

# 基于 GoogLeNet 和 ResNet 的深度融合神经网络在脉搏波识别中的应用<sup>①</sup>



张 选<sup>1</sup>, 胡晓娟<sup>2</sup>

<sup>1</sup>(中国科学技术大学 纳米技术与纳米仿生学院, 合肥 230026)

<sup>2</sup>(上海中医药大学 上海中医健康服务协同创新中心, 上海 201203)

通讯作者: 胡晓娟, E-mail: xjhu2013@sinano.ac.cn

**摘 要:** 为了提高脉搏波识别的准确率, 提出改进的深度融合神经网络 MIRNet2。首先, 经过主波提取、划分周期和制作 hdf5 数据集等, 获得 Caffe 可处理的数据集。其次, 提出由 Inception 模块和残差模块构成的融合网络 Inception-ResNet (IRNet), 包含 IRNet1、IRNet2 和 IRNet3。在此基础上, 改进 Inception 模块、残差模块和池化模块, 构造 Modified Inception-ResNet (MIRNet), 包含 MIRNet1 和 MIRNet2。与本文其它神经网络相比, MIRNet2 的分类性能最好, 特异性、灵敏度和准确率分别达到 87.85%、88.05% 和 87.84%, 参数量和运算量也少于 IRNet3。

**关键词:** 脉搏波; 识别; 卷积神经网络; Google inception net; 残差神经网络; 深度学习

引用格式: 张选, 胡晓娟. 基于 GoogLeNet 和 ResNet 的深度融合神经网络在脉搏波识别中的应用. 计算机系统应用, 2019, 28(10): 15-26. <http://www.c-s-a.org.cn/1003-3254/7110.html>

## Pulse Wave Recognition Using Deep Hybrid Neural Networks Based on GoogLeNet and ResNet

ZHANG Xuan<sup>1</sup>, HU Xiao-Juan<sup>2</sup>

<sup>1</sup>(School of Nano Technology and Nano Bionics, University of Science and Technology of China, Hefei 230026, China)

<sup>2</sup>(Shanghai Innovation Center of TCM Health Service, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China)

**Abstract:** To improve the accuracy of pulse wave recognition, MIRNet2 is proposed, which is a kind of modified deep hybrid neural networks. Firstly, processable data sets of Caffe are obtained by main pulse extraction, segmenting cycle and making hdf5 data sets. Secondly, deep hybrid neural networks are designed. Inception-ResNet (IRNet) is consisted of inception modules and residual modules, containing IRNet1, IRNet2 and IRNet3. Subsequently, Modified Inception-ResNet (MIRNet) composed of modified Inception modules, residual modules and pooling modules (or reduction modules) is proposed, including MIRNet1 and MIRNet2. Compared with other neural networks in the study, MIRNet2 is the best one, with the specificity of 87.85%, the sensitivity of 88.05% and the accuracy of 87.84%, respectively. In addition, parameters and operations of MIRNet2 are also less than that of IRNet3.

**Key words:** pulse wave; recognition; convolutional neural networks; Google inception net; residual neural networks; deep learning

## 1 引言

由于计算机硬件性能的提高, 深度学习及其应用成为目前研究的热点, 包括图像、语音等的识别。脉诊

是传统中医的重要组成部分, 客观化研究尚未成熟, 不同经验积累的医生对于疑难病症有不同的诊断结果。深度学习用于处理医学图像、脉搏波、心电图等, 有

① 基金项目: 中国科学院苏州纳米技术与纳米仿生研究所资助项目 (Y6AAJ11001)

Foundation item: The Project Owned by Suzhou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Sciences (Y6AAJ11001)

收稿时间: 2019-03-20; 修改时间: 2019-04-17; 采用时间: 2019-04-23; csa 在线出版时间: 2019-10-15

助于提升识别准确率,将推动智慧医疗的发展。

深度神经网络 (Deep Neural Networks, DNN) 包括卷积神经网络 (Convolutional Neural Networks, CNN)、递归神经网络 (Recurrent Neural Networks, RNN) 和强化学习 (Reinforcement Learning, RL)。CNN 在图像识别领域应用广泛, LeCun<sup>[1]</sup>等提出 LeNet, 包括卷积层、池化层、Sigmoid 激活函数, 多层感知机是最后的分类器, 奠定了 CNN 的基本组成。Krizhevsky<sup>[2]</sup>等提出 AlexNet, 使用线性修正单元 (Rectified Linear Unit, ReLU) 作为激活函数, 解决了网络加深时的梯度弥散问题; 使用 Dropout 随机忽略部分神经元, 避免过拟合; 使用重叠的最大池化和 CUDA 加速。Simonyan 等<sup>[3]</sup>研究了 CNN 深度与性能的关系, 提出 VGGNet, 反复堆叠 3×3 的卷积核和 2×2 的最大池化层, 构建了 16~19 层深的 CNN, 初步证明网络越深, 效果越好; Szegedy 等<sup>[4]</sup>提出 Google Inception Net (GoogLeNet), 是一种模块化的高效稀疏结构, 参数量和运算量比之前的神经网络少, 跨通道提取特征, 有利于提取多样性特征。He 等<sup>[5]</sup>提出残差神经网络 (Residual Neural Networks, ResNet), 网络深度可达上百层, 解决了网络加深时性能先升后降的退化问题, 使某些重要特征可直接传输到任一模块或层。这些经典网络是 CNN 发展的里程碑。

Liu 等<sup>[6]</sup>提出基于 GoogLeNet 的时序自适应神经网络, 由于 Inception 模块中多尺度的卷积核, 可用于识别视频帧之间的相关性动作, 运行时间较少且鲁棒性强, 表明 Inception 模块能较好地处理时序型数据; Khosravi<sup>[7]</sup>等提出改进的 Google Inception-V3, 用于识别肺癌、膀胱癌、乳腺癌的病理图, 准确率分别达到 100%、92%、95%, 对肿瘤异质性病理图分类性能很好, 能有效识别多种特征; Vesal 等<sup>[8]</sup>在 ImageNet 数据集上预训练 Google Inception V3 模型, 识别特定领域的特征, 然后, 利用迁移学习, 用预训练好的 Google Inception V3 模型识别乳腺癌病理图, 测试准确率达到 97.08%, 结合医学专家的分析, 尝试应用于临床。GoogLeNet 主要有以下优势: (1) 由不同的卷积层和一个池化层组成的 Inception 模块, 构建了一种高效的稀疏结构, 有利于提取丰富的纹理特征; (2) 常用多种大小的卷积核, 比如 1×1、3×3、5×5、7×7 等, 提取不同层次的高阶特征, 增加特征提取的多样性; (3) 模块中有不同的分支网络, 增加了网络宽度, 提高了网络的鲁棒性和表达能力, 使其对不同分辨率的图像有更好的泛化能力。

随着网络层数增加, CNN 的性能逐渐提升, 当达到一定层数时, 梯度发散问题会导致性能下降。ResNet 采用残差模块 (Residual Modules) 或跳跃连接 (Skip Connection) 解决了此问题 (如图 1), 使网络层数更多, 简化了网络的学习目标, 不同单元及层的信号可直接正向和反向传输给任一层, 加速网络训练和参数调优。Zhang 等<sup>[9]</sup>提出将中心损失函数用于 ResNet, 在手写体汉字集 ICDAR-2013 上取得 97.03% 的准确率; Lu 等<sup>[10]</sup>提出深度耦合 ResNet, 由一个主干网络和两个分支网络组成, 主干网络用于识别不同分辨率的人脸照片, 两个分支网络分别训练高分辨率图片和目标图片, 将其转化成特定分辨率的耦合图片。

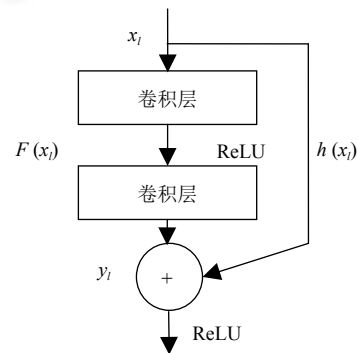


图 1 残差模块

$$F(x_l) = w * x_l + b \quad (1)$$

$$y_l = R(F) + h(x_l) \quad (2)$$

$$x_{l+1} = R(y_l) \quad (3)$$

式中,  $x_l$  是输入,  $w$  是权重,  $b$  是偏置,  $y_l$  是两个分支的和,  $R$  是 ReLU 函数,  $F(x_l)$  表示卷积运算,  $h(x_l)$  是对输入的某种简单变换,  $x_{l+1}$  是残差模块的最终输出。ReLU 是一种激活函数, 如式 (4), 有利于梯度传播和防止梯度弥散, 以防梯度在多层卷积之后而大幅衰减<sup>[2]</sup>。

$$R(x) = \max(0, x) \quad (4)$$

当  $x > 0$  时,  $R(x) = x$ , 其导数为 1; 当  $x \leq 0$  时,  $R(x) = 0$ , 其导数为 0。前向计算时, 只需输入值  $x$  和阈值 0, 即得到输出值; 反向计算时, 梯度为 1 或 0, 即梯度衰减较小或不变。与 Sigmoid 和 Tanh 等函数相比, 计算简单, 梯度衰减较小, 有利于加深网络。

研究者也探索了整合两种网络优势, 以构造融合网络。Habibzadeh 等<sup>[11]</sup>提出预训练的 Inception-ResNet, 用于四种主要白细胞的分类, Inception-V1-ResNet 获得 99.84% 的准确率; Szegedy 等<sup>[12]</sup>在 Inception 结构中

加入残差模块,发现残差模块可以加速 Inception 网络训练,但不能明显提高准确率; Zhuang 等<sup>[13]</sup>将 Google Inception V3 和不同的长短期记忆网络 (Long Short Term Memory, LSTM) 融合,构造 ConvDLSTM,在人群图片集上统一训练模型,将时间和空间信息统一在一个神经网络结构中,有效检测人群信息.网络结构的融合,不是简单的组合,而需要改进网络结构和参数设置等.

基于 DNN 的脉搏波识别,研究较少. Hu 等<sup>[14]</sup>用基本的 CNN(9L) 识别健康和亚健康的脉搏波信号,准确率达到 72.31%; Zhang 等<sup>[15]</sup>用 9 层的 CNN 识别弱脉、弦脉、深脉等 12 种脉搏,平均准确率达到 93.49%; Zhou 等<sup>[16]</sup>组合 DNN 和规则推理,在心电图数据集 MIT-BIH-AR 上检测室性早搏,准确率、特异性和灵敏度分别达到 99.41%、99.54% 和 97.59%. Jin 等<sup>[17]</sup>搭建由医生监控的云计算平台,该平台嵌入深度学习、规则推理和集成学习等,辅助诊断心脏疾病,输入病人的 12 导联 ECG 数据并给出诊断报告.有关脉搏波识别存在以下问题: (1) 基于深度学习的脉搏波识别研究较少,只使用一些基本的神经网络,而性能更好的 ResNet、深度融合神经网络、生成对抗网络 (Generative Adversarial Networks, GAN) 等并未使用; (2) DNN 用于脉搏波识别,需要调整网络结构、卷积核大小和数量、学习策略、损失函数等,将影响识别的准确率; (3) 目前已有的脉搏波识别算法,其准确率较低,难以达到临床应用的要求; (4) 脉搏波数据集缺乏标准,研究者使用不同的数据集,难以评测算法的性能.

本文提出由改进的 Inception 模块、残差模块和池化模块组成的深度融合神经网络 Modified Inception-ResNet (MIRNet),包括 MIRNet1 和 MIRNet2,其中 MIRNet2 的准确率达到 87.84%,性能最优,分析了 MIRNet2 的结构和原理.本文首先介绍数据集预处理的方法;其次,分析 Inception 模块和残差模块融合的原因,构建 IRNet 和 MIRNet 融合网络,探究融合网络的原理;此外,介绍参数量和运算量、准确率等网络性能评测指标.随后,介绍实验和分析结果,最后是结论与展望.

## 2 方法

主要阐述数据集预处理、IRNet 及 MIRNet 的结构、参数量、运算量计算及准确率等评价指标.实验流程,如图 2 所示,主要分为数据集预处理、算法识别和结果分析.

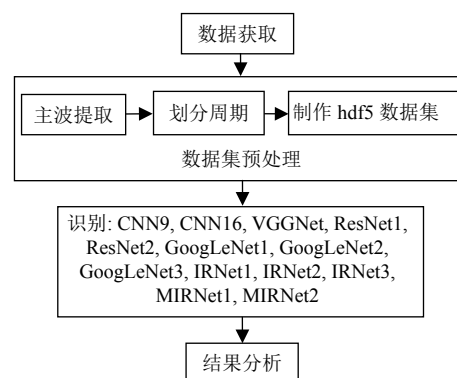


图2 脉搏波识别流程

### 2.1 数据集获取和预处理

#### 2.1.1 数据集获取

数据来源为上海中医药大学研发的 DDMX-100 脉象仪采集的 1125 个健康对象和 1158 个亚健康对象的脉搏波,采样频率为 200 Hz,持续 40 s,共计 8000 个采样点.根据世界卫生组织公布的区分健康和亚健康的 20 项指标,以调查问卷形式询问被调查者,得分大于 80 分者为健康对象,反之为亚健康对象<sup>[14]</sup>.

#### 2.1.2 主波提取

基于香农能量和希尔伯特变换的脉搏波主波提取算法过程:脉搏波原始信号首先经过带通滤波去除高频噪声(工频噪声)和低频噪声(基线漂移);然后,进行幅值归一化;归一化后的脉搏波再进行香农能量变换,经过低通平滑滤波得到香农能量包络线;包络线在希尔伯特变换之后减去均值,滤除信号因突变产生的干扰,在此基础上,找出香农能量包络线的峰值点.对比和分析大量的香农能量包络线峰值点和只经过带通滤波的脉搏波峰值点,发现二者的出现时间相差 0~0.25 秒.因此,香农能量包络线的峰值点的位置前后约 0.25 秒的范围对应着脉搏波主波位置所在的时间范围.在此范围内,可以找到脉搏波真正的主波位置<sup>[14]</sup>.

#### 2.1.3 划分周期

为了保证每个分段至少包含一个完整的脉搏波周期,其周期是 1.25 秒,因此选择 250 个点作为一个周期,以主波位置为基准,向前取 99 个点,向后取 150 个点,即可作为一个周期,如图 3 所示<sup>[14]</sup>.每条数据增加一列标签,0 和 1 分别代表健康和亚健康.共计 118 800 条  $1 \times 250$  周期,加标签 0 或 1 后,变为  $1 \times 251$ ,随机选择 60% 的数据为训练集,即 71 280 条,40% 为测试集,也是评价集,即 47 520 条.

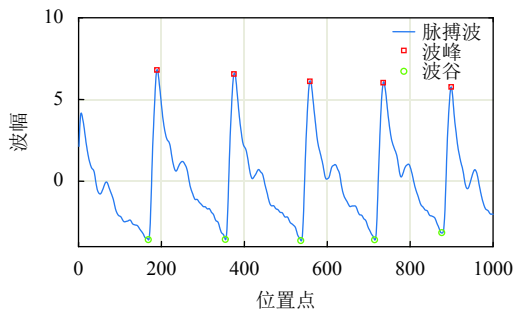


图3 脉搏波信号

由图3可知, (1) 脉搏波信号是典型的时间序列数据, 前后文相关性很大. (2) 升支、主波、重波前波、降中峡等是脉搏波具有生理意义的重要特征, 是脉诊的主要依据, 比如升支反映了心室收缩、快速射血、主动脉血量增加, 此时动脉血压升高, 出现陡峭而平滑的脉搏波曲线, 继而出现主波, 主波峰值点对应收缩压, 即收缩期的最高血压<sup>[14]</sup>. 但它们在各自周期内的波幅、出现时间(或出现位置点)等均不同. (3) 存在噪声, 比如在每个周期将结束时, 有轻微波动. 因此, 需要针对脉搏波数据的自身特征, 构建合适的网络结构.

#### 2.1.4 制作 hdf5 数据集

Caffe 作为深度学习的平台, 处理脉搏波数据时, 需将数据集转化成 Caffe 可读的 hdf5 格式的数据集.

## 2.2 IRNet 及 MIRNet 的结构和原理

### 2.2.1 网络融合的原因

GoogLeNet 中的 Inception 模块及其变种存在缺点: (1) 基本的 Inception 模块, 对网络性能提升的作用有限, 其变种则过于复杂, 参数量和运算量成为负担, 过拟合现象频发; (2) 网络宽度较宽, 深度不足, 宽度与深度失衡, 参数运算效率低. ResNet 中的残差模块亦有不足: (1) 虽然加深了网络, 提升了网络的分类准确率, 但参数量和运算量却快速增长, 比 Inception 模块及其变种增长快; (2) 网络结构加深, 但宽度较窄, 深度与宽度失衡, 特征提取的多样性比 Inception 模块差; (3) 如果残差模块过于复杂, 则跳跃连接带来的训练加速作用弱于参数量和运算量急剧增加带来的训练减速作用, 易导致训练中断或梯度爆炸. 在一定程度上, Inception 模块和残差模块可以优势互补.

此外, 根据脉搏波的特点: (1) 残差模块中的跳跃连接, 可认为是没有门机制的 LSTM, 门机制允许一部分数据经过简单特征提取后直接进入下一层; 然而,

LSTM 结构不能加深, 单个模块的性能优于简单的 CNN, 但其性能低于适当加深的 ResNet 和 GoogLeNet. 跳跃连接基本适合于脉搏波的第一个特点. (2) Inception 模块各分支的特征提取, 能获得多样的特征, 此基本适合于脉搏波的第二个特征.

鉴于以上初步分析, 提出由 Inception 模块和残差模块组成的融合网络 Inception-ResNet (IRNet).

另外, 研究了 CNN9、CNN16、VGGNet、ResNet 及 GoogLeNet 等网络用于脉搏波识别. CNN9 和 CNN16, 包含卷积层、池化层、全连接层及 Dropout 层, 共计 9 层和 16 层, 每个卷积层都有 BatchNorm、Scale、ReLU 等. VGGNet 中池化层数量少于 CNN16, 但总层数相近. ResNet1 和 ResNet2 使用不同的残差模块, 卷积核大小和数量也不同, ResNet1 由残差模块 res2(a, b)、res3(a, b)、res4(a, b) 和 res5(a, b) 组成; ResNet2 由残差模块 res2(a, b, c)、res3(a, b, c, d) 和 res4(a, b, c, d, e) 组成. GoogLeNet1、GoogLeNet2 和 GoogLeNet3 分别由 1、2、3 个 Inception 模块组成.

### 2.2.2 IRNet 结构

IRNet 是由基本的 Inception 模块和残差模块构成, 包括三个网络, 依次命名为 IRNet1、IRNet2 和 IRNet3, 主要思路是构建 Inception 模块和残差模块融合的神经网络, 经过实验, 调整各模块的排列顺序、连接方式和模块中的组成单元, 比如模块中的卷积层数量、池化层数量、卷积核大小、通道数量等. 三个融合网络的共同点是: (1) 网络中, 很多卷积层都包括 BatchNorm (BN)、ReLU(R) 和 Scale(S). BN 是一种有效的正则化方法, 加速大规模神经网络的训练, 一般和 Scale 同时使用<sup>[18]</sup>; (2) 网络的起始部分, 都用卷积核较大的卷积层, 降低参数量和运算量, 网络中的全连接层, 都用到 Dropout 层, 避免过拟合.

在实验中, 当两个 Inception 模块直接相连时, 网络的分类性能会下降; 当三个 Inception 模块直接相连时, 训练时会发生梯度爆炸. 如果残差模块过于复杂, 也会发生梯度爆炸. IRNet1 由一个 Inception 模块 (first Inception Module, IM1) 和一个残差模块[res3(a, b, c, d)]构成; IRNet2 由两个 Inception 模块 (IM1 和 Second Inception Module, IM2) 和一个残差模块[res3(a, b, c, d)]构成; IRNet3 由两个 Inception 模块 (IM1 和 IM2) 和两个残差模块[res3(a, b, c, d) 和 res4(a, b, c)]构成. 当第二个残差模块为 res4(a, b, c, d, e) 时, IRNet3 训练发生

中断;当继续增加 Inception 模块和残差模块数量,网络的分类性能会反而下降.残差模块 res3(a, b, c, d) 如图 4 所示, IM2\_output、res3a 等表示层或模块的名称, (1×4+2) 表示卷积核是 1×4 和步长是 2, 256、512 等表示输出通道数, 1 和 0 表示填充零和不填充零, R 表示 ReLU 层, Eltwise/SUM 表示特征拼接方式. IM1 如图 5 所示, pool1、IM1\_output、IM1\_conv3\_1 等表示该层名称, 8、16、32 等表示通道数, Concat 表示另一种特征拼接方式. Eltwise/SUM 是在分辨率层面拼接, 输入维度 (batch\_size, channel, w, h) 必须相同, 增加算法耗时; Concat 在通道数 (channel) 层面拼接, 计算相对简单.

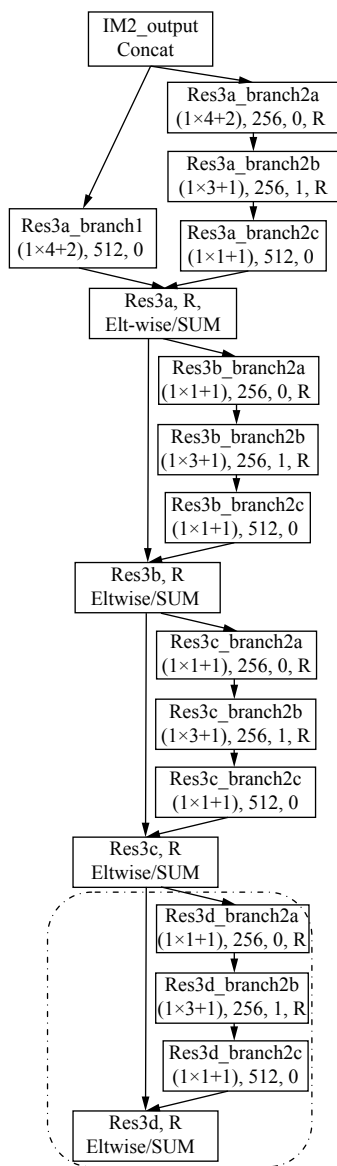


图 4 res3(a, b, c, d) 的结构图. res4(a, b, c) 和 res3 的结构相似, 但 res4 不含在虚线框中的“res3d”, 即, 没有 res4d

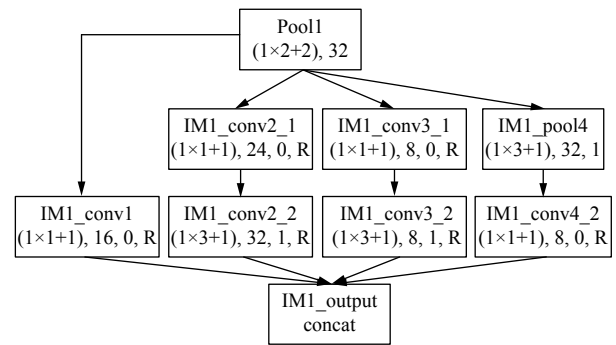


图 5 IM1, 和 IM2 结构相似, 区别在于卷积核大小和个数不同

在 IRNet 中, IRNet3 有更好的分类性能, 网络结构如图 6, BN 和 S 表示 BatchNorm 层和 Scale 层, fc7 和 fc8 表示全连接层 (fully-connected layer), drop 表示 Dropout 层, 随机丢弃 50% 的连接, IP 表示全连接层的通道数, loss 表示 SoftmaxWithLoss 层.

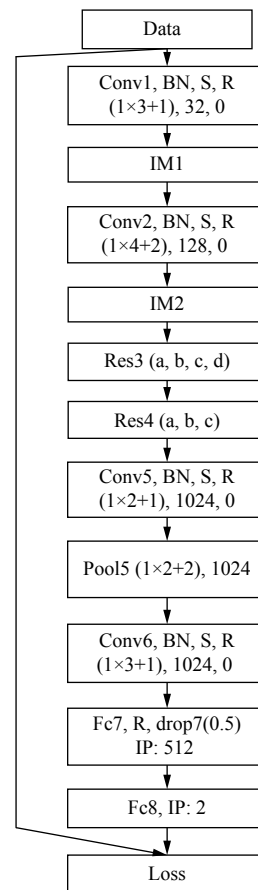


图 6 IRNet3 的结构图

### 2.2.3 MIRNet 的结构和原理

#### 2.2.3.1 MIRNet 结构

在实验中, IRNet3 的识别准确率与 ResNet2 相近,

但是 ResNet2 的参数数量和运算量更大, 其残差模块 res2 (a, b, c)、res3(a, b, c, d) 和 res4(a, b, c, d, e) 是参数数量和运算量急剧增长的主要因素. 在 IRNet3 基础上, 提出由改进的 Inception 模块、残差模块和池化模块组成的融合网络 MIRNet, 主要思路是: (1) 根据脉搏波第一特点, 残差模块的跳跃连接不可缺少, 但是要降低模块复杂度, 特征拼接方式由 Eltwise/SUM 改为 Concat, 弱化其带来的参数数量和运算量增加等负作用. (2) 根据脉搏波的第二特点, Inception 模块应更丰富, 增加模块中的分支, 使用较小的卷积核, 但是要保持高效的稀疏结构. 特征拼接方式为 Concat, 增加更多的特征表示, 保持相对较小的运算量. 这些改进有助于提取不同阶的抽象特征, 增强模块的表达能力. (3) 根据脉搏波第三特点, 在改进模块的分支网络中, 最后一层是 ReLU, 可以认为是加入特征提取权重, 随着网络的训练而自动调节各分支网络提取的特征; (4) 取消单独使用的池化层, 改为池化模块, 源于 (2) 的分析, 但和改进 Inception 模块不同的是, 其减小了特征面, 起到池化层的作用.

MIRNet 主要有两个改进网络, MIRNet1 和 MIRNet2, MIRNet1 主要作用是研究改进模块组成融合网络的可行性, MIRNet2 是主要改进的融合网络. 改进模块: (1) 利用残差模块的原理, 构造 pool3/conv3, 使池化层 pool3 和卷积层 conv3 并列为网络的两个分支; 构造 conv4\_1~conv4\_5 等模块, 精简原有的残差模块, 拼接方式改为 Concat, 主要是减少参数数量和运算量. (2) 在改进的 Inception 模块 (Modified Inception Modules, MIM1 和 MIM2) 中, 使用 1×1、1×2、1×3 等较小卷积核, 有利于提取更精细的特征. 1×1 卷积核改变层间连接数量, 对输出通道升维或降维, 增加特征的非线性变换, 提高网络的泛化能力, 有侧重地保留不同层次的高阶特征<sup>[9]</sup>; (3) 池化模块 (Reduction Modules, RM), 改变以往使用单一池化层的方式, 包含三个分支, 一支是池化层, 另外两支是卷积层, 使特征面分别池化和卷积, 减小特征面, 保证提取特征的多样性. MIRNet1 由 MIM1 和 RM1 组成, MIRNet2 在 MIRNet1 可行的基础上增加 MIM2 和 RM2. MIM2 和 RM2 分别如图 7 和图 8 所示, RM1\_concat、RM2\_pool1、MIM2\_conv3\_3 等表示该层的名称, 连接方式为 Concat. 图 9 是性能更优的 MIRNet2 的结构.

### 2.2.3.2 MIRNet 的原理

根据 MIRNet 的构造思路和网络结构, 模块是网

络的主要组成, 均可抽象成由不同的特征提取过程组成的特征提取模块.

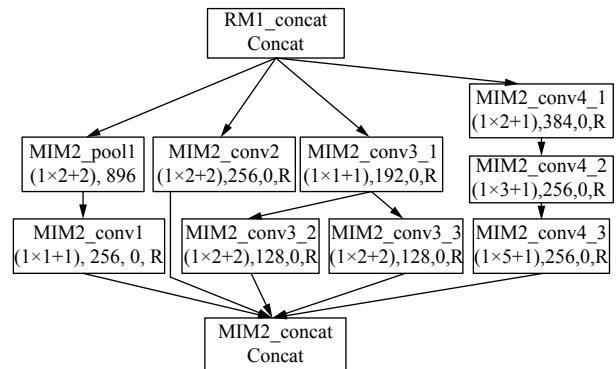


图 7 MIM2, 与 MIM1 相比, 有两个分支 MIM2\_conv3\_2 和 MIM2\_conv3\_3, 而 MIM1 只有一个分支 MIM1\_conv3\_2, 二者的卷积核大小和数量不同.

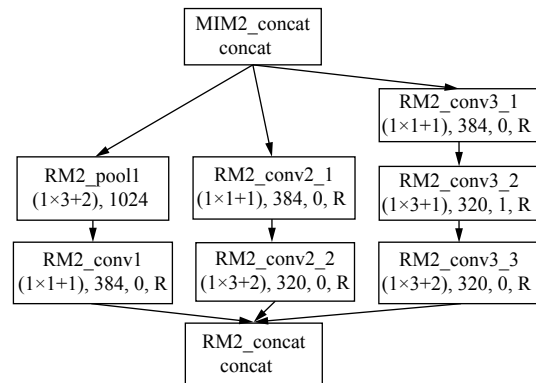


图 8 RM2, 与 RM1 相比, RM1 在池化层 RM2\_pool1 下面没有 RM2\_conv1, RM2\_conv1 的作用是减少层间连接数, 且二者卷积核大小和数量不同.

由式 (1)、(2)、(3) 可得, 在特征拼接时, 去掉 ReLU 函数, 即

$$x_{l+1} = y_l \quad (5)$$

$h(x_l)$  替换为卷积、ReLU 等操作函数, 即增加分支, 可得

$$x_L = [R_{11}(w_{11}x_l + b_{11}) + R_{12}(w_{12}R_{11} + b_{12})] + [R_{21}(w_{21}x_l + b_{21}) + R_{22}(w_{22}R_{21} + b_{22})] \quad (6)$$

$L$  表示某模块,  $x_L$  表示某模块的输出,  $x_l$  表示某模块的输入,  $R_{11}$  和  $R_{12}$  表示第 1 个分支中的两个卷积和 ReLU 操作, 同理  $R_{21}$  和  $R_{22}$  表示第 2 个分支中的两个卷积和 ReLU 操作,  $w$  和  $b$  分别表示权重和偏置.

假设, 某模块有  $n$  个分支, 每个分支中有  $m$  个卷积和 ReLU 函数, 则

$$x_L = \sum_{i=1}^n \left[ R_{i1}(w_{i1}x_i + b_{i1}) + \sum_{j=2}^m R_{ij}(w_{ij}R_{i,j-1} + b_{ij}) \right] \quad (7)$$

反向传播时, 损失函数 $S$ 的导数, 由链式求导原则得

$$\frac{\partial S}{\partial x_i} = \frac{\partial S}{\partial x_L} \frac{\partial x_L}{\partial x_i} = \frac{\partial S}{\partial x_L} \sum_{i=1}^n \left( r w_{i1} + r^2 w_{i1} w_{i2} + \dots + r^m \prod_{j=1}^m w_{ij} \right) \quad (8)$$

由式(8)得:

$$\frac{\partial S}{\partial x_i} = \frac{\partial S}{\partial x_L} \sum_{i=1}^n \left[ \sum_{j=1}^m \left( r^j \prod_{j=1}^m w_{ij} \right) \right] \quad (9)$$

式中,  $\frac{\partial S}{\partial x_L}$  表示损失函数对模块输出 $x_L$ 求导,  $\frac{\partial x_L}{\partial x_i}$  表示 $x_L$ 对各分支的输入 $x_i$ 求导,  $r$  是  $R$  的导数,  $r = \{0, 1\}$ ,  $w_{ij}$  表示第  $i$  分支中第  $j$  个卷积层的权重,  $r^j \prod_{j=1}^m w_{ij}$  表示第  $i$  分支中第  $j$  个卷积和 ReLU 等函数的导数,  $\sum_{j=1}^m \left( r^j \prod_{j=1}^m w_{ij} \right)$  表示第  $i$  分支中  $j$  个卷积和 ReLU 等函数的导数和. 式(7)和式(8)表明改进模块对特征不同程度地抽象变换.

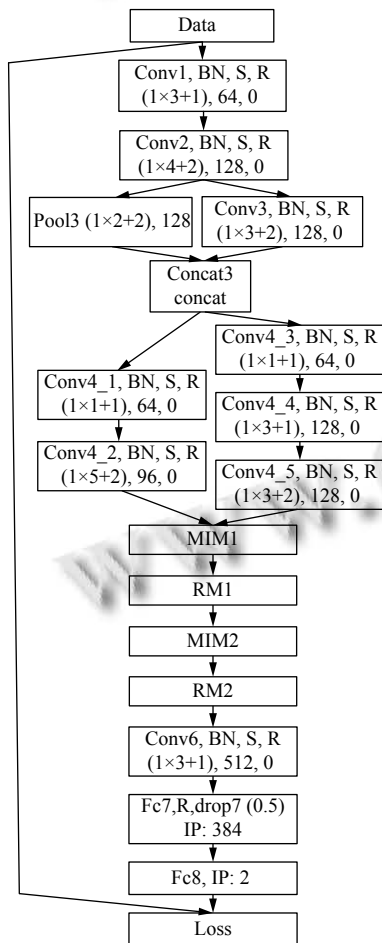


图9 MIRNet2 的结构图

MIRNet2 模块的分支数量多为 2、3、5 个, 比如 pool3/conv3、conv4\_1~conv4\_5、RM2 和 MIM2 等, 分支中的单元含 1、2、3 个卷积和 ReLU 等操作, 模块深度适当. 训练过程中, ReLU 自动调节梯度衰减, 当遇到脉搏波次要特征和噪声时, 导数为 0, 梯度不变, 损失函数未减小, 继续寻找使损失函数减小的梯度下降路径; 当遇到重要特征, 导数为 1, 梯度减小, 损失函数减小, 准确率增大, 学习到重要特征. 弱化噪声和次要特征, 使分支网络提取到重要特征. 各改进模块对特征进行不同程度地变换, 提取不同层次的高阶特征和前后文的相关性特征, 保持较低的数量和运算量. 但是, 模块中的分支也不应过多, 网络宽度和深度应平衡.

### 2.3 参数量和运算量计算

网络中, 池化层、BN、ReLU、Concat 等参数量和运算量很小, 相对于卷积层, 可忽略不计. 某卷积层的参数量  $P$ :

$$P = C * (K_1 * K_2 + 1) \quad (10)$$

某卷积层的运算量  $OP$ :

$$OP = C_1 * C_2 * W * H * (K_1 * K_2 + 1) \quad (11)$$

式中,  $C$  是该卷积层的通道数,  $C_1$ 、 $C_2$  是上下层的通道数;  $K_1$ 、 $K_2$  是卷积核大小, 1 表示一个偏置;  $W$  和  $H$  是特征面的宽和高.

由式(10)和式(11)可知, 参数量取决于通道数量和卷积核大小; 运算量取决于上下层的通道数量、卷积核大小和特征面大小.

### 2.4 性能评价

CNN 的分类性能通过准确率 (Accuracy,  $ACC$ )、特异性 (Specificity,  $SP$ )、灵敏度 (Sensitivity,  $SE$ ) 和约登指数 (Youden's index,  $\gamma$ ) 等评价.  $ACC$  表示算法识别的准确率, 是算法分类性能的综合指标;  $SP$  表示算法检测出阴性样本 (健康) 的能力;  $SE$  表示算法检测出阳性样本 (亚健康) 的能力;  $\gamma$  表示算法避免失败的能力, 其值越大表明算法避免失败的能力越强. 定义和计算如下:

真阳性 (True Positive,  $TP$ ): 算法检测正确的阳性样本数量;

真阴性 (True Negative,  $TN$ ): 算法检测正确的阴性样本数量;

假阳性 (False Positive,  $FP$ ): 算法检测错误的阳性样本数量;

假阴性 (False Negative,  $FN$ ): 算法检测错误的阴性样本数量.

$$SE = \frac{TP}{TP + FN} \quad (12)$$

$$SP = \frac{TN}{TN + FP} \quad (13)$$

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \quad (14)$$

$$\gamma = SE + SP - 1 \quad (15)$$

### 3 实验与结果分析

本节介绍了实验环境、网络参数和训练参数, 重点阐述实验结果与分析等.

#### 3.1 实验环境和参数设置

##### 3.1.1 实验环境

实验环境如表 1 所示.

表 1 实验环境

类别	名称	类型	性能
硬件	CPU	Intel Xeon	6 核, 1.7 GHz
	GPU	GTX TiTan X	5TFLOPs
软件	系统	Ubuntu 16.04	
	深度学习框架	Caffe	

注: 运算量单位是每秒万亿次浮点计算 TFLOPs

##### 3.1.2 参数设置

网络参数: (1) 使用较小的卷积核,  $kernel\_h$  为 1,  $kernel\_w$  为 1、3、5 等, 步长  $stride\_h$  为 1,  $stride\_w$  为 1 或 2. 随着卷积层数的增加, 通道数和层间连接数也会增加. 全连接层, 一般不应超过 1024 个连接节点, 否则会梯度爆炸. (2) 网络中的卷积层:  $convolution\_param$  的  $weight$  和  $bias$  初始化为“msra”和“false”. BatchNorm 层中,  $batch\_norm\_param$  { $use\_global\_stats$ :false}; Scale 层中,  $scale\_param$ { $bias\_term$ :false}, 初始化为 false, 方可在训练中不断优化  $batch\_norm\_param$  和  $scale\_param$  的值, 初始化为 True, 则不会优化. 全连接层  $inner\_product\_param$  中  $weight\_filler$  和  $bias\_filler$  分别初始化为“Xavier”和“constant”.

训练参数:  $batch\_size$  设为 10,  $test\_iter$  设为 500,  $test\_interval$  设为 400, 称为一代 (epoch),  $Max\_iter$  设为 35 000, 完成对测试集所有数据 8~9 次的测试, 共 87 代. 经实验分析, 学习策略  $lr\_policy$  设为“inv”,  $gamma$  是  $10^{-3}$ ,  $power$  是 0.75,  $base\_lr$  是  $10^{-5}$ ,  $momentum$

是 0.9,  $weight\_decay$  是  $5 \times 10^{-6}$ , 这组训练参数更适于改进后的网络.

$$lr = base\_lr * (1 + gamma * iter)^{-power} \quad (16)$$

式 (16) 中,  $lr$  是学习率,  $base\_lr$  是初始学习率,  $gamma$  是  $lr$  的衰减系数,  $iter$  是迭代次数,  $power$  是指数. 代入参数, 简化可得:

$$lr = \left[ 10^5 * (1 + 10^{-3} iter)^{0.75} \right]^{-1} \quad (17)$$

学习率与迭代次数的图像如图 10.

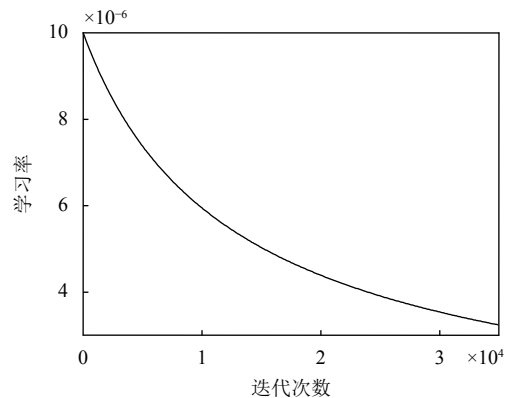


图 10 学习率与迭代次数曲线

由图 10 可知,  $lr$  随  $iter$  减小, 斜率先大后小,  $lr$  先快速减小, 随后缓慢减小.  $Loss$  随着学习率的减少而减小, 先快速找到极小值点, 再缓慢减小学习率找到最小值点, 即找到准确率最优点. 经过实验,  $base\_lr$  设为  $10^{-5}$ , 准确率随着迭代次数增大而增大, 大于  $10^{-5}$  时, 准确率较低且难以找到最优解, 此设置将更节省计算资源.

#### 3.2 结果与分析

经过实验, 各网络测试结果如表 2 所示, 选取相似网络中性能较好的 CNN9、CNN16、ResNet2、GoogLeNet2、IRNet3、MIRNet2 等, 绘制其训练过程中的测试准确率曲线, 如图 11、图 12 和图 13 所示.

CNN9: CNN(9L) 是 Hu 提出的网络模型, 用于脉搏波识别, 但是其只使用了基本的卷积层、池化层和全连接层. CNN9 相比于 CNN(9L), 每个卷积层中, 包含 BatchNorm、Scale 和 ReLU, 并在全连接层中含有 Dropout 层, 准确率提升了 10.8%. 其中, BatchNorm 和 Scale 的主要作用是加速网络训练, Dropout 的主要作用是避免过拟合, ReLU 有助于提升准确率. 在训练过程中, ReLU 通过调节梯度衰减, 突出脉搏波重要特征、次要特征和噪声之间的差别, 增强重要特征在分



类中的影响, 弱化次要特征和噪声, 提高了脉搏波识别的准确率. 在此过程中, 未损害重要特征的完整性, 只是增加对重要特征的高阶变换, 使其在高维空间线性可分.

表2 各网络的分类性能

神经网络	SP(%)	SE(%)	ACC(%)	$\gamma$ (%)
CNN(9L) <sup>[14]</sup>	75.64	68.99	72.31	44.63
CNN9	82.34	83.98	83.11	66.32
CNN16	85.06	84.21	84.94	69.27
VGGNet	83.91	77.46	83.31	61.37
ResNet1	85.36	84.75	85.53	70.11
<b>ResNet2</b>	<b>87.34</b>	<b>86.58</b>	<b>86.97</b>	<b>73.92</b>
GoogLeNet1	83.17	84.06	83.58	67.23
GoogLeNet2	85.79	83.54	85.45	69.33
GoogLeNet3	84.21	81.82	84.84	66.03
IRNet1	86.37	85.16	86.26	71.53
IRNet2	84.60	87.24	86.72	71.84
<b>IRNet3</b>	<b>87.24</b>	<b>87.65</b>	<b>86.98</b>	<b>74.89</b>
MIRNet1	87.28	86.33	86.91	73.61
<b>MIRNet2</b>	<b>87.85</b>	<b>88.05</b>	<b>87.84</b>	<b>75.90</b>

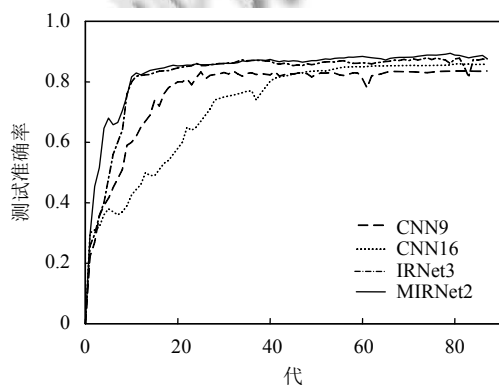


图11 CNN9、CNN16、IRNet3 和 MIRNet2 测试准确率曲线

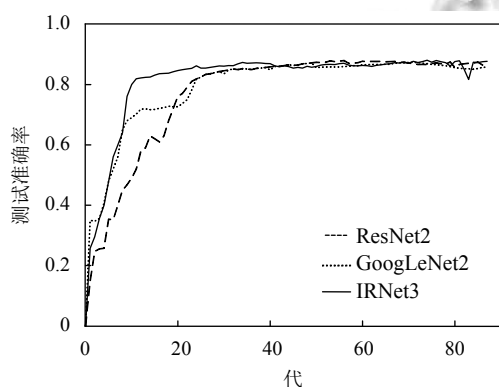


图12 ResNet2、GoogLeNet2 和 IRNet3 测试准确率曲线

CNN16 和 VGGNet: 相比于 CNN9, CNN16 结构更深, 准确率提高了 1.83%, 而 VGGNet 仅提升了 0.2%.

CNN16 比 VGGNet 的准确率高出 1.63%,  $\gamma$  高出 7.9%, 避免失败的能力提高. 图 11 所示, 训练过程中, CNN9 比 CNN16 更快达到稳定, 加深网络使训练速度减小, 但对网络分类性能的提升作用明显. CNN16 有更多的通道数、更小的卷积核, 网络结构和参数设置也优于 VGGNet.

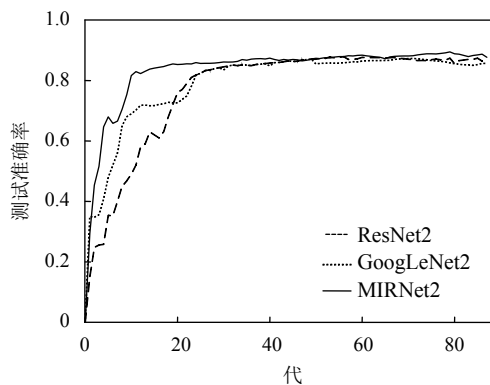


图13 ResNet2、GoogLeNet2 和 MIRNet2 测试准确率曲线

ResNet: 相比于 CNN16, ResNet1 由 4 个简单的残差模块组成, 网络层数更深, 准确率高出 0.59%,  $\gamma$  高出 0.84%. 相比于 ResNet1, ResNet2 由 3 个复杂的残差模块组成, 跳跃连接的作用更明显, 加深网络, 并未使网络臃肿, 准确率高出 1.44%,  $\gamma$  高出 3.81%, 算法避免失败的能力更高, 对阴性样本和阳性样本的测试性能更好. 但是, 在实验中, 继续增加残差模块, 使网络更深, 则导致网络训练中断.

GoogLeNet: 3 个改进网络, 分别是 GoogLeNet1、GoogLeNet2 和 GoogLeNet3, 依次使用了 1、2、3 个 Inception 模块. 表 2 中, GoogLeNet1 和 GoogLeNet3 的准确率都低于 CNN16, 仅 GoogLeNet2 的准确率高于 CNN16, 对比图 11 和图 12, GoogLeNet2 准确率达到稳定的时间快于 CNN16, 表明在网络深度接近时, 加宽能改善网络性能. 由表 2 和图 12 可知, ResNet1 和 ResNet2 的准确率都高于 GoogLeNet2, 尤其是 ResNet2 比 GoogLeNet2 高出 1.52%, 但其准确率达到稳定状态的速度慢于 GoogLeNet2. ResNet2 的网络结构更深, 残差模块的作用更明显, 加深比加宽对网络性能的提升作用更明显. 当两个以上的 Inception 模块直接相连时, 对脉搏波的分类准确率反而下降.

IRNet: 相比于 ResNet1 和 GoogLeNet2, IRNet1 和 IRNet2 的分类性能更优, 表明网络融合可以吸收各网

络的优势. IRNet 之间, IRNet3 最优, 准确率是 86.98%, 接近 ResNet2 的准确率, 但  $\gamma$  高出了 0.97%, 避免识别失败的能力较好. 图 12 中, 对比 ResNet2、GoogLeNet2 和 IRNet3, ResNet2 是三者中最深且复杂的, 准确率达到稳定状态的时间较长; GoogLeNet2 的结构相对简单, 特征提取效果在开始时优于 ResNet2, 但稳定后的准确率低于 ResNet2; IRNet3, 较快达到稳定的准确率, 得益于 Inception 模块的高效稀疏结构, 减少了参数量和运算量, 保证了特征提取的多样性. 如表 3 所示, res3(a,b,c,d) 和 res4(a,b,c) 是参数量和运算量增加的主要模块. 二者在稳定状态的准确率接近, 但 IRNet3 对样本的测试准确率较稳定, 提高了融合网络的鲁棒性. 当继续增加 Inception 模块和残差模块时, 训练发生中断.

表 3 IRNet3 的参数量和运算量

序号	IRNet3	参数量(个)	运算量(MFLOPs)
1	conv1	128	0.0317
2	IM1	272	1.6824
3	conv2	640	5.0381
4	IM2	1056	6.3713
5	res3(a,b,c,d)	13 568	112.1321
6	res4(a,b,c)	18 944	150.4707
7	conv5	3072	40.8945
8	conv6	4096	4.1943
总计		41 776	320.8151

注: 运算量单位是每秒百万次浮点计算 MFLOPs, 下同

MIRNet: 主要有两个改进网络 MIRNet1 和 MIRNet2. 如表 3、表 4、图 14 和图 15 所示, IRNet3 前 4 个层或模块的参数量和运算量分别是 2096 个和 13.1235 MFLOPs, MIRNet2 前 4 个层或模块的参数量和运算量分别是 3136 个和 14.9935 MFLOPs, 均大于 IRNet3, 是因为 MIRNet2 使用的卷积核比 IRNet3 更大, 这使得 MIRNet2 后面的层或模块的输入特征面减小, 为网络总参数量和总运算量降低奠定了基础. MIRNet2 的总参数量和总运算量分别是 IRNet3 的 0.483 倍和 0.230 倍, 残差模块是 IRNet3 参数量和运算量增大的主要原因. 然而, 这并未影响网络的分类性能, 如图 11 所示, 在训练起始阶段, MIRNet2 的准确率上升快于 IRNet3, 归因于改进模块 pool3/conv3 和 conv4\_1~conv4\_5 等极大地减少了参数量和运算量; 在整个训练过程中, MIRNet2 的准确率高于 IRNet3, 只在部分区域接近, 得益于改进的 Inception 模块(MIM1 和 MIM2) 及池化模块(RM1 和 RM2). 表 2 中, MIRNet2 的准确率是 87.84%, 分别比 IRNet3 和 ResNet2 高出 0.86% 和

0.87%,  $\gamma$  分别比 IRNet3 和 ResNet2 高出 1.01% 和 1.98%, 算法避免失败的能力有所提高, 测试性能更加稳定. 综合分析图 11 和图 13, MIRNet2 的准确率优于 ResNet2 和 GoogLeNet2, 表明改进的融合网络比使用单一核心模块的网络性能更优; MIRNet2 的准确率高于 IRNet3, 表明改进模块对提升融合网络的性能很重要, 网络融合不能仅借鉴原有的残差模块和 Inception 模块, 需要改进核心模块、网络结构等, 否则并不能充分发挥网络融合的优势. 改进的 Inception 模块及池化模块, 能提取不同层次的高阶特征和前后文的相关性特征, 保持了较少的参数量和运算量; ReLU 是改进模块中每个分支的最后一层, 训练中自动调节梯度衰减, 进而影响各分支网络的特征提取, 增加对特征的高阶变换, 而不损害重要特征的完整性, 使更多在低维空间线性不可分的特征在高维空间线性可分. 这些提高了融合网络对脉搏波的分类性能.

表 4 MIRNet2 的参数量和运算量

序号	MIRNet2	参数量(个)	运算量(MFLOPs)
1	conv1	256	0.0635
2	conv2	512	4.0305
3	pool3/conv3	512	3.9977
4	conv4_1~conv4_5	1856	6.9018
5	MIM1	1792	6.6519
6	RM1	3072	11.2722
7	MIM2	5120	22.3642
8	RM2	4992	16.9083
9	conv6	2048	1.4418
总计		20 160	73.6317

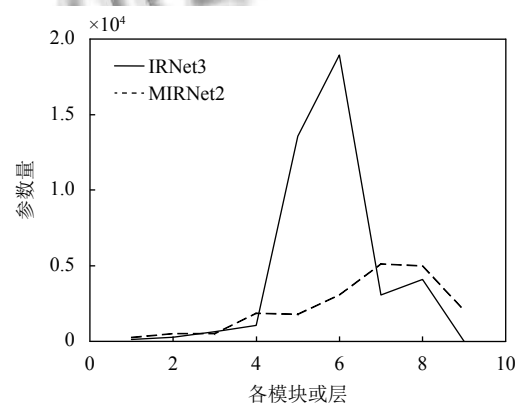


图 14 各模块或层的参数量

#### 4 结论与展望

为推动计算机辅助脉诊的临床应用, 需要提高脉搏波识别的准确率. 通过主波提取、划分周期和制作 hdf5 数据集, 获得 Caffe 可处理的 hdf5 格式的脉搏波

数据集. 研究了 CNN9、CNN16、ResNet1、ResNet2、GoogLeNet1、GoogLeNet2、GoogLeNet 3 及深度融合神经网络 IRNet1、IRNet2、IRNet3 和 MIRNet1 等, 根据脉搏波数据特点和其他网络的实验分析, 提出 MIRNet2, 其分类性能最优, 准确率和约登指数分别是 87.84% 和 75.90%, 改造的残差模块 pool3/conv3 和 conv4\_1~conv4\_5 大幅减少了参数量和运算量, 改进的 Inception 模块和池化模块, 能提取不同层次的高阶特征和前后文的相关性特征, 是 MIRNet2 性能提升的关键. 当网络加深时, 跳跃连接和层间稀疏连接有利于加速网络训练和避免梯度爆炸; 当网络层数接近时, 适当加宽网络, 有利于提取更丰富的特征. 但是, 要防止过多的层间连接和多个复杂模块直接相连, 否则将梯度爆炸或训练中断. MIRNet2 的分类性能优于现有的脉搏波识别方法, 相比于 Hu 提出的 CNN(9L) 的准确率和约登指数, 分别高出 15.53% 和 31.27%, 有较大提升, 将 MIRNet2 嵌入到云计算平台来识别健康和亚健康的脉搏波, 这是一个可行的应用场景<sup>[17]</sup>. 然而, 关于脉搏波标准数据集构建、GAN 及其融合网络用于脉搏波识别等问题尚未解决, 这些是未来需要深入研究的方向.

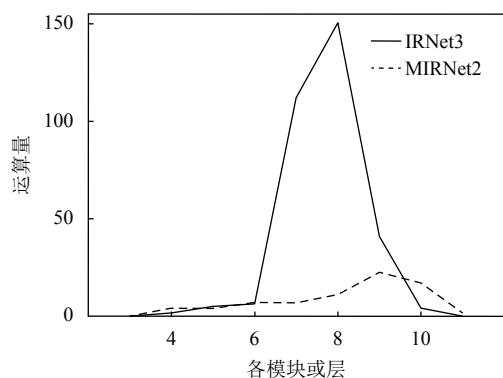


图 15 各模块或层的运算量

### 参考文献

- 1 LeCun YL, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278–2324. [doi: 10.1109/5.726791]
- 2 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097–1105.
- 3 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv pre-print arXiv: 1409.1556, 2015.
- 4 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 1–9. [doi: 10.1109/CVPR.2015.7298594]
- 5 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778. [doi: 10.1109/CVPR.2016.90]
- 6 Liu D, Wang ZW, Fan YC, *et al.* Learning temporal dynamics for video super-resolution: A deep learning approach. IEEE Transactions on Image Processing, 2018, 27(7): 3432–3445. [doi: 10.1109/TIP.2018.2820807]
- 7 Khosravi P, Kazemi E, Imielinski M, *et al.* Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. Ebiomedicine, 2018, 27: 317–328. [doi: 10.1016/j.ebiom.2017.12.026]
- 8 Vesal S, Ravikumar N, Davari A, *et al.* Classification of breast cancer histology images using transfer learning. Proceedings of the 15th International Conference on Image Analysis and Recognition. Póvoa de Varzim, Portugal. 2018. 812–819. [doi: 10.1007/978-3-319-93000-8\_92]
- 9 Zhang RY, Wang QQ, Lu Y. Combination of ResNet and center loss based metric learning for handwritten Chinese character recognition. Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition. Kyoto, Japan. 2017. 25–29. [doi: 10.1109/ICDAR.2017.324].
- 10 Lu Z, Jiang XD, Kot A. Deep coupled ResNet for low-resolution face recognition. IEEE Signal Processing Letters, 2018, 25(4): 526–530. [doi: 10.1109/LSP.2018.2810121]
- 11 Hahibzadeh M, Jannesari M, Rezaei Z, *et al.* Automatic white blood cell classification using pre-trained deep learning models: ResNet and inception. Proceedings Volume 10696, Tenth International Conference on Machine Vision. Vienna, Austria. 2017. 1–8. [doi: 10.1117/12.2311282]
- 12 Szegedy C, Ioffe S, Vanhoucke V, *et al.* Inception-V4, Inception-ResNet and the impact of residual connections on learning. 2017 31st AAAI Conference on Artificial Intelligence. San Francisco, CA, USA. 2017. 4278–4284.
- 13 Zhuang NF, Ye J, Hua KA. Convolutional DLSTM for crowd scene understanding. Proceedings of 2017 IEEE International Symposium on Multimedia. Taichung, Taiwan, China. 2018. 61–68. [doi: 10.1109/ISM.2017.19]
- 14 Hu XJ, Zhu HH, Xu JT, *et al.* Wrist pulse signals analysis based on deep convolutional neural networks. Proceedings of 2014 IEEE Conference on Computational Intelligence in

- Bioinformatics and Computational Biology. Honolulu, HI, USA. 2014. 1–7.
- 15 Zhang SR, Sun QF. Human pulse recognition based on convolutional neural networks. Proceedings of 2016 International Symposium on Computer, Consumer and Control. Xi'an, China. 2016. 366–369. [doi: [10.1109/IS3C.2016.101](https://doi.org/10.1109/IS3C.2016.101)]
- 16 Zhou FY, Jin LP, Dong J. Premature ventricular contraction detection combining deep neural networks and rules inference. Artificial Intelligence in Medicine, 2017, 79: 42–51. [doi: [10.1016/j.artmed.2017.06.004](https://doi.org/10.1016/j.artmed.2017.06.004)]
- 17 Jin LP, Dong J. Intelligent health vessel ABC-DE: An electrocardiogram cloud computing service. IEEE Transactions on Cloud Computing, 2018, 6: 1–14. [doi: [10.1109/TCC.2018.2825390](https://doi.org/10.1109/TCC.2018.2825390)]
- 18 Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning. Lille, France. 2015. 448–456.
- 19 Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the Inception architecture for computer vision. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 2818–2826. [doi: [10.1109/CVPR.2016.308](https://doi.org/10.1109/CVPR.2016.308)]