURVEDI S S, GUPTA K, PRASAD P S. Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet [C]//International conference on advanced machine learning technologies and applications. Singapore; Springer, 2020; 165–176.
- [16] CHOPADE A. Seven class classification of skin lesions by using incremental convolutional neural network in python [C]//Proceedings of the international conference on advances in electronics, electrical & computational intelligence (ICAEEC). Allahabad; IIIT, 2019; 457–465.