

文章编号 1004-924X(2024)24-3644-14

多尺度与三维交互特征优化的皮肤镜图像分类

王 迪¹, 吕晓琪^{1,2*}, 李 菁¹

(1. 内蒙古科技大学 数智产业学院, 内蒙古 包头市 014010;

2. 内蒙古工业大学 信息工程学院, 内蒙古 呼和浩特 010051)

摘要: 皮肤癌的早期诊断有助于增加病患的治愈机会, 减轻医疗系统的负担。针对皮肤镜图像分类存在的特征提取过程中会造成信息丢失与难以同时识别图像中独立的不同类型特征信息等问题, 提出了一种多尺度与三维交互特征优化网络(MTIFNet)。首先, 该网络应用多尺度空间自适应模块, 在训练过程中提取图像全局与局部上下文信息, 利用多尺度增强病灶周围模糊的像素点与像素点间的联系。其次, 设计了一个在不同维度之间建立联系、对信息进行交流和整合的三维交互特征优化模块。最后, 使用交叉熵损失衡量预测概率分布与真实类别分布之间的差异, 以优化模型的准确性。基于 ISIC 2018 与 ISIC 2017 数据集的实验结果表明, 该网络分别将准确率、精确率、召回率、特异性分别提升至 94.32%, 91.61%, 93.00%, 98.39% 和 98.57%, 98.20%, 98.47%, 99.13%。与目前流行的 ResNeSt, ConvNext, Fcanet 等分类网络相比, MTIFNet 具备更好的特征提取与交互能力, 有助于医生做出更准确的诊断和治疗决策。

关键词: 皮肤癌; 图像分类; 多尺度; 空间自适应; 三维交互

中图分类号: TP394.1; TH691.9 文献标识码: A doi: 10.37188/OPE.20243224.3644

Dermoscopic image classification based on multi-scale and three-dimensional interaction feature optimization

WANG Di¹, LÜ Xiaoqi^{1,2*}, LI Jing¹

(1. School of Digital and Intelligence Industry, Inner Mongolia University of Science and Technology, Baotou 014010, China;

2. School of Information Engineering, Inner Mongolia University of Technology, Hohhot 010051, China)

* Corresponding author, E-mail: lxiaoqi@imut.edu.cn

Abstract: Early diagnosis of skin cancer is crucial for improving patient outcomes and alleviating the burden on the healthcare system. However, the process of feature extraction in skin cancer image classification often results in information loss and challenges in simultaneously identifying independent types of features in the images. MTIFNet was proposed, which was a network that integrates three-dimensional spatial attention with information fusion. Initially, the network employed a multi-scale spatial adaptive module to extract both global and local contextual information from images during training. This module enhanced the connection between blurred pixels around lesions and the relationship in pixels at different scales. Subsequently, a three-dimensional interaction feature optimization module was introduced to facilitate connections across different dimensions, enabling the exchange and integration of information. Final-

收稿日期: 2024-04-03; 修订日期: 2024-07-15.

基金项目: 国家自然科学基金项目 (No. 61771266)

ly, cross-entropy loss was used to measure the difference between the predicted probability distribution and the true class distribution to optimize the accuracy of the model. The experimental results based on the ISIC 2018 and ISIC 2017 datasets indicate that the network has improved accuracy, precision, recall, and specificity to 94.32%, 91.61%, 93.00%, 98.39% and 98.57%, 98.20%, 98.47%, 99.13%, respectively. Compared to currently popular classification networks such as ResNeSt, ConvNext, and Fcanet, MTIFNet demonstrates superior capabilities in feature extraction and interaction, thereby assisting health-care professionals in making more precise diagnostic and treatment decisions.

Key words: skin cancer; image classification; multi-scale; spatial adaptation; three-dimensional interaction

1 引 言

皮肤癌是各肤色种族中常见的疾病,它影响着约 19 亿总人口中的近 33%。按平衡寿命年估计,皮肤病约占全球疾病问题的 1.79%^[1]。其中黑色素瘤是目前致死率最高的皮肤恶性肿瘤(全球死亡率约为 14%)。黑色素瘤的治愈率和预后治疗取决于早期的识别、是否发生淋巴结、肿瘤的大小等。早期诊断和治疗黑色素瘤可以降低其死亡率,并且 5 年生存率可以达到 95%^[2]。为了改善皮肤疾病的诊断效果,临床上引入皮肤镜^[3]检查以协助皮肤科医生进行诊断。利用皮肤镜减少皮肤表面的反射,增强对皮肤深层的视觉效果,并提供更多关于皮肤损伤的细节信息。常见的诊断皮肤镜图像的方法有 ABCD 准则、模式分析法、孟氏法和七点特征法,这些方法主要对应皮损区域各种形态学特征。然而,皮肤镜检查仍有不足,皮肤科医生对皮肤镜图像进行人工检查通常耗时耗力、主观性强,且图像中存在人眼难以区分的特征相似病变区域。因此,需要皮肤癌的自动识别系统,以协助医生诊断皮肤疾病^[4]。

为解决皮肤镜图像分类问题,一些研究者采用了传统的特征提取方法,如局部二值模式(Local Binary Patterns, LBP)和方向梯度直方图(Histogram of Oriented Gradients, HOG),来捕捉图像中的纹理和形状等特征。深度神经网络已经在医学图像分析领域取得了巨大进展,特别是在对象检测、自然图像分类以及图像分割等任务方面^[5]。与传统的特征工程方法相比,深度神经网络展现出更优越的性能,为医学图像分析带来了新的突破。Gu 等^[6]提出了用于跨域皮肤病识别新方法,从在 ImageNet 上预训练的完全监督的深度卷积神经网络分类器开始,本文通

过在两个皮肤疾病数据集上微调网络来探索两步渐进迁移学习技术。Jasil 等^[7]使用 DenseNet201 预训练架构的神经网络,该网络对七类皮肤镜图像进行分类,实现皮肤癌的早期检测。Sahib 等^[8]提出了一种可以去除图像中的伪影、增强图像并自动调整图像的对比度和亮度的预处理方式,同时引入了新的深度学习网络架构 Inception-Resnet 解决分类问题使分类更加稳健。李等^[9]提出一种改进 ConvNeXt 的算法,即 SE-SimAM-ConvNeXt 网络。降低皮肤镜图像中毛发、气泡等复杂背景的干扰,提升网络模型的抗干扰能力,同时解决样本分布不均衡的问题。皮肤镜图像具有复杂的特征和结构,简单的网络结构难以捕捉到图像中的高级特征信息,导致分类性能的不足。因此,为了更有效地实现皮肤镜图像分类,需要采用更复杂和深层的网络结构,以提高模型对图像特征的表达能力和分类准确性。Zhuang 等^[10]提出了一个简单而有效的框架,名为病变区域类注意力(CARE),通过将注意力嵌入到卷积神经网络的训练过程中来处理数据不平衡问题。所提出的注意力模块可以帮助卷积神经网络(Convolutional Neural Networks, CNN)关注罕见疾病的病变区域,从而帮助 CNN 更有效地学习其特征。Abdelhalim 等^[11]它将注意力机制与渐进式增长的生成对抗网络(Progressive Generative Adversarial Networks, PGGAN)相结合,获取皮损图像的全局特征,并引入两时间尺度更新规则来生成高细粒度的特征,同时增加生成对抗网络的稳定性。Pan 等^[27]提出了利用残差网络提取局部特征的皮肤病变诊断聚合算法和利用 Fisher Vector 编码将局部特征聚合到图像级表示的皮肤病变诊断算法。Yan 等^[28]提出了一种端到端黑色素瘤识别网络,

它加入了一个关注模块来关注中间层内皮肤病变的局部特征。目前,针对皮肤癌分类任务,研究重点主要集中在解决类别不平衡问题上,而忽视了对皮肤病灶本身细节特征的深入分析。尽管一些论文已经开始探索全局和局部特征的分析,尚缺乏一种全面且有效的方法来同时考虑皮肤病灶的全局特征和局部特征,从而实现更准确和细致的分类。且皮肤镜图像与自然图片存在色彩、光照等差异,在图像分类中,模型难以同时识别图像中独立的纹理、颜色和形状等不同类型的特征信息。

针对上述问题,本文提出了一种名为多尺度与三维交互特征优化的网络(Multi-scale and Three-dimensional Interaction Feature Networks, MTIFNet),旨在应用于皮肤癌图像分类任务。首先,采用了 3×3 的卷积核来替换 7×7 的卷积核,以降低模型的参数量。其次,在网络的不同层之间引入了多尺度空间自适应模块,首先通过空间自适应特征调控模块获取图像的全局上下

文信息,然后结合空洞卷积组合模块来获取局部上下文信息,有助于更准确地确定皮肤病变与周围健康组织的边界,进而识别病变的类型和严重程度。最后,为了弥补空间与维度信息独立的不足,在MTIFNet的Block结构中应用了三维交互特征优化模块。该模块使神经网络能够更好地理解数据的空间结构,并通过跨维度交互来关注不同空间位置,从而获取更有意义的特征表示,有助于更好地处理皮肤癌图像中不同类型特征之间的空间与维度信息的关联。此外,还应用交叉熵损失函数使训练过程稳定进行。

2 皮肤癌图像分类方法

2.1 网络模型

以 ResNeSt 模型^[12]为基础,本文提出了 MTIFNet 模型来实现皮肤镜的分类任务,其网络结构如图 1 所示。MTIFNet 的 Conv1 部分由三个 3×3 的卷积替换 7×7 卷积构成的卷积层将输

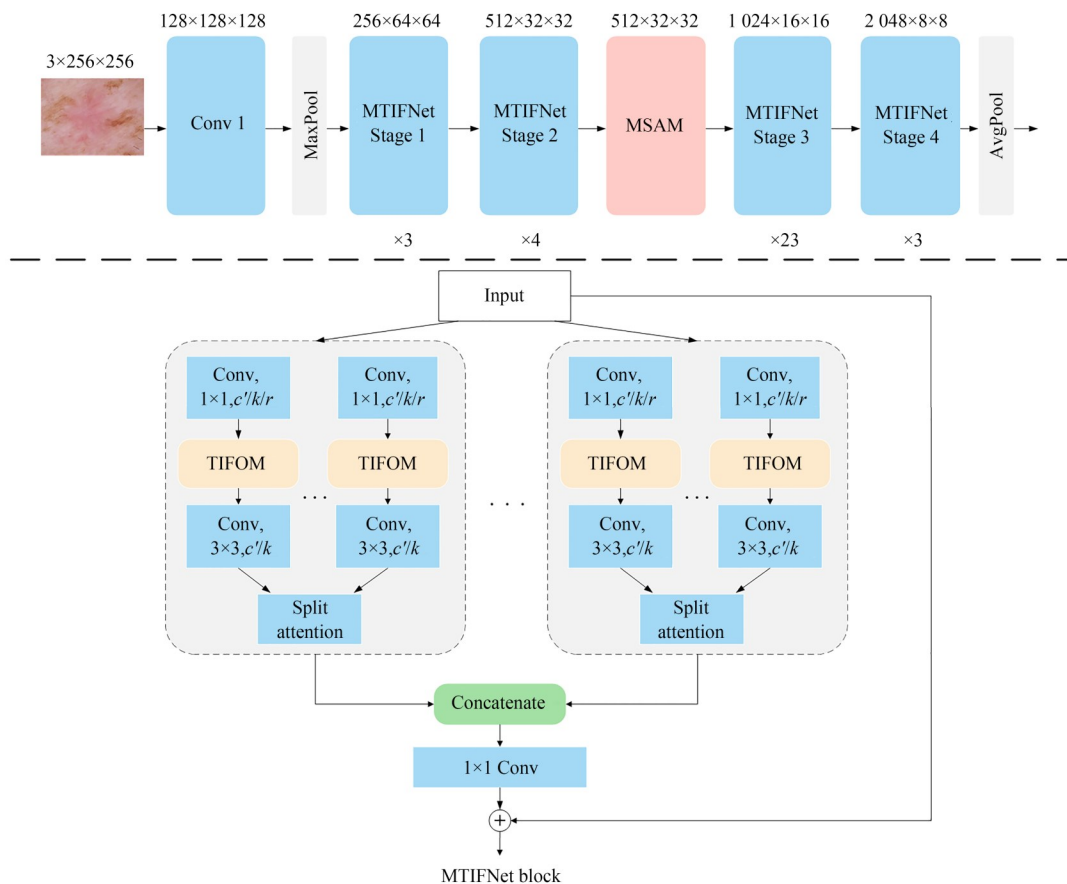


图 1 MTIFNet 整体结构图

Fig. 1 Overall network architecture of our MTIFNet model

入图像进行初步特征提取和处理,以获得更好的训练效果。再通过最大池化层将特征图的尺寸缩减至 64×64 ,由于最大池化只保留局部区域内的最大值,它可以帮助减少图像中的噪声,使得模型更关注皮肤病变的边缘轮廓、皮肤纹理等重要特征。在 MTIFNet Stage2 后应用多尺度空间自适应模块,它由两部分组成,先经过空间自适应特征调控模块以多尺度金字塔的形式获取全局上下文信息,后接空洞卷积组合模块捕获像素与像素间的局部上下文信息。

MTIFNet 网络的详细设置可见表 1。在 ResNeSt 中每个 Block 的特征首先由基数超参数 K 可以被划分为几个组,再使每个组中引入了一个新的超参数 R 它表示基数群中的分裂数,因此特征组的总数为 $G=K \times R$ 。它将通道关注应用于不同的网络分支,以利用它们在捕获跨特征交互和学习不同表征能力。与其不同的是,MTIFNet 在每个 MTIFNet Block 中的首个 1×1 卷积后加入三维交互特征优化模块,建立层层堆叠的 Block 中建立三维连接、增强跨维度交互能力。

表 1 MTIFNet 的详细设置

Tab. 1 Detailed settings of MTIFNet

Stage	Output size	MTIFNet
Conv1	$\frac{H}{2} \times \frac{H}{2}$	Conv $3 \times 3 \times 3$
Maxpool	$\frac{H}{4} \times \frac{H}{4}$	Conv 3×3
MTIFNet Stage1	$\frac{H}{4} \times \frac{H}{4}$	$\begin{bmatrix} \text{Conv } 1 \times 1 \\ \text{TIFOM} \\ \text{Conv } 3 \times 3 \\ \text{Conv } 1 \times 1 \end{bmatrix} \times 3$
MTIFNet Stage2	$\frac{H}{8} \times \frac{H}{8}$	$\begin{bmatrix} \text{Conv } 1 \times 1 \\ \text{TIFOM} \\ \text{Conv } 3 \times 3 \\ \text{Conv } 1 \times 1 \end{bmatrix} \times 4$
MSAM	$\frac{H}{8} \times \frac{H}{8}$	$\begin{bmatrix} \text{SAFRM} \\ \text{DWCCM} \end{bmatrix}$
MTIFNet Stage3	$\frac{H}{16} \times \frac{H}{16}$	$\begin{bmatrix} \text{Conv } 1 \times 1 \\ \text{TIFOM} \\ \text{Conv } 3 \times 3 \\ \text{Conv } 1 \times 1 \end{bmatrix} \times 23$
MTIFNet Stage4	$\frac{H}{32} \times \frac{H}{32}$	$\begin{bmatrix} \text{Conv } 1 \times 1 \\ \text{TIFOM} \\ \text{Conv } 3 \times 3 \\ \text{Conv } 1 \times 1 \end{bmatrix} \times 3$
Average pool, 7-d fc, softmax		

跨特征交互和跨通道交互允许模型从不同的角度和维度来考虑数据,从而生成更丰富和更具表征性的特征表示。最后一层平均池化,用于将特征图转换为一个固定长度的特征向量,作为输入给全连接层分类器,用于进行最终的分类任务。

2.2 多尺度空间自适应模块

皮肤疾病需要通过皮肤镜成像方可清晰地观察,皮肤病灶组织特征也较为复杂,所以相比于自然图像皮肤疾病的检测更加棘手。为了避免信息丢失设计了多尺度空间自适应模块 (Multi-scale Spatial Adaptation Module, MSAM) 如图 2 所示,MSAM 可以分为两个模块。空间自适应特征调控模块 (Spatial Adaptive Feature Regulation Module, SAFRM) 将特征矩阵按通道分成四等份进行多尺度特征提取,通过不同感野获取模型不同尺度的全局特征,为了同时捕获局部上下文信息设计空洞卷积组合模块 (DW Convolutional Combination Module, DWCCM),增强病变区域像素点与像素点间的联系,有助于了解皮肤病变在局部区域的分布情况以及可能的扩散路径,更准确地在皮肤镜图像中识别病变区域。

2.2.1 空间自适应特征调控模块

基于对 Sun 等^[13]提出的金字塔结构模型的思考,本文提出了用于从多尺度特征表示中学习长程相关性的空间自适应特征调控模块,以更有效地挖掘有用的特征用于提高皮肤癌图像分类准确性。为了简化模型结构并获取金字塔特征表示,首先对归一化后的输入特征进行通道分裂操作,生成四个部分分量。对第一部分进行 3×3 深度卷积处理,而其余部分则输入多尺度特征生成单元。对于给定输入特征 X ,这一过程可以表示为:

$$[X_0, X_1, X_2, X_3] = \text{Split}(X), \quad (1)$$

$$\text{DW_Conv}_{3 \times 3}(X_0), \quad (2)$$

$$\hat{X}_0 = \uparrow_p(\text{DW_Conv}_{3 \times 3}(\downarrow_{\frac{p}{2'}}(X_i))), 1 \leq i \leq 3, \quad (3)$$

其中: $\text{Split}(\cdot)$ 表示通道分裂操作, $\text{DW_Conv}_{3 \times 3}(\cdot)$ 表示核大小为 3×3 的深度卷积, $\uparrow_p(\cdot)$ 表示为了快速实现,通过最近的内插将特定级别的特征向上采样到原始分辨率 p ,而 $\downarrow_{\frac{p}{2'}}$ 表示将输入特征向下

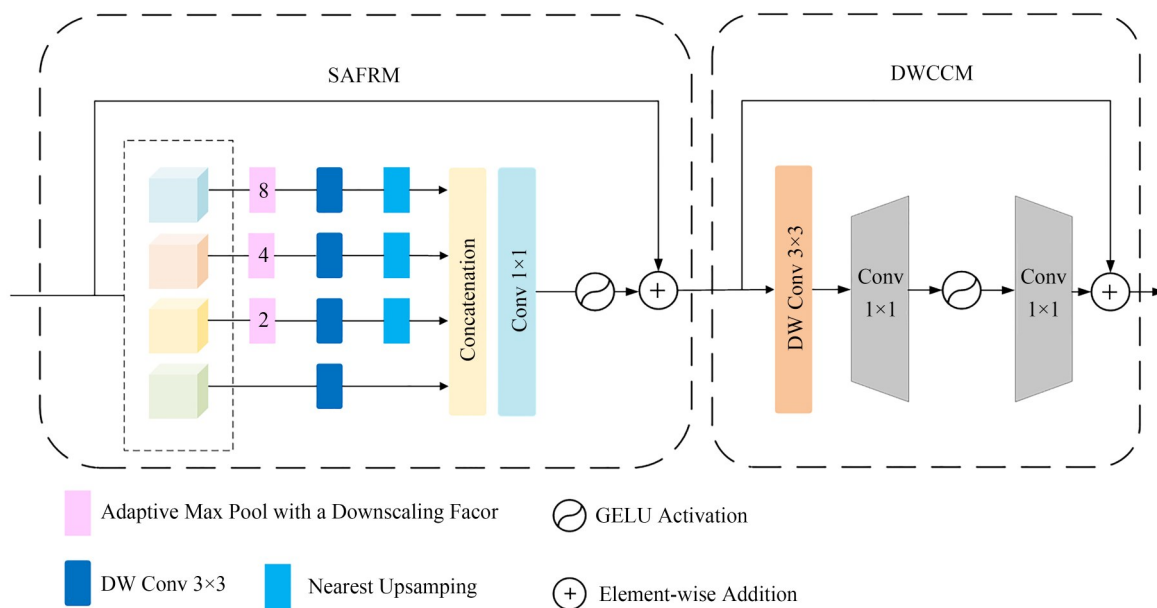


图2 多尺度空间自适应模块

Fig. 2 Multi-scale spatial adaptation module

采样到 $\frac{p}{2^i}$ 的大小。由于此模块的目标是选择识别特征以学习非局部交互作用,因此对输入特征应用自适应最大池操作,生成多尺度特征。结果表明,这种最大池操作有助于提高重建性能。接下来,本文将这些多尺度特征通过 1×1 卷积来聚合局部和全局关系:

$$\hat{X} = \text{Conv}_{1 \times 1}(\text{Concat}([\hat{X}_0, \hat{X}_1, \hat{X}_2, \hat{X}_3])), \quad (4)$$

其中: $\text{Concat}(\cdot)$ 表示沿通道维度的串联操作, $\text{Conv}_{1 \times 1}(\cdot)$ 表示 1×1 卷积。在得到精化表示后,本文通过 GELU 非线性激活函数对其进行归一化以估计注意力映射,并根据估计的注意力在元素乘积中自适应地调制 $\bar{X}$:

$$\bar{X} = \Phi(\hat{X}) \oplus \bar{X}, \quad (5)$$

其中: $\Phi(\cdot)$ 表示 GELU 激活函数, $\oplus$ 是元素相加。

2.2.2 空洞卷积组合模块

SAFRM 模块专注于探索全局上下文信息,捕捉整个图像中的重要特征,但皮肤镜图像包含大量相似的空间结构,为了紧密建立病灶周围临近像素与像素间的局部上下文信息,借鉴 Chen^[14] 的思想,本文提出了空洞卷积混合模块来增强局部空间的建模能力。DWCCM 模块通过使用局部感知和共享权重的卷积操作来捕捉图像中的局部上下文。这意味着网络会对每个像素点周

围的像素进行卷积操作,从而局部地获取特征和上下文信息。

在图2中可以看到 DWCCM 模块包含一个 3×3 深度可分离卷积和两个 1×1 卷积。其中,深度可分离卷积通过分解卷积操作,使得模型更加容易学习到输入数据的空间特征和通道间的相关性,且在保持较好性能的同时减少计算量和模型大小。本文使用两个连续的 1×1 卷积变换通道特征的增减探测不同的局部上下文信息。

2.3 三维交互特征优化模块

在皮肤癌分类任务中,图像包含多种不同类型的特征,如颜色、纹理和形状等。在 ResNeSt 网络中,因其利用特征的分组再分组的方法将通道关注应用于不同的网络分支,实现跨特征的交互能力,却忽略了空间维度的交互。借鉴 Misra^[15] 的思想,本文在 MTIFNet 中应用三维交互特征优化模块 (Three-dimensional Interaction Feature Optimization Module, TIFOM) 如图3所示,它是一个具有三重支路的网络结构,获取空间维度和通道维度之间的交互。使不同维度之间建立联系、进行信息交流和整合。受 Yang 等^[16] 启发在每条支路中使用一个可以在空间中有效生成真实 3-D 权重,创造优化能量函数来找出每个神经元的重要性。

三维交互特征优化模块它由三个并行分支

组成,其中两个负责捕获通道维度 C 与高度维度 H 或宽度维度 W 之间的跨维度交互。第三个分支类似于卷积注意力模块 (Convolutional Block Attention Module, CBAM), 用于构建空间注意力^[17]。最后将三个分支的输出结果经过简单平均进行聚合。

对于输入张量 $\chi \in R^{C \times H \times W}$, 首先将其分别传递给所提出的三元组注意模块中的三个分支。在第一个分支中, 建立宽度维度和通道维度的交互, 沿 W 轴逆时针旋转 90° 得到旋转的张量表示为 $\hat{\chi}_1$ 。在第二个分支中, 将高度维度和通道维度之间建立交互, 沿 H 轴逆时针旋转 90° 得到旋转张量表示为 $\hat{\chi}_2$ 。在最后一个分支中, 张量 χ 不改变表示为 χ_3 。通过三个分支分别捕获输入张量 (C, H) , (C, W) 和 (H, W) 维度之间的依赖关系, 再将三个分支的张量 $\hat{\chi}_1, \hat{\chi}_2, \chi_3$ 传递到池中并随后分别缩减为 $\hat{\chi}_1^*, \hat{\chi}_2^*, \chi_3^*$, Z 池表示为 Z_pool , 负责连接平均池化和最大池化特征, 使张量的第一个维

度减少到 2 维。使得该层在保留实际张量能够丰富表示的同时, 缩小其深度以使进一步的计算变得轻量级。其公式可表示为:

$$Z_pool = [MaxPool_{0d}(\chi)], [AvgPool_{0d}(\chi)]. \quad (6)$$

$\hat{\chi}_1^*, \hat{\chi}_2^*, \chi_3^*$ 通过由 7×7 的卷积核, 后接能量函数, 为了找到这些神经元最简单的实现是测量一个目标神经元与其他神经元之间的线性可分离性。根据这些神经科学发现, 本文为每个神经元定义以下能量函数:

$$e_i(w_i, b_i, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (w_i x_i + b_i))^2 + (1 - (w_i t + b_i))^2 + \lambda w_i^2, \quad (7)$$

其中: t 和 x_i 表示输入特征 $\chi \in R^{C \times H \times W}$ 的单个通道中的目标神经元和其他神经元。 i 表示空间维度上的索引, $M = H \times W$ 表示该通道上的神经元数量。 w_i 和 b_i 表示变换的权重和偏置, 并在方程式中添加正则化器。

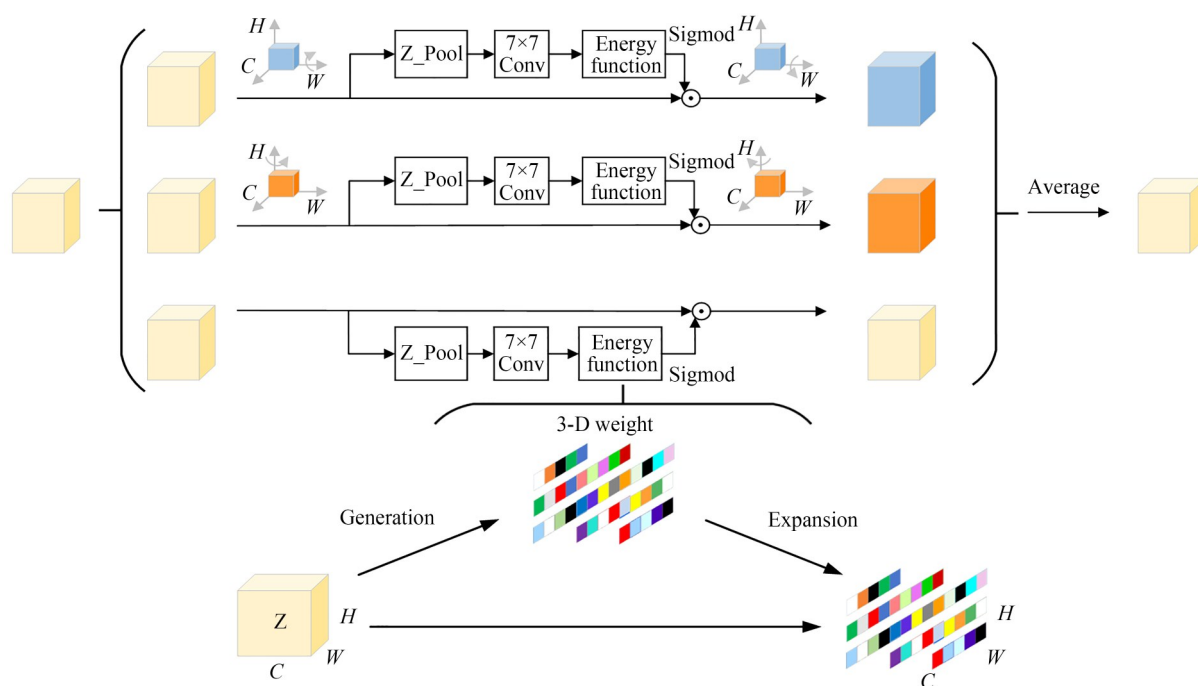


图 3 三维交互特征优化模块

Fig. 3 Three-dimensional interaction feature optimization module

推导出能量函数并找到了每个神经元的重要性。并且哺乳动物大脑中的注意力调节通常表现为对神经元反应的缩放效应。因此, 使用缩放运算符而不是加法来进行特征细化。能量函

数的整个细化阶段是:

$$\bar{X} = \sigma\left(\frac{1}{E}\right) \odot X, \quad (8)$$

其中: $\odot$ 表示逐元素乘法, $\sigma(\cdot)$ 表示 Sigmoid 激活

函数, E 将所有跨通道和空间维度进行分组。

最后经过张量批量归一化层然后传递到 Sigmoid 激活层来生成最终的注意力权重, 将其应用于输入张量 $\hat{\chi}_1^*, \hat{\chi}_2^*, \hat{\chi}_3^*$ 。并且第一个分支输出后沿 W 轴顺时针旋转 90° 以保持与输入 χ 相同的形状, 第二个分支沿 H 轴顺时针旋转 90° 以保留 χ 的原始输入形状。将三个分支的输出结果经过简单平均进行聚合。

综上所述, 对于输入张量 $\chi \in R^{C \times H \times W}$, 从三元组注意力中获得细化注意力应用的过程得到张量的过程可以用式(9)表示:

$$T = \frac{1}{3} (\overline{\hat{\chi}_1 \sigma(e_t(\text{Conv}_{7 \times 7}(\hat{\chi}_1^*)))}) + \overline{\hat{\chi}_2 \sigma(e_t(\text{Conv}_{7 \times 7}(\hat{\chi}_2^*)))}) + \overline{\hat{\chi}_3 \sigma(e_t(\text{Conv}_{7 \times 7}(\hat{\chi}_3^*)))}) \quad (9)$$

其中: $e_t(\cdot)$ 表示能量函数, $\text{Conv}_{7 \times 7}(\cdot)$ 表示核大小为 7×7 的卷积。

2.4 损失函数

皮肤镜图像数据在临床诊疗过程产生, 无法像自然图像拥有海量样本数, 且黑色素瘤等恶性肿瘤仅发于少部分患者, 这使得所采集的数据会造成数据集不均衡的影响的结果。交叉熵损失函数(Cross-Entropy Loss Function)在分类任务中可以帮助平衡数据集, 使得模型更加关注少数类别, 从而提高整体分类性能, 具体如下:

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(\hat{y}_{ij}) \quad (10)$$

其中: $L(y, \hat{y})$ 表示交叉熵损失函数, C 表示类别的数量, N 表示样本数量, y_{ij} 表示样本 i 属于类别 j 的真实标签, $\hat{y}_{ij}$ 表示模型预测样本 i 属于类别 j 的概率。

3 实验与分析

3.1 实验环境

实验在 Ubuntu16.04 LTS 操作系统上完成, 采用 Python 3.8 编程环境和 Pytorch 1.10 框架。基本配置为: Intel(n) Xeon E5-2698 V4 CPU 系列, Nvidia V100 GPU, 显存容量 32 GB。实验中的训练迭代次数为 100 次, 批次样本数设置为 64, 优化器为 AdamW。同时, 在训练过程中使用验证集保存最优模型, 避免过拟合现象。

3.2 数据集

本文采用 ISIC 2018 数据集^[18]进行实验, 来自国际皮肤成像协会(The International Skin Imaging Collaboration, ISIC)2018 年皮肤镜图像分类公开挑战赛。该数据集由国际权威临床中心提供, 经专业皮肤科医生进行注释和标记。图 4 可以看到, ISIC 2018 数据集由尺寸为 600×450 的 10 015 张皮肤镜图像组成分别是: 黑色素瘤(MEL)、黑色素细胞痣(NV)、基底细胞癌(BCC)、光化性角化病(AKIEC)、良性角化病(BKL)、皮肤纤维瘤(DF)、血管病变(VASC)。实验采用 ISIC 2017 数据集用作泛化实验, 包含 2 750 张尺寸为 $1\,022 \times 767$ 的图像。数据集中有 3 个类别的图像: 良性痣(Benign nevi)、脂溢性角化病(Seborrheic keratosis)和黑色素瘤(Melanoma)。泛化实验为二分类, 良性痣与脂溢性角化病为良性样本, 黑色素瘤为恶性样本。本实验将皮肤镜图像按 7:2:1 的比例分为训练集、测试集与验证集, 同时固定测试集, 将剩余图像随机划分十次分为训练集与验证集。实验数据为同等条件下 10 次实验得到结果的平均值与标准差。

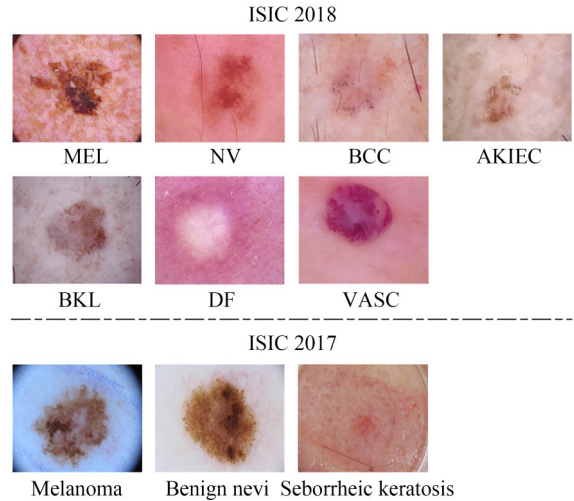


图 4 数据集类别实例

Fig. 4 Example of dataset categories

3.3 评价指标

为评价网络对皮肤镜图像的分类效果, 本实验使用准确率(Accuracy, Acc)、精确率(Precision, Pre)、召回率(Recall, Rec)和特异性(Specificity, Spe)共同对分类结果进行评估。其次, 为

了评估模型的计算效率,使用浮点运算次数(FLOPs)、与单次前向传播时间(Time)。其中,准确率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,特异性反映了模型预测正确的所有负样本占实际所有负样本的概率。它们的表达式为:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \quad (11)$$

$$Precision = \frac{TP}{TP + FP}, \quad (12)$$

$$Recall = \frac{TP}{TP + FN}, \quad (13)$$

$$Specificity = \frac{TN}{TN + FP}, \quad (14)$$

其中:TP为真阳性(truepositive),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(falsepositive),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(trueneegative),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(falsenegative),代表实例在分类器中的预测结果为负,而实际结果为正。Accuracy, Precision、Recall和 Specificity 值越高,说明模型分类效果越好。

3.4 迁移学习效果对比分析

实验采用迁移学习的方法,使用初始模型在ImageNet数据集上已训练好的模型参数,将这些参数迁移到皮肤镜图像分类任务中,使其能够捕捉图像间通用的特征和模式,从而提高模型的性能。实验对未改进的ResNeSt模型和预训练ResNeSt模型的准确率和交叉熵损失进行对比,如图5所示。应用了预训练后的ResNeSt对比最初的模型的准确率有明显提高,且交叉熵损失能够更快速、稳定地收敛。

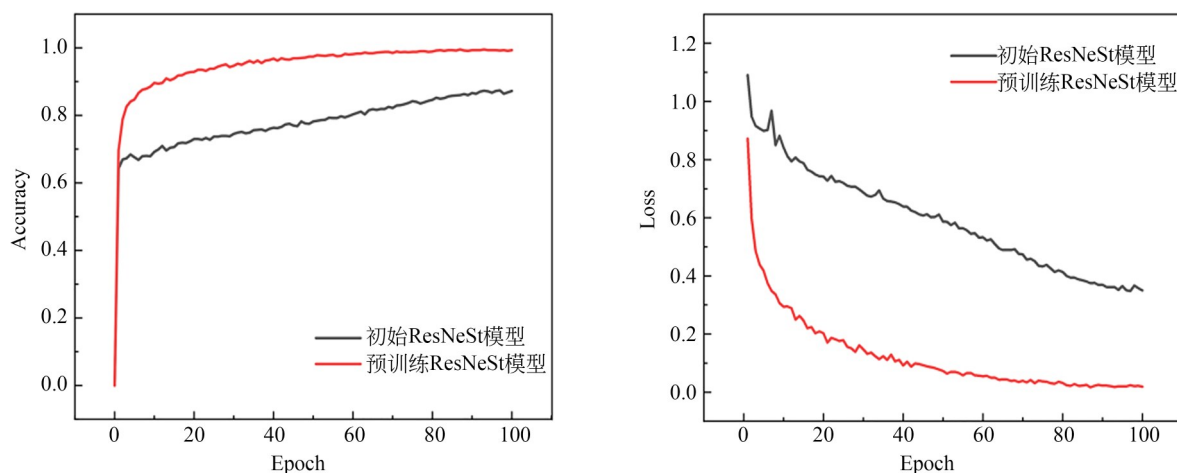


图5 迁移学习对实验的影响

Fig. 5 Impact of transfer learning on experiments

3.5 消融实验结果和分析

3.5.1 MSAM模块的有效性实验

将构建的多尺度空间自适应模块应用在Densenet121, Fcanet50, ResNet101, ConvNext和ResNeSt101网络中进行分类实验,验证该模块在各种网络架构下的通用性和稳健性。从表2可以看出,MASM模块对这些卷积网络的分类任务均有积极作用,对比各网络使用该模块前后的Accuracy指标分别提升了1.14%, 0.40%, 0.35%, 0.88%, 1.29%。FLOPs, Time在使用

了MASM模块后均有增加,各网络对比下来ResNeSt101的计算量与计算时间增加较少。其中ResNeSt101未使用模块的基准性能最高,同时使用模块后Accuracy, Precision, Recall, Specificity的提升几乎最多达到了93.85%, 91.23%, 90.70%, 98.27%。

为检验MSAM模块的位置对皮肤镜图像的分类效果的影响,分别将其置于MTIFNet网络的4个Stage后。根据表3可观察该模块置于各位置的模型效果。实验得出结论,当该模块置于

表 2 不同模型使用 MSAM 的效果

Tab. 2 Comparison of effects of different models using MSAM

Method	MSAM	Acc	Pre	Rec	Spe	FLOPs/G	Time/s
Densenet121	✓	85.57±0.47	76.35±0.93	72.15±0.56	96.42±0.47	239.55	0.560 7
		86.71±0.41	76.93±0.58	72.10±0.57	97.02±0.43	265.34	0.575 4
Fcanet50	✓	90.71±0.35	83.47±0.63	86.79±0.93	97.36±0.34	43.03	0.568 7
		91.11±0.61	86.86±0.62	86.90±2.11	97.73±0.78	249.93	0.821 6
ResNet101	✓	90.30±0.65	84.08±0.90	84.56±0.40	97.44±0.49	81.84	0.722 7
		90.65±0.62	85.63±0.61	84.64±0.93	97.90±0.46	94.79	0.577 1
ConvNext	✓	92.07±0.46	88.00±0.96	88.04±0.49	97.97±0.36	160.67	0.583 4
		92.95±0.50	90.39±0.79	87.98±0.88	97.99±1.01	163.90	0.858 7
ResNeSt101	✓	92.56±0.35	89.02±0.96	86.33±1.92	98.00±0.44	107.04	0.575 5
		93.85±0.56	91.23±2.53	90.70±2.65	98.27±0.18	119.98	0.609 0

第二个 Stage 后时 Accuracy, Precision, Recall, Specificity 分别提升至 93.85%, 91.23%, 90.70%, 98.27%。各 Satge 的 FLOPs 相差不

大、Stage2 的单次前向传播时间对比 Stage4 缩减了 0.201 2 s, 可以看出当 MSAM 置于第二个 Stage 后时网络的分类的效果最好。

表 3 MSAM 位置的消融实验

Tab. 3 Ablation experiment of MSAM position

Method	Acc	Pre	Rec	Spe	FLOPs/G	Time/s
Stage1	92.66±0.32	88.92±0.51	85.69±1.58	97.96±0.24	120.04	0.578 7
Stage2	93.85±0.56	91.23±2.53	90.70±2.65	98.27±0.18	119.98	0.609 0
Stage3	92.45±0.41	89.29±3.30	85.77±1.37	97.87±0.15	119.96	0.671 9
Stage4	92.70±0.31	88.68±1.49	86.05±1.19	98.07±0.39	119.94	0.810 2

3.5.2 TIFOM 模块的有效性实验

在 MTIFNet 中, TIFOM 模块帮助模型有效地融合皮肤癌图像的不同类型的特征, 提高分类性能。表 4 中, 可以看出在网络中单独使用 Triplet Attention 与 Simam Attention 分类效果均提升了 0.44%, 0.45%。相比于 Triplet Attention 使网络实现跨维度交互能力和利用 Simam Attention 生成 3-D 权重找出每个神经元的重要性, 本

文提出的三维交互特征优化模块同时具备他们的优点, 且 FLOPs 并未显著增加, 对比 Triplet Attention, Simam Attention 的单次前向传播时间还略微减少。在网络中利用三重支路结构并在每条支路中创造优化能量函数来找出每个神经元的重要性, 结果证明 Accuracy 较初始模型提升了 0.95%, 模块的性能较未改变的注意力机制得到了提升。

表 4 TIFOM 的有效性验证

Tab. 4 Validation of TIFOM

Method			Acc	Pre	Rec	Spe	FLOPs/G	Time/s
Base	Triplet Attention	SimAM Attention						
✓			92.56±0.35	89.02±0.96	86.33±1.92	98.00±0.44	107.04	0.575 5
✓	✓		93.00±0.51	89.18±1.70	87.03±0.96	98.10±0.11	107.30	0.807 9
✓		✓	93.01±0.55	89.35±1.18	87.15±1.75	97.97±0.39	107.04	0.876 7
✓	✓	✓	93.51±0.52	89.88±1.33	89.57±2.67	98.20±0.21	107.30	0.811 9

为了更直观地了解三维交互特征优化模块与其他注意力在不同类别上的表现情况,使用混淆矩阵对他们的分类性能进行全面的评估。如

图 6 所示,混淆矩阵可以清晰地识别出分类器在不同类别上所犯的错误类型,可以看出皮肤癌分类难度在于如何区分黑色素瘤与黑色素细胞痣。

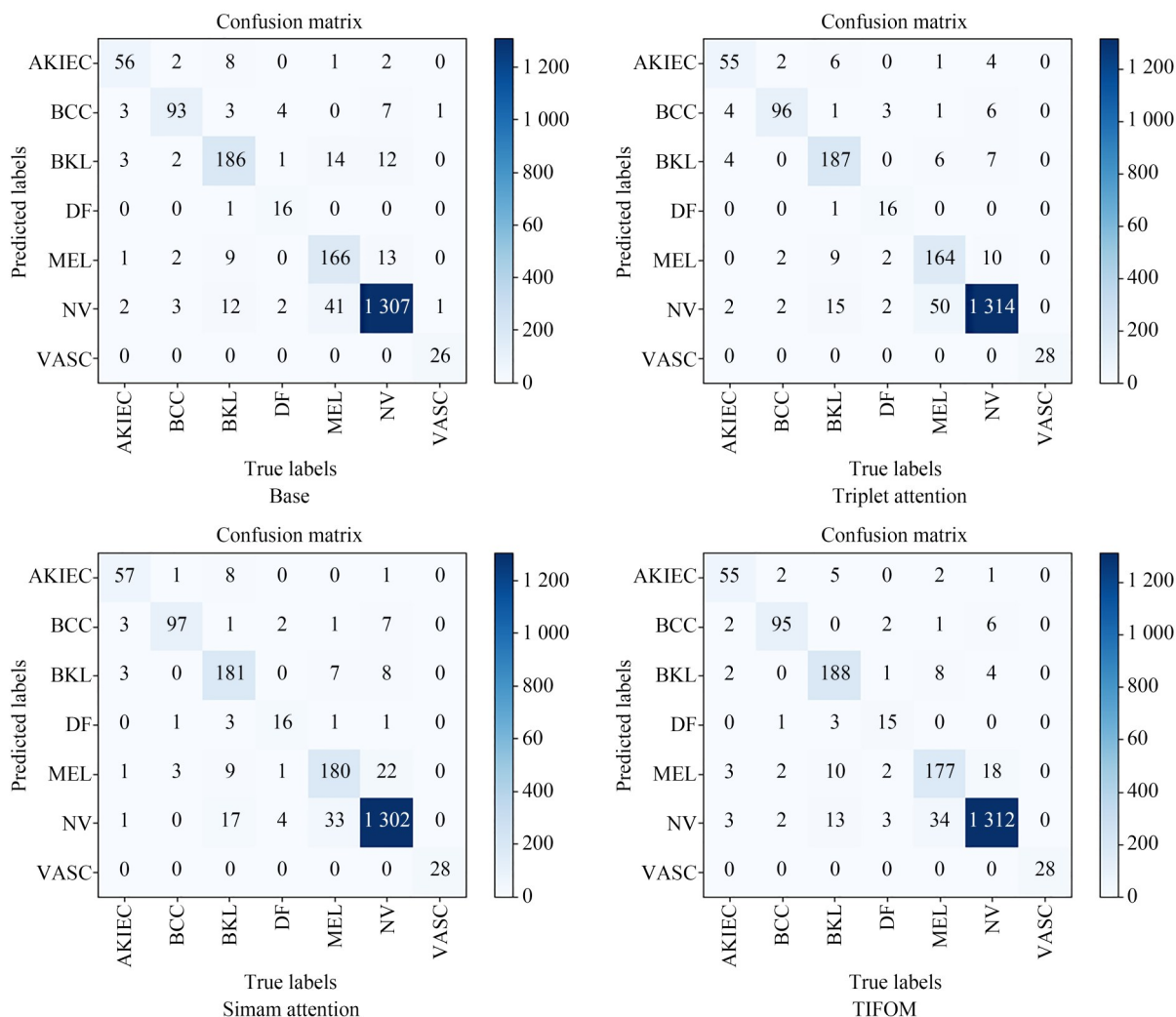


图 6 模块改进前后的混淆矩阵

Fig. 6 Confusion matrix before and after module improvements

3.5.3 整体网络的消融实验

设置五组实验,首先,分别在基础网络中加入 SAFRM 模块与 MSAM 模块,在对比的同时证明 DWCCM 模块补充了 SAFRM 模块无法捕捉图像局部特征上下文信息的短板。其次,在基础网络中加入 TIFOM 模块与同时使用 MSAM 与 TIFOM 模块的 MTIFNet 网络完成整个消融实验,来分析改进部分对 MTIFNet 网络效果的影响。从表 5 可以看出,补充了 DWCCM 的 MSAM 模块的 Accuracy, Precision, Recall, Speci-

ficity 分别提升了 0.91%, 1.53%, 1.0%, 0.32%, 在基础网络中应用 MSAM 模块、TIFOM 模块后, Accuracy, Precision, Recall, Specificity 分别提升了 1.29%, 2.21%, 4.37%, 0.27% 与 0.95%, 0.86%, 3.24%, 0.20%, 可以看出这两个模块对网络的分类性能均有效果。相较最初的基线网络,本文提出的 MTIFNet 的 FLOPs 增加了 3.2G, Time 增加了 0.3618 s, 计算效率并未大幅度地降低。本文的分类网络 MTIFNet 将 MSAM 模块与 TIFOM 模块结合起来, Accura-

cy, Precision, Recall, Specificity 达到最优,分别提升了 1.76%,2.59%,6.67%,0.39%。

表 5 整体网络的消融
Tab. 5 Ablation experiments of the whole network

Method				Acc	Pre	Rec	Spe	FLOPs/G	Time/s
Base	SAFRM	DWCCM	TIFOM						
✓				92.56±0.35	89.02±0.96	86.33±1.92	98.00±0.44	107.04	0.575 5
✓	✓			92.94±0.35	89.70±1.09	89.70±3.27	97.95±0.16	109.21	0.526 5
✓	✓	✓		93.85±0.56	91.23±2.53	90.70±2.65	98.27±0.18	119.98	0.609 0
✓			✓	93.51±0.52	89.88±1.33	89.57±2.67	98.20±0.21	107.30	0.811 9
✓	✓	✓	✓	94.32±0.45	91.61±1.29	93.00±0.60	98.39±0.17	120.24	0.937 3

为了直观地显示模块的有效性,本文比较基础网络和改进网络的可视化图像通过类别激活映射(Class Activation Map,CAM)和导向反向传播(Guided-Backpropagation)的可视化方法以及结合 CAM 和 Guided-Backpropagation 的方式。改进网络在 CAM 和 Guided-Backpropagation 生成的图像中,更清晰地突出了与皮肤癌分类相关的关键特征区域。由于基底细胞癌会导致患者局部

皮肤出现轻度脱屑、粉红色或肤色边缘隆起的鳞屑性斑块、持续扩散生长这些较难与正常皮肤区分的特征。所以图 7 选取基底细胞癌作为可视化图像可以有效地测试模型的表现情况。可以看出从基础网络至优化的网络中,模型关注的有效病理区域增广,MTIFNet 模型关注病变区域较其他模型生成的与分类相关的区域匹配度更高,致使在皮肤镜图像分类任务中识别准确的病变特征。

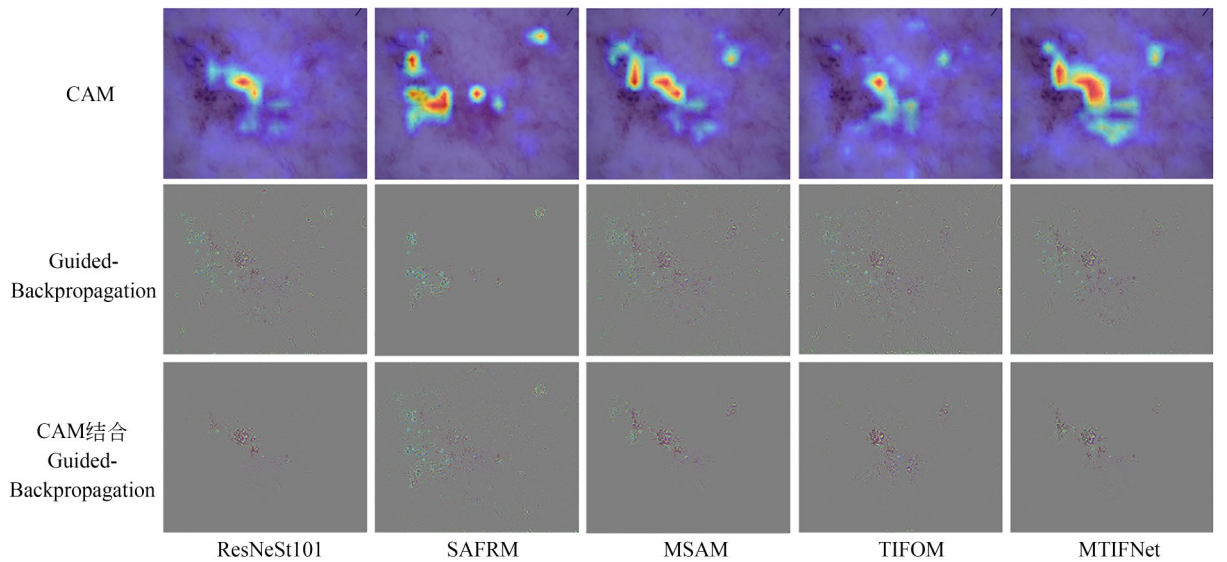


图 7 消融实验的可视化对比图
Fig. 7 Visual comparison chart of the ablation experiment

3.6 与经典模型的对比实验

将本文算法与近年的皮肤镜图像处理算法对比,结果如表 6 可知。MTIFNet 的分类指标 Accuracy, Precision, Recall, Specificity 分别达到了 94.32%,91.61%,93.00%,98.39%,与其他算法定原文的数据比较,本文提出的 MTIFNet 模型在

皮肤镜图像分类领域表现最优。对比 Le 等^[19]提出的方法,本文提出的 MTIFNet 模型 Accuracy, Precision, Recall 较其提升了 1.32%,3.61%,7%,与其提出的类加权损失函数和焦点损失来克服数据不平衡的方法比较,MTIFNet 模型不止解决了类不平衡且具有更好的信息交互能力。

表 6 ISIC 2018数据集上不同模型的分类结果

Tab.6 Classification results of different models on the ISIC 2018 dataset

Method	Acc	Pre	Rec	Spe
Le ^[19]	93.00	88.00	86.00	—
Liu ^[20]	93.14	47.70	81.84	92.30
Datta ^[21]	93.40	93.70	—	—
Cai ^[22]	93.81	88.64	90.14	98.36
Yang ^[23]	94.10	82.43	84.43	97.86
Ours	94.32	91.61	93.00	98.39

3.7 MTIFNet模型泛化实验

为验证 MTIFNet 模型在其他数据集上的效果,本文选用 ISIC 2017 数据集进行性能评估,同时选取近年的算法与本文算法比较结果。由表 7 可知,本文提出的算法在 ISIC 2017 数据集分类实验中的 Accuracy, Precision, Recall, Specificity 分别达到了 98.57%, 98.20%, 98.47%, 99.13%, 分类效果对比其他算法最佳。相比于 Khoulood 等^[26]引入了两种新的深度学习网络架构先分割再分类。在皮肤癌分割阶段中可能会错过一些微小的病变或是图像中不够明显的区域,这会影响到分类器对于特定区域的特征提取,而本文提出的算法弥补了模型对皮肤癌变区域特征过程中信息丢失的缺憾。

4 结 论

本文提出了一种针对皮肤癌图像的多尺度与三维交互特征优化网络(MTIFNet)。该方法

表 7 ISIC 2017数据集上不同模型的分类结果

Tab.7 Classification results of different models on the ISIC 2017 dataset

Method	Acc	Pre	Rec	Spe
Zhang ^[24]	86.80	—	87.80	86.70
Al-Masni ^[25]	81.57	—	75.33	80.62
Datta ^[21]	90.40	—	91.60	83.30
Kim ^[29]	89.50	91.30	49.60	99.20
Khoulood ^[26]	96.97	95.71	95.15	97.87
Ours	98.57	98.20	98.47	99.13

首先使用多尺度空间自适应模块,在训练过程中从图像中提取全局与局部的上下文信息。其次,该方法设计了一个三维交互特征优化模块,它能够同时进行跨维度的交互,并提供更有意义的特征表征能力,使得空间位置的获取更加精准。弥补了原网络只能特征交互的缺陷,实现了同时识别图像中独立的纹理、颜色和形状等不同类型的特征信息。使用交叉熵损失函数在皮肤癌分类中衡量预测概率分布与真实类别分布之间的差异,以优化模型的准确性。实验结果中,所提出的 MTIFNet 网络在 ISIC 2018 数据集中准确率、精确率、召回率和特异性分别提升了 1.76%, 2.59%, 6.67%, 0.39%, 满足了医学图像分类在临床应用中需要高精度度的要求。本模型在训练中完全依赖皮肤镜图像标记数据,但医疗领域数字资源有限,未来的研究将着重于探索如何充分利用有限的标记数据,并降低对数据标注的依赖性。

参考文献:

[1] ARORA G, DUBEY A K, JAFFERY Z A, *et al.* A comparative study of fourteen deep learning networks for multi skin lesion classification (MSLC) on unbalanced data[J]. *Neural Computing and Applications*, 2023, 35(11): 7989-8015.

[2] 兰雪梅,姜祎群. 人工智能在黑素瘤诊疗中的应用[J]. *中国皮肤性病学杂志*, 2023, 37(8): 855-860.

LAN X M, JIANG Y Q. Application of artificial intelligence in the diagnosis and treatment of melanoma [J]. *The Chinese Journal of Dermatovenereology*, 2023, 37(8): 855-860. (in Chinese)

[3] ARGENZIANO G, SOYER H P. Dermoscopy of pigmented skin lesions: a valuable tool for early diagnosis of melanoma [J]. *The Lancet Oncology*, 2001, 2(7): 443-449.

[4] 李大湘,杨福杰,刘颖,等. 融入交叉注意力编码的皮肤病变分割网络[J]. *光学精密工程*, 2024, 32(4): 609-621.

LI D X, YANG F J, LIU Y, *et al.* Skin lesion segmentation network with cross-attention coding [J]. *Opt. Precision Eng.*, 2024, 32(4): 609-621. (in Chinese)

[5] 宋建丽,吕晓琪,谷宇. 语义流引导采样结合注意力机制的脑肿瘤图像分割[J]. *光学精密工程*,

- 2024, 32(4): 565-577.
- SONG J L, LÜ X Q, GU Y. Brain tumor image segmentation based on semantic flow guided sampling and attention mechanism [J]. *Opt. Precision Eng.*, 2024, 32(4): 565-577. (in Chinese)
- [6] GU Y Y, GE Z Y, BONNINGTON C P, *et al.* Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification [J]. *IEEE Journal of Biomedical and Health Informatics*, 2020, 24(5): 1379-1393.
- [7] GODLIN JASIL S P, ULAGAMUTHALVI V. Skin lesion classification using pre-trained densenet201 deep neural network [C]. 2021 *3rd International Conference on Signal Processing and Communication (ICPSC)*. Coimbatore, India. IEEE, 2021: 393-396.
- [8] SINGH S K, ABOLGHAEMI V, ANISI M H. Skin cancer diagnosis based on neutrosophic features with a deep neural network [J]. *Sensors*, 2022, 22(16): 6261.
- [9] 李建威, 吕晓琪, 谷宇. 基于改进 ConvNeXt 的皮肤镜图像分类方法 [J]. *计算机工程*, 2023, 49(10): 239-246, 254.
- LI J W, LÜ X Q, GU Y. Dermoscopy image classification method based on improved ConvNeXt [J]. *Computer Engineering*, 2023, 49(10): 239-246, 254. (in Chinese)
- [10] ZHUANG J X, CAI J B, ZHANG J G, *et al.* Class attention to regions of lesion for imbalanced medical image recognition [J]. *Neurocomputing*, 2023, 555: 126577.
- [11] ABDELHALIM I S A, MOHAMED M F, MAHDY Y B. Data augmentation for skin lesion using self-attention based progressive generative adversarial network [J]. *Expert Systems with Applications*, 2021, 165: 113922.
- [12] ZHANG H, WU C R, ZHANG Z Y, *et al.* ResNeSt: split-attention networks [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. New Orleans, LA, USA. IEEE, 2022: 2735-2745.
- [13] SUN L, DONG J X, TANG J H, *et al.* Spatially-adaptive feature modulation for efficient image super-resolution [C]. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France. IEEE, 2023: 13144-13153.
- [14] CHEN J R, KAO S H, HE H, *et al.* Run, Don't walk: chasing higher flops for faster neural networks [C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada. IEEE, 2023: 12021-12031.
- [15] MISRA D, NALAMADA T, ARASANIPALAI A U, *et al.* Rotate to attend: convolutional triplet attention module [C]. 2021 *IEEE Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, HI, USA. IEEE, 2021: 3138-3147.
- [116] GUO N, JIANG M Y, GAO L J, *et al.* Simam: a simple, parameter-free attention module for convolutional neural networks [C]. *International Conference on Machine Learning*. PMLR, 2021: 11863-11874.
- [17] ZHANG Y F, FENG W, WU Z Y, *et al.* Deep-learning model of ResNet combined with CBAM for malignant-benign pulmonary nodules classification on computed tomography images [J]. *Medicina*, 2023, 59(6): 1088.
- [18] CODELLA N, ROTEMBERG V, TSCHANDL P, *et al.* Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by The International Skin Imaging Collaboration (ISIC) [EB/OL]. 2019: *arXiv*: 1902.03368. <http://arxiv.org/abs/1902.03368>
- [19] LE D N T, LE H X, NGO L T, *et al.* Transfer Learning with Class-Weighted and Focal Loss Function for Automatic Skin Cancer Classification [EB/OL]. 2020: *arXiv*: 2009.05977. <http://arxiv.org/abs/2009.05977>
- [20] LIU Q D, YU L Q, LUO L Y, *et al.* Semi-supervised medical image classification with relation-driven self-ensembling model [J]. *IEEE Transactions on Medical Imaging*, 2020, 39(11): 3429-3440.
- [21] DATTA S K, SHAIKH M A, SRIHARI S N, *et al.* Soft attention improves skin cancer classification performance [C]. 4th International Workshop, IMIMIC 2021, and 1st International Workshop, TDA4MedicalData 2021 France, September 27, 2021: 13-23.
- [22] CAI G, ZHU Y, WU Y, *et al.* A multimodal transformer to fuse images and metadata for skin disease classification [J]. *The Visual Computer*,

- 2023, 39(7): 2781-2793.
- [23] YANG G, LUO S H, GREER P. A novel vision transformer model for skin cancer classification[J]. *Neural Processing Letters*, 2023, 55(7): 9335-9351.
- [24] ZHANG J P, XIE Y T, XIA Y, *et al.* Attention residual learning for skin lesion classification [J]. *IEEE Transactions on Medical Imaging*, 2019, 38(9): 2092-2103.
- [25] AL-MASNI M A, KIM D H, KIM T S. Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification [J]. *Computer Methods and Programs in Biomedicine*, 2020, 190: 105351.
- [26] KHOULOUD S, AHLEM M, FADEL T, *et al.* W-net and inception residual network for skin lesion segmentation and classification [J]. *Applied Intelligence*, 2022, 52(4): 3976-3994.
- [27] PAN Y S, XIA Y. Residual Network based Aggregation Model for Skin Lesion Classification[EB/OL]. 2018: *arXiv*: 1807.09150. <http://arxiv.org/abs/1807.09150>
- [28] YAN Y Q, KAWAHARA J, HAMARNEH G. Melanoma recognition via visual attention[C]. *International Conference on Information Processing in Medical Imaging*. Cham: Springer, 2019: 793-804.
- [29] KIM S, GAIBOR E, HAEHN D. Web-based Melanoma Detection [EB/OL]. 2024: *arXiv*: 2403.14898. <http://arxiv.org/abs/2403.14898>

作者简介:



王 迪(2002—),女,黑龙江绥化人,硕士研究生,2022年于牡丹江师范学院获得学士学位,主要从事智能图像处理、深度学习及图像分类方面的研究。E-mail: wangdi020506@163.com

通讯作者:



吕晓琪(1963—),男,内蒙古包头人,博士,教授,博士生导师,1984年于内蒙古大学获得学士学位,1989年于西安交通大学获得硕士学位,2003年于北京科技大学获得博士学位,主要从事智能信息处理、医学图像处理、数字化医疗相关技术等方面的研究。E-mail: lxiaoqi@imut.edu.cn