Modeling and Simulation 建模与仿真, 2024, 13(2), 1814-1829
Published Online March 2024 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2024.132170

Hans汉斯

基于改进MobileNetV2的中耳炎影像分类诊断模型

胡江婧，张学典

上海理工大学医用光学技术集仪器教育部重点实验室，上海

收稿日期：2024年1月19日；录用日期：2024年3月21日；发布日期：2024年3月29日

摘　要

中耳炎诊断是有效防止耳道疾病进一步恶化的重要前提，针对现有中耳炎相关研究中存在的数据集量少、网络参数量大、模型识别精度低等问题，提出一种基于改进MobileNetV2模型的中耳炎分类方法。首先，在MobileNetV2网络的倒置残差结构中嵌入坐标注意力机制，增强网络对中耳炎影像特征细化能力；其次，设计改进注意特征融合模块替换原始特征简单相加过程，加强模型在跨通道背景下对不同尺度特征的提取能力；同时采用HardSwish激活函数替换原始ReLU6函数，提升模型鲁棒性；最后，减少模型瓶颈层通道数，简化模型结构。实验结果表明，所提出CIH-MobileNetV2模型在中耳炎数据集上的识别准确率和F1 Score达到91.05%和89.06%，相较于原始MobileNetV2模型，分别提高了2.31%和3.69%，参数量较初始模型减少了43%。与经典网络AlexNet、GoogleNet、VGG16、ResNet50、MobileNetV3、ShuffleNetV2等网络相比，有更高的识别准确率和F1值，因此，该研究所提出模型能够较好对中耳炎类型进行分类，为中耳炎诊断提供有效帮助。

关键词

中耳炎分类，MobileNetV2，坐标注意力，特征融合，HardSwish

Otitis Media Image Classification and Recognition Model Based on Improved MobileNetV2

Jiangjing Hu, Xuedian Zhang

Key Laboratory of Medical Optical Instrument and Equipment of Ministry of Education, University of Shanghai for Science and Technology, Shanghai

Received: Jan. 19th, 2024; accepted: Mar. 21st, 2024; published: Mar. 29th, 2024

文章引用: 胡江婧, 张学典. 基于改进 MobileNetV2 的中耳炎影像分类诊断模型[J]. 建模与仿真, 2024, 13(2): 1814-1829. DOI: 10.12677/mos.2024.132170

Abstract

Otitis media recognition is an important prerequisite for effectively preventing the deterioration of otological diseases. Aiming at the problems of small amount of dataset, large number of network parameters, low model recognition accuracy, and excessive computational volume in the existing otitis media recognition research, we propose an otitis media recognition model based on improved MobileNetV2. First, the coordinate attention mechanism is embedded in the inverted residual structure of the MobileNetV2 network to enhance the network's ability to refine the otitis media image features. Second, Iterative Attention Feature Fusion module is used to replace the simple summation of the original features, which strengthens the model's capability of extracting features of different scales in the context of cross-channel. At the same time, the HardSwish activation function is used to replace the original ReLU6 function to improve the robustness of the model. Finally, the number of channels in the bottleneck layer of the model is reduced to simplify the model structure. The experimental results show that the recognition accuracy and F1 Score of the proposed CIH-MobileNetV2 model on the otitis media dataset reach 91.05% and 89.06%, which are improved by 2.31% and 3.69%, respectively. Compared with the original MobileNetV2 model, the number of parameters is reduced by 43%. Compared with the classical networks AlexNet, GoogleNet, VGG16, ResNet50, MobileNetV3, ShuffleNetV2, etc., there are higher recognition accuracy and F1 value, therefore, the proposed model of the institute is able to classify the type of otitis media better and provide an effective help for otitis media diagnosis.

Keywords

Otitis Media Classification, MobileNetV2, Coordinate Attention, Feature Fusion, HardSwish

1. 引言

中耳炎(OM)是鼓膜后方中耳腔和颞骨的气腔与咽鼓管覆盖的黏膜发生感染而导致的一种炎症[1]。其症状包括耳剧痛、耳溢液、鼓膜穿孔、耳道化脓感染等，常见的中耳炎类型有急性中耳炎(AOM)和慢性化脓性中耳炎(CSOM)。急性中耳炎(AOM) [2]是与上呼吸道感染密切相关的炎症，临床表现为耳膜鼓包、耳道红肿、分泌物增多、耳胀。慢性化脓性中耳炎(CSOM) [3]是一种持续性、隐匿性的中耳炎症状表现，会持续 2 周以上向外耳道排出分泌物，并常伴有鼓膜穿孔症状。阻塞性耳垢(EarWax)通常积聚在外耳道。一般来说，耳垢可以防止鼓膜被灰尘和其他微小粒子引发感染，但外耳道阻塞性耳垢长时间堆积后未脱落会引起耳道发炎感染，伴随疼痛、耳鸣和听力下降症状，严重时会造成继发性急性中耳炎[4]。

耳镜成像、声反射术、超声评估、鼓室测量等方式对中耳炎疾病的检测起了很大的作用[5]。临床上通常使用电子耳镜检查耳膜的状态，该设备包含小型照相机、卤素光源、低功率放大镜，以及连接到计算机实现存储图像[6]。医生通过获得的鼓膜图像进一步对患者进行评估诊断，后续诊断准确率完全依靠于医生的诊治经验，相关研究表明，在临床检测中，188 名耳鼻喉科医生诊断中耳炎的准确率仅为 73% [7]，且不同医生对相同患者的诊断结果相差较大。若不及时治疗，严重时会造成面部神经麻痹、听力损失、

颅内感染和认知障碍等并发症[8]；而过度诊断则会导致抗生素滥用现象，并造成患者用药不当的医疗事故[9]。因此，中耳炎的准确诊断在耳道疾病治疗过程中至关重要。

目前，随着人工智能和计算机视觉技术的不断革新，机器学习方法开始广泛应用于医学图像分析[10]。鉴于人工诊断过程中，其准确性受限于医生的个体医疗技能水平，越来越多的研究者开始使用机器学习技术来实现中耳炎诊断。Mironica 等人[11]采用包括 K-最近邻(KNN)、多层感知机(MLP)以及支持向量机(SVM)等在内的六种不同算法，在所采集的耳镜图像上进行实践。实验结果表明，支持向量机(SVM)表现出最佳的分类效果，其准确率达到了 72.04%。Myburgh 等人[12]提出了一种利用图像处理技术和决策树(DTs)的中耳炎诊断模型，用于区分五个不同类别，包括外耳道阻塞性耳垢、正常鼓膜、化脓性中耳炎(AOM)、中耳炎溢液和慢性化脓性中耳炎，其采用特定的特征提取算法来分析鼓膜图像，并将相关特征输入到决策树中。实验结果显示，该模型达到了 80.60%的准确率。Shie 等人[13]设计了一种用于中耳炎的计算机辅助诊断系统。其使用了不同类型的滤波器来提取特征，针对四类中耳炎(正常、化脓性中耳炎(AOM)、粘液性中耳炎(OME)和慢性化脓性中耳炎(COM))，准确率为 88.06%。Kruvilla 等人[14]提出了一种基于词汇表和语法集的自动化算法，这些词汇表和语法集分别对应于鼓膜的数值测量和决策规则，来用于确定中耳炎的诊断类别。该算法在正常、化脓性中耳炎(AOM)和粘液性中耳炎(OME)类别的分类中取得了 89.9%的准确率。Barasan 等人[15]提出了一种基于 Faster R-CNN 和预训练的卷积神经网络结合的分类模型，用于区分正常和异常的鼓膜。使用 Faster R-CNN 自动确定鼓膜图像中鼓膜的位置。该过程生成一个覆盖鼓膜的补丁，而不是完整的鼓膜图像。得到图像补丁在 VGG-16 上训练，最终得到 90.48%的分类准确率。

上述研究工作中，传统机器学习方法通常依赖于冗长的人工特征提取的过程，这个过程繁琐且耗时，由于耳内镜下鼓膜图像的特征较复杂，人为特征提取极大程度受到主观因素影响。而基于深度学习的经典卷积神经网络的参数量和计算复杂度较高[16]，模型过大时难以部署在移动设备上运行。因此，能否针对耳镜下中耳炎数字图像的复杂特征，利用轻量化卷积神经网络克服现有方法的不足，同时提高中耳炎诊断的准确率，是本文研究的重点。

针对上述问题，本文提出了一种基于轻量级卷积神经网络 MobileNetV2 的中耳炎诊断模型 CIH-MobileNetV2。首先，对于现有耳内镜下鼓膜数字影像数据集规模有限与深度学习网络所需训练数据量大的矛盾，采用图像增强技术扩充数据集样本。其次，融入轻量化坐标注意力机制 CA (Coordinate Attention)，既能获得图像通道信息，又保留了特征的方向信息，增强了网络模型对数字耳镜影像特征的分析能力；同时利用改进注意力特征融合模 IAFF (Iterative Attentional Feature Fusion)来替换网络结构中简单的特征层相加功能，获取跨通道下不同大小尺度的特征；其次，使用 HardSwish 激活函数替换原始 ReLU6 函数，提升了模型的鲁棒性；最后，对模型通道数进行缩减，减少了模型参数量，降低模型复杂度。实验结果表明，本文所提出的方法有效提高了对中耳炎不同类别诊断的准确率，同时减轻了对移动设备的性能要求，为轻量型中耳炎识别模型部署在移动设备上的研究提供了相应参考。

2. 模型方法

2.1. MobileNetV2 模型

MobileNet 网络[17]是专门针对移动端设备和资源受限环境下而设计的轻量化神经网络架构。其采用了深度可分离卷积结构，将传统 3 × 3 卷积结构分离成 3 × 3 逐通道卷积和 1 × 1 逐点卷积，在不降低网络精度的前提下，大幅度减少了模型参数量，训练时间更短。

MobileNetV2 [18]的模型结构如下表 1 所示，其在提出的 V1 版本上进行了改进，除了保留深度可分离卷积结构以外，主要特点表现为瓶颈层中的倒置残差结构(Inverted Residual, IR)，倒残差结构在 3 × 3

逐通道卷积模块前引入 1 × 1 逐点卷积，以提升逐通道卷积模块的特征表达能力，同时采用残差连接实现特征复用，有效提升训练效率。为了避免 ReLU 激活函数所导致的低维特征丢失问题，瓶颈层中最后一个卷积层采用线性激活函数替代 ReLU 激活函数。倒置残差模块如图 1 所示。

Table 1. Structure of MobileNetV2 network
表 1. MobileNetV2 网络结构

输入	结构层	扩张系数	卷积核个数	结构层数	卷积步长
224 × 224 × 3	conv2d	-	32	1	2
112 × 112 × 32	bottleneck	1	16	1	1
112 × 112 × 16	bottleneck	6	24	2	2
56 × 56 × 24	bottleneck	6	32	3	2
28 × 28 × 32	bottleneck	6	64	4	2
14 × 14 × 64	bottleneck	6	96	3	1
14 × 14 × 96	bottleneck	6	160	3	2
7 × 7 × 160	bottleneck	6	320	1	1
7 × 7 × 320	conv2d 1 × 1	-	1280	1	1
7 × 7 × 1280	avgpool 7 × 7	-	-	1	-
1 × 1 × 1280	conv2d 1 × 1	-	k	-	-

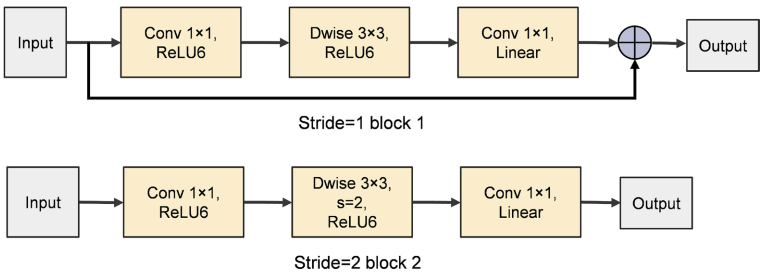


Figure 1. Structure diagram of the inverted residuals module
图 1. 倒置残差结构模块图

2.2. CIH-MobileNetV2 模型

针对耳镜下中耳炎影像诊断任务，本文设计了 CIH-MobileNetV2 模型，由于本文实验所用中耳炎影像数据集背景复杂，存在病变区域分布不均、病变区域面积大小各异等特点，且原始 MobileNetV2 网络在数据集上表现效果较次，为提高模型在分类任务上的性能表现，本文实验对初始模型进行一系列改进，改进后模型内部构造如图 2 所示。

与原始 MobileNetV2 相比，本研究首先在倒置残差模块中 1 × 1 降维卷积 Linear 层后引入坐标注意力机制，增强网络在图像训练过程中对目标区域的特征提取能力；其次采用迭代注意力特征融合模块替换原始瓶颈层中特征相加机制，以此融合跨层级不同尺度的特征，避免遗失小区域特征信息；然后将深度可分离卷积模块中 ReLU6 激活函数替换成 HardSwish 函数，保留训练样本中正负特征信息，以提升模型的鲁棒性；最后，缩减模型通道数，降低模型参数量，以便模型更利于部署在移动设备中。

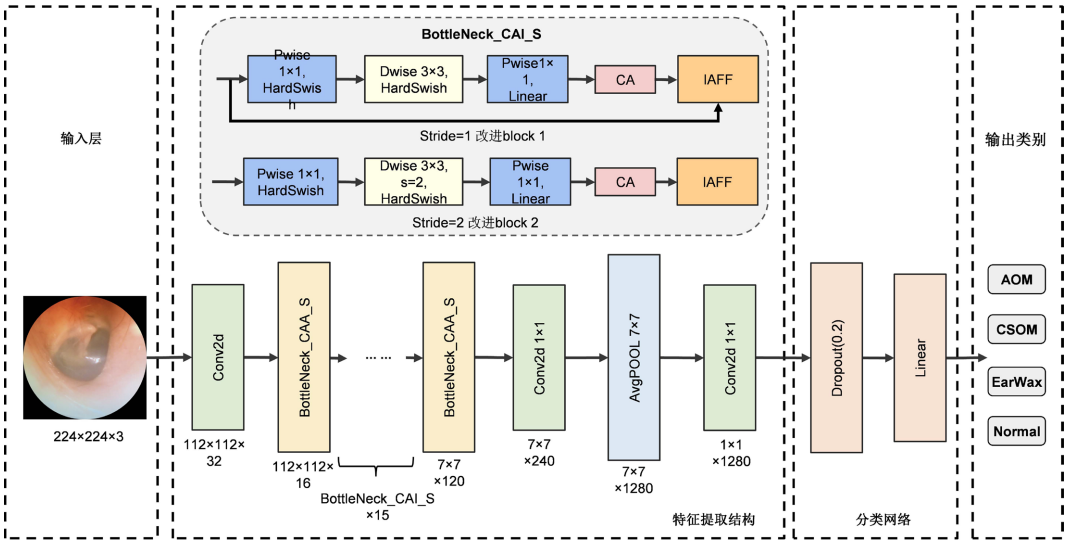


Figure 2. Structural diagram of the CIH-MobileNetV2 model
图 2. CIH-MobileNetV2 模型结构图

2.2.1. CA 注意力机制

在神经网络中，卷积核仅关注输入图像的局部特征，而忽略全局表征对局部的影响。在网络中引入注意力机制不仅能提高对图片特征的提取能力，同时使网络更加关注到目标区域，提升任务效能。

CA (Coordinate attention)注意力是 Hou [19]等人针对移动网络设计提出的坐标注意力机制，旨在将位置信息嵌入通道注意力中，从而增强网络模型特征学习的表达能力，获取目标更详细的特征信息，同时也避免了产生重大的计算开销。如图 3 所示，CA 注意力机制作用到网络模型分为两个步骤，第一步为坐标信息嵌入，第二步为坐标注意力生成。

在坐标信息嵌入过程中，为了避免二维全局集合造成的位置信息损失，坐标注意力机制将通道注意分解为两个并行的一维特征编码过程，使用尺寸为(H, 1)或(1, W)的平均池化核对每个通道的水平方向和垂直方向进行编码，得到沿 X 方向和 Y 方向的感知特征图，如公式(1)、(2)所示。其中 $Z_c^h(h)$ 和 $Z_c^w(w)$ 分别表示经过一维全局平均池化后第 c 通道在高度 h 处以及在宽度为 w 处特征输出张量。

$$Z_c^h(h)=\frac{1}{W}\sum_{0\le i\le w}X_c(h,i) \tag{1}$$

$$Z_c^w(w)=\frac{1}{H}\sum_{0\le j\le H}X_c(j,w) \tag{2}$$

通过上述两种变换可分别获得沿两个空间方向的聚合特征，生成一对方向感知特征图。在坐标注意力生成阶段，将获取的宽度方向和高度方向的特征图进行 Concat 拼接操作，通过 1 × 1 卷积变换函数 F1，送入 Sigmoid 非线性激活函数生成大小为 $C/r\times1\times(H+W)$ 的融合特征图 f，如公式(3)所示。

$$f=\delta\left(F_1\left(\left[z^h,z^w\right]\right)\right) \tag{3}$$

按照空间维度将得到的特征 f 分解成两个不同方向张量 $f_h\in RC/r\times H$ 和 $f_w\in RC/r\times W$，再使用 1×1 卷积的线性变换和分别将通道数量恢复至原始规模 C，经过 Sigmoid 激活函数后分别得到水平方向注意力权重 g^h 与垂直方向的注意力权重 g^w，如公式(4) (5)所示。

$$g^h=\delta\left(F_h\left(f^h\right)\right) \tag{4}$$

$$g^w = \delta\left(F_w\left(f^w\right)\right) \tag{5}$$

最终将水平方向输出权重 g^h 和垂直方向注意力权重 g^w，与输入特征 X 在相应的坐标位置进行加权融合，得到最终的坐标注意力输出特征图 y_c，如公式(6)所示。

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{6}$$

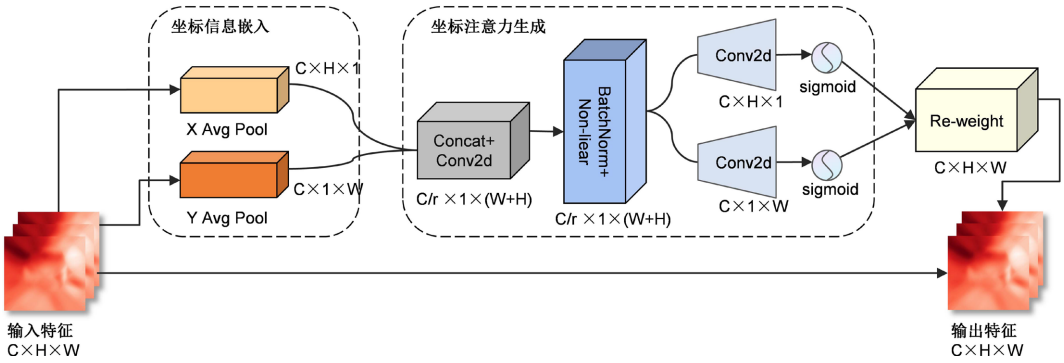


Figure 3. The structural of CA module
图 3. CA 模型结构

2.2.2. IAFF 特征融合模块

简单的特征融合通常采用求和(summation)或连接(concatenation)的线性计算对来自不同层级或分支的特征进行组合，这种融合方式会造成跨层后特征权值质量下降，导致模型表现受限。为了更好地融合跨层级架构中语义和尺度不一致的特征，以及避免丢失小目标上下文信息，本研究引入可迭代注意特征融合模块 IAFF (Iterative Attentional Feature Fusion) [20]，以进一步提高神经网络的表征能力。

为了缓解尺度变化和小目标所带来的问题，在多尺度通道注意模块(MS-CAM)中，如图 4(a)所示，其中 r 为通道缩减比。其关键思想是通过改变空间池化的大小，在多个尺度上实现通道注意融合。MS-CAM 结构在获得全局上下文特征同时引入局部特征分支，通过逐点卷积(PWConv)模块作为局部上下文聚合器，同时较普通 1 × 1 卷积而言，大幅度减少了参数量，最终得到的注意权重通过加权平衡全局特征与局部细微特征信息。在通道注意力模块中结合局部和全局的上下文特征。初始输入特征 X 经过瓶颈结构运算得到的全局特征通道注意力 $g(X)$ 和局部特征通道注意力 $L(X)$，得到的细化特征 $X' \in R^{C\times H\times W}$，相应公式如下：

$$L(X) = B\left(PWConv_2\left(\sigma\left(B\left(PWConv_1(X)\right)\right)\right)\right) \tag{7}$$

$$X' = X \otimes M(X) = X \otimes \left(L(X) \oplus g(X)\right) \tag{8}$$

其中 $M(X) \in R^{C\times H\times W}$ 表示通过 MS-CAM 模块产生的注意权重，B 表示批标准化(Batch Normalization)，σ 表示为 ReLU 激活函数。

如图 4(b)所示，普通 AFF 模块在 MS-CAM 模块的基础上，将基于注意力的特征融合从同层场景推广到跨层场景，包括短跳连接和长跳连接，以及涵盖多个特征内部的初始整合。在 AFF 模块中，对于两个特征图 $X, Y \in R^{C\times H\times W}$，默认情况下 Y 是具有更大感受野的特征图。计算公式为式(9)，其中 Z 为融合后的特征，$M(X+Y)$ 为融合权重，结构图中虚线表示为 $1-M(X+Y)$。初始对输入的两个特征 X，Y 进行元素级简单相加特征融合，经过 sigmoid 激活函数后，融合特征权值范围在 0~1 之间，同时经过减 1 计算，使得网络在后续训练过程中在 X 和 Y 之间进行软选择或加权平均，来确定不同特征的权重。

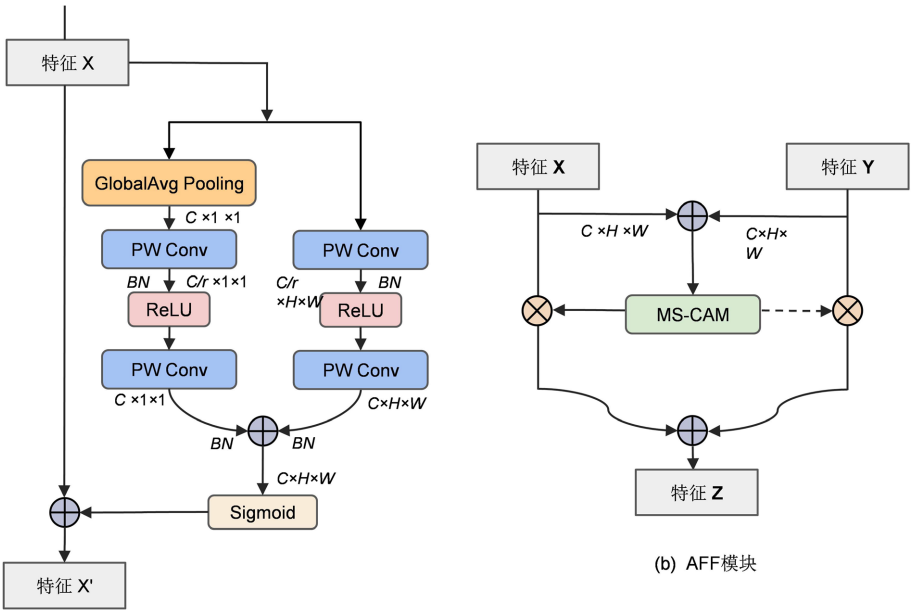


Figure 4. The structural of MS-CAM module
图 4. MS-CAM 模块及 AFF 模块结构图

$$Z = M\left(X \oplus Y\right) \otimes X + \left(1 - M\left(X \oplus Y\right)\right) \otimes Y \tag{9}$$

在实际场景应用中，初始特征融合仅关注到对应输入特征简单相加，而如何集成初始特征特性，保证特征融合质量对最终输出权重产生的影响，本研究最终采用可迭代注意特征融合模块替换瓶颈层特征简单 concat 模式，即在普通 AFF 基础上，使用额外 AFF 融合模式来生成更细粒度的初始特征输入，具体公式如下所示，IAFF 结构如图 5 所示。

$$Z = X \oplus Y = M\left(X + Y\right) \otimes X + \left(1 - M\left(X + Y\right)\right) \otimes Y \tag{10}$$

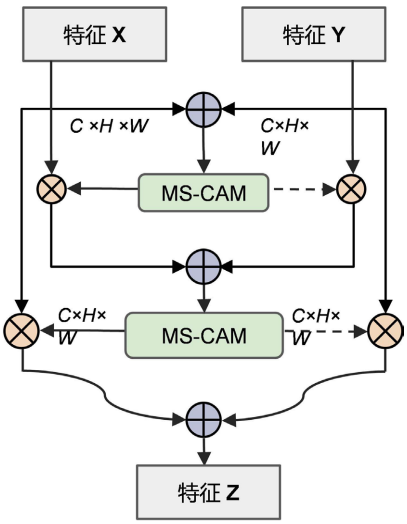


Figure 5. The structural of IAFF module
图 5. IAFF 模型结构图

2.2.3. HardSwish 激活函数

ReLU 激活函数是网络训练中最常用的激活函数[21]，在加速网络收敛的同时解决了梯度消失的问题，缓解了过拟合现象的发生。公式表现如式(10)。ReLU6 函数是 ReLU 函数的一种改进，公式如式(11)，原始正值输出范围被抑制为 6，这使得在移动端设备低精度计算中具有较好的数值分辨率。但在具体任务训练中，随着网络学习的不断深入，Float16 所表示的数值范围受限，当激活函数输出值超过其数值时，精确度则大幅度下降，导致信息丢失；同时函数的输入值为负时，其输出梯度始终为 0，会造成神经元无法更新参数、直接失活的现象，从而影响模型的学习能力。

$$\mathrm{ReLU}(x)=\max(0,x) \tag{11}$$

$$\mathrm{ReLU6}(x)=\min(\max(0,x),6) \tag{12}$$

为了克服上述 ReLU 函数与 ReLU6 函数存在的缺点，本文使用 HardSwish 激活函数[22]替换模型中原有 ReLU 及 ReLU6 函数，函数图像如图 6。HardSwish 是 Swish 函数的近似体，计算模式简洁，是无上限的轻量级非线性激活函数，能够提高推理性能，降低损失能耗；且在零值附近的梯度表现更佳平滑，在训练中能够避免梯度爆炸，增强模型训练的稳定性和可靠性。具体计算公式如下。

$$\mathrm{HardSwish}(x)=\begin{cases}0 & x\le -3;\\ x & x\ge 3;\\ x\cdot\dfrac{x+3}{6} & \text{其他}\end{cases} \tag{13}$$

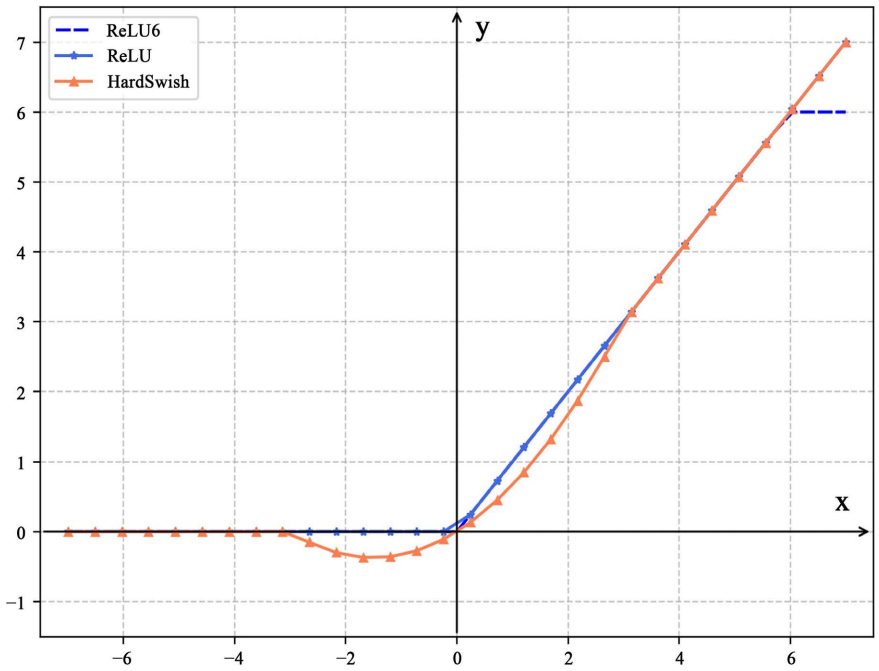


Figure 6. Comparison the images of ReLU and ReLU6 with HardSwish activation functions
图 6. ReLU 及 ReLU6 与 HardSwish 激活函数的对比图

2.2.4. 模型通道数调整

上述所提出的改进策略能有效提升模型的表达能力，但在一定程度上增大了参数量，实际医用中会

造成相应系统开销。在 MobileNetV3 模型中，研究者将初始卷积通道数减少为原来的一半，为了防止减少过多通道数导致模型精度大幅度降低，因此，本研究将瓶颈层中超过 64 的通道数缩减少 1/4，其他通道数不变，得到的新瓶颈结构记为 Bottleneck_CAI_S，改进后模型 CIH_MobileNetV2 的结构表如下表 2 所示。

Table 2. Structure of CIH_MobileNetV2 network
表 2. CIH_MobileNetV2 网络结构

输入	结构层	扩张系数	卷积核个数	结构层数	卷积步长
224 × 224 × 3	conv2d	-	32	1	2
112 × 112 × 32	Bottleneck_CAI_S	1	16	1	1
112 × 112 × 16	Bottleneck_CAI_S	6	24	2	2
56 × 56 × 24	Bottleneck_CAI_S	6	32	3	2
28 × 28 × 32	Bottleneck_CAI_S	6	64	4	2
14 × 14 × 64	Bottleneck_CAI_S	6	72	3	1
14 × 14 × 72	Bottleneck_CAI_S	6	120	3	2
7 × 7 × 120	Bottleneck_CAI_S	6	240	1	1
7 × 7 × 240	conv2d 1 × 1	-	1280	1	1
7 × 7 × 1280	avgpool 7 × 7	-	-	1	-
1 × 1 × 1280	conv2d 1 × 1	-	4	-	-

3. 实验及结果分析

3.1. 数据集

本文所使用数据集建立在 CTG ANALYSIS 研究小组公布的开源鼓膜影像数据集 Tympanic_Membrane 的基础上[23]，该数据集共包含 956 幅耳镜影像，获取于 2018 年 10 月至 2019 年 06 月期间在 Özel Van Akdamar 医院接受治疗的中耳炎患者处，该数据集中共包括 9 类耳镜下鼓膜图像，但由于数据集中某些类别样本数量过少(10 张以内)，且存在黑暗、模糊等非有效视觉特征，故本文只关注急性中耳炎(AOM)、慢性化脓性中耳炎(CSOM)、耳内道阻塞性耳垢(EarWax)以及正常鼓膜(Normal)的分类。

针对原数据集图像数据不足以及部分类别分类有误的情况，本研究在耳科临床专家的进一步分类筛选下，整合了 Viscanio 等人与智利大学临床医院耳鼻喉科合作创建并公开的鼓膜影像数据集 Datos [24]，构建出全新的中耳炎影像数据集。该数据集中共包括 1517 张图片，其中急性中耳炎(AOM)图像共 119 张，慢性化脓性中耳炎(CSOM)图像共 283 张，外耳道阻塞性耳垢(EarWax)图像共 360 张，正常鼓膜(Normal)图像共 755 张。数据集详细情况如表 3 所示，数据集样本图像展示为图 7 所示。

Table 3. The category distribution of dataset
表 3. 数据集类别分布

标签	类别	数量
AOM	急性中耳炎	119
CSOM	慢性化脓性中耳炎	283
EarWax	外耳道阻塞性耳垢	360
Normal	正常鼓膜	755
共计		1517

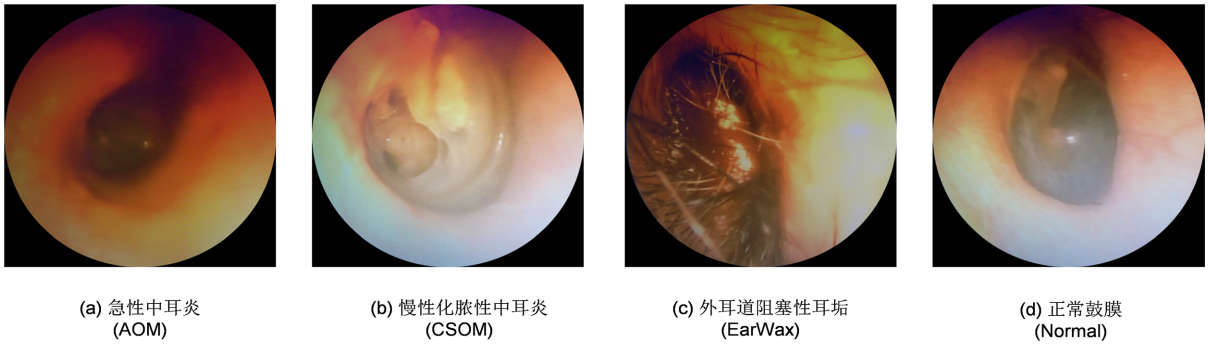


Figure 7. Images of tympanic membrane samples and category labels. (a) Acute otitis media (AOM); (b) Chronic purulent otitis media (CSOM); (c) Obstructive earwax of the external auditory canal (EarWax); (d) Normal tympanic membrane (Normal)
图 7. 部分鼓膜样本图像及类别标签。(a) 急性中耳炎(AOM)；(b) 慢性化脓性中耳炎(CSOM)；(c) 耳内道阻塞性耳垢(EarWax)；(d) 正常鼓膜(Normal)

3.2. 数据预处理

本文将实验数据集按照 8:2 的比例随机划分为训练集与测试集，并且在训练集中随机抽取 20%作为模型的验证集，用于评估模型的泛化能力。为了防止由于数据集规模较小且样本间不平衡造成模型训练过程中出现过拟合现象，以及提高模型分类的准确性。本研究对训练集中不同类别图像进行相应数据增强操作，由于耳镜影像具有一定的特征表达性，为避免对数据集中鼓膜形态结构造成干扰，本研究仅针对训练集影像进行水平、垂直翻转，旋转 90 度，旋转 180 度，旋转 270 度，水平方向平移，中心裁剪以及随机光照、随机对比度、随机饱和度等多种增强方式，样本增强部分示例表现如图 8 所示。

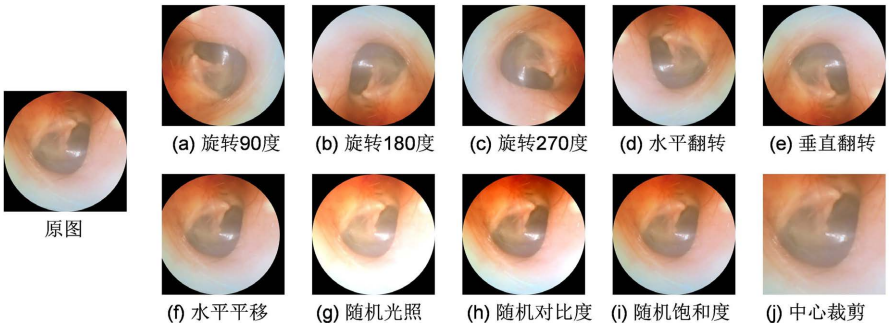


Figure 8. Image examples of data augmentation (a) Rotate 90 degrees; (b) Rotate 180 degrees; (c) Rotate 270 degrees; (d) Horizontal flip; (e) Vertical flip; (f) Horizontal translate; (g) Random brightness; (h) Random contrast; (i) Random saturation; (j) Center clipping
图 8. 数据增强示例。(a) 旋转 90 度；(b) 旋转 180 度；(c) 旋转 270 度；(d) 水平翻转；(e) 垂直翻转；(f) 水平平移；(g) 随机光照；(h) 随机对比度；(i) 随机饱和度；(j) 中心裁剪

3.3. 实验环境及参数设置

本文实验环境基于 Ubuntu18.04 操作系统，使用 Pytorch1.10 深度学习框架、Python3.9 编译环境、Anaconda 操作软件、CUDA 架构以及 PyCharm 编译器。硬件平台配置 CPU 为 Intel(R) Core(TM) i5-8200@1.30 GHz，内存为 64GB，GPU 为 2 张 NVIDIA RTX 2080Ti 12GB。

在实验过程中，所有样本图片统一设置成 224 × 224 像素大小输入网络进行训练，使用 Adam 优化器来更新模型权重以及偏差参数，初始学习率设置为 0.0001，动量参数为 0.9，批处理大小 batch size 设为 16，训练周期为 100 轮。相应超参数如下表 4 所示。

Table 4. The hyperparameters of experiments
表 4. 实验超参数设置

超参数	具体数值
Image Input size	224 × 224
Batch size	16
Epoch	100
Learning rate	0.0001
Momentum	0.9

3.4. 评价指标

本文实验为中耳炎多分类任务，考虑到实验数据集存在样本数据量不平衡的问题，为了更加客观地评估各模型在该数据集上的性能表现，本实验选择混淆矩阵、准确率 Accuracy、精确率 Precision、召回率 Recall 以及 F1 Score 作为模型的综合评价指标[25]，计算公式如下。

$$\text{Accurary} = \frac{\text{TP}_i + \text{TN}_i}{\text{TP}_i + \text{TN}_i + \text{FP}_i + \text{FN}_i} \times 100\% \tag{14}$$

$$\text{Precision} = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \times 100\% \tag{15}$$

$$\text{Recall} = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \times 100\% \tag{16}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \tag{17}$$

其中，TP_i 表示样本被正确预测为第 i 类，实际真实类别也为第 i 类的样本量；TN_i 表示除去第 i 类的其他类别样本被正确预测为对应类别的样本量；FP_i 表示非第 i 类的样本被预测为第 i 类的样本量；FN_i 表示真实类别为第 i 类但被错误预测为其他类别的样本量。

3.5. 消融实验及结果

为了验证在 MobileNetV2 模型上相应改进方法对提高模型表现能力的有效性，在同一数据集下对不同改进方法的模型进行消融实验。其中，当仅使用坐标注意力机制时，将 CA 注意力模块嵌入到 MobileNetV2 瓶颈层的倒残差结构中；当仅使用 IAFF 特征融合结构时，将原倒残差结构中简单特征相加部分进行替换；使用 HardSwish 激活函数替换原有 ReLU6 函数；以及缩减模型通道数的情况。实验中，使用测试准确率(Accuracy)、F1 Score 以及模型参数量作为评价指标，具体结果如表 5 所示。

Table 5. Results of ablation experiments
表 5. 消融实验结果

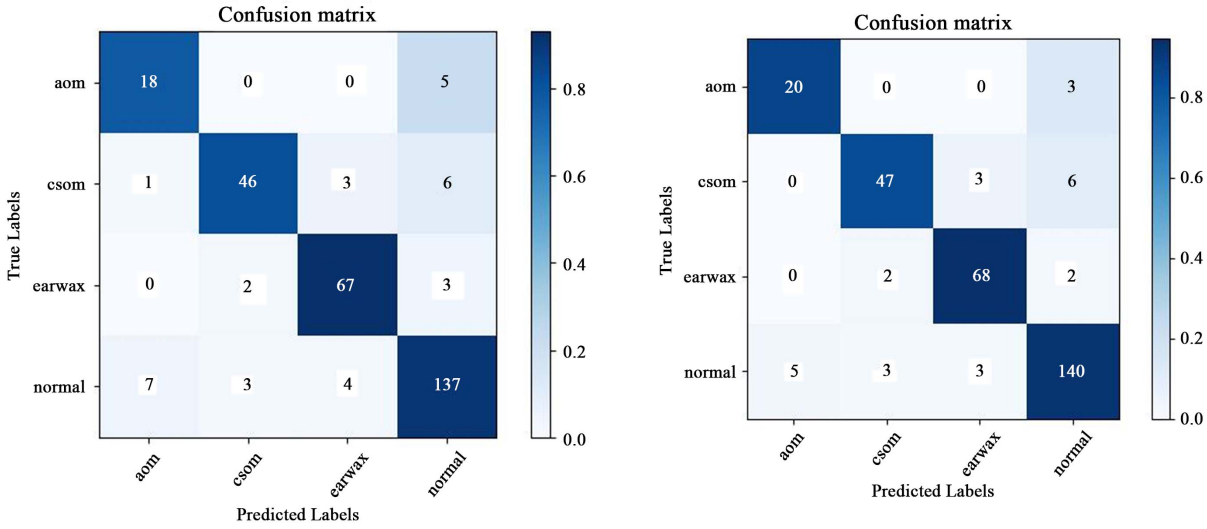
模型	改进因素				Acc/%	F1/%
	模型 通道数调整	CA 注意力机制	IAFF 特征融合	HardSwish 激活函数		
MobileNetV2	—	—	—	—	88.74	85.49
	√	—	—	—	87.74	83.31

续表

	—	√	—	—	90.72	88.54
MobileNetV2	—	—	√	—	89.73	87.31
	—	—	—	√	89.40	86.88
CIH-MobileNetV2	√	√	√	√	91.05	89.06

注：“—”表示在原始 MobileNetV2 网络中未引入该方法，“√”表示在原始 MobileNetV2 中使用该改进方法。

上述结果表明，在初始 MobileNetV2 模型倒置残差结构中嵌入 CA 注意力机制，训练所得模型在测试数据上的准确率以及 F1 Score 分别提升了 1.98 和 3.05 个百分点；使用 IAFF 迭代注意特征融合结构替换原始特征简单相加方式，所得模型相应准确率提高了 0.99 个百分点，F1 Score 提高了 1.82 个百分点；通过将 ReLU6 函数调整为 HardSwish 激活函数，所得准确率以及 F1 Score 分别提升了 0.66 和 1.39 个百分点。最终，通过嵌入坐标注意力机制、优化模型特征融合方式替换激活函数以及的模型改进策略，改进后模型在中耳炎数据集上的准确率与 F1 Score 表现较未改进 MobileNetV2 模型而言，分别提升了 2.31 和 3.57 个百分点。为了更直观地展示本研究提出的 CIH-MobileNetV2 在中耳炎分类任务上的有效性，测试数据集在初始 MobileNetV2 和改进后 CIH-MobileNetV2 模型上混淆矩阵详细情况如下图 9 所示。



(a) 初始MobileNetV2混淆矩阵　　(b) 改进后CIH-MobileNetV2混淆矩阵

Figure 9. The confusion matrix of the model on the test set. (a) Initial MobileNetV2 confusion matrix; (b) Improved CIH-MobileNetV2 confusion matrix

图 9. 模型分别在测试集上所得混淆矩阵。(a) 初始 MobileNetV2 混淆矩阵；(b) 改进后 CIH-MobileNetV2 混淆矩阵

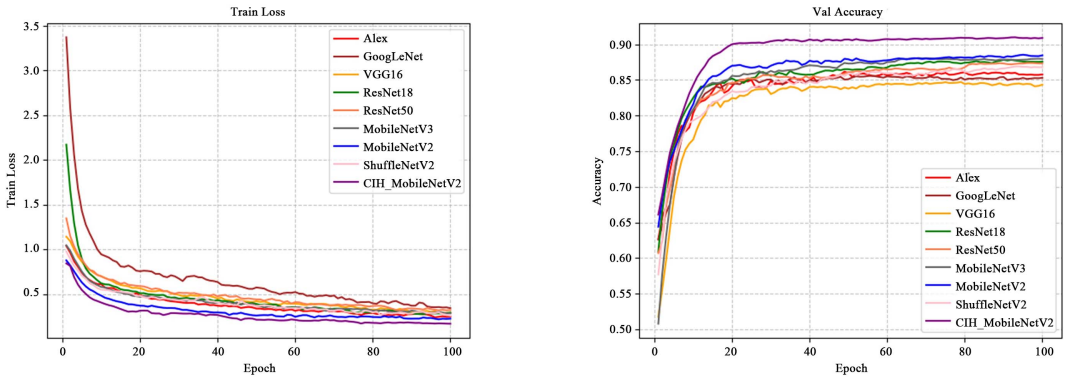
3.6. 不同模型对比实验及结果

为了更好的验证 CIH-MobileNetV2 模型在本文任务上的有效性，将改进后的 CIH-MobileNetV2 模型与目前深度学习领域图像分类任务中经典网络 AlexNet、GoogleNet、VGG16、ResNet18、ResNet50、ShuffleNetV2 进行对比实验。在相同中耳炎数据集以及运行环境下，以测试集准确率(Accuracy)、精确率(Precision)、召回率(Recall)以及 F1 Score、模型参数量、FLOPs 作为评估指标。训练过程中，各分类模型在训练集上的损失值及在验证集上的准确值变化如图 10 所示，相应指标数值如下表 6 所示。

Table 6. System resulting data of standard experiment
表 6. 标准试验系统结果数据

模型	Accuracy /%	Precision /%	Recall /%	F1 /%	参数量 /M	FLOPs /G
AlexNet	86.09	82.48	80.03	81.00	14.58	0.311
GoogleNet	85.09	81.67	82.04	81.42	5.97	1.59
VGG16	84.76	83.24	78.26	80.46	134.27	15.5
ResNet18	88.11	84.02	86.10	84.71	11.69	1.824
ResNet50	88.07	86.24	83.43	84.72	23.51	4.12
MobileNetV3	88.41	85.92	85.62	85.76	4.21	0.233
MobileNetV2	88.74	85.18	86.05	85.49	2.25	0.326
ShuffleNetV2	87.41	85.23	85.30	84.89	2.48	0.305
CIH-MobileNetV2	91.05	89.04	89.51	89.18	1.27	0.238

由上表数据可得，改进后模型同轻量化模型 ShuffleNetV2 相比，在测试集中准确率和 F1 Score 分别高出3.64%和4.78%，同经典分类模型AlexNet、GoogleNet、VGG16、ResNet50相比在准确率(Accuracy)、精确率(Precision)、召回率(Recall)以及 F1 Score 有明显的提升，且参数量大幅度下降，在确保模型性能表现的同时，更好地符合模型轻量化的要求。训练过程中，各模型在训练集上的损失值及在验证集上的准确值如图 10 所示，训练得到的模型在测试集上得到的相关混淆矩阵如下图 11 所示。



(a) 各模型在训练集上的损失值　　(b) 各模型在验证集集上的准确值

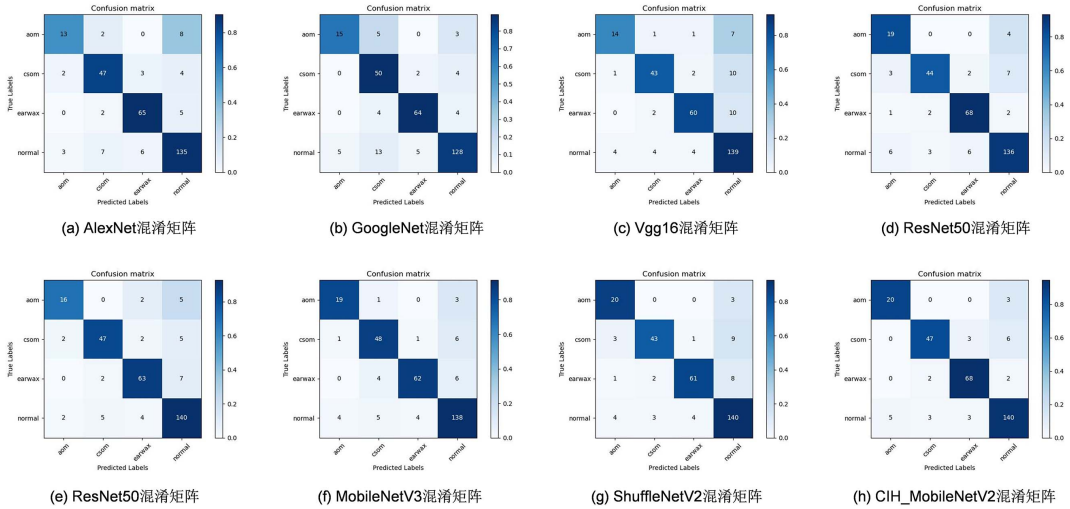
Figure 10. Improved model iteration curve. (a) The loss of each model on the training set; (b) The exact values of each model on the validation set
图 10. 改进模型迭代曲线。(a) 各模型在训练集上的损失值；(b) 各模型在验证集上的准确值

为了更好的体现本研究提出的中耳炎诊断模型在实际分类中的准确性，将所训练完成的模型权重对测试集内图像进行测试，具体表现如下图 12，可以得出，本研究所设计模型能够得到较好的诊断结果，能为中耳炎诊断提供有效帮助。

4. 结论

本文研究以中耳炎病变类别为对象，提出了一种融合 CA 注意力机制以及 IAFF 特征融合模块的中耳

炎诊断模型 CIH-MobileNetV2，并引入 HardSwish 激活函数替换原有瓶颈层中 ReLU6 激活函数，最后缩小卷积结构内部通道数，以此减少模型参数量，降低模型大小，节约了计算成本。



(a) AlexNet混淆矩阵 (b) GoogleNet混淆矩阵 (c) Vgg16混淆矩阵 (d) ResNet50混淆矩阵

(e) ResNet50混淆矩阵 (f) MobileNetV3混淆矩阵 (g) ShuffleNetV2混淆矩阵 (h) CIH_MobileNetV2混淆矩阵

Figure 11. The confusion matrix representation of each model on the test set
图 11. 各模型在测试集上的混淆矩阵表现

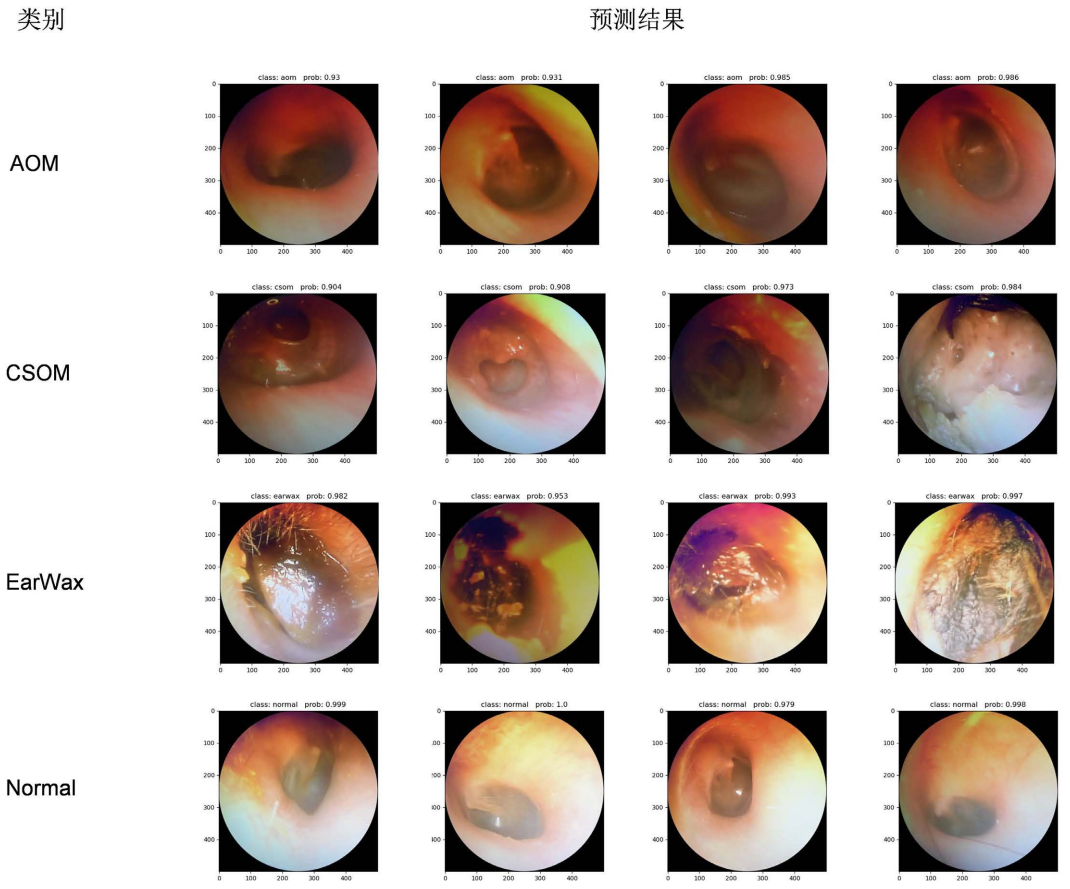


Figure 12. The predict results of CIH-MobileNetV2 on the test set
图 12. 改进 CIH-MobileNetV2 模型在测试集上预测结果

实验结果表明，本文提出改进方法能兼顾高精度以及轻量化两大要求，在对比经典分类网络AlexNet、GoogleNet、VGG16、ResNet18、ResNet50 中，改进的 CIH-MobileNetV2 网络在性能表现上效果更佳，分类准确率达到了 91.05%，F1 Score 为 89.18%，从而验证了本文方法的有效性。

后续工作中，将着眼于扩大数据规模，从而提升中耳炎识别模型的泛化能力。并且在实际临床场景中，识别准确率会受到耳镜下数字影像背景环境的影响，因此，在复杂背景中，在模型轻量化的前提下，进一步提高模型准确率是未来研究的目标。

基金项目

本研究得到了国家重点研发计划(2021YFB2802300)的工作支持。

参考文献

[1] Bock, S. and Weiß, M. (2019) A Proof of Local Convergence for the Adam Optimizer. 2019 *International Joint Conference on Neural Networks* (*IJCNN*), Budapest, 14-19 July 2019, 1-8. https://doi.org/10.1109/IJCNN.2019.8852239

[2] Kim, S.H., Kim, J.R., Song, J.J. and Chae, S.W. (2020) Trend and Patterns in the Antibiotics Prescription for the Acute Otitis Media in Koreanchildren. *International Journal of Pediatric Otorhinolaryngology*, **130**, Article ID: 109789. https://doi.org/10.1016/j.ijporl.2019.109789

[3] Aduda, D.S., Macharia, I.M., Mugwe, P., Oburra, H., Farragher, B., Brabin, B. and Mackenzie, I. (2013) Bacteriology of Chronic Suppurative Otitis Media (CSOM) in Children in Garissa District, Kenya: A Point Prevalence Study. *International Journal of Pediatric Otorhinolaryngology*, **77**, 1107-1111. https://doi.org/10.1016/j.ijporl.2013.04.011

[4] Cömert, Z. and Kocamaz, A.F. (2018) Open-Access Software for Analysis of Fetal Heart Rate Signals. *Biomedical Signal Processing and Control*, **45**, 98-108. https://doi.org/10.1016/j.bspc.2018.05.016

[5] Marom, T., Kraus, O., Habashi, N. and Tamir, S.O. (2019) Emerging Technologies for the Diagnosis of Otitis Media. *Otolaryngology*: *Head and Neck Surgery*, **160**, 447-456. https://doi.org/10.1177/0194599818809337

[6] Goggin, L.S., Eikelboom, R.H. and Atlas, M.D. (2007) Clinical Decision Support Systems and Computer-Aided Diagnosis in Otology. *Otolaryngology*: *Head and Neck Surgery*, **136**, s21-s26. https://doi.org/10.1016/j.otohns.2007.01.028

[7] Sorrento, A. and Pichichero, M.E. (2001) Assessing Diagnostic Accuracy and Tympanocentesis Skills by Nurse Practitioners in Management of Otitis Media. *Journal of the American Academy of Nurse Practitioners*, **13**, 524-529. https://doi.org/10.1111/j.1745-7599.2001.tb00019.x

[8] Myburgh, H.C., Jose, S., Swanepoel, D.W. and Laurent, C. (2018) Towards Low Cost Automated Smartphone-and Cloud-Based Otitis Media Diagnosis. *Biomedical Signal Processing and Control*, **39**, 34-52. https://doi.org/10.1016/j.bspc.2017.07.015

[9] Nyquist, A.C., Gonzales, R., Steiner, J.F. and Sande, M.A. (1998) Antibiotic Prescribing for Children with Colds, Upper Respiratory Tract Infections, and Bronchitis. *JAMA*, **279**, 875-877. https://doi.org/10.1001/jama.279.11.875

[10] 俞益洲, 石德君, 马杰超, 等. 人工智能在医学影像分析中的应用进展[J]. 中国医学影像技术, 2019, 35(12): 1808-1812. https://doi.org/10.13929/J.1003-3289.201909150

[11] Mironică, I., Vertan, C. and Gheorghe, D.C. (2011) Automatic Pediatric Otitis Detection by Classification of Global Image Features. 2011 *E-Health and Bioengineering Conference* (*EHB*), Iasi, 24-26 November 2011, 1-4.

[12] Myburgh, H.C., Van Zijl, W.H., Swanepoel, D., Hellström, S. and Laurent, C. (2016) Otitis Media Diagnosis for Developing Countries Using Tympanic Membrane Image-Analysis. *eBioMedicine*, **5**, 156-160. https://doi.org/10.1016/j.ebiom.2016.02.017

[13] Shie, C.K., Chang, H.T., Fan, F.C., Chen, C.J., Fang, T.Y. and Wang, P.C. (2014) A Hybrid Feature-Based Segmentation and Classification System for the Computer Aided Self-Diagnosis of Otitis Media. 2014 36*th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, 26-30 August 2014, 4655-4658.

[14] Kuruvilla, A., Shaikh, N., Hoberman, A. and Kovačević, J. (2013) Automated Diagnosis of Otitis Media: Vocabulary and Grammar. *Journal of Biomedical Imaging*, **2013**, Article ID: 327515. https://doi.org/10.1155/2013/327515

[15] Başaran, E., Cömert, Z. and Çelik, Y. (2020) Convolutional Neural Network Approach for Automatic Tympanic Membrane Detection and Classification. *Biomedical Signal Processing and Control*, **56**, Article ID: 101734. https://doi.org/10.1016/j.bspc.2019.101734

[16] Lemley, J., Bazrafkan, S. and Corcoran, P. (2017) Deep Learning for Consumer Devices and Services: Pushing the

Limits for Machine Learning, Artificial Intelligence, and Computer Vision. *IEEE Consumer Electronics Magazine*, **6**, 48-56. https://doi.org/10.1109/MCE.2016.2640698

[17] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Adam, H., *et al.* (2017) Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv: 1704.04861.

[18] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L. C. (2018) Mobilenetv2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4510-4520. https://doi.org/10.1109/CVPR.2018.00474

[19] Hou, Q., Zhou, D. and Feng, J. (2021) Coordinate Attention for Efficient Mobile Network Design. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 13708-13717. https://doi.org/10.1109/CVPR46437.2021.01350

[20] Dai, Y., Gieseke, F., Oehmcke, S., Wu, Y. and Barnard, K. (2021) Attentional Feature Fusion. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, 3-8 January 2021, 3559-3568. https://doi.org/10.1109/WACV48630.2021.00360

[21] Glorot, X., Bordes, A. and Bengio, Y. (2011) Deep Sparse Rectifier Neural Networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, 11-13 April 2011, 315-323.

[22] Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Adam, H., *et al*. (2019) Searching for MobileNetV3. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 1314-1324. https://doi.org/10.1109/ICCV.2019.00140

[23] Zafer, C. (2020) Fusing Fine-Tuned Deep Features for Recognizing Different Tympanic Membranes. *Biocybernetics and Biomedical Engineering*, **40**, 40-51. https://doi.org/10.1016/j.bbe.2019.11.001

[24] Viscaino, M., Maass, J.C., Delano, P.H., Torrente, M., Stott, C. and AuatCheein, F. (2020) Computer-Aided Diagnosis of External and Middle Ear Conditions: A Machine Learning Approach. *PLOS ONE*, **15**, e0229226. https://doi.org/10.1371/journal.pone.0229226

[25] Doyle, S., Hwang, M., Shah, K., Madabhushi, A., Feldman, M. and Tomaszeweski, J. (2007) Automated Grading of Prostate Cancer Using Architectural and Textural Image Features. 2007 4*th IEEE International Symposium on Biomedical Imaging*: *From Nano to Macro*, Arlington, 12-15 April 2007, 1284-1287. https://doi.org/10.1109/ISBI.2007.357094