

融合多维依赖建模与空间自适应特征学习的 皮肤病变分类方法

冯浩, 魏立臻, 张依林, 刘甲辉, 张丽艳*

大连交通大学轨道智能工程学院/电子与通信工程系, 辽宁 大连

收稿日期: 2026年6月15日; 录用日期: 2026年7月7日; 发布日期: 2026年7月20日

摘要

皮肤癌作为全球致死率极高的恶性肿瘤之一, 其早期精准分类对于临床辅助诊断与治疗至关重要。针对现有卷积神经网络在处理形态多变、大小不一的皮肤病变图像时, 存在多尺度特征捕获不足以及对辨别细粒度病灶聚焦不充分等挑战, 本文提出了一种融合分层上下文聚合与多核特征协同学习的改进型深度学习分类框架。首先, 针对病灶区域多尺度的特性, 设计了分层上下文聚合模块(HCAM), 通过空间金字塔聚合与深度细化机制, 动态提取并融合跨尺度的全局上下文特征。其次, 构建了多核特征融合模块(MKFFM), 利用多路并行且感受野递增(1×1 至 7×7)的深度可分离卷积, 自适应地捕捉不同尺度的局部纹理特征, 突破了单一卷积核信息提取的局限性。此外, 本文引入了基于空间注意力引导的全局-局部双路特征聚合机制, 并在训练阶段结合CutMix混合数据增强策略, 有效抑制了复杂皮肤背景的干扰, 显著增强了模型在数据不平衡及小样本场景下的泛化能力与鲁棒性。基于ISIC 2018等公开数据集的五折交叉验证实验结果表明, 该模型显著提升了各类皮肤病变的分类精确度与F1值, 为计算机辅助皮肤病诊断提供了一种兼具高鲁棒性与解释性的技术方案。

关键词

皮肤病变分类, 分层上下文聚合(HCAM), 多核特征融合(MKFFM), 空间注意力机制, CutMix数据增强

Skin Lesion Classification via Integrated Multi-Dimensional Dependency Modeling and Spatial Adaptive Feature Learning

Hao Feng, Lizhen Wei, Yilin Zhang, Jiahui Liu, Liyan Zhang*

Department of Electronic Communications/School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian Liaoning

*通讯作者。

文章引用: 冯浩, 魏立臻, 张依林, 刘甲辉, 张丽艳. 融合多维依赖建模与空间自适应特征学习的皮肤病变分类方法[J]. 图像与信号处理, 2026, 15(3): 363-374. DOI: 10.12677/jisp.2026.153032

Abstract

As one of the most fatal malignant tumors globally, the early and accurate classification of skin cancer is crucial for clinical auxiliary diagnosis and treatment. To address the challenges of insufficient multi-scale feature capture and inadequate focus on discriminative fine-grained lesions in existing convolutional neural networks when processing skin lesion images with variable morphologies and sizes, this paper proposes an improved deep learning classification framework integrating hierarchical context aggregation and multi-kernel feature collaborative learning. First, targeting the multi-scale characteristics of lesion regions, a Hierarchical Context Aggregation Module (HCAM) is designed to dynamically extract and fuse cross-scale global context features through spatial pyramid aggregation and depthwise refinement mechanisms. Second, a Multi-Kernel Feature Fusion Module (MKFFM) is constructed, which utilizes multi-branch parallel depthwise separable convolutions with increasing receptive fields (from 1×1 to 7×7) to adaptively capture local texture features at different scales, breaking through the limitations of single convolution kernel information extraction. Furthermore, this paper introduces a global-local dual-branch feature aggregation mechanism guided by spatial attention, and combined with the CutMix mixed data augmentation strategy during the training phase, it effectively suppresses the interference of complex skin backgrounds and significantly enhances the generalization ability and robustness of the model under data imbalance and small-sample scenarios. The five-fold cross-validation experimental results based on public datasets such as ISIC 2018 demonstrate that the proposed model significantly improves the classification accuracy and F1-score of various skin lesions, providing a highly robust and interpretable technical solution for computer-aided dermatology diagnosis.

Keywords

Skin Lesion Classification, Hierarchical Context Aggregation Module (HCAM), Multi-Kernel Feature Fusion Module (MKFFM), Spatial Attention Mechanism, CutMix Data Augmentation

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,皮肤恶性肿瘤的发病率呈全球性上升趋势,已成为严重威胁人类生命健康的公共卫生问题之一[1]。据统计,美国每年因皮肤癌导致的死亡人数超过 10,000 人。临床研究表明,皮肤癌的致死率与其确诊阶段密切相关,若能在早期发现并及时进行病灶切除,患者的五年存活率可高达 95% [2]。在常规临床检查中,活检方法由于存在创伤性,通常作为最终确诊手段;而皮肤镜[3] (Dermoscopy)技术作为一种无创光学放大成像技术,能够消除皮肤表面反射光,使医生清晰观察到表皮深处的微观结构与色素分布,已成为早期筛查皮肤癌的重要手段。然而,由于不同类型的皮肤病变在形态、纹理、色泽上存在极高的视觉相似性(类间差异小),且同类病变在不同发病阶段表现各异(类内差异大),传统依赖皮肤科医生主观视觉经验的诊断方式不仅耗时费力,且诊断准确率易受医生主观经验差异的限制。

随着人工智能的飞速发展,深度学习技术在计算机视觉尤其是医学图像分析领域取得了突破性进展[4]。深度卷积神经网络(CNN)凭借其强大的非线性特征表征能力,已被广泛应用于皮肤镜图像的自动分

类任务中,取得了阶段性的显著成果[5]。在大规模图像识别中,迁移学习(Transfer Learning)通过微调预训练模型参数,有效缓解了医学图像数据稀缺带来的训练难题[6]。例如,Gu 等[7]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;Celebi 等人提出了基于图像处理的分类方法[8],他们利用欧几里得距离变换将皮肤病图像划分为不同区域,然后使用支持向量机[9]分类。该方法在小样本条件下取得了一定效果,但依赖于专家知识,特征提取过程繁琐;Ballerni 等[10]则结合 KNN 分类器对病变的色泽、质地特征进行量化处理,在 5 类常见皮肤病的识别中取得了超过 74%的精度。

尽管现有的深度学习方法在皮肤癌诊断中取得了积极进展,但仍面临以下严峻的挑战:首先是病变尺度多变性。皮肤病灶在图像中所占比例差异巨大,传统 CNN 采用单一固定尺寸的卷积核,难以同时兼顾大尺度病变的全局结构与小尺度病变的局部细粒度特征;其次是复杂背景与边界模糊性。皮肤镜图像中常包含毛发、气泡、血管等大量背景噪声,且病灶边缘往往呈现渐变或不规则状态,导致模型容易提取无效的背景干扰特征;最后是临床数据的不平衡性。现实中采集的恶性肿瘤(如黑色素瘤)样本远少于良性病变(如色素痣),小样本数据极易导致复杂模型陷入过拟合。

针对上述挑战,本文提出了一种名为分层上下文与多核特征融合的网络(Hierarchical Context and Multi-kernel Feature Fusion Networks, HCMFNet)。该方法以 EfficientNetV2-S 为骨干网络,通过引入多尺度特征挖掘与注意力机制,显著提升了模型对复杂皮肤病损图像的辨别与泛化能力,将皮肤镜图像精准划分为 AKIEC、BCC、BKL、DF、MEL、NV、VASC 七大类。

2. 相关工作

2.1. 网络模型

以 EfficientNetV2-S 模型为基础,本文提出了 HCMFNet 模型来实现皮肤镜的分类任务,其网络结构如图 1 所示。该网络主要由特征提取骨干网络、分层上下文聚合模块(HCAM)、多核特征融合模块(MKFFM)、基于空间注意力的全局-局部双分支聚合模块以及分类器(Classifier)五个核心部分组成。本文提出的分层上下文与多核特征融合网络(HCMFNet)首先采用预训练的 EfficientNetV2-S 提取深层语义特征,随后通过分层上下文聚合模块(HCAM)与多核特征融合模块(MKFFM)分别实现跨尺度全局上下文获取与多尺度局部纹理精炼。在特征融合阶段,网络构建了全局-局部双分支架构,利用空间注意力掩码抑制背景噪声并强化辨别性病灶特征的表达,最后将融合后的多维特征送入包含全连接层、批量归一化及 Dropout 结构的分类器,实现对 AKIEC、BCC 等七类皮肤病变的精准分类。

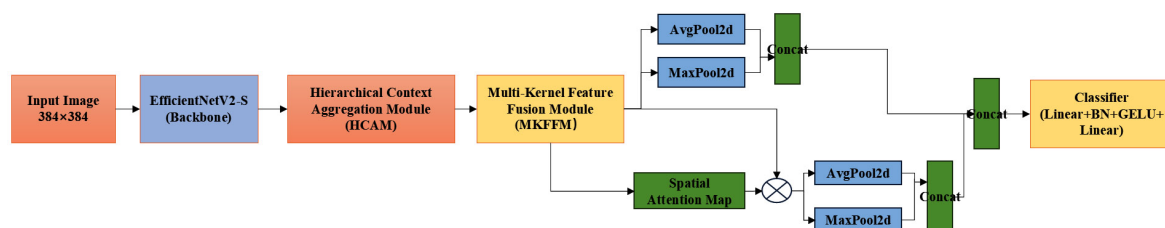


Figure 1. Overall architecture diagram of HCMFNet

图 1. HCMFNet 整体结构图

2.2. 分层上下文聚合模块

皮肤镜图像中的病灶组织特征极其复杂,且病变区域与周围健康组织的边界往往较为模糊,这使得传统的特征提取过程容易造成关键信息的丢失。为了更准确地确定皮肤病变范围并识别多尺度的病理特

征, 基于对 Sun 等[11]提出的金字塔结构模型的思考以及借鉴 Chen [12]的思想, 提出了分层上下文聚合模块(Hierarchical Context Aggregation Module, HCAM)。如图 2 所示, HCAM 主要由两部分串联构成: 负责提取全局多尺度信息的空间金字塔聚合器(Spatial Pyramid Aggregator, SPA)和负责捕获局部像素级上下文联系的深度级联细化器(Depthwise Refinement, DR)。相比于依赖空洞卷积易引发“网格效应”且缺乏特征后置精炼的经典 ASPP 模块, 本研究所设计的分层上下文聚合模块(HCAM)展现出更严谨的拓扑优势。其空间金字塔聚合器(SPA)彻底舍弃空洞卷积, 先将特征图沿通道均匀切分为四个独立分支(每路通道数为 $C/4$), 除首个分支外, 其余分支利用自适应最大池化将空间分辨率动态压缩至原图的 $1/2$ 、 $1/4$ 和 $1/8$, 并在降采样空间内前置应用 3×3 深度卷积与批量归一化(BN)层, 随后通过最近邻插值恢复分辨率、拼接并级联 1×1 聚合卷积与 GELU 激活, 在消除网格效应的同时实现由局部到全局的非线性上下文关联; 在此基础上, HCAM 在 SPA 之后无缝级联了深度级联细化器(DR), 利用由 3×3 深度卷积与双层 1×1 逐点卷积构成的瓶颈残差网络, 对聚合后的多尺度特征进行像素级近邻依赖的主动精炼, 确保了宏观全局病灶形态与微观边缘细节在特征流中的高效复现。

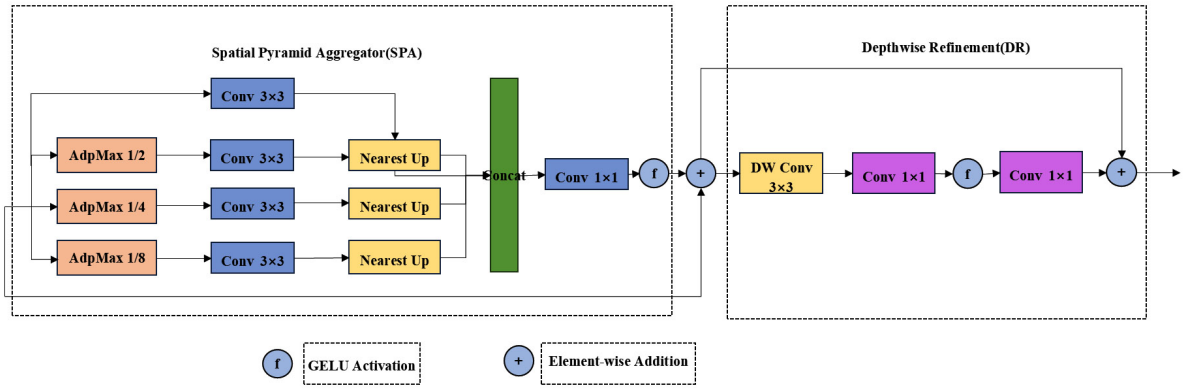


Figure 2. Hierarchical context aggregation module (HCAM)

图 2. 分层上下文聚合模块

2.2.1. 空间金字塔聚合器(Spatial Pyramid Aggregator, SPA)

病变在皮肤表面的分布具有多尺度特性。为了从不同感受野获取图像的全局上下文信息并减少计算冗余, SPA 模块采用了通道分组与空间金字塔池化相结合的策略。

首先, 对输入的特征图 X_0, X_1, X_2, X_3 沿通道维度均匀分割为四个独立的分支, 每个分支的通道数为 $C/4$ 。对于第一个分支 X_0 , 直接应用 3×3 的深度可分离卷积(Depthwise Convolution)以保留原始分辨率下的空间特征。对于其余三个分支 $X_i (i \in \{1, 2, 3\})$, 为了促使模型学习长程相关性并过滤局部噪声, 模块引入了不同缩放因子的自适应最大池化(Adaptive Max Pooling), 将特征图分辨率分别下采样至原图的 $1/2$ 、 $1/4$ 和 $1/8$ 。随后, 对下采样后的特征进行 3×3 深度可分离卷积, 并利用最近邻插值(Nearest Upsampling)将特征图恢复至原始分辨率大小。该过程可由下式表示:

$$X'_0 = \text{DWConv}_{3 \times 3}(X_0) \quad (1)$$

$$X'_i = \text{Up}\left(\text{DWConv}_{3 \times 3}\left(\text{AdpMaxPool}_{1/2^i}(X_i)\right)\right), \quad i \in \{1, 2, 3\} \quad (2)$$

其中, $\text{DWConv}_{3 \times 3}(\cdot)$ 表示 3×3 深度可分离卷积, $\text{AdpMaxPool}(\cdot)$ 表示自适应最大池化, $\text{Up}(\cdot)$ 表示最近邻上采样操作。在获取多尺度特征表示后, 将四个分支的输出在通道维度上进行拼接(Concatenation), 并通过 1×1 卷积实现跨通道的局部与全局关系聚合。最后, 经过 GELU 激活函数处理, 并通过残差连接

(Residual Connection)与原始输入 X 相加, 得到 SPA 模块的输出:

$$Y_{SPA} = X + \delta\left(\text{Conv}_{1 \times 1}\left(\text{Concat}\left([X'_0, X'_1, X'_2, X'_3]\right)\right)\right) \quad (3)$$

其中, $\text{Concat}(\cdot)$ 表示拼接操作, $\delta(\cdot)$ 表示 GELU 激活函数。

2.2.2. 深度级联细化器 (Depthwise Refinement, DR)

虽然 SPA 模块有效地提取了全局多尺度特征, 但皮肤镜图像中病灶内部及边缘存在大量相似且渐变的空间结构。为了进一步增强局部空间的建模能力, 紧密建立病灶周围相邻像素间的局部上下文联系, 本文在 SPA 之后引入了深度级联细化器(DR)。

DR 模块主要通过局部感知和权重共享机制来捕获图像细节。如图二右侧所示, 该模块首先接收 Y_{SPA} , 并依次经过一个 3×3 的深度可分离卷积和两个连续的 1×1 逐点卷积(Pointwise Convolution)。深度可分离卷积能够在保持空间特征相关性的同时大幅降低计算量, 而夹在两次 1×1 卷积之间的 GELU 激活函数则增强了特征变换的非线性表达能力, 进而探测不同的局部上下文信息。最后, 通过残差连接将细化后的特征与输入融合:

$$Y_{out} = Y_{SPA} + \text{Conv}_{1 \times 1}\left(\text{PWConv}_{1 \times 1}\left(\text{DWConv}_{3 \times 3}\left(Y_{SPA}\right)\right)\right) \quad (4)$$

通过 SPA 与 DR 的级联结构, HCAM 模块不仅能够从宏观上把握病灶的多尺度全局形态, 还能在微观上细化像素间的纹理边界, 从而为后续的皮肤病变精准分类提供了更具判别性的上下文特征。

2.3. 多核特征融合模块

皮肤癌病灶在图像中往往呈现出极大的尺度差异, 且边缘形态不规则。传统的单一尺寸卷积核难以同时捕获病灶的局部微观细节(如色素网络、微小毛细血管)和全局宏观结构(如整体不对称性、大面积色斑)。GoogLeNet [13]通过引入 Inception 模块实现多尺度特征提取, 将不同尺寸的卷积操作并联融合, 有效提升了网络对局部与全局信息的兼容能力, 并极大地降低了参数数量。因此本文设计了多核特征融合模块(Multi-Kernel Feature Fusion Module, MKFFM)。该模块通过并行引入多种感受野的深度可分离卷积, 在有效控制模型参数数量的前提下, 实现了多尺度空间特征的自适应提取与融合。MKFFM 的具体结构如图 3 所示。传统的 Inception 模块虽然通过并行多分支实现了多尺度特征提取, 但其采用标准卷积核, 在网络深层会导致参数数量和计算量呈爆炸式增长, 且全通道混合计算容易造成各通道间辨别性特征的混淆(Inter-channel Interference)。与之相比, 本研究设计的 MKFFM 模块在两方面进行了核心改进: 首先是通道解耦与轻量化自适应, 模块前置了通道均匀切分(C/4)策略, 将特征空间解耦到独立的四个并行分支中, 全部分支均采用深度可分离卷积(Depthwise Separable Convolution)。这使得尺寸达 5×5 和 7×7 的超大感受野分支能够以极低的计算代价捕获大面积大尺度病损(如 NV、MEL)的宏观轮廓特征, 突破了单一卷积核提取信息的局限。随后, 在各分支空间滤波后紧跟逐点卷积(PWConv)与 GELU 激活函数, 强制进行跨通道的特征线性平滑组合与非线性交互增强。这种设计在实现多尺度多维依赖高效建模的同时, 相比经典 Inception 能够显著抑制皮肤镜中毛发、气泡等非病灶背景噪声的横向传播, 大幅提升了模型在复杂干扰下对细粒度病理特征的鉴别能力。

对于给定的输入特征图 $X \in R^{C \times H \times W}$, 为了降低计算复杂度并促使网络学习多样化的特征表达, 首先将输入特征沿通道维度均匀切分为 4 个子特征图: X_1, X_2, X_3, X_4 , 每个子特征图的通道数为 $C/4$ 。

随后, 将这 4 个子特征图分别送入四个并行的卷积分支中。为了提取不同空间尺度的上下文信息, 四个分支分别采用了尺寸为 1×1 、 3×3 、 5×5 和 7×7 的深度卷积(Depthwise Convolution, DWConv)。

深度卷积在各个通道上独立进行空间滤波, 不仅大幅减少了参数量, 还有效避免了不同通道间信息的过度干扰。在深度卷积之后, 每个分支均接入一个 1×1 的逐点卷积(Pointwise Convolution, PWConv)以实现跨通道的信息交互, 并依次通过批量归一化(Batch Normalization, BN)和 GELU 激活函数以增强非线性表达能力。四个分支的特征变换过程分别由式(5)~(8)所示。

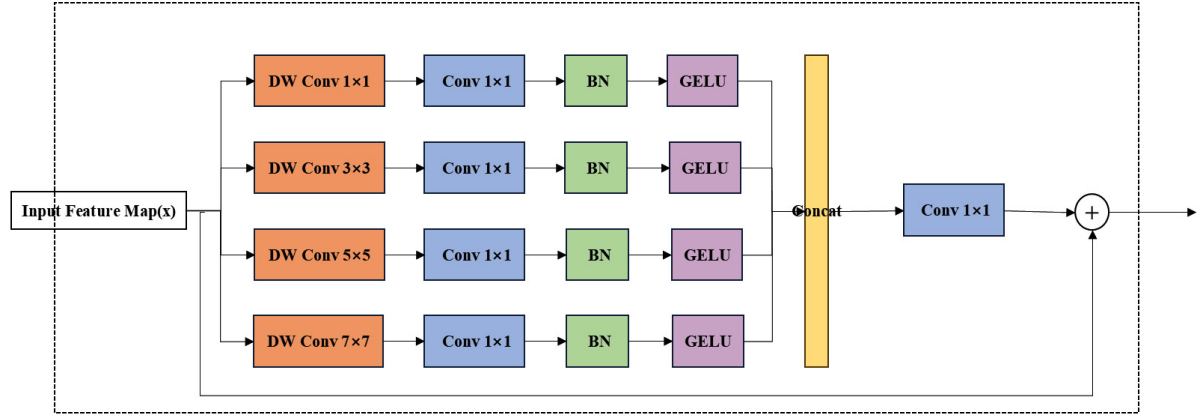


Figure 3. Multi-kernel feature fusion module (MKFFM)

图 3. 多核特征融合模块

$$Y_1 = \delta \left(\beta \left(\text{PWConv}_{1 \times 1} \left(\text{DWConv}_{1 \times 1} (X_1) \right) \right) \right) \quad (5)$$

$$Y_2 = \delta \left(\beta \left(\text{PWConv}_{1 \times 1} \left(\text{DWConv}_{3 \times 3} (X_2) \right) \right) \right) \quad (6)$$

$$Y_3 = \delta \left(\beta \left(\text{PWConv}_{1 \times 1} \left(\text{DWConv}_{5 \times 5} (X_3) \right) \right) \right) \quad (7)$$

$$Y_4 = \delta \left(\beta \left(\text{PWConv}_{1 \times 1} \left(\text{DWConv}_{7 \times 7} (X_4) \right) \right) \right) \quad (8)$$

其中, $\text{DWConv}_{k \times k}(\cdot)$ 表示卷积核大小为 $k \times k$ 的深度卷积, $\beta(\cdot)$ 表示批量归一化操作, $\delta(\cdot)$ 表示 GELU 激活函数。需要注意的是, 分支 1 中的 1×1 深度卷积主要用于保持通道特征的独立映射, 而分支 2、3、4 则分别负责捕获局部、中等距离以及全局的结构信息。

在并行多尺度特征提取完成后, 将四个分支的输出结果在通道维度上进行拼接(Concatenation), 恢复至原有的 C 个通道。为了将这些来自不同感受野的特征进行深度融合, 模块采用了一个 1×1 的聚合卷积 (Feature Aggregation) 对拼接后的特征进行通道维度的线性组合。最后, 为了防止网络加深带来的梯度消失问题并保留原始特征信息, 引入了残差连接, 将聚合后的特征与原始输入 X 相加, 得到该模块的最终输出:

$$Y_{out} = X + \text{Conv}_{1 \times 1} \left(\text{Concat}([Y_1, Y_2, Y_3, Y_4]) \right) \quad (9)$$

综上所述, MKFFM 模块利用通道拆分与多核并行策略, 使得模型能够同时兼顾皮肤镜图像中的高频细节与低频轮廓, 极大提升了网络对复杂且多变皮肤病灶的特征表征能力。

2.4. 损失函数

在皮肤镜图像的临床采集中, 各类皮肤病变的发病率差异巨大(例如色素痣样本极多, 而皮肤纤维瘤

或血管病变样本极其稀少)。这种长尾分布(Long-tailed Distribution)和严重的类别不平衡问题,使得传统模型在训练时容易被多数类主导,从而对少数恶性肿瘤类别的敏感度大幅下降。因此,本文从数据供给侧与损失正则化两个维度进行了联合优化,提出了一种基于加权随机采样(Weighted Random Sampling)与 CutMix [14]混合损失协同的优化策略。

2.4.1. 基于类别频率的加权随机采样

为了从根本上消除类别极度不平衡对损失函数梯度的负面影响,本文在数据加载阶段前置了均衡策略。通过统计训练集中各类别的样本总数,为每个类别的样本分配动态采样权重,其计算公式为:

$$W_i = \frac{1}{\max(N_i, 1)} \quad (10)$$

其中, W_i 表示第 i 类皮肤病变样本的动态采样权重; N_i 代表训练集中该类别的样本总数。

随后,利用该权重对训练集进行有放回的加权随机采样。这一策略强制网络在每一个训练批次(Mini-batch)中都能稳定、充分地观察到少数类样本。由于输入批次的类别分布已趋于均衡,模型在优化时可直接采用标准的多分类交叉熵损失(Cross-Entropy Loss),无需额外引入复杂的损失权重项。标准交叉熵损失定义如下:

$$L_{CE}(\hat{y}, y) = -\sum_{c=1}^C y_c \log(\hat{y}_c) \quad (11)$$

其中, C 为类别总数, y_c 为真实标签的独热编码指示值, $\hat{y}_c$ 为网络输出的预测概率。

2.4.2. 基于 CutMix 的正则化损失构建

医学图像的数据量往往具有局限性,使得深层网络容易陷入过拟合。为了进一步增强模型的泛化能力,并迫使网络从更广阔的空间区域学习具有判别性的病灶特征,本文在训练过程中引入了基于 CutMix 的数据增强与损失函数重构机制。在损失计算阶段,由于输入图像包含了两个不同类别的局部特征,模型的最终损失不能再由单一的交叉熵决定,而是根据拼接面积比例 λ 构建一平滑的组合损失。本文最终优化的重构损失函数 L_{total} 定义为:

$$L_{total} = \lambda L_{CE}(\hat{y}, y_a) + (1 - \lambda) L_{CE}(\hat{y}, y_b) \quad (12)$$

$\hat{y}$ 为模型对拼接后图像的预测输出。通过这种正则化损失策略,不仅有效缓解了模型对图像背景或局部单一特征的过度依赖,还通过生成连续的软标签(Soft Labels)提升了网络决策边界的平滑度与鲁棒性。

2.4.3. 采样机制与算法层损失的优劣对比及增强协同机理讨论

在面对皮肤镜图像严重的长尾分布时,本研究在数据加载阶段前置选择了数据层面的加权随机采样(Weighted Random Sampling)机制,而非算法层面的加权损失函数(如 Class-Weighted CE)。这一选择主要在于算法层加权通过人为放大少数类权重项调节梯度,在训练初期极易引发梯度流剧烈波动并使模型陷入局部最优,同时也会相对稀释多数类的惩罚权重从而削弱其基本盘特征的泛化面;而加权随机采样能稳定构建类别均衡的 mini-batch 供给流,使标准交叉熵损失的梯度轨迹更加平滑自然。此外,该采样机制与训练中引入的 CutMix 数据增强存在显著的正向协同效应:若直接在原始长尾数据上应用 CutMix,由于多数类被选中的概率极高,生成的混合图像绝大多数将呈现为多数类内部的无效拼接;而在通过采样平衡 mini-batch 后, CutMix 能够大幅增加“将少数类(如 MEL、AKIEC)的判别性病灶斑块剪切并粘贴至多数类常规皮肤背景中”的概率,这不仅强制网络学会在复杂背景下精准剥离并辨识细粒度病损,还通过生成连续的软标签(Soft Labels)有效平滑了高度相似类别间的临床决策边界,实现了抑制模型过拟合与缓解长尾不平衡的双重理论增益。

3. 实验与分析

3.1. 实验环境

本实验环境为 windows 操作系统，在 AMD R9-7945HX 处理器和 NVIDIA GeForce RTX 4060 (8 GB 显存)的设备上进行训练，使用 Pytorch 框架。网络训练的批次大小(Batch Size)设置为 16，总训练轮数(Epochs)设定为 50，为确保评估结果的严谨性与客观性，本文实验严格遵守 5 折交叉验证(Stratified 5-Fold Cross-Validation)机制，划分数据集时保持了各类别的原始比例，最终模型的性能评估取 5 次测试结果的平均值。模型优化器选用具有良好收敛特性的 AdamW 算法，初始学习率设定为，权重衰减(Weight Decay)系数设置为 1×10^{-4} 。同时，采用余弦退火学习率调度策略(CosineAnnealingLR)，最小学习率下限设为 1×10^{-6} ，以避免模型在训练后期陷入局部最优。损失函数采用标准的交叉熵损失(Cross-Entropy Loss)。

3.2. 数据集介绍

本次实验采用 ISIC2018 提供的 HAM10000 数据集[15]。如表 1 所示，该数据集包含 10015 张经临床金标准(如病理活检等)严格确认的皮肤镜图像，共涵盖 7 类常见的皮肤色素性病变：黑色素细胞痣(NV)、黑色素瘤(MEL)、良性角化病(BKL)、基底细胞癌(BCC)、光化性角化病及上皮内癌(AKIEC)、血管性病变(VASC)与皮肤纤维瘤(DF)。

Table 1. Classification and quantity of lesions in the dataset

表 1. 数据集病变分类及数量

皮肤癌名称	中文名称	图片数量	百分比(%)
AKIEC	光化性角化病和上皮内癌	327	3.27
BCC	基底细胞癌	514	5.13
BKL	良性角化病	1099	10.97
DF	皮肤纤维瘤	115	1.15
MEL	黑色素瘤	1113	11.11
NV	黑色素细胞痣	6705	66.95
VASC	血管性病变	142	1.42

3.3. 评价指标

为了客观且定量地评估所提深度学习模型的分类性能，本文选用混淆矩阵(Confusion Matrix)衍生的多项标准评价指标进行综合测度。在定义评价指标前，首先引入混淆矩阵的四个基本元素：真正例(True Positive, TP)、假正例(False Positive, FP)、真负例(True Negative, TN)与假负例(False Negative, FN)。基于此，本文采用以下四个关键指标来验证模型的有效性：

准确率(Accuracy)准确率表征模型对全部样本分类的总体正确程度，即正确分类的样本数占样本总数的比例。其计算公式如式所示：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (13)$$

精确率(Precision, P)精确率旨在量化模型预测结果的“查准能力”，即在模型所有预测为正类的样本中，实际为正类样本所占的比率。其计算公式如式(3)所示：

$$\text{Precision} = \frac{TP}{TP + FP} \quad (14)$$

召回率(Recall, R)召回率旨在量化模型对正类样本的“查全能力”，即在所有真实的总体正类样本中，被模型正确预测为正类的比率。其计算公式如式(4)所示：

$$\text{Recall} = \frac{TP}{TP + FN} \quad (15)$$

F1 值(F1-score)由于精确率与召回率往往存在相互制约的关系，本文引入 F1 值作为二者的调和平均数，用于综合衡量模型的整体分类性能。特别是在面对类别不平衡的数据集时，F1 值能够更客观地反映模型的有效性与鲁棒性。其计算公式如式(5)所示：

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

3.4. 实验结果和分析

3.4.1. 对比实验

为全面验证本文所提模型 HCMFNet 的有效性，本节将其与经典卷积神经网络(ResNet 系列、ShuffleNet-V2)、主流 Transformer 架构(Swin-tiny, ViT-Base)以及近年来针对皮肤病变领域专门设计的先进(SOTA)模型在 ISIC 2018 数据集上进行了横向对比实验，具体结果如表 2 所示。在 ISIC 2018 数据集上的对比实验表明，本文提出的 HCMFNet 整体性能优势显著。其总体准确率(Accuracy)达到 88.8%，相较于通用网络 ResNet50 (82.4%)以及近年领域专用模型 Zhang (86.1%)均有明显提升；同时，其精确率(Precision)高达 90.2%，比次优的 STFL (85.1%)提升了 5.1 个百分点，断层领先所有对比模型，有力验证了 HCAM 与 MKFFM 模块在抑制背景噪声、精准锁定辨别性病灶特征方面的卓越能力；在衡量长尾分布泛化力的 F1 值(F1-Score)上，HCMFNet 达到 81.8%，与前沿顶尖模型 STFL (81.9%)处于同等高水平。然而，HCMFNet 75.6%的召回率略逊于 STFL (85.2%)和 CAGNet (78.3%)，这表明超高精确率的门槛在极端长尾约束下，一定程度上压缩了模型对边界极度模糊的少数类难例的决策激进性，未来将通过学习以进一步突破瓶颈。

Table 2. Comparative experimental results on the ISIC2018

表 2. ISIC2018 数据集对比实验结果

Model	Accuracy	F1-Score	Precision	Recall
Resnet34	81.7	67.2	69.0	66.8
Resnet50	82.4	70.1	71.3	69.6
Resnet101	81.9	68.7	68.9	69.2
Swin-tiny	80.0	61.9	62.3	62.4
ShuffleNet-V2	80.3	68.9	73.7	66.9
ViT-Base	77.8	56.9	60.9	55.0
CAGNet [16]	85.3	77.9	79.3	78.3
Zhang [17]	86.1	79.5	83.2	76.6
STFL [18]	85.5	81.9	85.1	85.2
HCMFNet	88.8	81.8	90.2	75.6

3.4.2. 消融实验

为进一步探究所提出模型中各关键网络结构及训练策略的有效性，本小节围绕层次上下文聚合模块(HCAM)、多核特征融合模块(MKFFM)、CutMix 数据增强策略以及空间注意力掩码生成器(Spatial Attention Map)进行了系统的消融实验分析。具体方法为：在基线模型(Baseline，仅保留 EfficientNetV2-S 主干提取全局特征)的基础上，采用逐步添加(Incremental)的策略依次引入各改进模块，并通过对比各项性能指标(如 Accuracy、Precision、Recall、F1-score 等)的变化，评估各模块在皮肤病变分类任务中的独立贡献与协同增益作用。相关实验结果如表 3 所示。

Table 3. Overall ablation experimental results
表 3. 整体消融实验结果

EfficientNetV2-S	HCAM	MKFFFM	Cutmix	SPA	Accuracy	F1-Score	Precision	Recall
✓					84.9	75.6	86.5	69.5
✓	✓				84.7	76.3	81.9	72.1
✓	✓	✓			84.6	76.3	84.5	70.5
✓	✓	✓	✓		85.3	77.0	85.2	71.5
✓	✓	✓	✓	✓	88.9	81.8	90.1	75.6

在消融实验中，基线模型(Accuracy84.9%, F1-Score75.6%)表现尚可但召回率偏低；逐步引入改进模块后，HCAM 与 MKFFM 的结合显著提升了模型的特征表达能力与平衡性(Precision 回升至 84.5%)，而 Cutmix 数据增强进一步稳固了模型的泛化能力(Accuracy 达 85.3%)；最终，空间注意力(SPA)模块的加入带来了最为显著的性能飞跃，使得完整模型的 Accuracy (88.9%)、F1-Score (81.8%)、Precision (90.1%)和 Recall (75.6%)均达到最优，充分证明了多模块协同策略在皮肤病变分类任务中的卓越有效性。

3.4.3. 混淆矩阵

本研究在独立测试集上评估了融合多尺度特征与局部注意力机制的改进网络，取得了 88.82%的综合准确率。如图 4 混淆矩阵结果表明，该模型在特征典型且样本充足的黑色素细胞痣(NV, 0.99)和基底细胞癌(BCC, 0.84)上表现卓越，同时也能有效识别良性角化病(BKL, 0.78)和血管病变(VASC, 0.71)，验证了分层上下文聚合(HCAM)、多核特征融合(MKFFM)以及空间注意力机制在捕捉关键病灶特征方面的有效性。然而，尽管训练中引入了加权随机采样策略来缓解数据不平衡，少数类向多数类偏移的误判趋势依然存在。总体而言，本模型整体性能优异且具备较高的临床辅助价值，但部分难例类别间的混淆也提示，未来研究需引入细粒度特征对比学习或临床元数据进行多模态融合，以进一步突破高度相似皮肤病变鉴别的瓶颈。

4. 结论

针对皮肤镜图像中病变形态复杂且类别极度不平衡的临床挑战，本研究提出了一种基于多尺度特征融合与局部注意力机制的改进型 EfficientNetV2 皮肤癌分类模型。通过创新性地引入分层上下文聚合模块(HCAM)与多核特征融合模块(MKFFM)，网络有效提取并融合了病灶的全局多尺度上下文信息；同时，空间注意力机制的加入使得模型能够动态抑制无关背景，精准聚焦并放大局部辨别性病灶特征。此外，结合严格的数据增强、CutMix 混合策略以及基于频次倒数的加权随机采样(Weighted Random Sampler)，本方法在网络端到端训练中成功缓解了长尾分布带来的“多数类主导”问题。严谨的五折交叉验证与独立测试集评估表明，该模型在保持较高计算效率的同时，显著提升了复杂皮肤病变的分类准确率与泛化能力，为临床皮肤癌的早期筛查与智能辅助诊断(CAD)提供了一种高效、可靠的新方案。

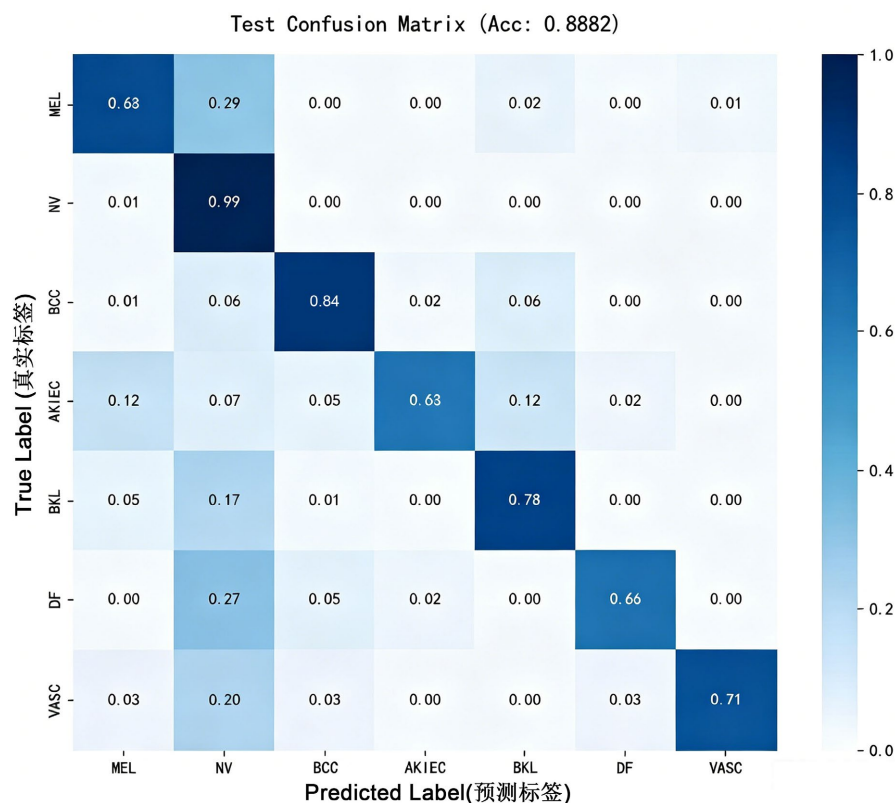


Figure 4. Confusion matrix

图 4. 混淆矩阵

参考文献

- [1] Dachani, S.R., Kaleem, M., Mujtaba, M.A., Mahajan, N., Ali, S.A., Almutairy, A.F., *et al.* (2024) A Comprehensive Review of Various Therapeutic Strategies for the Management of Skin Cancer. *ACS Omega*, **9**, 10030-10048. <https://doi.org/10.1021/acsomega.3c09780>
- [2] 兰雪梅, 姜伟群. 人工智能在黑素瘤临床中的应用[J]. 中国皮肤性病学杂志, 2023, 37(8): 855-860.
- [3] Argenziano, G. and Soyer, H.P. (2001) Dermoscopy of Pigmented Skin Lesions—A Valuable Tool for Early. *The Lancet Oncology*, **2**, 443-449. [https://doi.org/10.1016/s1470-2045\(00\)00422-8](https://doi.org/10.1016/s1470-2045(00)00422-8)
- [4] 鲁硕, 张魁星. 基于深度学习的皮肤癌图像分类研究进展[J]. 计算机系统应用, 2026, 35(2): 23-39.
- [5] 朱文豪. 基于卷积神经网络和注意力机制的皮肤癌图像分类及分割研究[D]: [硕士学位论文]. 衡阳: 南华大学, 2024.
- [6] 庞德鹏. 基于迁移学习的图像识别案例分析[J]. 电子技术, 2024, 53(4): 64-67.
- [7] Gu, Y., Ge, Z., Bonnington, C.P. and Zhou, J. (2020) Progressive Transfer Learning and Adversarial Domain Adaptation for Cross-Domain Skin Disease Classification. *IEEE Journal of Biomedical and Health Informatics*, **24**, 1379-1393. <https://doi.org/10.1109/jbhi.2019.2942429>
- [8] Celebi, M.E., Kingravi, H.A., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., *et al.* (2007) A Methodological Approach to the Classification of Dermoscopy Images. *Computerized Medical Imaging and Graphics*, **31**, 362-373. <https://doi.org/10.1016/j.compmedimag.2007.01.003>
- [9] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1023/a:1022627411411>
- [10] Ballerini, L., Fisher, R.B., Aldridge, B. and Rees, J. (2013) A Color and Texture Based Hierarchical K-NN Approach to the Classification of Non-Melanoma Skin Lesions. In: Celebi, M. and Schaefer, G., Eds., *Color Medical Image Analysis*, Springer, 63-86. https://doi.org/10.1007/978-94-007-5389-1_4
- [11] Sun, L., Dong, J., Tang, J. and Pan, J. (2023) Spatially-Adaptive Feature Modulation for Efficient Image Super-resolution.

-
- 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 13144-13153. <https://doi.org/10.1109/iccv51070.2023.01213>
- [12] Chen, J., Kao, S., He, H., Zhuo, W., Wen, S., Lee, C., *et al.* (2023) Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 12021-12031. <https://doi.org/10.1109/cvpr52729.2023.01157>
- [13] Szegedy, C., Wei Liu,, Yangqing Jia,, Sermanet, P., Reed, S., Anguelov, D., *et al.* (2015) Going Deeper with Convolutions. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 1-9. <https://doi.org/10.1109/cvpr.2015.7298594>
- [14] Nidhi, and Verma, B. (2024) A Lightweight Convolutional Swin Transformer with Cutmix Augmentation and CBAM Attention for Compound Emotion Recognition. *Applied Intelligence*, **54**, 7793-7809. <https://doi.org/10.1007/s10489-024-05598-5>
- [15] Tschandl, P., Rosendahl, C. and Kittler, H. (2018) The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. *Scientific Data*, **5**, Article No. 180161. <https://doi.org/10.1038/sdata.2018.161>
- [16] 刘业辉. 基于深度学习的皮肤癌分类算法研究[D]: [硕士学位论文]. 南宁: 广西大学, 2025.
- [17] Zhang, Y., Chen, J., Wang, K. and Xie, F. (2023) ECL: Class-Enhancement Contrastive Learning for Long-Tailed Skin Lesion Classification. In: Greenspan, H., *et al.*, Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2023*, Springer, 244-254. https://doi.org/10.1007/978-3-031-43895-0_23
- [18] 韩硕, 袁伟理, 杜泽宇. 基于自反馈阈值学习的半监督皮肤癌诊断模型[J]. 河北大学学报(自然科学版), 2024, 44(4): 441-448.