基于注意力机制的轻量级双卷积手指静脉识别网络

赵冰艳[①] 梁义怀*[①] 张中霞[②] 张文政[②③]

[①](西南交通大学信息科学与技术学院 成都 611756)

[②](中国电子科技网络信息安全有限公司 成都 610041)

[③](保密通信全国重点实验室 成都 610041)

摘 要：基于深度学习的指静脉识别方法已被广泛应用于生物特征识别领域，然而现有模型普遍存在复杂度与分类性能失衡的问题，难以在内存受限和计算资源稀缺环境下高效完成识别任务。针对上述问题，该文提出了一种融合注意力机制的轻量化双通道卷积神经网络模型。该模型设计有双分支协同架构，旨在分别提取核心特征与辅助特征，从而丰富特征集合并增强网络对远程依赖特征的捕捉能力。通过设计一种并行双重注意力机制，以促进融合特征间的信息交互，引导模型聚焦于高价值信息，学习更具区分度的特征表示。实验结果显示，此模型在USM, HKPU和SDUMLA 3个公开数据集上的识别准确率分别达到99.70%, 98.33%和98.27%，比现有先进方法分别提升7.94%, 21.66%和28.62%，而参数量减少11.35%～60.19%，表明提出的双卷积模型实现了网络规模与识别准确率之间的有效平衡。

关键词：指静脉识别；双卷积模型；特征表示；轻量化；并行双重注意力机制

中图分类号：TN02;TP389 **文献标识码**：A **文章编号**：1009-5896(2026)03-1128-11

DOI: 10.11999/JEIT250380 **CSTR**: 32379.14.JEIT250380

1 引言

随着信息技术的飞速发展，个人身份隐私的安全防护已提升至前所未有的重要地位。生物特征识别技术凭借其高效性与安全性，成为身份验证领域的主流方法。该技术体系主要分为两大研究方向。(1)生理特征识别：以人体固有生理构造为识别对象，典型特征包括人脸、指纹、掌纹和虹膜等。(2)行为特征识别：基于个体行为模式进行身份鉴别，常见模态涵盖声纹、步态和签名等。

手指静脉识别作为新一代生理特征识别技术，凭借其活体识别特性、基于内部特征的唯一性与永久性、识别稳定性高及难以伪造等优势[1]，脱颖而出。此外，其还具有采集过程非接触式、设备微型化及支持多指共同识别等独特亮点，正逐步成为智能时代身份认证技术的重要演进方向。其识别流程如图1所示。

当前，深度学习(Deep Learning, DL)技术正处于快速发展阶段，为人们的实际生活带来诸多变革。其中，研究者正积极探索并应用基于卷积神经网络(Convolutional Neural Network, CNN)的前沿技术于指静脉生物识别领域。CNN已被验证能够从感兴趣区域(Region Of Interest, ROI)中高效提取更具表征力的特征[2]，这一发现极大地推动了手指静脉识别领域技术的发展，促使深度学习技术逐渐取代传统算法。近年来，一系列以CNN为基础的研究方法进一步印证其在指静脉识别领域所具有的巨大潜力与价值。文献[3]介绍了一种名为DeepVein的指静脉验证方法，该方法基于深度卷积神经网络(Deep convolutional Neural Network, DNN)架构，能够精确匹配手指静脉模式。文献[4]则通过增加卷积层的数量来强化网络的特征提取能力，但这却增加了网络复杂度。此外，研究者提出了指静脉网络(FV-Net)[5]，以解决静脉模式识别中由于平移和旋转导致的不对齐问题。随后，文献[6]提出了名为FV-DDC的识别算法，该算法不仅有效应对了手指变形带来的挑战，还针对性地解决了数据集规模受限的问题。文献[7]提出了一种改进的合并卷积神经网络(Merge CNN)，该网络采用短路径的CNN结构来高效提取特征，文献[8]提出一种双线性网络(MRFBCNN)来提取手指静脉图像的高阶特征，通过引入维度交互注意力机制，增强了特征图通道与空间之间的相关性，进一步提升了手指静脉识别的准确性。

但是，这些深度学习算法在指静脉识别领域应用时仍面临3大突出问题：首先，手指静脉在管径大小、脉络布局及埋藏深度等方面展现出多样的形态学特征，这些方法未能充分考虑到指静脉独特的生理构造，局限于在同质化范畴内探索特征；其

收稿日期：2025-05-07；改回日期：2026-01-06；网络出版：2026-01-15

*通信作者： 梁义怀 liangyh@swjtu.edu.cn

基金项目：中央高校基本科研业务费(2682024CX031)

Foundation Item: The Fundamental Research Funds for the Central Universities(2682024CX031)

次，对于长度尺度下特征间的内在联系以及图像背景中潜藏的身份信息缺乏足够的关注，难以提取丰富的判别特征，且对细粒度信息的敏感度不足；此外，大多方法因网络参数庞大、训练周期长，在处理手指静脉这类小规模数据集时，常导致过拟合问题。

鉴于当前研究现状，构建一种聚焦于指静脉独特特征模式，且能够有效兼顾识别精度和模型复杂度的卷积神经网络架构至关重要。为此，本文提出了一种新型的基于注意力机制的轻量级双卷积指静脉识别网络(Lightweight Dual-Convolutional finger vein recognition Network based on Attention mechanism, ALDCNet)，主要贡献包括：(1)设计了具有不同形状和尺寸的卷积核，用以提取主要静脉形态特征和背景强度变化信息，实现分类特征有效融合，进而提高特征的判别性；(2)构建了一种轻量级注意力机制，旨在通过极小的额外参数开销，增强网络对静脉血管特征的敏感捕捉能力，并引导其专注于具有高贡献度的关键信息；(3)ALDCNet不仅实现了轻量化，还具备合理的网络深度，这一设计策略有效地降低了过拟合和欠拟合的双重风险。实验结果表明，本研究通过创新性地结合两种卷积网络模式挖掘图像信息，能够高效地捕捉区分度更强的手指静脉特征，在识别任务中表现出优良的性能。

2　方法

2.1　网络结构

本文提出的ALDCNet网络通过系统性地优化网络参数量、适配数据集特性、平衡识别精度-计算效率以及克服单通道CNN的表征局限性，构建了3级递进式特征学习框架。其整体架构如图2所示，该网络主要由双卷积特征提取(Dual-Convolutional Feature Extraction, DCFE)模块、并行双重注意力(Parallel Dual Attention, PDA)模块和特征优化分类(Feature Optimization Classification, FOC)模块构成，形成“特征提取-动态校准-决策融合”的完整流程。

首先，DCFE模块采用双分支轻量化架构，通过矩形卷积分支(DCFE-Rectangular Convolutional branch, DCFE-RC)与方形卷积分支(DCFE-Square Convolutional branch, DCFE-SC)的异构核协同机制，在保持低通道数特征流的前提下，实现多尺度感受野的互补性特征表征。此设计有效缓解了传统单通道CNN在静脉纹理细节捕捉与全局结构建模间的矛盾，通过核形状差异形成特征响应的多样性，不仅压缩了参数规模，同时能够减少原始特征信息损失并增强网络的特征表达能力。

其次，PDA模块通过构建通道注意力子模块(PDA-Channel)和空间注意力子模块(PDA-Spatial)联合机制动态校准通道权重与空间依赖性建模，实现关键特征强化和冗余信息抑制的双重优化，并进一步提升特征层级的判别性表达能力和对静脉微观结构的敏感度。

最后，FOC模块采用3级卷积压缩结构(Conv1, Conv2, Conv3)进行特征降维与语义增强，通过双全连接层(FC1, FC2)实现多层级特征信息的深度融合，最终完成分类决策并输出识别结果。此模块

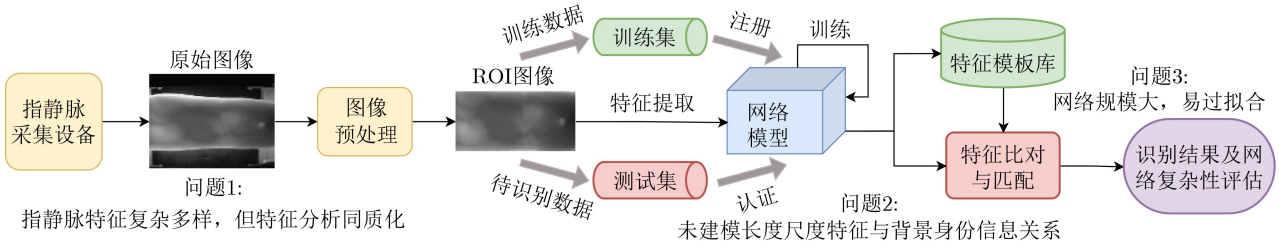


图 1 手指静脉识别流程图和现有方法的问题

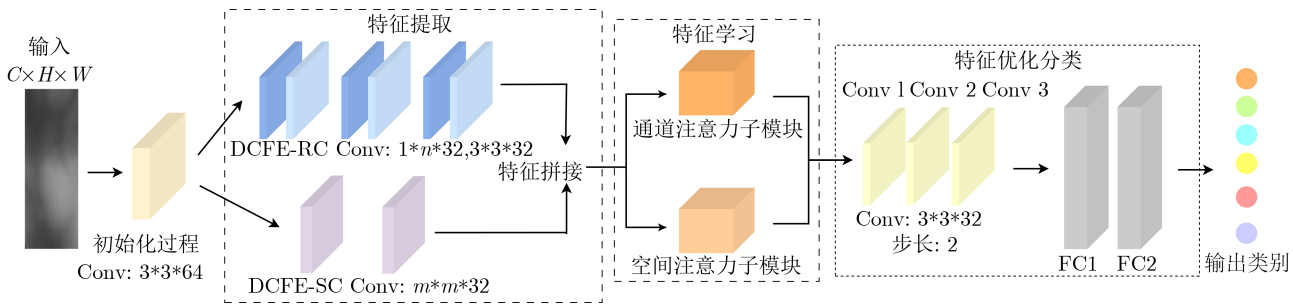


图 2 ALDCNet算法整体结构

通过渐进式降维形成多尺度层级化表征，配合自适应特征重校准策略，在分类器前形成具有强判别性的紧致特征表达，可有效解决同类静脉图像因局部相似性导致的误分类问题。

总体而言，ALDCNet通过其多样化和多层次的特征表征能力，能够更有效地适应特征相似度较高的手指静脉图像识别任务，并取得更高的识别准确率。

2.2 双卷积特征提取模块

指静脉识别中，静脉作为典型的长纹理载体，其特征提取的核心在于对方向性长纹理的高效捕捉与敏感表征。现有基于DL的指静脉识别网络普遍依赖方形卷积核进行特征提取，而Chai等人[9]的研究表明在基于CNN的指静脉生物识别任务中，矩形卷积核的性能普遍超越了方形卷积核，这一优势主要与手指静脉ROI图像的矩形形状驱动高度相关。

矩形卷积核通过非对称尺寸设计，在水平方向构建长感受野，从根源上主动适配手指静脉的长条形解剖结构特征，沿静脉走向精准捕捉连续延伸的长纹理及分支弯曲轨迹，在垂直方向保留短感受野，通过空间维度压缩抑制无关噪声(如皮肤纹理)，形成各向异性感受野，实现方向性特征定向强化；而方形卷积核以均匀感受野同步处理水平与垂直方向，更适合提取各向同性特征。矩形卷积核从形状层面主动“预适配”特征提取，避免了方形卷积核均匀处理导致的方向性信息稀释。

从宏观拓扑结构观察，不同个体的指静脉呈现具有类间区分度的纵横交织骨架；聚焦到微观纹理层面时，个体特有的脉管分支模式与细节走向则构成高区分度的生物特征标识。

受此形状适配性原则与生理结构特性的联合启发，本文基于指静脉的特征分布设计双卷积特征提取模块，通过异构卷积核的协同工作实现多粒度特征建模，从宏观上捕捉脉管全局走向，同时在微观层面解析局部纹理特征，形成多样化和具有互补性的融合特征。

如图2所示，网络前端采用3∗3卷积层(通道数64)对归一化后的指静脉ROI进行初步特征编码，随后通过DCFE模块进行深度特征提炼。如图3所示，DCFE模块采用双分支异构设计。(1)矩形卷积分支(DCFE-RC)：由3个级联模块串行构成，每个模块包含1个$1\times n(n\in\{3,5,7\})$的矩形卷积层和3×3的方形卷积层，重点提取具有方向敏感性的显著血管结构特征。(2)方形卷积分支(DCFE-SC)：采用$m\times m(m\in\{5,7,9\})$的两个方形卷积层，侧重于挖掘背景强度分布中的身份判别信息以及局部纹理细节特征。

为了最大限度保留两条分支提取的不同类型特征，将其输出的特征直接进行通道拼接，得到辨别性更强的融合特征。特征提取的公式为：$\mathbf{RC}^{l,p}=\sigma\left(\sum_{i=0}^{c}k_1^{l,p}*\boldsymbol{x}_i^p+b^{l,p}\right)$，$\mathbf{MF}^{l,p}=\sigma\left(\sum_{i=0}^{c}k_2^{l,p}*\mathbf{RC}^{l,p}+b^{l,p}\right)$，$\mathbf{SF}^{l,q}=\sigma\left(\sum_{i=0}^{c}k_3^{l,q}*\boldsymbol{x}_i^q+b^{l,q}\right)$，$\mathbf{MSF}=\mathrm{concat}\left(\mathbf{MF}^{l,p}|\mathbf{SF}^{l,q}\right)$，其中，$\boldsymbol{x}_i^p$，$\boldsymbol{x}_i^q$分别表示第$p,q$个卷积层的第$i$个通道特征映射，$k_1^{l,p},k_2^{l,p},k_3^{l,q}$分别表示大小为$1*n$，$3*3$，$m*m$的卷积核，$c$代表特征映射的数量，$*$代表卷积操作，$\mathbf{RC}^{l,p}$为第$p$个卷积层中第$l$个通道的卷积结果，第$l$个输出映射的偏差为$b^{l,p}$和$b^{l,q}$，$\mathbf{MF}^{l,p}$和$\mathbf{SF}^{l,q}$分别是DCFE-RC和DCFE-SC的卷积结果，$\sigma(\cdot)$代表非线性变换，$\mathrm{concat}(|)$代表通道层面的级联，**MSF**为DCFE模块最终的融合特征。

图4展示了DCFE-RC和DCFE-SC输出特征图的可视化结果。DCFE-RC热图呈广泛扩散型激活，高激活区契合静脉主干走向与轮廓，形成连续拓扑映射，表明矩形卷积设计侧重捕捉静脉全局脉络，维持形态连贯性。DCFE-SC热图则表现为高度局部化的激活特征，精准定位静脉边缘、背景纹理及末端细微结构，其大面积低激活区亮度均匀稳定，背景鉴别权重增强，证明方形卷积在强化边缘锐度与背景信息判别上有效。

DCFE模块以低通道数的上下两个轻量化网络分支作为特征提取的骨架，实现了特征响应图的互补性集成；通过核形状差异设计实现了对特征的多

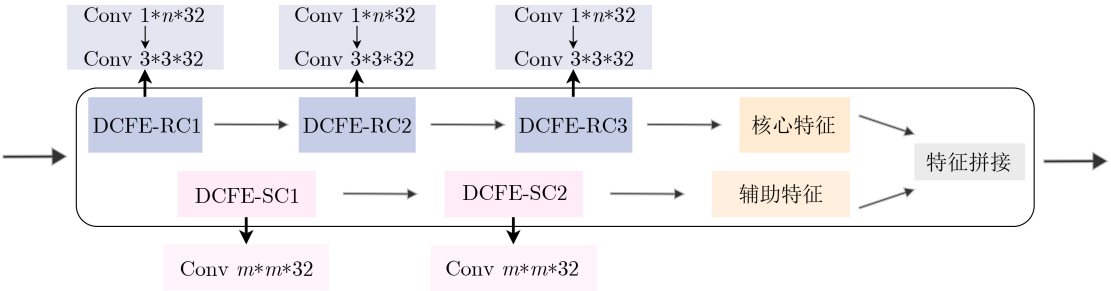


图 3 DCFE模块结构

样化提取，此外，这种复合感受野机制相比传统单尺度设计，对静脉管径变化和图像噪声具有更强的鲁棒性，同时有效压缩了参数规模。

但由于DCFE-RC和DCFE-SC的关注目标存在本质差异，得到的融合特征不可避免地产生信息冗余与跨域干扰，导致特征空间熵值升高，即不同通道中包含部分重复或无关信息：一是核心特征污染，即DCFE-RC提取的静脉拓扑结构中混入少量背景噪声，会模糊血管边界；二是辅助特征干扰，表现为DCFE-SC捕获的背景身份信息中残留显著血管特征，导致特征空间分布混乱。这种特征纯净度的下降会直接影响后续识别任务的精度与效率。

2.3 并行双重注意力模块

为解决前述问题，本文使用注意力机制来进一步权衡不同类型特征的重要程度，从而优化特征表示的纯净度与判别性。同时，考虑到轻量级网络在特征提取能力上通常弱于大型网络的问题，故采用一种简易高效的注意力模块来提升网络性能尤为必要。

本文通过分析指静脉识别任务中血管结构的关键表征需求，并为了进一步提升网络对静脉血管的敏感性，需要在网络架构中嵌入专门的注意力增强模块。为此，本文在ALDCNet中设计了一种灵活轻量的并行双重注意力模块——PDA模块。

如图5所示，给定一个特征向量$\boldsymbol{X}\in R^{C\times H\times W}$作为输入，首先，PDA-Channel模块和PDA-Spatial模块对其应用GAP(Global Average Pooling)操作，这一步骤获得尺寸为$1\times1\times C$和$H\times W\times1$的特征图。随后，对重塑后的特征向量进行卷积操作，再应用sigmoid函数分别生成信道映射$\boldsymbol{C}_{\mathrm{M}}$和空间映射$\boldsymbol{S}_{\mathrm{M}}$。其公式为$\boldsymbol{C}_{\mathrm{M}}=\sigma(f^{1}(\mathrm{Avg}_{\mathrm{C}}(\boldsymbol{X}_{\mathrm{in}})))$，$\boldsymbol{S}_{\mathrm{M}}=\sigma(f^{2}(\mathrm{Avg}_{\mathrm{F}}(\boldsymbol{X}_{\mathrm{in}})))$。最后，将输入数据与两个特征映射进行同位特征值与权值相乘，得到该模块的输出特征$\boldsymbol{X}_{\mathrm{out}}$。其公式为$\boldsymbol{X}_{\mathrm{out}}=\boldsymbol{X}_{\mathrm{in}}\odot\boldsymbol{C}_{\mathrm{M}}\odot\boldsymbol{S}_{\mathrm{M}}$，其中，$\boldsymbol{X}_{\mathrm{in}}$为输入特征，$\mathrm{Avg}_{\mathrm{C}}(\cdot)$和$\mathrm{Avg}_{\mathrm{F}}(\cdot)$分别是沿通道方向和特征点位置方向进行GAP操作，$f^{1}(\cdot)$和$f^{2}(\cdot)$为卷积运算的组合符号，$\sigma(\cdot)$代表sigmoid激活函数，$\odot$表示对应位置元素相乘。

PDA-Channel通过自适应权重分配强化静脉拓扑相关通道的核心特征表达，在抑制背景噪声的同时保留背景信息通道的有效身份特征，并弱化冗余静脉特征干扰。PDA-Spatial则构建像素级依赖关系建模，提升判别性区域的特征响应强度，并通过空间注意力掩膜动态抑制非判别性区域，实现特征空间分布重校准。二者以并行结构融合，可更好地保留前一模块得到的结构信息，以充分利用通道间的特征关系，通过可微分注意力图实现联合特征校准，在抑制背景干扰与通道冗余的同时，将DCFE块输出的融合特征优化为更精炼的关键判别特征，不仅提升了特征纯净度，还通过减少无效特征参与后续计算，降低了识别任务的计算复杂度。此外，轻量级的PDA块避免了模型引入注意力机制带来的参数增长问题，提高了模型整体对数据特征的敏感度，优化了模型的准确率。

与传统的双分支网络仅聚焦于特征的线性提取与拼接不同，DCFE块与PDA块共同构建“解耦-融合-纯化”3段式流程：解耦阶段分离静脉拓扑与背景身份信息，融合阶段保留异构特征判别性差异，纯化阶段通过动态权重分配抑制冗余信息、强化关键区域响应。此模式使模型更高效地解析指静脉复杂特征，增强了对异构特征的表征能力。

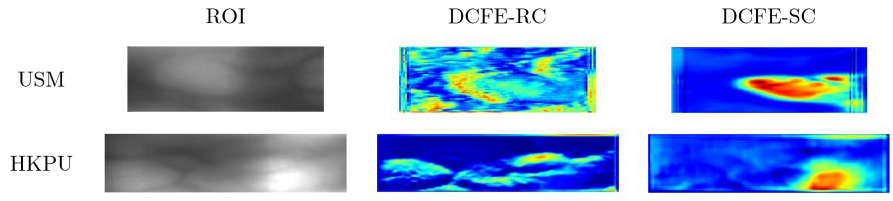


图 4 矩形卷积和方形卷积作用下输出特征的热图

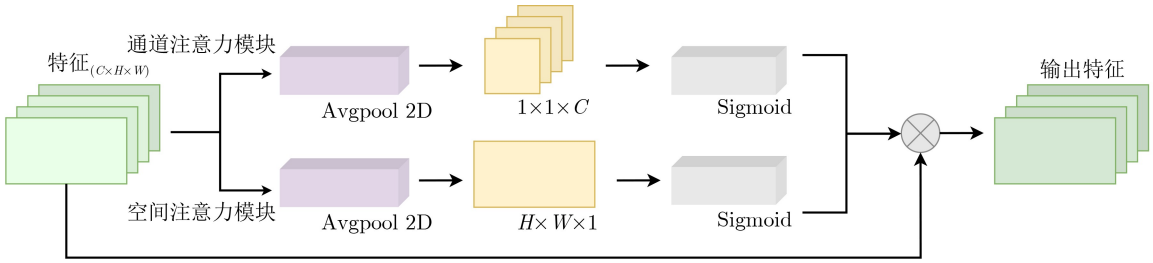


图 5 PDA模块结构

2.4 优化分类模块

基于特征层次化学习原理，本模块通过同构层堆叠策略实现了特征表达能力随网络深度的非线性增长；并构建双线性全连接层分类架构实现特征高效整合与精准分类，有效缓解了浅层架构中常见的欠拟合现象以及网络过深造成的过拟合问题，使模型在有限参数条件下达到更优的判别性能。

如图2所示，在特征优化阶段，PDA模块输出的原始特征映射依次通过3个级联的卷积模块(Conv1～Conv3)，每个模块均采用3*3空间滤波器配合步长2的采样策略，实施渐进式空间降维。此设计在保持静脉纹理核心判别信息完整性的同时，通过层级化感受野扩张机制(有效感受野从3*3逐级扩展至7*7)，实现了多尺度静脉特征的序列式提纯。此过程模拟人类视觉系统的层次化认知机制，构建出由局部血管细节到整体静脉网络结构的层次化特征表示体系，形成从具体到抽象的层次化表征。

在分类决策阶段，全连接层FC1采用特定维度的特征映射，通过ReLU激活进行非线性特征融合，形成紧凑的身份判别特征；全连接层FC2作为分类器，采用Softmax激活函数将特征映射到类别概率空间，实现端到端的身份识别，得到最终的分类结果。

3 实验和分析

3.1 实验平台及相关参数

本文实验环境为Windows10系统、处理器为Intel i7-9750H @2.60 GHz、显卡为NVIDIA GeForce GTX 1650、内存为32 GB，提出的方法在Pytorch框架下实现。模型训练阶段，采用交叉熵损失函数来衡量损失，运用自适应矩估计(Adam)优化算法来优化损失。在此过程中，学习率设为0.000 1，批处理大小(batch size)为20，训练周期(epoch)则设为15。鉴于模型本身非常轻量，整个训练流程耗时不超过半小时即可完成。

3.2 实验数据集

为了对ALDCNet模型进行评估和分析，本文在3个公开的数据集上进行了实验，数据集概况如下所示。

(1)马来西亚理工大学指静脉数据集(FV-USM)[10]，收集了123名受试者左右手的食指和中指，共计492个手指类别，图像总数为5 904张。本文使用该数据集自带的ROI图像(尺寸为300×100)进行实验。

(2)香港理工大学指静脉数据集(FV-HKPU)[11]，获取受试者156人，本文使用文献[11]中算法提取出前210个类别的ROI图像(尺寸为81×333)进行实验，总计2 520张图像样本。

(3)山东大学机器学习与数据挖掘实验室指静脉数据集SDUMLA[6]，采集了106名受试者的双手食指、中指和无名指静脉图像，形成636个手指类别，共计3 816张原始图像。本文通过自定义ROI提取算法，将关键静脉区域裁剪为150×300像素的标准化图像。

实验过程中，针对FV-USM, FV-HKPU和SDUMLA这3个数据集，其测试集与训练集的划分比例依次为2:10, 2:10和1:5。

3.3 评价指标

手指静脉识别属于多分类问题，准确率(Accuracy, Acc)是众多图像分类任务中常用的指标。该指标定义为：在所有参与分类的样本中，被模型正确分类的样本数量与总样本数量之比。评价指标的值越高，代表模型性能越好。其公式为：$\mathrm{Acc}=\dfrac{N_{\mathrm{C}}}{N}$，式中，$N_{\mathrm{C}}$为正确的测试集分类个数，$N$为测试总数。

在生物识别技术体系中，等错误率(Equal Error Rate, EER)是衡量系统性能的核心指标，其定义为错误接受率(False Acceptance Rate, FAR)与错误拒绝率(False Rejection Rate, FRR)达到平衡状态时的系统误判概率，该指标数值越小，意味着模型区分能力越强。其公式为：$\mathrm{FAR}=\dfrac{N_{\mathrm{FA}}}{N_{\mathrm{I}}}$，$\mathrm{FRR}=\dfrac{N_{\mathrm{FR}}}{N_{\mathrm{G}}}$，$\mathrm{EER}=\mathrm{FAR}=\mathrm{FRR}$，其中，$N_{\mathrm{FA}}$为被错误接受的冒名顶替次数，$N_{\mathrm{FR}}$为被错误拒绝的合法用户次数，$N_{\mathrm{I}}$为冒名顶替的总尝试次数，$N_{\mathrm{G}}$为合法用户的总尝试次数。

3.4 实验结果

3.4.1 网络识别性能分析

为验证本文所设计网络在手指静脉识别任务中的性能表现，并评估不同组合的矩形卷积核和方形卷积核对识别效果的影响，分别选取了3种不同尺寸的矩形卷积核(1*3, 1*5, 1*7)和方形卷积核(5*5, 7*7, 9*9)，在3个公共数据集上各进行了9组实验。实验结果参考表1。

表1的结果显示，在USM数据集上，当矩形卷积核尺寸为1*5，方形卷积核为5*5的组合时，网络的识别精度最高，在HKPU数据集上，1*3和9*9的卷积核组合效果最优，而采用1*5和9*9的卷积核组合时，模型在SDUMLA数据集上达到最佳性能。这一差异主要归因于ROI提取方法和数据集采集设备及过程的不同，会直接影响数据集中特征的分布方式，则在不同的数据集上实现最优效果

表 1　不同卷积核组合在3个数据集上的识别准确率(%)

矩形卷积核	方形卷积核(USM)			方形卷积核(HKPU)			方形卷积核(SDUMLA)		
	5∗5	7∗7	9∗9	5∗5	7∗7	9∗9	5∗5	7∗7	9∗9
1∗3	99.49	99.59	99.39	97.14	97.62	**98.33**	97.64	97.48	97.80
1∗5	**99.70**	99.39	99.49	97.38	97.62	98.10	97.32	97.16	**98.27**
1∗7	99.29	99.29	98.78	97.86	97.86	97.62	97.48	97.64	97.80

注：粗体表示所有方案中性能指标的最高值。

时，通常需要在卷积核的形状和尺寸上进行相应的调整。ALDCNet在3个数据集上的最佳识别准确率分别达到99.70%, 98.33%和98.27%，表明了本文方法的识别有效性和精确度。

通过Grad-CAM(Gradient-weighted Class Activation Mapping)热图技术，本文对ALDCNet模型最终输出的判别性特征进行了系统性可视化分析。如图6所示，热图高亮区完整覆盖主静脉轨迹，精准捕捉细微分支静脉走向，且其静脉边缘渐变区

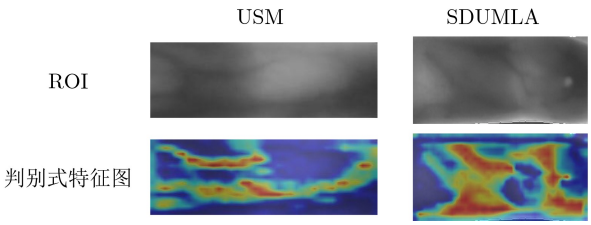


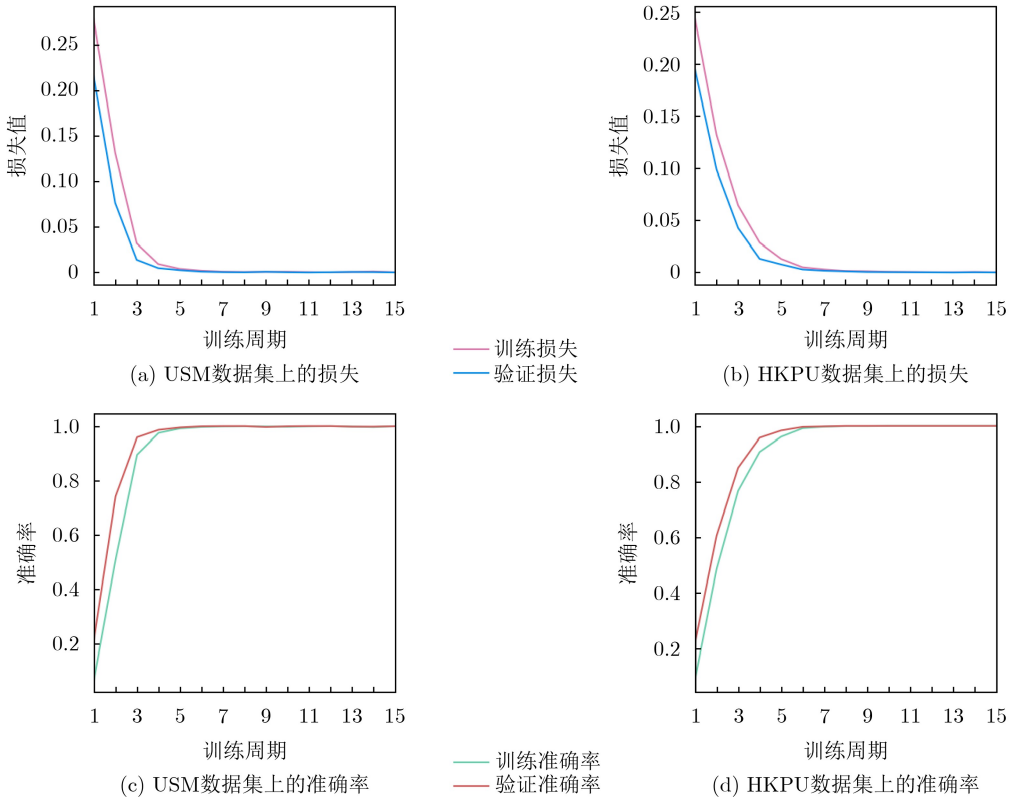
图 6 Grad-CAM热图

与实际静脉在形态(如曲率、分支角度)以及密度分布(如灰度值梯度)上表现出高度一致性。该热图通过空间定位、细节捕捉、形态匹配等维度展现了模型对静脉结构从主干到分支的层级化解析能力，充分验证了模型从复杂图像中精准提取关键判别特征的有效性。

3.4.2　可靠性和稳健性分析

图7展示了在两个公共数据集上，对ALDCNet模型进行训练时所获得的损失值和准确率变化曲线。从图中可观察到，损失曲线在训练过程中呈现出平稳且正常的收敛态势，表明模型在训练过程中误差逐渐减小，拟合效果良好。准确率曲线则呈现出持续上升的趋势，反映模型对数据的识别与分类能力在不断提升，说明了此方案具有一定的可靠性和稳健性。

图8显示了模型在两个数据集上针对随机选取的20个类别所生成的混淆矩阵可视化结果。此矩阵



(a) USM数据集上的损失　(b) HKPU数据集上的损失

(c) USM数据集上的准确率　(d) HKPU数据集上的准确率

图 7 两个公共数据集的损失曲线和准确率曲线

能够直观反映模型性能，并计算各类别准确性。通过对角线上数值表示各类别的预测正确率，非对角线上数值反映误分类比例，从而全面评估了模型的分类性能。结果显示，ALDCNet模型分类准确率达到100%，进一步验证了本文方案的有效性和可行性。

3.4.3 不同方法对比

为了进一步评估ALDCNet模型的识别能力，本研究通过对比具有典型代表性的卷积神经网络、Transformer/ViT架构及当前前沿算子方法，以验证本文模型在指静脉识别任务上的性能优势。各方法的识别准确率结果和网络方法的EER值如表2所示。同时，为全面评估模型效率，本研究进一步对比了卷积神经网络方法的计算复杂度，包括模型参数量和单样本推理时间两项核心指标，相关参数的具体配置情况详见表3。

表2的对比结果显示，本文网络在3个数据集上均展现出了最优性能。这是由于本文方法不仅聚焦于提取关键的长距离特征，还兼顾了背景中的细粒度身份信息。通过融合这些特征，在最大限度保留原始特征的同时，增强了特征的层次性和个性化。而其他网络方法仅关注到局部纹理模式的显性表达或单一维度的特征聚合，未能有效建模长距离依赖关系与背景中的隐性身份关联。基于Transformer/ViT的方法则过度依赖于图像增强策略和大规模数

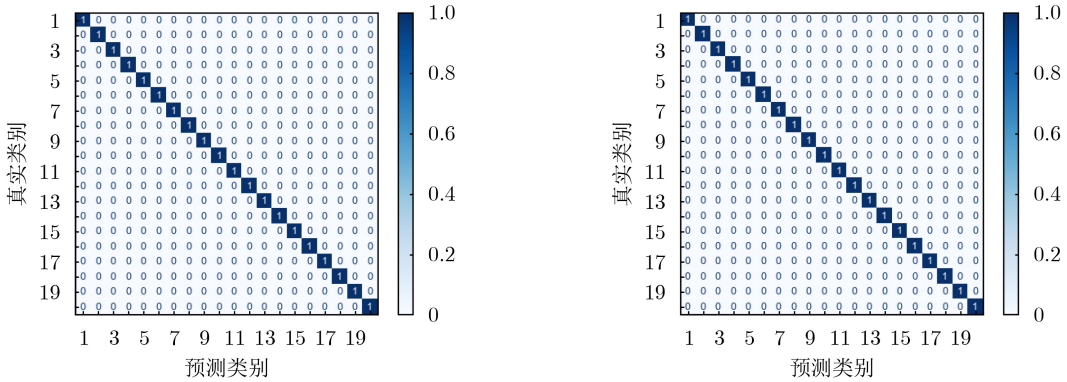


图 8 混淆矩阵图

表 2 与其他方法性能指标的对比(%)

方法	Acc(USM)	EER(USM)	Acc(HKPU)	EER(HKPU)	Acc(SDUMLA)	EER(SDUMLA)
AlexNet[12]	97.26	0.10	86.90	3.99	92.45	1.10
VGG16[13]	95.93	0.31	76.67	4.78	78.14	1.73
Liu等人[14]	97.97	0.20	77.14	4.24	69.65	4.56
LALBP[15]	79.47	*	73.10	*	76.89	*
MMNBP[16]	83.29	*	75.16	*	80.00	*
ViT-Cap[17]	98.33	0.28	96.18	1.66	93.79	1.30
Lu等人[18]	91.76	2.46	90.98	3.89	94.12	0.94
ALA Net[19]	97.74	0.34	97.86	0.32	94.50	0.53
NLNet[20]	98.98	0.48	96.22	0.55	94.30	1.13
ALDCNet	**99.70**	**0.08**	**98.33**	**0.28**	**98.27**	**0.35**

注：粗体表示所有方法中性能指标的最高值。

表 3 模型参数配置的对比

方法	准确率(%)	卷积核数	操作大小	卷积层数	参数量(M)	时间(ms)
AlexNet[12]	97.26	64,192,384,256×2	11∗11,5∗5,3∗3,3∗3	5	2.454	**1.01**
VGG16[13]	95.93	64×2,128×2,256×3,512×3,512×3	3∗3,3∗3,3∗3,3∗3,3∗3	13	14.714	3.00
Liu等人[14]	97.97	64,64,64,64,64	5∗5,5∗5,5∗5,5∗5,5∗5	5	0.412	1.22
ALDCNet	**99.70**	32×2,32×3,32×3,32×3,64	3∗3,1∗5,5∗5,3∗3,3∗3	12	**0.164**	1.66

注：粗体表示最优值。

据训练，当输入数据未经特定预处理时，模型性能显著降低。此外，本文方法中的并行双重注意力机制能够从不同维度更精准地定位高区分度特征，从而进一步提升了识别效果。

表3详细列出了本文方法与神经网络方案在USM数据集上应用的网络模型参数配置。通过对比分析，可以发现ALDCNet的参数量最小，仅有164 k，且模型处理数据集中单个ROI图像的平均时间仅需1.66 ms，实现了最高的准确率，其高效性本质上是源于网络整体的低通道轻量化骨架设计，以及在特征处理阶段对纯度的优化。

计算复杂度分析验证了本研究架构的合理性——通过参数效率与特征纯度的协同优化，在毫秒级响应约束下实现了识别性能的新平衡。进一步说明ALDCNet不仅具有合理的网络深度，而且在参数少、计算量小的情况下也能捕获更抽象且全局性的特征表征，同时避免了网络复杂度的增加。相反，层数过少的网络模型则易局限于对简单局部特征的捕捉。综上所述，这些实验数据证明了本文网络的精炼性、有效性和高效性。

3.4.4 消融实验

为检验本文提出的核心模块是否能有效提高模型的性能，本文设计了6组消融实验，测试结果见表4。其中，“无PDA”表示未使用并行双重注意力机制，“无DCFE-RC”表示未使用矩形卷积分支，“无DCFE-SC”表示未使用方形卷积分支。

表 4　消融实验(%)

模型	Acc (USM)	EER (USM)	Acc (HKPU)	EER (HKPU)	Acc (SDUMLA)	EER (SDUMLA)
无PDA	99.09	0.20	97.38	0.52	96.54	0.47
无DCFE-RC	99.29	0.10	97.38	0.48	96.69	0.64
无DCFE-SC	99.39	0.10	97.86	0.48	97.01	0.63
ALDCNet	**99.70**	**0.08**	**98.33**	**0.28**	**98.27**	**0.35**

注：加粗数值表示最优值。

由表4可知，在3个公共数据集上，通过在神经网络中引入并行双重注意力模块，识别准确率分别提升了0.61%, 0.95%和1.73%，这一显著提升的原因在于PDA能够自适应地学习每个通道的重要程度，并进一步关注在空间上最有区分力的位置，使模型聚焦于最具分类属性的特征；采用DCFE-RC和DCFE-SC模块在3个数据集上分别提高了0.41%与0.31%, 0.95%与0.47%和1.58%与1.26%的精度。此外，在4个对比模型中，EER指标呈现相应的动态变化趋势，而ALDCNet模型在所有实验场景下的EER值均稳定保持最低水平。总体而言，矩形卷积分支相较于方形卷积分支展现出了更突出的效果，这得益于其应用矩形卷积核能够更精准地提取图像中的核心特征，这些特征具有更强的区分性；而方形卷积分支采用方形卷积核，侧重提取图像背景中的少量身份信息和局部信息，发挥辅助补充作用。消融实验直观地证明了3个核心部分的必要性，均能使模型性能得到明显提升。

4　结束语

本文设计了一种基于注意力机制的轻量级双通道卷积神经网络，此网络的核心组件包括双卷积特征提取模块、并行双重注意力模块和特征优化分类模块3个部分。通过低通道的DCFE策略挖掘长距离静脉结构特征和图像背景信息的融合特征，大大减小模型参数和提高网络区分个体的能力；通过灵活轻量的PDA模块帮助网络更高效地捕捉指静脉区域的关键特征，解决了大多数注意力机制虽能提升性能，却伴随大量参数增加不利于网络轻量化的难题。FOC模块通过堆叠式结构实现特征降维、优化与决策分类，有效缓解了欠拟合和过拟合问题。研究结果表明，相较于现有的轻量级手指静脉识别方法，ALDCNet网络在USM, HKPU和SDUMLA 3个数据集上实现更高识别精度的情况下，参数量减少了11.35%～60.19%。此方法在识别精度与参数量之间实现了良好的平衡，将本文的轻量模型应用于算力有限的小型设备上，将有效节约身份识别所需资源。

在本研究中，不同数据集经ROI提取后获得的静脉图像尺寸存在差异，而当前模型未充分考虑到多尺寸输入的兼容性，导致跨数据集部署的通用性受限。为突破此局限，未来研究将重点开发迁移能力更强的轻量化模型架构，并探索其在生物识别技术领域的安全性与应用广度。

参 考 文 献

[1] LI Yantao, FAN Chao, QIN Huafeng, *et al*. Deep reinforcement learning-based feature extraction and encoding for finger-vein verification[J]. *IEEE Transactions*

on Emerging Topics in Computational Intelligence, 2025, 9(1): 522–536. doi: 10.1109/TETCI.2024.3398022.

[2] TAN Xiao. Advanced feature extraction algorithms for deep learning in image recognition[C]. 2024 International Conference on Power, Electrical Engineering, Electronics and Control (PEEEC), Athens, Greece, 2024: 1055–1060. doi: 10.1109/PEEEC63877.2024.00195.

[3] HUANG Houjun, LIU Shilei, ZHENG He, *et al*. DeepVein: Novel finger vein verification methods based on deep convolutional neural networks[C]. 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), New Delhi, India, 2017: 1–8. doi: 10.1109/ISBA.2017.7947683.

[4] 何鑫, 陈迅. 基于改进卷积神经网络的指静脉识别[J]. 计算机工程与设计, 2019, 40(2): 562–566. doi: 10.16208/j.issn1000-7024.2019.02.046.

HE Xin and CHEN Xun. Finger vein recognition based on improved convolution neural network[J]. *Computer Engineering and Design*, 2019, 40(2): 562–566. doi: 10.16208/j.issn1000-7024.2019.02.046.

[5] HU Hui, KANG Wenxiong, LU Yuting, *et al*. FV-Net: Learning a finger-vein feature representation based on a CNN[C]. 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 2018: 3489–3494. doi: 10.1109/ICPR.2018.8546007.

[6] REN Hengyi, SUN Lijuan, REN Jinting, *et al*. FV-DDC: A novel finger-vein recognition model with deformation detection and correction[J]. *Biomedical Signal Processing and Control*, 2025, 100: 107098. doi: 10.1016/j.bspc.2024.107098.

[7] BOUCHERIT I, ZMIRLI M O, HENTABLI H, *et al*. Finger vein identification using deeply-fused convolutional neural network[J]. *Journal of King Saud University-Computer and Information Sciences*, 2022, 34(3): 646–656. doi: 10.1016/j.jksuci.2020.04.002.

[8] WANG Kaixuan, CHEN Guanghua, and CHU Hongjia. Finger vein recognition based on multi-receptive field bilinear convolutional neural network[J]. *IEEE Signal Processing Letters*, 2021, 28: 1590–1594. doi: 10.1109/LSP.2021.3094998.

[9] CHAI Tingting, LI Jiahui, PRASAD S, *et al*. Shape-driven lightweight CNN for finger-vein biometrics[J]. *Journal of Information Security and Applications*, 2022, 67: 103211. doi: 10.1016/j.jisa.2022.103211.

[10] MOHD ASAARI M S, SUANDI S A, and ROSDI B A. Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics[J]. *Expert Systems with Applications*, 2014, 41(7): 3367–3382. doi: 10.1016/j.eswa.2013.11.033.

[11] KUMAR A and ZHOU Yingbo. Human identification using finger images[J]. *IEEE Transactions on Image Processing*, 2012, 21(4): 2228–2244. doi: 10.1109/TIP.2011.2171697.

[12] FAIRUZ S, HABAEBI M H, and ELSHEIKH E M A. Finger vein identification based on transfer learning of AlexNet[C]. 2018 7th International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, 2018: 465–469. doi: 10.1109/ICCCE.2018.8539256.

[13] NADIR C, ATTALLAH B, and BRIK Y. Finger vein based CNN algorithms for human recognition[C]. 2022 International Conference of Advanced Technology in Electronic and Electrical Engineering, M'sila, Algeria, 2022: 1–6. doi: 10.1109/ICATEEE57445.2022.10093105.

[14] LIU Wenjie, LI Weijun, SUN Linjun, *et al*. Finger vein recognition based on deep learning[C]. 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), Siem Reap, Cambodia, 2017: 205–210. doi: 10.1109/ICIEA.2017.8282842.

[15] 黄艳国, 杨训根, 周满国. 基于感兴趣区域的改进型LBP手指静脉识别[J]. 传感器与微系统, 2023, 42(4): 143–147. doi: 10.13873/J.1000-9787(2023)04-0143-05.

HUANG Yanguo, YANG Xungen, and ZHOU Manguo. Improved LBP finger vein recognition based on RoI[J]. *Transducer and Microsystem Technologies*, 2023, 42(4): 143–147. doi: 10.13873/J.1000-9787(2023)04-0143-05.

[16] 付华, 李涛, 司南楠. 基于MMNBP的手指静脉识别方法[J]. 传感器与微系统, 2019, 38(5): 45–48. doi: 10.13873/J.1000-9787(2019)05-0045-04.

FU Hua, LI Tao, and SI Nannan. Finger vein recognition method based on MMNBP[J]. *Transducer and Microsystem Technologies*, 2019, 38(5): 45–48. doi: 10.13873/J.1000-9787(2019)05-0045-04.

[17] LI Yupeng, LU Huimin, WANG Yifan, *et al*. ViT-Cap: A novel vision transformer-based capsule network model for finger vein recognition[J]. *Applied Sciences*, 2022, 12(20): 10364. doi: 10.3390/APP122010364.

[18] LU Zhiying, WU Runhao, and ZHANG Jianfeng. Finger-vein feature extraction method based on vision transformer[J]. *Journal of Electronic Imaging*, 2022, 31(4): 043010. doi: 10.1117/1.JEI.31.4.043010.

[19] HUANG Yiwei, MA Hui, and WANG Mingyang. Axially enhanced local attention network for finger vein recognition[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 5020210. doi: 10.1109/TIM.2023.3291785.

[20] GUO Zishuo, MA Hui, and LIU Junbo. NLNet: A narrow-channel lightweight network for finger multimodal

recognition[J]. *Digital Signal Processing*, 2024, 150: 104517. doi: 10.1016/j.dsp.2024.104517.

赵冰艳：女，硕士生，研究方向为手指静脉图像识别.
梁义怀：男，讲师，研究方向为数据安全与隐私保护、群智感知、联邦学习.
张中霞：女，博士，研究方向为信息安全、图像处理.
张文政：男，高级工程师，研究方向为密码学理论和技术.

责任编辑：廖海贝

Lightweight Dual Convolutional Finger Vein Recognition Network Based on Attention Mechanism

ZHAO Bingyan[①]　LIANG Yihuai[①]　ZHANG Zhongxia[②]　ZHANG Wenzheng[②③]

[①](*School of Information Science and Technology, Southwest Jiaotong University, Chengdu* 611756, *China*)

[②](*China Electronics Technology Cyber Security Co., Ltd., Chengdu* 610041, *China*)

[③](*National Key Laboratory of Security Communication, Chengdu* 610041, *China*)

Abstract:

Objective　Finger vein recognition is an emerging biometric authentication technology valued for its physiological uniqueness and advantages in in vivo detection. However, mainstream deep learning recognition frameworks still face two challenges. High-precision recognition often depends on complex network structures, which increase parameter counts and hinder deployment in memory-limited embedded devices and edge scenarios with constrained computing resources. Model compression can reduce computational cost but often weakens feature representation, creating a conflict between recognition accuracy and efficiency. To address these issues, a lightweight dual convolutional model integrated with an attention mechanism is proposed. A parallel heterogeneous convolution module and an attention guidance mechanism are designed to extract diverse image features and improve recognition accuracy while preserving a lightweight network structure.

Methods　The proposed architecture adopts a three-level collaborative mechanism comprising feature extraction, dynamic calibration, and decision fusion. A dual convolutional feature extraction module is constructed using normalized ROI images. This module combines heterogeneous convolution kernels. Rectangular convolution branches with different shapes capture venous topological structures and diameter orientations, whereas square convolution branches employ stacked square kernels to extract local texture details and background intensity distributions. These branches operate in parallel with reduced channel numbers and generate complementary responses through kernel shape diversity. This design reduces parameter scale while improving feature discrimination. A parallel dual attention mechanism is then applied to achieve two-dimensional calibration through joint optimization of channel attention and spatial attention. Channel attention adaptively assigns weights to enhance discriminative venous texture features, whereas spatial attention constructs pixel-level dependency models that focus on effective discriminative regions. A parallel concatenation fusion strategy preserves structural information without introducing additional parameters and improves sensitivity to critical features. Finally, a three-level progressive feature optimization structure is implemented. A convolutional compression module with stride 2 nests multi-scale receptive fields and progressively refines primary features during dimensionality reduction. Two fully connected layers then perform feature space transformation. The first layer applies ReLU activation to form sparse representations, and the final layer applies Softmax for probability calibration. This structure balances shallow underfitting and deep overfitting while maintaining efficient forward inference.

Results and Discussions　The effectiveness and robustness of the proposed network are evaluated on three public datasets, namely USM, HKPU, and SDUMLA. Recognition accuracy is assessed using the Acc metric. Experimental results (Table 1) show strong recognition performance. Feature visualization heatmaps (Fig. 4, Fig. 6) confirm that the network extracts complete and discriminative venous features. Training visualizations (Fig. 7, Fig. 8) show stable loss and accuracy trends, achieving 100% classification performance and demonstrating training reliability and robustness. Quantitative comparisons (Tables 2 and 3) indicate that the

proposed method effectively addresses the trade-off between model complexity and classification performance and achieves superior results across all three datasets. Ablation studies (Table 4) further verify the effectiveness of the proposed modules and show significant improvements in finger vein recognition performance.

Conclusions A lightweight dual convolutional neural network with an attention mechanism is proposed. The network consists of three core modules: a dual convolutional feature extraction module, a parallel dual-attention module, and a feature optimization classification module. During feature extraction, long-range venous features and background information are jointly encoded through a low-channel parallel design, which substantially reduces parameter counts while improving inter-individual discrimination. The attention module efficiently captures critical venous features without the parameter expansion commonly observed in conventional attention mechanisms. The feature optimization classification module applies progressive feature recalibration, which reduces underfitting and overfitting during stacked dimensionality reduction. Experimental results show recognition accuracies of 99.70%, 98.33%, and 98.27% on the USM, HKPU, and SDUMLA datasets, corresponding to an average improvement of 2.05% over existing state-of-the-art methods. Compared with representative lightweight finger vein recognition approaches, the proposed method reduces parameter scale by 11.35%～60.19%, achieving a balance between model lightening and performance improvement.

Key words: Finger vein recognition; Dual convolutional model; Feature representation; Lightweight; Parallel dual attention mechanism