# Improved MobileNet's Pneumonia X-Ray Image Lightweight Detection Model

HOU Zhongyuan[1*], HUANG Yuhan[2]
(1.Educational Technology and Information Center, Shangqiu Medical College, Shangqiu Henan 476000, China;
2.School of Medical Technolog, Shangqiu Medical College, Shangqiu Henan 476000, China)

**Abstract**: In response to the demand for rapid detection of pneumonia medical images, lightweight pneumonia image detection network based on MobileNetV3 is improved to ensure model accuracy and improve model inference speed. The algorithm is designed with three different modules: feature learning, feature fusion, and prediction. Firstly, feature learning module introduces depthwise separable convolution to reduce model complexity and obtain more useful feature information from the original features with minimal inference cost. Secondly, conflicts between multi-level feature maps are resolved through feature fusion to achieve detection of pneumonia lesion areas. To improve the convergence speed of training, an adaptive Bayesian HyperBand hyperparameter optimization algorithm is proposed, which automatically adjusts parameters based on the search state to improve search efficiency and convergence speed. At the same time, by actively learning data optimization strategies, data quality and model performance can be improved. Experimental verification shows that the accuracy and F-score of the improved model are 98.57% and 0.968, respectively, with a relatively small number of model parameters and an FPS of 85.3. The improved lightweight pneumonia detection model achieves a balance between detection accuracy and lightweight, making it suitable for deployment in resource constrained environments.

**Key words**: pneumonia detection; medical image processing; lightweight; real-time; deep convolutional neural networks; MobileNet
**EEACC**: 7510 **doi**: 10.3969/j.issn.1004-1699.2026.09.021

# 改进 MobileNet 的肺炎 X 射线影像轻量化检测模型

侯中原[1*], 黄宇涵[2]
(1.商丘医学高等专科学校教育技术与信息中心, 河南 商丘 476000;
2.商丘医学高等专科学校医学技术学院, 河南 商丘 476000)

**摘 要**: 针对肺炎医学影像快速检测的需求, 改进了 MobileNetV3 的轻量化检测网络, 保证模型精度不降且提高模型推理速度。算法设计了特征学习、特征融合、预测三个不同模块, 首先特征学习引入深度可分离卷积降低模型复杂度, 以很小推理代价从原始特征得到更多有用特征信息; 其次通过特征融合解决多层次特征映射之间的冲突, 实现对肺炎病灶区域检测。为提升训练收敛速度, 提出了自适应贝叶斯 HyperBand 超参数优化算法, 根据搜索状态自动调整参数, 提高搜索效率和收敛速度; 同时通过主动学习的数据优化策略, 提高数据质量和模型性能。实验验证, 改进模型准确率与 F-score 分别为 98.57%和 0.968, 且模型参数量较少, FPS 达到 85.3。改进模型在检测精度和轻量化上达到均衡, 适合在资源受限环境下的部署应用。

**关键词**: 肺炎检测; 医学图像处理; 轻量化; 实时性; 浓度卷积神经网络; MobileNet
**中图分类号**: TP181; R563 **文献标识码**: A **文章编号**: 1004-1699(2026)09-2091-09

肺炎是可由多种细菌、病毒和真菌引起的常见呼吸道感染病症, 已在全球造成了巨大的健康负担。根据世界卫生组织的统计[1-2], 2018 年全球约有 140 万儿童因肺炎而丧生, 2019 年冠状病毒病(COVID-19)席卷全球, 夺去了数千万人的生命, 引起了全球广泛关注[3-4]。胸部 X 线成像作为诊断肺炎的常用方法, 能够看到肺部病变的特征, 如肺实变和密度增高影等。然而解读胸部 X 射线图像需要放射科医生长久的经验积累, 因此如何快速准确地诊断肺炎至关重要。

近年来, 基于深度学习的肺炎检测技术受到了广泛关注。研究人员开始将卷积神经网络(Convolutional Neural Network, CNN)应用于医学图像处理领域, 利用深度学习算法提高了肺炎的诊断效率和准确性[5-8]。Karaddi 等[9]利用 CNN 将肺炎疾病分为 5 类, 使用多个预训练模型进行训练, 最后选择性能最好的模型进行分类。Singh 等[10]提出了基于注意力感知的卷积神经网络实现对肺炎医学图像的分类检测。但是这些网络架构并没有考虑模型的轻量化需求。Yi 等[11]提出了一种可扩展和可解释的深

**收稿日期**: 2024-05-22 **修改日期**: 2024-09-05

度卷积神经网络,用于胸部 X 射线图像肺炎识别,模型从图像中提取有用的特征,然后分为正常和肺炎两类。Wu 等[12]引入 Res2Net 得到肺炎的多尺度特征,然后提出模糊非最大抑制算法,通过融合重叠的检测框得到一个更鲁棒的预测框,提高精度。Erdem 等[13]提出了多深度学习的方法,使用不同卷积神经网络模型进行训练,利用数据增强和迁移学习等手段对模型进行微调。Sharma 等[14]基于 VGG16 设计了肺炎检测模型,探索了融合 VGG16 网络和其他机器学习算法对肺炎检测性能的影响。Athar 等[15]对 AlexNet 进行预训练,采用迁移学习和对抗性训练来提高胸部 X 光片中肺炎识别的准确性。Wu 等[16]提出了一种基于肺炎检测的无锚定目标检测框架,采用数据增强方案对胸部 X 射线图像进行预处理,然后采用无锚定目标检测框架进行肺炎检测。Singh 等[17]基于 ViT 提出了一个肺炎检测模型,该模型融合自注意机制和 Transformer 结构,在自建数据集上达到了一个很好的检测效果。但是肺炎影像的差异性使得检测难度增大,同时大部分深度学习检测模型需要庞大的计算资源和内存,难以在资源受限的设备上进行实时检测。因此设计轻量化肺炎检测模型可以降低复杂度,减少资源使用率。影响深度学习模型训练效果的重要步骤是超参数调优,选择正确的超参数将有助于用更少训练时间达到更高精度。近年来超参数优化算法也有很多研究[18-20],其中,网格搜索[21]被广泛使用,然而该方法仅适用于超参数数量较少的模型优化。随机搜索算法(Random Search,RS)在超参数的指定范围内随机选择[22],以确保了对搜索空间的公平探索。这种预算策略可以有效地用于配置生成。除此之外 Hyperband 算法通过连续减半将其预算均匀分配给所有选定的配置[23],设定有限的预算,从初始化配置开始,使用最大预算运行,选择性能更好的配置结束,但是对参数依赖性较大。贝叶斯 Hyperband 优化(Bayesian Optimization Hyperband,BOHB)是一种将 Bayesian 和 Hyperband 结合的方法[24]。BOHB 取代了 Hyperband 使用的随机搜索,通过贝叶斯优化从搜索空间中选择配置,但是 BOHB 算法相对复杂,计算开销大且对资源利用效率欠佳,因此将设计新的超参数调优方法,提高搜索效率和收敛速度。

所以本研究提出了改进轻量化肺炎检测模型,基于 MobileNet 模型,设计了参数少运行速度快的轻量化特征提取模块,以很小的推理代价从原始特征中发掘到更多有用的信息;通过设计的特征融合模块来解决颈部多层次特征映射之间的冲突,并执行所有层次的特征的融合。为了提升训练过程中模型收敛速度,提出了 ABHB 超参数优化算法,通过自适应调整算法参数的方法,根据搜索进展和超参数空间的特点动态调整算法参数,提高算法的收敛速度和性能。同时为了提升数据质量,将主动学习和数据增强相结合,以减少标记数据的需求,并提高目标检测模型的性能。

## 1 改进的肺炎检测模型

提出的改进肺炎检测模型结构如图 1 所示,该模型包含三个关键组成模块,分别为特征学习、特征

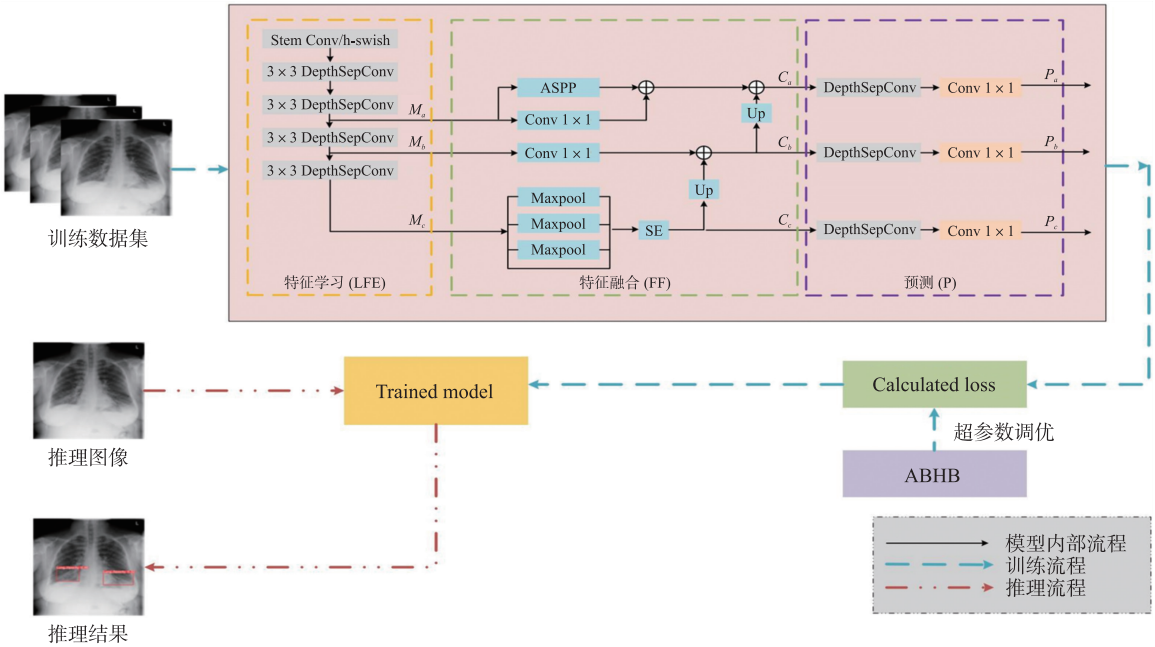



**图 1 改进肺炎检测模型结构**

融合、预测,共同构建了一个高效且准确的肺炎检测系统。当影像输入到检测模型中,特征学习模块 LFE 对 X 光影像进行学习,通过特征学习模块学习到浅层纹理信息和深层的语义信息。选择三个不同特征层 $\boldsymbol{M}_i(i=a,b,c)$ 作为输入,输入到特征融合模块 FF 中,对学习到的不同层次信息进行融合,将非邻近层的信息进行交互。特征融合模块可以扩大感受野增加语义信息,同时又能很好地保留浅层纹理信息。该模块也将产生三个输出 $P_i(i=a,b,c)$,输入到预测层,预测层从大、中、小三个不同维度进行胃炎病灶检测,实现对肺炎病变的快速识别和定位。

### 1.1　特征学习模块 LFE

基于 MobileNetv3 的深度可分离卷积和逐点卷积[25],设计了轻量化特征提取模块(Lightweights Feature Extraction,LFE),特征学习模块 LFE 的结构如图 2 所示。通过简单操作生成更多的特征图,应用一系列线性变化,以很小的推理代价从原始特征中发掘到更多有用的信息,包含丰富语义信息的同时又不丢失纹理信息,进而搭建更轻量级的模型。

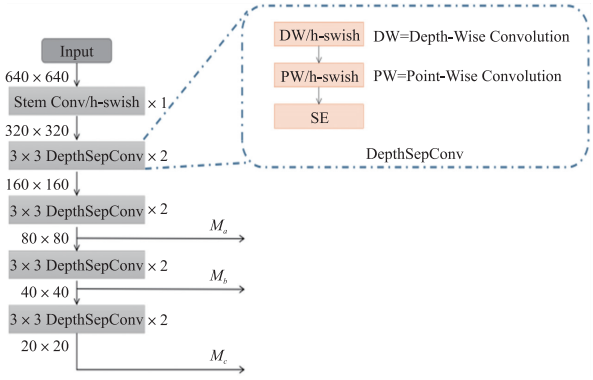



**图 2　LFE 模块网络结构**

当图片输入到网络中,连续经过 5 次的特征图变化,特征图由 640×640 变成 20×20,然后分别使得 $\boldsymbol{M}_a$、$\boldsymbol{M}_b$、$\boldsymbol{M}_c$ 三层特征图作为输入传输到下一个模块中。在 LFE 模块中,通过深度可分离卷积模块(Depth-Wise Separable Convolutions,DepthSepConv)实现对特征的学习。DepthSepConv 模块包含一个深度卷积(Depth-Wise Convolution,DW)模块和一个点卷积(Point-Wise Convolution,PW)模块,此处 DW 和 PW 的作用是为了减少参数量,降低计算成本。

在 DepthSepConv 模块中,引入了注意力模块(Squeeze Excitation,SE)[26],SENet 模块主要关注主干网络中卷积操作的信道间的特征融合,可以提高检测精度,但是相应也会增加一些计算量。该模块通过关注信道之间的关系,自动学习不同信道特征的重要性。通过压缩(Squeeze)和激励(Excitation)操作,扩大对特征的学习能力。在 DepthSepConv 模块中,采用全局平均池化,将信道上的整个空间特征编码为局部特征,提供特征转换效率,在减少计算量的同时,保证精度不会降低很多。将 SE 模块置于末端,自动获取到每个特征通道的重要程度,依照重要程度去提升有用的特征并抑制对当前任务用处不大的特征。

本模块中激活函数选用 H-swish 函数[27],采用该函数是为了提高模型性能而不影响推理时间,其公式为:

$$\text{h-swish}[x]=x\frac{\text{ReLU6}(x+3)}{6} \tag{1}$$

### 1.2　特征融合模块 FF

在肺炎图像检测中,小目标问题是一个常见的挑战。小目标通常指图像中占比较小的病灶或病变区域,这些区域在图像中的像素数较少,但它们的检测对于早期诊断和治疗至关重要。由于小目标在大尺度特征图中只占很小的比例,而特征图的缩放或降维可能会导致小目标的信息丢失。当主干网络的输出特征出现降维,则不可避免地导致微小目标的信息丢失。为了解决这些问题,提出了一个轻量级的特征融合(Feature Fusion,FF)模块来解决颈部多层次特征映射之间的冲突,并执行所有层次的特征的融合。如图 3 所示。

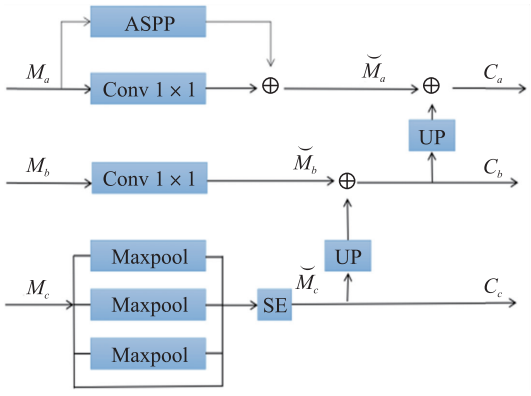



**图 3　FF 模块网络结构**

$\boldsymbol{M}_a$、$\boldsymbol{M}_b$、$\boldsymbol{M}_c$ 分别代表不同层级的特征,感受野也不同,有基于浅层的纹理信息,有基于深层的语义信息,具有不同感受域的特征的直接叠加可能会引起一些混淆,导致精度下降,因此本文采取级联操作,对不同层级的特征进行融合。

$\boldsymbol{M}_c$ 特征层经过多级池化操作,进行多重感受野的融合,提升模型对空间布局和物体变性的鲁棒性。然后经过 SE 模块提升对重要语义信息的关注度,在医学图像上的映射使模型侧重于对病灶区域的提

取表达,然后得到 $\hat{\boldsymbol{M}}_c$ 特征图,即为 $\boldsymbol{C}_c$ 特征图,然后将 $\hat{\boldsymbol{M}}_c$ 进行上采样;$\boldsymbol{M}_b$ 经过卷积模块得到 $\hat{\boldsymbol{M}}_b$,与 $\hat{\boldsymbol{M}}_c$ 经过上采样后的特征图进行特征拼接,输出 $\boldsymbol{C}_b$ 特征图;$\boldsymbol{M}_a$ 特征层经过一个 ASPP 模块和卷积模块处理,采取 Resnet 方式连接,提取语义信息的同时保留纹理信息,得到 $\hat{\boldsymbol{M}}_a$,与 $\boldsymbol{M}_c$ 等高层特征提取方式不同,对 $\boldsymbol{M}_a$ 特征的提取侧重于从低级特征中生成多尺度特征来目标小目标的信息,浅层 $\boldsymbol{M}_a$ 的特征图保留了病灶的精细边界细节,加强对小目标病灶的特征提取。与经过上采样处理后的 $\boldsymbol{M}_b$ 进行拼接,最后输出 $\boldsymbol{C}_a$。$\boldsymbol{C}_a$、$\boldsymbol{C}_b$、$\boldsymbol{C}_c$ 为特征融合阶段的三个输出。

### 1.3 预测模块

在预测阶段,引入深度可分离卷积来减少参数的数量,对三路特征图分别采取串联操作,这样可以用更少的参数和更深的网络深度实现感受野的扩展。最后通过一个过渡层(1×1 卷积)进一步集成特征。结构如图 4 所示。

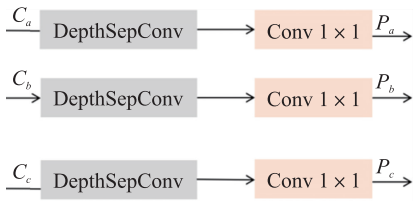



**图 4 预测模块网络结构**

### 1.4 损失函数

改进损失函数由三部分组成:置信度损失 $l_{\text{obj}}$、分类损失 $l_{\text{cls}}$、位置损失 $l_{\text{box}}$。公式如下:

$$\text{Loss}=l_{\text{obj}}+l_{\text{cls}}+l_{\text{box}} \tag{2}$$

置信度损失 $l_{\text{obj}}$ 为:

$$l_{\text{obj}}=\sum_{i=0}^{s^2}\sum_{j=0}^{B}l_{ij}^{\text{obj}}\left[\hat{C}_i\log(C_i)+(1-\hat{C}_i)\log(1-\hat{C}_i)\right]-\lambda_{\text{noobj}}\sum_{i=0}^{s^2}\sum_{j=0}^{B}l_{ij}^{\text{noobj}}\left[\hat{C}_i\log(C_i)+(1-\hat{C}_i)\log(1-\hat{C}_i)\right] \tag{3}$$

置信度损失用于衡量模型对目标是否存在的预测准确性,使用交叉熵损失函数来计算。其中 $C$ 为置信度分数,其中 $l^{\text{obj}}$、$l^{\text{noobj}}$ 分别代表存在目标、不存在目标的情况。

分类损失 $l_{\text{cls}}$ 为:

$$l_{\text{cls}}=\sum_{i=0}^{s^2}l_{ij}^{\text{obj}}\sum_{c\in\text{classes}}\{\hat{P}_i(c)\log[P_i(c)]+[1-\hat{P}_i(c)]\log[1-P_i(c)]\} \tag{4}$$

分类损失用于衡量模型对目标进行正确分类的能力。其中 $c$ 表示目标类别,$P(c)$ 表示对目标的预测值。在位置损失 $l_{\text{box}}$ 设置中,选择 CIoU 来衡量位置损失:

$$l_{\text{box}}=L_{\text{CIoU}}=I\text{-IoU}+\frac{\rho^2(b,b^{\text{gt}})}{c^2}+\alpha v \tag{5}$$

$$v=\frac{4}{\pi^2}\left(\arctan\frac{\omega^{\text{gt}}}{h^{\text{gt}}}-\arctan\frac{\omega}{h}\right)^2$$

$$\alpha=\frac{v}{(1-\text{IoU})+v} \tag{6}$$

式中:$b$ 和 $b^{\text{gt}}$ 表示预测框和目标框的中心的,$\rho$ 表示欧氏距离,$c$ 为两个检测框最小外接矩形框的对角线长度。$\alpha$ 是一个正的权衡参数,$v$ 衡量长宽比的一致性。与 GIoU 相比,CIoU 考虑了边界框纵横比的尺度信息,从重叠面积、中心点距离和纵横比三个角度测量,使预测框回归效果更好。

### 1.5 超参数调优算法

针对 BOHB 计算复杂开销大的问题,设计了自适应贝叶斯 Hyperband 优化(Adaptive Bayesian Hyperband,ABHB)算法,该算法引入自适应采样策略,根据超参数空间的特点和搜索进展动态调整采样策略;同时基于验证损失的阈值选择连续减半过程中进入下一轮迭代的配置数量。该阈值动态更新,通过自适应变化,以从搜索空间中选择更多潜在的最佳配置。通过这种自适应调整算法参数的方法,根据搜索进展和超参数空间的特点动态调整算法参数,以提高算法的适应性和性能。算法的伪代码如下。

Input:$R,M_0,H,S,T,\eta$;

Output:Optinal hyperparameter configuration

初始化 HT(不为空);设置 $s_{\max}=\log_\eta(R)$

for $s\in\{s_{\max},s_{\max-1},\cdots,0\}$:

if $s=s_{\max}$:

确定 HP(从 $H$ 中随机采样 $n$ 个样本)

else:

确定 HP(获得超参数配置):get_hyperparameter_configuration($n$)

end if:

$n=\lceil\frac{B\eta^s}{R(s+1)}\rceil,R\eta^{-s}$

for $i\in1,2\cdots s$:

If $i=0$:

$n_i=\lfloor n\eta^{-i}\rfloor$ *and* $r_i=r\eta^i$

else:

$n_i=\text{count(HP)}$ and $r_i=\left|\frac{R}{n_i}\right|$

end if:

for $j$ to $r_i$

$h^*=\text{argmax}_h S(h,M_{j-1})$

评估:$f(h^*)$

更新 HT:$(h^*, f(h^*)_{k+1})$
选择新的 $M$
$L$=run_then_retun_val_loss$(h, r_i)$:$h \in$ HP
确定新的 HP
for $h \in$ HP
if $L[h, r_i, l_i]$<set_threshold$(T)$
将 $h$ 添加到更新后的 HP 中
输出最新配置

其中,$r$ 为单个超参数组合实际所能分配的预算;$R$:单个超参数组合所能分配的最大预算;$M_0$ 为使用默认超参数构建的初始化模型;$H$ 为超参数搜索空间;$T$ 为验证损失的最小阈值初始化;$s_{\max}$用来控制总预算的大小;$\eta$ 用于控制每次迭代后淘汰参数设置的比例;get_hyperparameter_configuration$(n)$采样得到 $n$ 组不同的超参数设置;run_then_retun_val_loss$(h, r_i)$表示根据指定的参数设置和预算计算 valid loss;$L$ 表示在预算为 $r_i$ 的情况下各个超参数设置的验证误差。

本文提出的算法对最小阈值 $T$ 进行了改进,每次确定阈值 $T$ 时,该值是上一次迭代阈值 $T_{i-1}$和本次迭代中使得损失最小的阈值 $T_{\text{small}}$的平均值。公式如下:

$$T_i = (T_{i-1} + T_{\text{small}})/2 \tag{7}$$

通过这种策略,可以保证每次迭代过程可以实现阈值的自适应动态变化,算法可以更快地向搜索空间中的最佳区域移动。

## 2　基于主动学习的数据优化策略模型训练

为了提升数据集质量,将主动学习和数据增强相结合,减少标记数据的需求,提升数据集质量,进而提高目标检测模型的性能。图 5 所示为引入主动学习对数据进行优化的流程,使用数据增强方法对原始数据进行扩充,再通过主动学习算法来选择最具信息量的样本进行标记,通过迭代训练,选择高质量样本投入训练,从而提高模型性能。

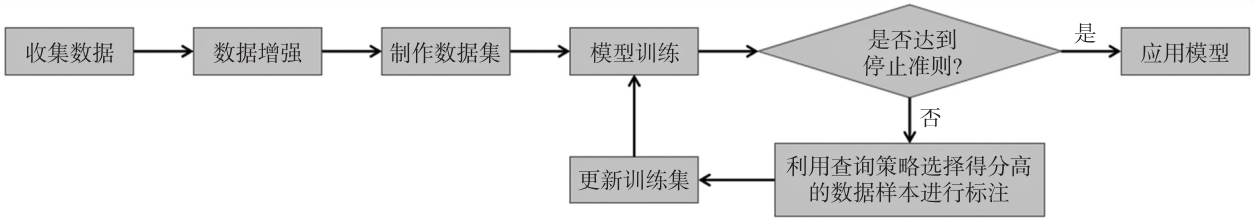



**图 5　基于主动学习的数据优化策略**

基于主动学习的数据优化策略定义为 $R=(D, L, P, Q, U)$,其中 $D$ 为检测器(即目标检测模型),$L$ 为已标注样本,$P$ 为监督者,$Q$ 为查询函数,$U$ 为未标记样本,首先在未标记的数据集中选取少量样本进行标注,然后输入到提出模型中进行训练,使用训练好的模型对未标注样本进行预测评估,通过最小置信度查询策略,对于每个数据点选择概率的最大值从小到大进行排序,根据查询策略得出未标记样本的得分情况,根据得分选择需要人工标记的数据。将重新标记的数据加入到训练集中重新训练,不断优化迭代新模型。当迭代条件达到预先设定条件后(本实验中设置迭代次数为 5),停止训练,得到最新模型。

$$x_{\text{LC}}^* = \text{argmax}_x [1 - P_\theta(\hat{y}|x)] = \text{argmin}_x P_\theta(\hat{y}|x) \tag{8}$$

$$\hat{y} = \text{argmax}_y P_\theta(y|x) \tag{9}$$

式中:$\hat{y}$ 是模型预测最大概率的类别,$P_\theta(y|x)$是最大概率的类别对应的概率。

该方法具有较强的适应性,能够根据任务需求动态地选择标记样本和生成扩充数据,使模型更适应不同的数据场景。并且通过增加数据集的多样性提高数据质量,有助于降低过拟合风险,控制模型的泛化能力,有效改进了目标检测效果。

## 3　实验

### 3.1　数据集

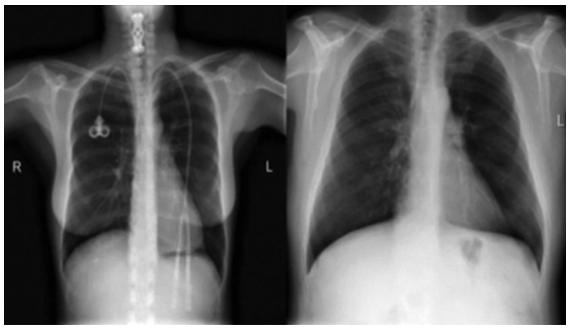
RSNA Pneumonia Detection Challenge 数据集是由 RSNA(Radiological Society of North America)提供的用于肺炎检测挑战的数据集,数据集包含数千张胸部 X 光图像(CXR),这些图像来自不同的患者,包括正常的胸部 X 光片,如图 6(a)所示,和患有肺炎的 X 光片,如图 6(b)所示。每张 X 光图像都有相应的标注,标注信息包括是否存在肺炎病灶,以及病灶的位置。

### 3.2　实验设置

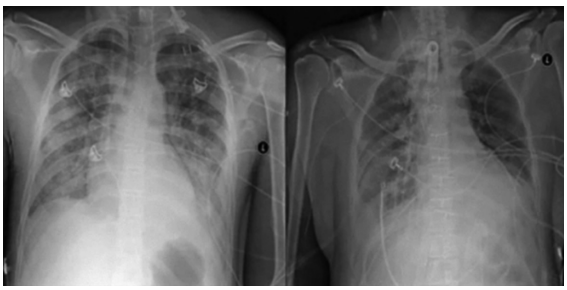
该实验的环境设置均一致,GPU 为 NVIDIA GeForce GTX 2080-SUPPER,12 GB 的显存;操作系统为 Windows;CUDA 版本是 v9.0,cuDNN 版本是 v7.6.5;PyTorch 版本为 1.8。

### 3.3　评价指标

本文采用的算法评价指标选用平均精确均值(mean Average Precision,mAP)和 $F_1$-score。mAP 和 $F_1$-score 由精度(Precision)或查全率(Recall)计算



(a) 正常胸部X光片



(b) 患有肺炎的X光片

**图 6　数据集示例**

得出,其公式如下。

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{10}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{11}$$

$$\text{AP}=\int_0^1 p(r)\,\mathrm{d}r \tag{12}$$

$$\text{mAP}=\frac{\sum_{n=1}^{N}\text{AP}(n)}{N} \tag{13}$$

$$F_1\text{-Score}=(1+\beta^2)\times\frac{\text{Precision}\times\text{Recall}}{\beta^2\times\text{Precision}+\text{Recall}} \tag{14}$$

式中:TP 为预测结果为正样本,实际也为正样本;FP 为预测结果为正样本,实际为负样本;TN 为预测结果为负样本,实际为负样本;FN 为预测结果为负样本,实际为正样本。

### 3.4　实验结果与分析

为了验证提出的 ABHB 超参数调优算法性能,和其他超参数调优方法进行对比实验,优化的超参数如表 1 所示。针对目标检测模型中超参数调优方法的比较实验,结果如图 7 所示,ABHB 表现出了最佳的性能,其收敛速度较快且达到了最优结果。ABHB 在搜索过程中采用了更有效的策略,更适合医学图像目标检测任务。相比之下,RS 优化方法和 BO 优化方法虽然能够找到较好的超参数组合,但需要更多的迭代次数才能收敛到最优结果。Hyperband 和 BOHB 在一定程度上提高了搜索效率,但仍然不及 ABHB。这些结果提示了在选择超参数调优方法时,应当综合考虑搜索效率和性能表现,以及特定任务的需求。

**表 1　调优超参数列表**

| 超参数 | 超参数详解 |
|---|---|
| lr0 | 初始学习率 |
| momentum | 动量 |
| weight_decay | 优化器权重衰减 |
| warmup_epochs | 预热 epoch |
| warmup_momentum | 预热动量 |
| warmup_bias_lr | 预热初始偏置 |
| Dropout | Dropout 概率 |
| hsv_h | 图像 HSV-Hue 增强 |
| hsv_s | 图像 HSV-Saturation 增强 |
| hsv_v | 图像 HSV-Value 增强 |
| Degrees | 图像旋转度 |
| Scale | 图像尺寸 |
| Shear | 图像拼接 |
| Flipud | 图像上下翻转 |
| Fliplr | 图像左右翻转 |
| Mixup | 图像混合 |

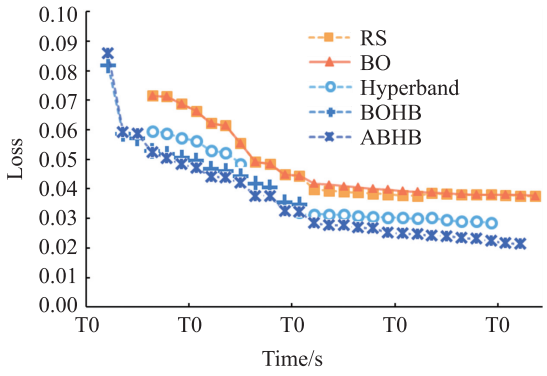



**图 7　超参数调优方法对比**

为了验证本文提出的算法性能,本文先对不同部分进行消融实验,如图 8 所示,对四种不同模式进行实验,图中 A 表示基础模型(RtDPD)训练;B 表示基础模型+融合主动学习的数据优化;C 表示基础模型+改进超参数调优模式;D 表示基础模型+改进超参数调优 ABHB+融合主动学习的数据优化。

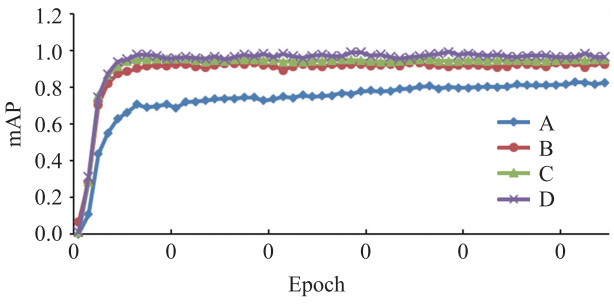



**图 8　不同模块性能对比**

可以看出,只使用基础模型进行病灶检测时,虽然精度也能达到较高水平,但是收敛速度慢,收敛过程不稳定。当对数据增强方式和超参数调优方式进行改进时,训练过程中收敛速度更快,且实验精度更高。当使用提出的检测模型 RtDPD 时(融合主动学习策略和超参数调优算法 ABHB)性能达到最佳,AP 值达到了 98.57%。该模型检测效果如图 9 所示,从图中可以看出,可以对病灶区域进行了精准检测。对于病灶区域明显的图像,如图 9(a2)、(a3)、(c1)等,模型可以有效检测。对病灶区域较小的图像,如图 9(a1)、(b3)、(d3)等,也可以有效检测,说明本文提出的模型具有较好的检测性能。

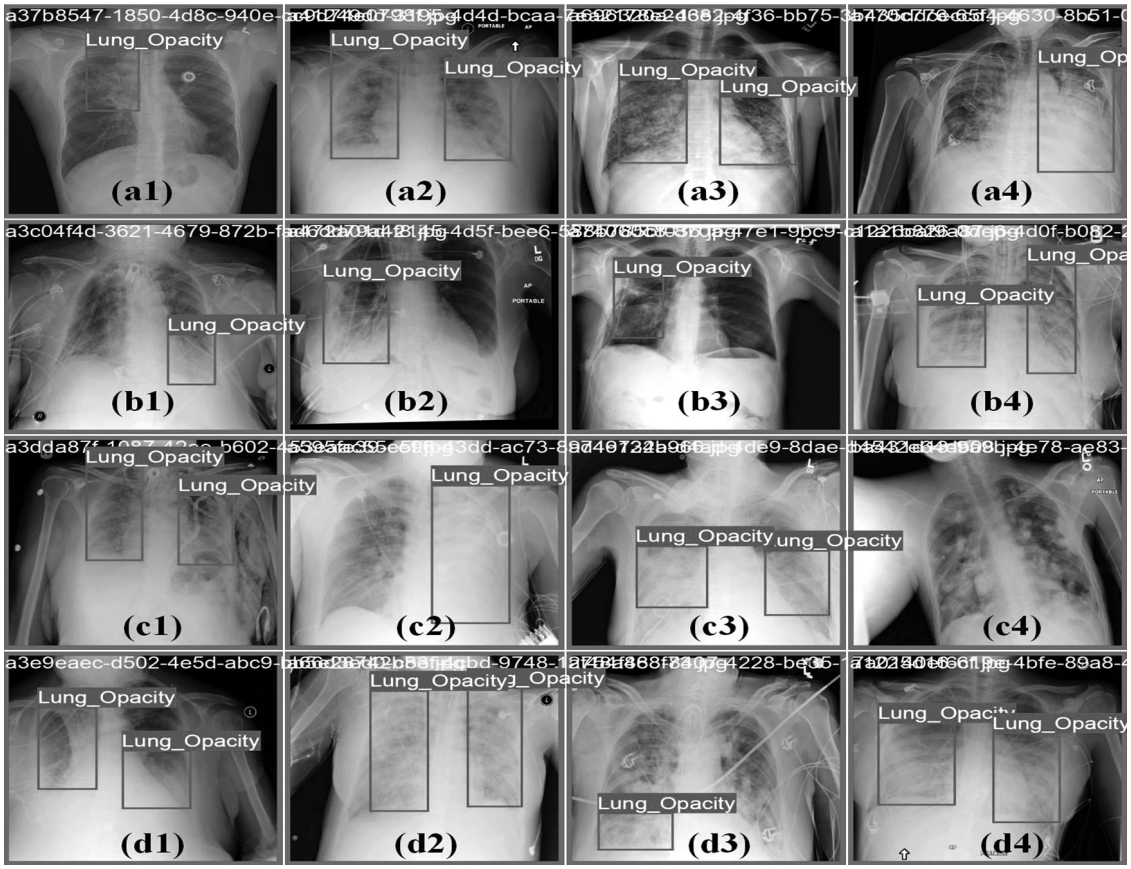



**图 9　模型检测效果图**

针对肺炎的轻量化医学目标检测问题,评估了多种经典和新型的基于深度学习,肺炎检测模型的性能,如“Quaternion Residual Network”[28]、“Vision Transformer”[29]。实验结果如表 2 所示,提出的模型在 mAP 和 $F_1$-score 上均达到了 98.57% 和 0.968 的高水平,且参数数量为 175005,FPS 较 MobileNetV3 稍差,但优于其他模型,MobileNetV3 在 mAP 指标上表现不佳,说明本模型在精度和轻量化设计上达到平衡。Vision Transformer 模型虽然准确率最高(97.61%),但参数数量达到了 85 800 194,较高的参数数量可能限制了其在资源受限环境下的实际应用。而其他模型如 VGG、ResNet、Inception、YOLO、MobileNet 等在准确率和 F-score 上表现良好,但参数数量相对较大,不利于轻量化部署。综上所述,本文提出的目标检测模型在肺炎检测任务上取得了显著的性能优势,具有较高的精度和 F-score,并且参数数量较少,适合在资源受限环境下部署和应用。

**表 2　不同方法对比实验**

| Model | mAP | $F_1$-score | parameters | FPS |
|---|---|---|---|---|
| VGG16 | 92.14 | 0.923 | 50 178 | 59.9 |
| VGG19 | 90.22 | 0.901 | 50 178 | 60.3 |
| ResNet50 | 82.37 | 0.828 | 215 008 | 51.8 |
| ResNet101 | 75.96 | 0.761 | 215 008 | 51.3 |
| ResNet152 | 87.19 | 0.871 | 215 008 | 52.5 |
| Inception | 90.71 | 0.905 | 102 402 | 60.3 |

续表 1 CEC 测试函数基本信息

| Model | mAP | $F_1$-score | parameters | FPS |
|---|---|---|---|---|
| YOLOv4 | 89.26 | 0.894 | 200 706 | 66.2 |
| YOLOv5 | 92.62 | 0.925 | 180 200 | 79.3 |
| GhostNet | 92.94 | 0.931 | 750 325 | 68.5 |
| InceptionResNet | 90.70 | 0.902 | 200 905 | 77.6 |
| MobileNetV1 | 91.82 | 0.917 | 100 354 | 85.9 |
| MobileNetV2 | 88.78 | 0.888 | 163 074 | 84.1 |
| MobileNetV3 | 91.83 | 0.918 | 188 162 | 90.9 |
| Quaternion Residual Network | 93.75 | 0.941 | 560 769 | 50.2 |
| Vision Transformer | 97.61 | 0.953 | 85 800 194 | 21.4 |
| **Ours** | **98.57** | **0.968** | **175 005** | **85.3** |

## 4 总结

针对目标检测模型参数量多、计算复杂的问题，提出改进 MobileNetV3 的轻量化肺炎检测模型。该模型通过设计轻量化特征提取模块 LFE、特征融合模块 FF 和预测模块，对病灶特征进行学习和检测，在保证准确性的同时，以 175 005 的参数量达到了 85.3 FPS 的推理速度，实现模型轻量化。同时设计了超参数优化算法 ABHB 以及基于主动学习的数据优化方法，减少模型收敛时间，提升训练效率。实验结果表明，模型在 mAP 和 F-score 上均表现出显著优势，在参数量和推理速度上具有明显优势，适合在资源受限环境下的部署和应用。

本研究突出了在医学图像处理领域实现轻量化模型设计的重要性，并通过实验验证了模型的有效性和可行性。未来的工作方向会进一步优化模型结构以提高性能，并探索在其他医学图像识别任务中的应用潜力。

**参考文献：**


[1] gonzalez-Peris S, Campins M, Garcia-Garcia J J, et al. Necrotizing Pneumonia Due to Streptococcus Pneumoniae in Children During the Period of Non-Systematic Use of PCV13 in Catalonia, Spain[J]. Enfermedades Infecciosas y Microbiología Clínica, 2021, 39(10): 486-492.

[2] Papazian L, Klompas M, Luyt C E. Ventilator-Associated Pneumonia in Adults: A Narrative Review[J]. Intensive Care Medicine, 2020, 46(5): 888-906.

[3] Duployez C, Le Guern R, Tinez C, et al. Panton-Valentine Leukocidin-Secreting Staphylococcus Aureus Pneumonia Complicating COVID-19 [J]. Emerging Infectious Diseases, 2020, 26(8): 1939.

[4] Fu L, Wang B, Yuan T, et al. Clinical Characteristics of Coronavirus Disease 2019 (COVID-19) in China: A Systematic Review and Meta-Analysis[J]. Journal of Infection, 2020, 80(6): 656-665.

[5] Yi R, Tang L, Tian Y, et al. Identification and Classification of Pneumonia Disease Using a Deep Learning-Based Intelligent Computational Framework[J]. Neural Computing and Applications, 2023, 35(20): 14473-14486.

[6] Bhatt H, Shah M. A Convolutional Neural Network Ensemble Model for Pneumonia Detection Using Chest X-Ray Images[J]. Healthcare Analytics, 2023, 3: 100176.

[7] Celik G. Detection of Covid-19 and Other Pneumonia Cases from CT and X-Ray Chest Images Using Deep Learning Based on Feature Reuse Residual Block and Depthwise Dilated Convolutions Neural Network[J]. Applied Soft Computing, 2023, 133: 109906.

[8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[9] Karaddi S H, Sharma L D. Automated Multi-Class Classification of Lung Diseases from CXR-Images Using Pre-Trained Convolutional Neural Networks [J]. Expert Systems with Applications, 2023, 211: 118650.

[10] Singh S, Rawat S S, Gupta M, et al. Deep Attention Network for Pneumonia Detection Using Chest X-Ray Images[J]. Computs, Materials and Continna, , 2023, 74: 1673-1691.

[11] Yi R, Tang L, Tian Y, et al. Identification and Classification of Pneumonia Disease Using a Deep Learning-Based Intelligent Computational Framework[J]. Neural Computing and Applications, 2023, 35(20): 14473-14486.

[12] Wu H L, Lu H J, Ping M Z, et al. A Deep Learning Method for Pneumonia Detection Based on Fuzzy Non-Maximum Suppression[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2024, 21(4): 902-911.

[13] Erdem E, Aydin T. Detection of Pneumonia with a Novel CNN-Based Approach[J]. Sakarya University Journal of Computer and Information Sciences, 2021, 4(1): 26-34.

[14] Sharma S, Guleria K. A Deep Learning Based Model for the Detection of Pneumonia from Chest X-Ray Images U-

sing VGG-16 and Neural Networks [J]. Procedia Computer Science,2023,218:357-366.

[15] Athar A,Asif R N,Saleem M,et al. Improving Pneumonia Detection in Chest X-Rays Using Transfer Learning Approach(AlexNet) and Adversarial Training[C]//2023 International Conference on Business Analytics for Technology and Security(ICBATS),Dubai,United Arab Emirates,IEEE,2023:1-7.

[16] Wu L h,Zhang J,Wang Y L,et al. Pneumonia Detection Based on RSNA Dataset and Anchor-Free Deep Learning Detector[J]. Scientific Reports,2024,14(1):1929.

[17] Singh S,Kumar M,Kumar A,et al. Efficient Pneumonia Detection Using Vision Transformers on Chest X-Rays [J]. Scientific Reports,2024,14(1):2487.

[18] Bergstra J,Bengio Y. Random Search for Hyper-Parameter Optimization[J]. Journal of Machine Learning Research,2012,13(2):281-305.

[19] Hutter F,Hoos H H,Leyton-Brown K. Sequential Model-Based Optimization for General Algorithm Configuration [C]//5th International Conference on Learning and Intelligent Optimization,Rome,Italy,2011:507-523.

[20] Snoek J,Larochelle H,Adams R P. Practical Bayesian Optimization of Machine Learning Algorithms[C]//26th International Conference on Neural Information Processing Systems,Lake Tahoe,NV,USA,2012:2951-2959.

[21] Bergstra J,Yamins D,Cox D. Making a Science of Model Search:Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures[C]//30th International Conference on Machine Learning, Atlanta, GA, USA, 2013:115-123.

[22] Bergstra J,Bardenet R,Bengio Y,et al. Algorithms for Hyper-Parameter Optimization [C]//25th International Conference on Neural Information Processing Systems, Granoda,Spain,201125462554.

[23] Li L,Jamieson K,DeSalvo G,et al. Hyperband:A Novel Bandit-Based Approach to Hyperparameter Optimization [J]. Journal of Machine Learning Research,2017,18(1):6765-6816.

[24] Falkner S,Klein A,Hutter F. BOHB:Robust and Efficient Hyperparameter Optimization at Scale[C]//35th International Conference on Machine Learning. PMLR,2018:1437-1446.

[25] Nascimento M G D,Prisacariu V ,Fawcett R. Dsconv:Efficient Convolution Operator[C]//2019 IEEE/CVF International Conference on Computer Vision,Seoul,South Korea,2019:5147-5156.

[26] Hu J,Shen L,Sun G. Squeeze-and-Excitation Networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition,Salt Lake City,UT,USA,2018:7132-7141.

[27] Howard A,Sandler M,Chen B,et al. Searching for MobileNetv3[C]//2019 IEEE/CVF International Conference on Computer Vision,Seoul,South Korea,2019:1314-1324.

[28] Singh S,Tripathi B K,Rawat S S. Deep Quaternion Convolutional Neural Networks for Breast Cancer Classification [J]. Multimedia Tools and Applications,2023,82(20):31285-31308.

[29] Singh S,Kumar M,Kumar A,et al. Efficient Pneumonia Detection Using Vision Transformers on Chest X-Rays [J]. Scientific Reports,2024,14(1):2487.

**侯中原**(1991—),男,硕士研究生,讲师,主要研究方向为计算机视觉与深度学习应用,742862175@qq.com,18203705959;

**黄宇涵**(1989—),女,硕士研究生,讲师,主要研究方向为药物基础,940615710qq@.com。