



面向掌纹掌静脉识别网络轻量化的非对称双模态融合方法

林孙旗, 徐家梦, 郑瑜杰, 王翀, 王军

引用本文:

林孙旗, 徐家梦, 郑瑜杰, 王翀, 王军. 面向掌纹掌静脉识别网络轻量化的非对称双模态融合方法[J]. 智能系统学报, 2024, 19(5): 1190–1198.

LIN Sunqi, XU Jiameng, ZHENG Yujie, et al. An asymmetric bimodal fusion method for lightweight palm print and palm vein recognition network[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(5): 1190–1198.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202212031>

您可能感兴趣的其他文章

基于级联宽度学习的多模态材质识别

Cascade broad learning for multi-modal material recognition

智能系统学报. 2020, 15(4): 787–794 <https://dx.doi.org/10.11992/tis.201908021>

深度学习的双人交互行为识别与预测算法研究

Human interaction recognition and prediction algorithm based on deep learning

智能系统学报. 2020, 15(3): 484–490 <https://dx.doi.org/10.11992/tis.201812029>

基于增强AlexNet的音乐流派识别研究

Music genre recognition research based on enhanced AlexNet

智能系统学报. 2020, 15(4): 750–757 <https://dx.doi.org/10.11992/tis.201909032>

基于小样本学习的LCD产品缺陷自动检测方法

An automatic small sample learning-based detection method for LCD product defects

智能系统学报. 2020, 15(3): 560–567 <https://dx.doi.org/10.11992/tis.201904020>

基于卷积神经网络的盲文音乐识别研究

Research on braille music recognition based on convolutional neural networks

智能系统学报. 2019, 14(1): 186–193 <https://dx.doi.org/10.11992/tis.201805002>

基于改进卷积神经网络的多标记分类算法

A multi-label classification algorithm based on an improved convolutional neural network

智能系统学报. 2019, 14(3): 566–574 <https://dx.doi.org/10.11992/tis.201804056>

DOI: 10.11992/tis.202212031

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20240828.0921.002>

面向掌纹掌静脉识别网络轻量化的非对称 双模态融合方法

林孙旗¹, 徐家梦², 郑瑜杰¹, 王翀^{1,2}, 王军²

(1. 宁波大学 信息科学与工程学院, 浙江 宁波 315211; 2. 中国矿业大学 信息与控制工程学院, 江苏 徐州 221116)

摘要: 深度学习已在掌纹掌静脉领域广泛应用, 但随着任务使用场景的不断微型化、终端化, 现有的深度学习模型往往难以在算力匮乏、内存有限的边缘设备上顺利部署。本文基于知识蒸馏方法提出了轻量化的掌纹掌静脉识别网络。根据模态特征提取复杂程度, 为掌纹与掌静脉模态分别选用不同的网络深度。在常规知识蒸馏方法中引入新设计的模态特征损失函数, 强化教师模型对各模态特征提取的指导作用。实验结果表明, 该方法有效协调了模型大小与性能, 为边缘计算环境下的生物特征识别技术提供了一种有效的解决方案。

关键词: 深度学习; 生物特征识别; 掌纹掌脉识别; 多模态网络; 知识蒸馏; 模型压缩; 卷积神经网络; 类激活图
中图分类号: TP30 **文献标志码:** A **文章编号:** 1673-4785(2024)05-1190-09

中文引用格式: 林孙旗, 徐家梦, 郑瑜杰, 等. 面向掌纹掌静脉识别网络轻量化的非对称双模态融合方法 [J]. 智能系统学报, 2024, 19(5): 1190-1198.

英文引用格式: LIN Sunqi, XU Jiameng, ZHENG Yujie, et al. An asymmetric bimodal fusion method for lightweight palm print and palm vein recognition network[J]. CAAI transactions on intelligent systems, 2024, 19(5): 1190-1198.

An asymmetric bimodal fusion method for lightweight palm print and palm vein recognition network

LIN Sunqi¹, XU Jiameng², ZHENG Yujie¹, WANG Chong^{1,2}, WANG Jun²

(1. Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China; 2. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: Deep learning has been widely used in palm print and palm vein recognition. However, with the continuous miniaturization and terminalization of task usage scenarios, it is often challenging to deploy current deep-learning models successfully on edge devices that suffer from limited computational power and memory constraints. In this study, we propose a lightweight palm print and palm vein recognition network based on knowledge distillation. First, we select different network depths for the palm print and palm vein modalities according to the complexity of their feature extraction. We introduce a novel modality feature loss function into the traditional knowledge distillation method to enhance the guiding role of the teacher model in the feature extraction of each modality. The experimental results demonstrate that this method effectively balances model size with performance and offers a viable solution for biometric recognition technologies within an edge computing environment.

Keywords: deep learning; biometrics; palm print and vein recognition; multimodal network; knowledge distillation; model compression; convolutional neural network; class activation map

收稿日期: 2022-12-31. 网络出版日期: 2024-08-28.

基金项目: 科技部科技创新 2030 —“新一代人工智能”重大项目 (2020AAA0107300); 宁波市自然科学基金项目 (20221JCGY010068); 中国创新挑战赛 (宁波) 项目 (2022T001).

通信作者: 王翀. E-mail: wangchong@nbu.edu.cn.

©《智能系统学报》编辑部版权所有

当今社会, 信息技术快速发展, 寻求一种便捷、高效、准确的个人身份认证方案成为该领域研究者亟待解决的难题。掌纹识别和掌静脉识别作为一种新兴的生物识别方式, 因其良好的非入

侵性、活体鉴别性、稳定性被认为是最有效的个人身份认证方案之一^[1-2]。随着近年来深度学习不断发展,神经网络的性能得到了大幅提升,逐步有相关研究者将深度神经网络应用到掌纹识别和掌静脉识别任务中。

2018 年, Zhong 等^[3]提出了一种端到端的掌纹识别方法,其中应用了 2 个参数共享的 VGG-16 网络用于提取图像的卷积特征,将 2 张掌纹图片分别输入特征提取网络,后使用顶层网络输出两图间的特征相似度用于识别任务,在 PolyU 数据集中最高获得 99.6% 的准确率。Zhao 等^[4]于 2019 年设计了一种联合深度卷积的特征表示方法 (joint deep convolutional feature representation, JDCFR) 用于多光谱掌纹识别任务,即设计了一个 16 层的卷积神经网络用于提取和融合各光源下图像的特征,并引入基于协作表示的分类算法作为图像的特征表示与分类方法。2022 年, Ahmed 等^[5]将掌静脉与掌背部静脉图像作为任务输入,使用主成分分析技术 (principal component analysis, PCA) 提取图像特征并用 K 最近邻 (K-nearest neighbors, K-NN) 算法分类器进行图像匹配操作,最终在 CASIA 多光谱数据集中获得 97.6% 的分类准确率。Lou 等^[6]在掌静脉识别任务中,引入了基于 VGG-16 的迁移学习方法加快训练速度,并融合注意力机制与多任务损失函数提高图像匹配精度,最后使用 K 均值聚类 (K-means clustering) 方法确定图像匹配阈值,并在验证集中取得 98.89% 的正确率。

虽然,随着硬件算力的发展,深度神经网络的训练已并非难事,许多大规模神经网络模型也已在各领域取得领先的成绩。但是,本文中掌纹掌静脉融合识别任务的应用场景一般是算力有限、内存小的边缘设备,这使得具有高计算复杂度、高内存需求的深度神经网络模型难以直接部署在此类设备上。同时,多模态网络也容易出现模态平衡问题^[7],即因不同模态以不同速率拟合与泛化,导致模型难以同时学到多模态特征知识,从而影响了模型精度。

为解决模型部署与模态平衡的问题,本文提出面向掌纹掌静脉识别网络轻量化的非对称双模态融合方法。首先设计了一个掌纹掌静脉识别的深度网络框架,通过更换特征提取网络的方式,分析各特征提取网络的性能。然后,采用知识蒸馏方法将大模型的知识传授给小模型,并引入模态特征蒸馏损失函数强化教师模型对学生模型中各模态特征提取网络的指导作用。同时,为缓解

模态平衡问题,本文调整了各模态特征提取网络的复杂度去平衡不同模态的训练进度,并使用梯度加权类激活图观测各模态对模型决策过程的贡献度,以验证本文方法对模态平衡问题的改善。最后,实验结果证明本文提出方法有效缓解了模态平衡问题,并实现了模型的压缩与识别精度的提高。

1 掌纹掌静脉识别与知识蒸馏方法

1.1 掌纹掌静脉识别的网络设计

1.1.1 残差网络结构

在残差网络出现之前,研究者们发现,在神经网络模型收敛后,继续堆叠层数,模型的训练损失反而会开始增大,精度降低,这被称为模型的退化问题。为解决这一问题, He 等^[8]提出了基于残差学习 (residual learning) 的残差网络 (residual network, ResNet), 根据模型的深度不同,使用不同的残差网络块,在增加模型深度的同时,防止了模型退化问题的出现。

ResNet 网络在构建 50 层以下的尺寸版本时采用基础模块 (building block), 达到 50 层后,采用瓶颈模块 (bottleneck block) 的设计减小计算量。两残差块的结构如图 1 所示,其中残差块的前向传播过程可以表示为

$$\mathbf{x}_{l+1} = \mathbf{h}_l + F(\mathbf{x}_l, \mathbf{W}_l) \quad (1)$$

式中: \mathbf{h}_l 为恒等映射, $F(\mathbf{x}_l, \mathbf{W}_l)$ 为残差映射。如式 (1) 所示,残差块通过在输入和输出间建立一条直接传输的恒等映射支路,巧妙地避过了使用网络拟合恒等映射的问题,即使主路网络未学习知识,也可通过恒等映射支路传输的特征图,缓解模型退化问题。同时,恒等映射支路也保证了在主路网络的梯度为 0 时,因支路的梯度仍为 1,整个网络的梯度并不会消失,使得整个网络的梯度可以被有效传导。综上,因 ResNet 网络具有良好的性能,本文选用 ResNet 系列网络作为模态特征提取网络。

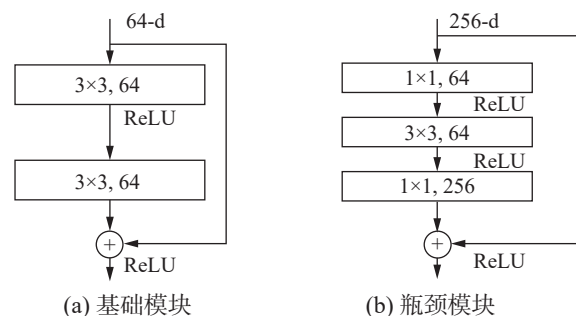


图 1 2 种残差块结构

Fig. 1 Different residual blocks

1.1.2 卷积块注意力模块

为提高网络对关键特征的提取能力, 本文将卷积块注意力模块 (convolutional block attention

module, CBAM)^[9] 引入网络。CBAM 由通道注意力 (channel attention) 与空间注意力 (spatial attention) 构成, 其结构如图 2 所示。

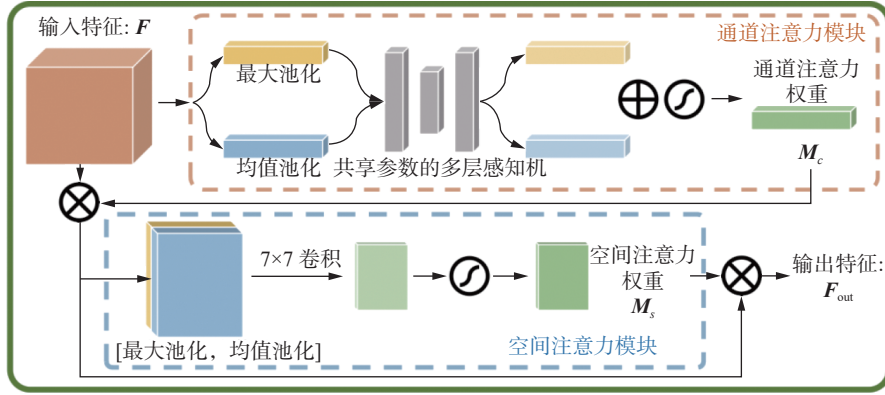


图 2 卷积块注意力模块结构

Fig. 2 Convolutional block attention module structure

通道注意力部分如图 2 中黄色框所示, 输入特征首先分别进行最大池化和平均池化操作聚合空间信息, 得到 2 组特征图, 后经过 1 个多层感知机输出 2 个 $1 \times 1 \times C$ 的通道注意力图, 再将 2 个注意力图相加激活, 最终得到通道注意力图, 该过程的表达式为

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) = \sigma(W_1(W_0(F_{\text{avg}})) + W_1(W_0(F_{\text{max}})))$$

空间注意力部分如图 2 中蓝色框所示, 首先在通道维度对输入的特征图进行最大池化和平均池化操作, 再将 2 张生成的特征图进行维度拼接, 最后利用大小为 7×7 的卷积核生成空间注意力图, 该过程的表达式为

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\text{max}}^s]))$$

具体应用上, 本文将卷积块注意力模块添加在每个模态的特征提取网络之后, 以强化模型在

通道与空间维度的特征提取能力。

1.1.3 选择性核网络

选择性核网络 (selective kernel networks, SKNets)^[10] 是一种轻量但高效的特征提取网络, 它引入了一种选择性核的操作, 通过多核信息聚合的方式, 实现神经元自适应调节感受野大小的功能。选择性核由 3 个算子组成, 分别为 Split、Fuse 和 Select。Split 算子使用不同卷积核大小的卷积操作, 提取不同感受野大小下的特征信息。Fuse 算子将 Split 算子生成的数个特征信息整合, 并生成各特征信息的选择权重矩阵。Select 算子根据选择权重矩阵融合不同卷积核大小生成的特征信息。最终自适应生成融合多感受野信息的特征图。

1.1.4 掌纹掌静脉识别的网络设计

针对掌纹掌静脉识别任务的特点, 本文设计了一种双路输入、单路输出的掌纹掌静脉识别网络, 架构细节如图 3 所示。

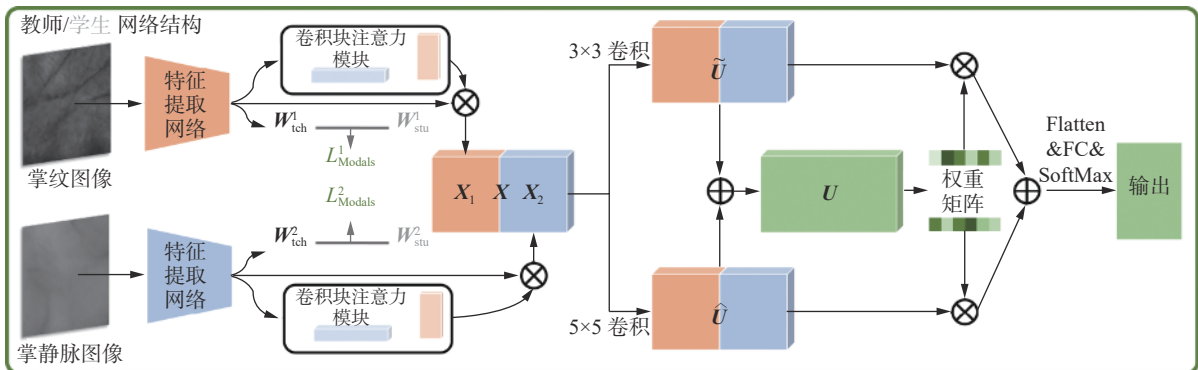


图 3 掌纹掌静脉识别网络框架

Fig. 3 Framework of palm print and vein recognition network

首先, 将掌纹、掌静脉的图像各自输入权重不共享的模态特征提取网络, 输出 2 张模态特征图 $\mathbf{W}_{\text{teacher/student}}^1$ 、 $\mathbf{W}_{\text{teacher/student}}^2$, 其中 $\mathbf{W}_{\text{teacher/student}}^i$ 表示教师或学生模型的第 i 个模态经特征提取网络输出的模态特征图, 各特征的提取网络可根据任务需求自行选择。

然后, 2 个模态特征图分别经过卷积块注意力模块, 强化模型对通道维与空间维的特征捕捉, 生成 2 个强化后的模态特征图 \mathbf{X}_1 、 \mathbf{X}_2 , 两者横向拼接成组合特征图 \mathbf{X} 。随后, 组合特征图 \mathbf{X} 分别经卷积核大小为 3×3 、 5×5 的自适应卷积操作, 提取不同感受野的特征信息, 并使用注意力机制生成选择权重矩阵, 指导不同卷积核下特征信息的融合, 生成全局特征图。

最后, 将全局特征图展平, 使用全连接层与 Softmax 层输出最终分类结果。

1.2 知识蒸馏方法

近年来, 随着深度神经网络的不断发展, 深度学习在人工智能的各个领域都已经取得重大突破。为了应对各种复杂的学习任务, 深度网络模型也逐步被设计得更深、更复杂。如 LeNet^[11]、AlexNet^[12] 到 Bert^[13]、GPT-3^[14], 参数量与运算量都大幅增长。但随着前沿技术不断下沉, 在算力有限、内存资源少的边缘设备上部署高效神经网络模型已成为不可忽视的实际需求, 模型压缩的概念应运而生。当前主流模型压缩技术包括剪枝^[15-16]、量化^[17-18]、知识蒸馏^[19-20] 等, 本部分着重介绍所使用到的知识蒸馏技术及本文提出的非对称双模态融合蒸馏方法。

1.2.1 知识蒸馏方法基本原理

2015 年, Hinton 等^[21] 在前人的基础上, 提出了知识蒸馏方法, 即在有监督学习任务中, 使用教师模型的输出作为额外的监督信息, 可将教师模型的知识一定程度上迁移至学生模型中, 可达到提高学生模型精度, 同时减少模型大小的效果。

在大部分神经网络任务中, 最后一层通常使用 Softmax 层将前一层的输出映射到 (0,1) 的区间内, 从而达到类似概率输出的结果, 但这也导致错误类的概率被压缩成接近零的值。Hinton 等^[21] 认为教师模型输出的正确类与错误类间的关系信息可作为一项有效的监督信息, 他们将这种信息称为暗知识 (dark knowledge), 并通过引入温度系数 T 来软化 Softmax 输出的分类信息, 公式表示为

$$p_i(\mathbf{z}_i, T) = \exp(\mathbf{z}_i / T) / \sum_{j=0}^k \exp(\mathbf{z}_j / T) \quad (2)$$

式中: T 为引入的温度系数, 用于控制概率输出的软化程度, 当 $T=1$ 时, 式 (2) 为常规的 Softmax 函数, 随着 T 的上升, 正确类与错误类间的差距会逐步减少, 直至完全消除。在合适的温度系数下, 学生模型能更好地接受到教师模型输出的类间关系。

除引入温度系数软化教师模型输出的硬标签外, Hinton 等^[21] 还提出, 在学生模型的训练过程中, 加入一定比例的正确数据标签, 即加入 $T=1$ 时 Softmax 函数输出的硬标签, 可以使学生模型达到更好的学习效果。因此, 传统的知识蒸馏方法中设置了 2 个不同的目标函数, 第 1 个目标函数使用交叉熵损失来计算教师模型和学生模型软化后的输出之间的差异, 也称蒸馏损失。第 2 个目标函数是学生模型输出的预测值与真实值之间的损失。2 种目标函数分别表示为

$$L_{\text{soft}}(p(\mathbf{t}, T), p(\mathbf{s}, T)) = \sum_{i=0}^k -p_i(\mathbf{t}_i, T) \log(p_i(\mathbf{s}_i, T))$$

$$L_{\text{hard}}(\mathbf{y}, p(\mathbf{z}, 1)) = \sum_{i=0}^k -y_i \log(p_i(\mathbf{z}_i, 1))$$

式中: \mathbf{t} 表示教师模型的输出, \mathbf{s} 、 \mathbf{z} 表示学生模型的输出, T 表示所使用的温度系数, \mathbf{y} 表示真实值。因此, 知识蒸馏方法的总损失可表示为

$$L_{\text{KD}} = \lambda L_{\text{soft}}(p(\mathbf{t}, T), p(\mathbf{s}, T)) + (1 - \lambda) L_{\text{hard}}(\mathbf{y}, p(\mathbf{z}, 1)) \quad (3)$$

式中: λ 是超参数, 可根据实际情况动态调整。常规知识蒸馏方法的框架如图 4 所示。

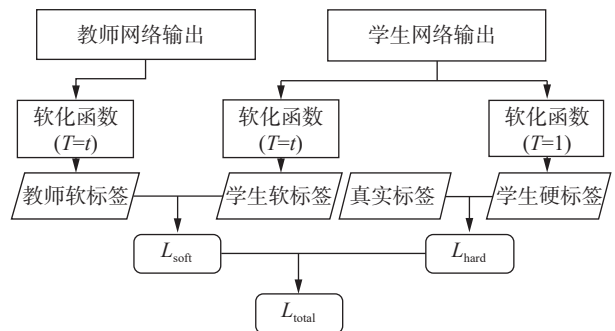


图 4 知识蒸馏基本框架

Fig. 4 Basic frame of knowledge distillation

在式 (3) 的规范下, 学生网络的输出逐步向教师的输出靠近, 当更轻量的学生网络实现了和教师网络相近的网络性能, 便实现了神经网络的轻量化。

1.2.2 非对称双模态融合方法

在多模态任务中, 由于每个模态的特征提取难度不同, 容易存在因易训练的模态过早拟合而

导致其他模态训练不足的问题,同时,对不同模态采用相同的特征提取网络也容易出现不必要的模型参数冗余,加大模型部署难度。

因此,本文提出了面向掌纹掌静脉识别网络轻量化的非对称双模态融合方法,根据模态的特征提取复杂度不同,在特征提取网络的选用阶段,选择与对应模态性能匹配的非对称特征提取网络,以减少模型中冗余参数并平衡模态间的学习速率。同时,为进一步缓解各模态间的训练平衡问题,本文受 FitNets^[22] 知识蒸馏方法中中间层特征一致性的启发,设计了一种模态蒸馏损失函数,用以强化教师模型的模态特征提取网络对学生模型对应部分的指导作用,使学生模型更有针对性地学习到教师模型中各模态特征提取模块的知识。

具体方法上,本文将教师模型中特征提取模块输出的特征图定义为 $\mathbf{W}_{\text{tch}}^j$, j 表示第 j 个模态的输出,同理将学生模型中对应模态输出的特征图定义为 $\mathbf{W}_{\text{stu}}^j$,并使用合适的温度系数 T_2 对两特征图进行软化,减小两图间的差距,后使用交叉熵损失函数表示学生模型输出的模态特征图与教师模型输出的模态特征图的差距,使学生模型学会教师模型的特征表达。模态蒸馏损失函数定义为

$$L_{\text{Modals}}(p(\mathbf{W}_{\text{tch}}^j, T_2), p(\mathbf{W}_{\text{stu}}^j, T_2)) = \sum_{j=0}^m (-p_i(\mathbf{W}_{\text{tch}}^j, T_2) \log(p_i(\mathbf{W}_{\text{stu}}^j, T_2)))$$

并将 L_{Modals} 加入到常规的知识蒸馏损失函数中,总损失 L_{total} 可被定义为

$$L_{\text{total}} = L_{\text{KD}} + \mu L_{\text{Modals}}(p(\mathbf{W}_{\text{tch}}^j, T), p(\mathbf{W}_{\text{stu}}^j, T))$$

最后,使用加入了模态蒸馏损失函数的知识蒸馏方法对非对称学生模型进行训练,并通过最小化损失函数 L_{total} 完成对学生模型的迭代优化。

2 实验设计

2.1 实验平台

本实验运行在 Windows 11 操作系统,使用的 GPU 为 Nvidia GeForce RTX 3090,显存为 24 GB,所使用的深度神经网络框架为 PyTorch 1.12.1。

2.2 实验数据集选用

本实验采用的数据集为 CASIA 多光谱掌纹数据集^[23] 和同济非接触掌纹掌静脉数据集^[24],以下简称 CASIA 数据集和 Tongji 数据集。CASIA 数据集是由中国科学院自动化研究院组织建立,使用 6 种光源(波长为 460、630、700、850、940 nm

和白光)采集了 100 人的 7 200 张掌纹图像,并以“编号_左/右手_光波段_采样阶段”的格式分别保存为 8 位灰度 JPEG 格式图像。同济非接触掌纹掌静脉数据集收集自 300 位志愿者,并分 2 次收集数据,共计收集 12 000 ($300 \times 2 \times 10 \times 2$) 张高质量的非接触图像。部分数据集如图 5 所示。

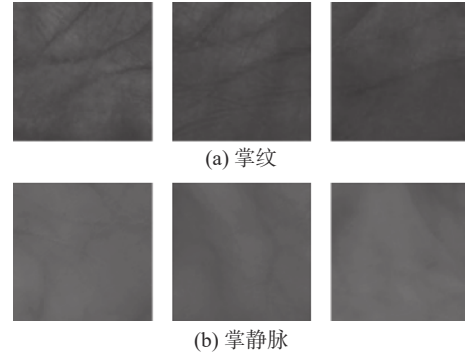


图 5 CASIA 数据集部分图像

Fig. 5 Partial image of the CASIA dataset

2.3 评价标准

2.3.1 梯度加权类激活图

梯度加权类激活图 (gradient-weighted class activation mapping, Grad-CAM)^[25] 是一种神经网络解释技术,用于可视化卷积神经网络的决策过程。它通过计算源图像中每个像素对最终输出结果的贡献度,来表示网络对于某个特定类别的感兴趣区域。对于给定图像,它首先通过网络的前向传播得到最终输出结果。后计算第 k 个通道的特征图 \mathbf{A}^k 相对于目标类别 y^c 的梯度均值,表示第 k 个通道的特征图对于最终输出结果的贡献度 α^c ,公式为

$$\alpha_k^c = \frac{1}{i \times j} \sum_i \sum_j \frac{\partial y^c}{\partial \mathbf{A}_{ij}^k}$$

式中 i 、 j 分别指特征图的高度与宽度。最后,使用这些通道贡献度对特征图逐通道进行加权求和并取 ReLU 激活函数筛去负值,以显示网络对目标类别的分类依据,公式为

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c \mathbf{A}^k \right)$$

本文选用 Grad-CAM 方法,并将相关分类依据像素以类激活图的方式进行可视化,图中红色代表模型感兴趣程度高,反之蓝色代表程度低,用以验证本文所提方法对模态平衡问题的改善。

2.3.2 基本模型评估指标

本文使用准确率 (accuracy, Acc)、平均精确率 (average precision, APre)、平均召回率 (average re-

call, ARecall)、平均 F_1 分数 (average F_1 , AF_1) 及模型大小作为模型评估指标, 验证模型的性能^[26]。精确率指被正确预测为该类别的样本数占所有被预测为该类别样本数的比值。召回率指被正确分类为该类别的样本数占所有正确样本数的比值。 F_1 分数指对精确率与召回率的综合评估。

3 实验结果分析

3.1 掌纹掌静脉识别实验

根据上文设计的掌纹掌静脉识别网络, 本文使用不同的特征提取网络, 在相同训练集和验证集的情况下, 从验证集准确率、模型文件大小、推理速度方面来衡量各个模态特征提取网络的性能, 为后续知识蒸馏实验中教师模型与学生模型的选取做好准备。实验结果如表 1 所示。

表 1 各特征提取网络性能对比实验

Table 1 Comparative experiment on the performance of each feature extraction network in CASIA dataset

掌纹与掌静脉 特征提取网络	Acc/%	ARecall/%	模型大小/MB
ResNet8+ResNet8	76.00	76.01	38.9
ResNet14+ResNet14	93.83	92.01	88.5
ResNet18+ResNet18	99.67	99.17	89.7
ResNet34+ResNet34	99.33	98.84	170.7
ResNet50+ResNet50	99.00	98.33	188.9
ResNet101+ResNet101	99.87	99.83	341.4

从表 1 可以看出, 在验证集识别准确率方面, 随着 ResNet 网络层数的增加, 其验证集准确率呈现出先升后平稳的趋势。具体而言, 当网络层数小于 18 层时, 随着层数的增加, 准确率呈现上升趋势; 当网络层数大于 18 层时, 准确率在一定范围内波动, 最终在所有模态特征提取网络均采用 ResNet101 架构的情况下达到 99.87% 的峰值。在模型大小上, 随着网络深度的增加, 模型大小也在快速增大, 最深的 ResNet101 模型比其他模型大 1.80~8.77 倍。

为更好地验证本文方法, 以 CASIA 数据集为例, 本文采取以下 2 项原则选择学生与教师模型。首先, 在学生模型的选择上, 应尽可能保持模型性能与模型大小的统一, 不能以极大幅度的模型性能下降换取模型大小的减小。由表 1 可知, 当特征提取网络为 ResNet8 时, 模型大小较 ResNet14 及更深网络皆有极大减小, 性能也没有过大差距, 可作为学生模型。同时, 因模态特征

提取复杂度的不同, 选用性能更强的 ResNet14 网络, 用以提取复杂模态的图像特征。

其次, 在教师模型的选取方面, ResNet18 及更深的特征提取网络在验证集精度方面相差不大, 都可作为知识蒸馏方法中教师模型的选项, 但众所周知, 更深的网络结构可以提取更高维的图像特征, 指导学生模型学习更抽象的特征, 故应尽可能选用更深的网络作为知识蒸馏方法中的教师模型。但同时为防止知识蒸馏领域中常出现的, 因教师模型与学生模型性能差距过大导致的知识传递效率低下问题, 实验所选择的教师网络与学生网络性能也不能差距过大。因此, 本文选用性能优异且网络能力与学生模型相差不大的 ResNet34 网络作为教师模型的模态特征提取网络, 指导学生模型 ResNet8 与 ResNet14 的学习。以同样的原则, 在 Tongji 数据集上选择了 ResNet14 作为教师模型的模态特征提取网络, ResNet10 与 ResNet6 作为学生模型的模态特征提取网络。值得注意的是, 由于 Tongji 数据集相对简单, 本文调整了该数据集所使用的 ResNet 网络的通道数, 如将 ResNet8 网络所使用 3 类残差块的通道数从 128、256、512 降为 16、32、64, 以验证本文方法的有效性。

3.2 本文方法相关实验

3.2.1 模态蒸馏损失函数的有效性实验

实验选择了 CASIA 与 Tongji 2 个主流掌纹掌静脉数据集作为基准。将由常规知识蒸馏方法训练的学生模型性能与本文提出的非对称融合蒸馏方法训练的学生模型进行对比, 以验证所提方法的有效性。在使用相同的教师模型与学生特征提取网络并保证其他变量一致的情况下, 实验结果如表 2、3 所示。从表 2 中可见, 在非对称双模态融合蒸馏方法的知识蒸馏框架下训练的学生模型, 相较于常规知识蒸馏方法下使用相同特征提取网络训练的学生模型, 其性能都有了不同程度的提升。如当掌纹掌静脉的特征提取网络皆为 ResNet8 时, 其模型的准确率提升了 2.83 百分点。并且在掌纹特征提取网络为 ResNet8, 掌静脉特征提取网络为 ResNet14 时, 模型达到最高准确率 98.63%。同时, 虽然学生模型的最高精度仍与教师模型有一定差距, 但当掌纹掌静脉特征提取网络为 ResNet8 与 ResNet14 时, 模型大小与教师模型相比, 减小了近 60%。非对称网络的性能提升是由于网络平衡了不同模态的训练进程, 同时缓解了模型对单一模态的过度依赖。

表 2 CASIA 数据集上不同方法下各特征提取网络配置的实验结果

Table 2 Results of network configurations for feature extraction under different frameworks in CASIA dataset

实验序号	方法名称	掌纹特征提取网络	掌静脉特征提取网络	Acc/%	APre/%	ARecall/%	AF ₁ /%	模型大小/MB
1	常规知识	ResNet8	ResNet8	87.33	88.12	87.69	86.22	38.9
2	蒸馏方法 ^[21]	ResNet14	ResNet14	94.83	92.89	93.34	92.05	88.5
3	非对称	ResNet8	ResNet8	90.16	88.69	89.01	87.00	39.2
4		ResNet14	ResNet14	96.00	96.20	96.51	95.83	88.8
5		ResNet8	ResNet14	98.33	98.63	98.18	98.10	64.0
6	双模态融合蒸馏方法	ResNet14	ResNet8	95.83	96.56	96.01	95.59	64.0

表 3 Tongji 数据集上不同方法下各特征提取网络配置的实验结果

Table 3 Results of network configurations for feature extraction under different frameworks in Tongji dataset

实验序号	方法名称	掌纹+掌静脉特征提取网络	Acc %
1	教师网络	ResNet14+14	99.78
2	常规训练	ResNet6+6	30.11
3		ResNet10+10	84.89
4	常规知识	ResNet6+6	50.44
5	蒸馏方法 ^[21]	ResNet10+10	90.61
6	非对称	ResNet6+6	95.00
7		ResNet10+10	98.94
8		ResNet6+10	95.28
9	双模态融合蒸馏方法	ResNet10+6	93.22

在表 3 中同样可以观察到, 本文方法训练的 4 种学生模型相对于常规知识蒸馏方法训练的模型有不同程度的性能提升。同样值得关注的是, 当使用非对称网络提取模态特征时, 对比实验 6、9 可以发现当模型对掌纹模态的特征提取能力更强时, 性能不升反降。反之, 对比实验 6、8 可以发现, 强化对掌静脉模态的特征提取能力对总模型的性能提升有益。

以上实验结果说明经非对称双模态融合蒸馏方法训练的学生模型可以更高效地从教师模型中学到有益知识, 充分验证本文方法的有效性。

3.2.2 模态特征提取网络选择方法的有效性验证

为进一步探讨本方法对模态平衡现象的缓解, 本文使用了梯度类激活图观察模型对各模态的依赖程度。根据掌纹与掌静脉的图像特点并结合表 1 中对各特征提取网络的性能分析, 本部分实验在 CASIA 数据集上将 ResNet8 与 ResNet14 网络排列组合, 分别作为掌纹、掌静脉图像的特征提取网络, 并分别在常规知识蒸馏方法与本文

提出的知识蒸馏方法下验证针对模态特点选用特征提取网络的可行性, 实验结果如图 6 所示。

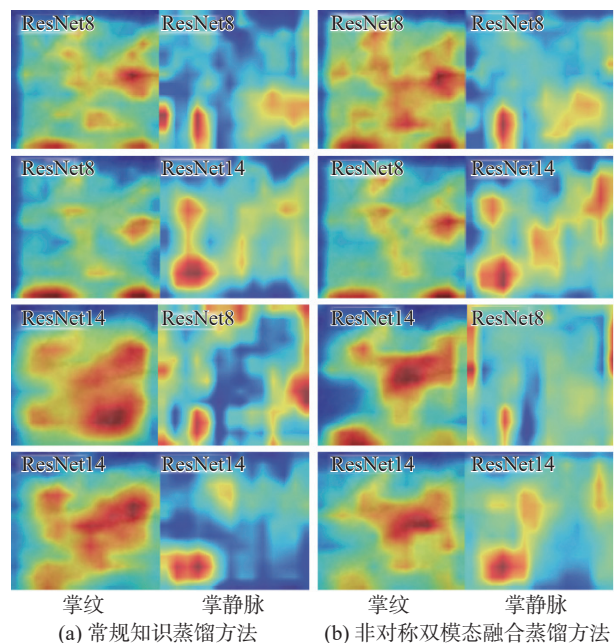


图 6 CASIA 数据集上不同方法与不同实验配置下类激活图

Fig. 6 Class activation diagram under different frameworks and different experimental configurations in CASIA dataset

由图 6 可见, 在正常知识蒸馏方法下, 当掌纹、掌静脉的特征提取网络能力相同时, 类激活图的红色大部分处在掌纹部分, 而掌静脉部分出现大面积深蓝色区域, 即模型分类结果更多依靠掌纹模态, 由此可推断出, 对不同模态使用相同特征提取网络并不能很好地挖掘多模态数据的优势。而纵向对比图 6(a), 当为掌纹与掌静脉模态选用 ResNet8 与 ResNet14 作为特征提取网络时, 原本类激活图中掌静脉部分的一些深蓝色块被青绿色替代并且掌纹部分的黄色被弱化为青绿色, 这说明在模型决策过程中, 掌静脉模态的特征被进一步利用, 本文提出的非对称特征提取网络有助于模态平衡问题的缓解。同时, 横向对比图 6(a) 和图 6(b), 在相同特征提取网络配置下, 经过本文

提出的知识蒸馏方法同样可以一定程度上缓解模型对单一模态的过度依赖。

3.2.3 模型性能对比

将本文的实验结果与相关识别网络与 CASIA 数据集下的实验结果进行对比, 如表 4 所示。由表 4 可见, 经本文所提模型轻量化方法训练的掌纹掌静脉识别网络与其他现有方法相比, 在减少模型规模的情况下, 仍具有良好识别性能。

表 4 模型性能对比
Table 4 Model performance comparison %

方法名称	所用数据	Acc
JCLSR ^[27]	掌纹	98.94
Hassan 等 ^[28]	掌静脉	98.00
Ahmed 等 ^[5]	掌静脉+掌背静脉	97.60
本文方法(ResNet101)	掌纹+掌静脉	99.87
本文方法(非对称融合蒸馏, 教师模型)	掌纹+掌静脉	99.33
本文方法(非对称融合蒸馏, 学生模型)	掌纹+掌静脉	98.33

4 结束语

针对掌纹掌静脉识别任务中深度学习模型部署难、模态学习不平衡的问题, 本文提出了面向掌纹掌静脉识别网络轻量化的非对称双模态融合方法。本文根据掌纹掌静脉的模态特点, 选用了具有性能差异的特征提取网络, 用于平衡 2 个模态的学习速率, 实验证明, 非对称的模态特征提取网络有助于缓解模态平衡问题。此外, 本文还设计了一种模态特征蒸馏损失函数, 用于加强教师模型对学生模型中模态特征提取网络的指导作用。对比实验表明, 通过本文提出的知识蒸馏方法训练的学生模型, 相较于传统方法训练的模型, 以更小的计算资源实现了更高的性能, 最终实现模型的轻量化。同时, 该方法也一定程度上改善了模态平衡问题。本文对于非对称模态特征提取网络的选择上仍缺乏量化标准。在未来的工作中, 需要制定更加规范化的模态特征提取网络选择方法, 并探索更高效的多模态网络轻量化方法, 以加快掌纹掌静脉识别模型在边缘设备上的部署。

参考文献:

- [1] 李倩颖, 阮秋琦. 分辨率 LBP 的掌纹特征提取 [J]. 智能系统学报, 2010, 5(6): 482–486.
LI Qianying, RUAN Qiuqi. Palmprint feature extraction based on multiresolution LBP[J]. CAAI transactions on intelligent systems, 2010, 5(6): 482–486.
- [2] 孙波. 基于深度学习的掌静脉识别算法研究与识别系统 [D]. 桂林: 桂林电子科技大学, 2022.
SUN Bo. Research on palmar vein recognition algorithm and recognition system based on deep learning[D]. Guilin: Guilin University of Electronic Technology, 2022.
- [3] ZHONG Dexing, YANG Yuan, DU Xuefeng. Palmprint recognition using Siamese network[C]//Chinese Conference on Biometric Recognition. Cham: Springer, 2018: 48–55.
- [4] ZHAO Shuping, ZHANG B, PHILIP CHEN C L. Joint deep convolutional feature representation for hyperspectral palmprint recognition[J]. Information sciences, 2019, 489: 167–181.
- [5] AHMED M A, ROUSHDY M, SALEM A B M. Multi-modal technique for human authentication using fusion of palm and dorsal hand veins[C]//New Approaches for Multidimensional Signal Processing. Singapore: Springer, 2022: 63–78.
- [6] LOU Jiashu, ZOU Jie, WANG Baohua. Palm vein recognition via multi-task loss function and attention layer [EB/OL]. (2022–11–11)[2022–12–31]. <http://arxiv.org/abs/2211.05970>.
- [7] WANG Weiyao, TRAN D, FEISZLI M. What makes training multi-modal classification networks hard? [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 12692–12702.
- [8] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [9] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//European Conference on Computer Vision. Cham: Springer, 2018: 3–19.
- [10] LI Xiang, WANG Wenhui, HU Xiaolin, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 510–519.
- [11] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.
- [12] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84–90.
- [13] DEVLIN J, CHANG Mingwei, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[EB/OL]. (2018–10–11)[2022–12–31]. <https://arxiv.org/abs/1810.04805>.

- [14] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. *Advances in neural information processing systems*, 2020, 33: 1877–1901.
- [15] HAN Song, MAO Huizi, DALLY W J. Deep compression: compressing deep neural networks with pruning, trained quantization and huffman coding[EB/OL]. (2015–10–01) [2022–12–31]. <http://arxiv.org/abs/1510.00149>.
- [16] HE Yihui, ZHANG Xiangyu, SUN Jian. Channel pruning for accelerating very deep neural networks[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 1398–1406.
- [17] CHENG Jian, WU Jiaxiang, LENG Cong, et al. Quantized CNN: a unified approach to accelerate and compress convolutional networks[J]. *IEEE transactions on neural networks and learning systems*, 2018, 29(10): 4730–4743.
- [18] MERONE M, GRAZIOSI A, LAPADULA V, et al. A practical approach to the analysis and optimization of neural networks on embedded systems[J]. *Sensors*, 2022, 22(20): 7807–7822.
- [19] 邵仁荣, 刘宇昂, 张伟, 等. 深度学习中知识蒸馏研究综述[J]. *计算机学报*, 2022, 45(8): 1638–1673.
SHAO Renrong, LIU Yuang, ZHANG Wei, et al. A survey of knowledge distillation in deep learning[J]. *Chinese journal of computers*, 2022, 45(8): 1638–1673.
- [20] 黄震华, 杨顺志, 林威, 等. 知识蒸馏研究综述[J]. *计算机学报*, 2022, 45(3): 624–653.
HUANG Zhenhua, YANG Shunzhi, LIN Wei, et al. Knowledge distillation: a survey[J]. *Chinese journal of computers*, 2022, 45(3): 624–653.
- [21] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. (2015–03–09) [2022–12–31]. <http://arxiv.org/abs/1503.02531>.
- [22] ROMERO A, BALLAS N, KAHOU S E, et al. FitNets: hints for thin deep nets[EB/OL]. (2014–12–19) [2022–12–31]. <http://arxiv.org/abs/1412.6550>.
- [23] SUN Zhenan, TAN Tieniu, WANG Yunhong, et al. Ordinal palmprint representation for personal identification [representation read representation][C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005: 279–284.
- [24] ZHANG Lin, CHENG Zaixi, SHEN Ying, et al. Palmprint and palmvein recognition based on DCNN and a new large-scale contactless palmvein dataset[J]. *Symmetry*, 2018, 10(4): 78–39.
- [25] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. *International journal of computer vision*, 2020, 128(2): 336–359.
- [26] 吴晓昱. 电子废弃物拆机塑料的近红外光谱识别[D]. 上海: 上海交通大学, 2020.
WU Xiaoyu. Identification of electronic waste dismantling plastics by near infrared spectroscopy[D]. Shanghai: Shanghai Jiao Tong University, 2020.
- [27] ZHAO Shuping, ZHANG B. Joint constrained least-square regression with deep convolutional feature for palmprint recognition[J]. *IEEE transactions on systems, man, and cybernetics: systems*, 2022, 52(1): 511–522.
- [28] HASSAN N F, ABDULRAZZAQ H I. Pose invariant palm vein identification system using convolutional neural network[J]. *Baghdad science journal*, 2018, 15(4): 503–510.

作者简介:



林孙旗, 硕士研究生, 主要研究方向为深度学习、模型轻量化。E-mail: sunqi1209@gmail.com。



王翀, 副教授, 主要研究方向为模型轻量化, 零样本、小样本目标检测, 视频异常检测。主持国家自然科学基金项目 1 项。发表学术论文 50 余篇。E-mail: wangchong@nbu.edu.cn。



王军, 教授, 博士生导师, 主要研究方向为智能机器人与无人系统、生物特征识别、机器视觉。主持科技部科技创新 2030—“新一代人工智能”重大项目, 获得国家级教学科研奖 1 项、省部/学会级教学科研奖 5 项。获授权发明专利数十项, 发表学术论文 60 余篇, 出版专著教材 6 部。E-mail: jrobot@126.com。