



DOI:10.12404/j.issn.1671-1815.2401917

引用格式: 英特扎尔·艾山江, 伊力哈木·亚尔买买提. 基于知识蒸馏的低质量人脸图像识别算法[J]. 科学技术与工程, 2025, 25(2): 695-703.

Yingtezhaer Aishanjiang, Yilihamu·Yaermaimaiti. Knowledge distillation based algorithm for low quality face image recognition[J]. Science Technology and Engineering, 2025, 25(2): 695-703.

## 基于知识蒸馏的低质量人脸图像识别算法

英特扎尔·艾山江, 伊力哈木·亚尔买买提\*

(新疆大学电气工程学院, 乌鲁木齐 830017)

**摘要** 针对基于统一特征子空间的低质量人脸图像识别算法存在对低质量人脸的鲁棒性较差、特征表示能力有限等缺点, 提出了一种基于知识蒸馏的低质量人脸图像识别算法。首先, 将 ResNeXt 网络作为骨干特征提取网络, 并引入双通道注意力模块构建具有注意力机制的教师-学生知识蒸馏框架。其次, 采用教师网络的输出特征作为标签知识, 将有效的识别特征传递给学生网络, 采用注意力图特征作为中间层知识, 弥补输出层知识信息单一的不足, 通过结合两种知识蒸馏的方式丰富特征知识以保证教师网络模型知识信息的多样性。然后, 将标签知识蒸馏损失、注意力图蒸馏损失和识别损失的加权平均融合作为网络总的损失函数, 确保学生网络模型具有更好的学习能力。最后, 在 AgeDB-30 和 CPLFW 测试集不同质量图像下进行测试, 消融实验结果表明, 相比无蒸馏的通用人脸识别模型, 经过两种知识蒸馏的模型, 在识别准确率上分别获得了 2.25%、11.33%、24.64% 和 2.8%、10.58%、27.85% 的提升。对比实验表明, 与其他主流算法相比, 本文所提算法在准确率上也获得了不同程度的提升。

**关键词** 低质量人脸图像; 知识蒸馏; 注意力机制; ResNeXt

**中图分类号** TP391.41; **文献标志码** A

## Knowledge Distillation Based Algorithm for Low Quality Face Image Recognition

YINGTEZHAER Aishanjiang, YILIHAMU·Yaermaimaiti\*

(School of Electrical Engineering, Xinjiang University, Urumqi 830017, China)

**[Abstract]** Aiming at the shortcomings of low-quality face recognition algorithms based on unified feature space, such as poor robustness to low-quality faces and limited feature representation capability, a low-quality face image recognition algorithm based on knowledge distillation was proposed. First, the ResNeXt network was used as the backbone feature extraction network, and the two-channel attention module was introduced to construct a teacher-student knowledge distillation framework with an attention mechanism. Secondly, the output features of the teacher network were adopted as labeled knowledge, and the effective recognition features were passed to the student network. And the attention graph features were adopted as the intermediate layer knowledge to solve the lack of single knowledge information in the output layer, and the feature knowledge was enriched by combining two kinds of knowledge distillation to ensure the diversity of knowledge information in the teacher network model. Then, the weighted average of labeled knowledge distillation loss, attention graph distillation loss, and recognition loss were fused as the total network loss function to ensure that the student network model has a better learning ability. Finally, tested under different quality images in AgeDB-30 and CPLFW test sets, the results of the ablation experiments show that compared to the generic face recognition model without distillation, the model with two types of knowledge distillation gains 2.25%, 11.33%, 24.64% and 2.8%, 10.58%, 27.85% improvement in recognition accuracy, respectively. Comparative experiments show that the algorithm proposed in this paper also obtains different degrees of improvement in accuracy compared to other mainstream algorithms.

**[Keywords]** low quality face images; knowledge distillation; attention mechanism; Resnext

随着计算机视觉和模式识别技术的发展, 人脸识别技术广泛应用于识别和验证个人身份, 监控安全, 人机交互等领域。传统的人脸识别方法有基于

特征脸(eigenface)<sup>[1]</sup>的方法、基于贝叶斯(Bayes)<sup>[2]</sup>的方法、基于局部二值模式(local binary patterns, LBP)<sup>[3]</sup>的方法和基于主成分分析(principal compo-

收稿日期: 2024-03-18; 修订日期: 2024-10-31

基金项目: 国家自然科学基金(62362063, 61866037)

第一作者: 英特扎尔·艾山江(1998—), 女, 维吾尔族, 新疆博乐人, 硕士研究生。研究方向: 模式识别与智能系统。E-mail: yingteze\_xju@163.com。

\* 通信作者: 伊力哈木·亚尔买买提(1978—), 男, 维吾尔族, 新疆乌鲁木齐人, 硕士, 教授, 硕士研究生导师。研究方向: 人工智能, 模式识别、人脸识别。E-mail: 65891080@qq.com。

投稿网址: www.stae.com.cn

ment analysis, PCA)<sup>[4]</sup>的方法等,但这些方法在特征提取方面有局限性,对光照变化、视角变化非常敏感。随着深度学习的兴起,人脸识别的研究重点转向了基于深度学习的方法。Schroff等<sup>[5]</sup>提出了FaceNet,采用大规模的人脸图像数据集来训练网络,使用三元组损失(triplet loss)函数来优化网络参数,解决了类内差异和类间差异不明显的问题。Deng等<sup>[6]</sup>提出了一个加性角度边距损失(additive angular margin loss, ArcFace),以获取用于人脸识别的高判别度特征。对于低质量人脸图像识别,一个重要的潜在应用领域是视频监控。在现实环境中采集到的人脸图像,由于拍摄条件的复杂性都是在大的对峙、不同照明条件和不同视角下所拍摄,导致这些图片中的人脸一般都很小或清晰度较低。但现有的人脸识别算法主要针对高清晰度人脸,当目标对象变为低质量人脸图像时模型性能急剧下降。

针对低质量人脸图像识别,目前存在两种解决方案,分别为基于超分辨率重建的方法与基于统一特征子空间的方法。前者从低分辨率(low-resolution, LR)图像中重建出高分辨率(high-resolution, HR)人脸图片,从而提高后续人脸识别系统的识别率。Ma等<sup>[7]</sup>提出了一种基于双循环网络迭代协作的深度人脸超分辨率重建方法,其中每个重复步骤中的恢复分支利用先验信息来生成超分辨率图像,但增加了迭代超分辨过程带来的内存和计算成本。Liu等<sup>[8]</sup>提出了一种基于渐进式上采样的级联递归卷积网络,在网络的每个阶段并行提取人脸的深度特征和解析先验,并进行整合,以提高网络的深度表示能力。后者通过寻找不同质量图像之间存在的某种联系,将两种特征映射到一个公共特征子空间,提高其特征在子空间分布的紧凑性来解决该问题。Lu等<sup>[9]</sup>提出了基于卷积神经网络(convolutional neural network, CNN)的一种新颖的模型(deep coupled ResNet, DCRN),首先通过类似ResNet的主干网络提取具有不同分辨率的面部图像所共享的区分特征。然后通过分支网络学习耦合映射,将特定分辨率的高质量图像和相应的低质量图像的特征投影到一个公共子空间中,最小化它们之间的距离。Zangeneh等<sup>[10]</sup>提出了一种使用深度卷积神经网络的低质量人脸图像识别的新颖耦合映射方法。网络结构由两个分支组成,通过非线性变换将不同质量的人脸图像映射到公共特征空间,使所有人脸属于同一个人的距离比不属于同一个人的距离更近。尽管这些研究工作在一定程度上提升了识别性能,但在实际工程应用中仍有较大的提升空间。

目前,许多研究人员将知识蒸馏<sup>[11]</sup>应用于该领域。知识蒸馏由Hinton等<sup>[12]</sup>于2015年首次提出,其最初目的是用于模型压缩<sup>[13]</sup>。其基本思想是通过从一个大的模型(教师)把知识转移给另一个小的模型(学生),利用特定的损失函数来约束小的模型从而引导小的模型也能够达到大的模型的性能效果。Xu等<sup>[14]</sup>将知识蒸馏方法应用到低质量图像分类任务中,证实了从教师网络到学生网络的简单logit蒸馏能够显著提升低质量图像分类性能,甚至优于超分辨率方法。Massoli等<sup>[15]</sup>提出的F-KD方法通过传递教师网络的中间层表示,解决了低质量人脸识别问题。然而,这些研究仅选择一种知识进行蒸馏,知识形式单一,未能充分发挥知识蒸馏的优势。

基于知识蒸馏的思想,现提出一种用于低质量人脸图像识别算法。首先,采用ResNeXt网络作为骨干特征提取网络,并引入双通道注意力模块,构建了具有注意力机制的“教师-学生”知识蒸馏框架。其次,利用教师网络的输出特征作为标签知识缩短教师-学生网络输出之间的分布差异,同时将注意力图特征作为中间层知识,以增强教师-学生网络中注意力图之间的相似性,通过结合两种知识进行蒸馏,保证了知识信息的多样性。最后,将标签知识蒸馏损失、注意力图蒸馏损失和识别损失的加权平均融合作为网络总的损失函数,保证了学生网络模型具有更好的特征表示能力。

## 1 基于知识蒸馏的低质量人脸图像识别算法

### 1.1 ResNeXt网络结构

ResNeXt<sup>[16]</sup>是一种深度神经网络架构,它在ResNet的基础上进行了改进和扩展。ResNeXt的设计思想是通过增加网络的宽度来增强模型的表达能力,从而进一步提升网络的性能。ResNeXt的核心组件是残差模块(residual block),它由多个分支组成,每个分支都包含了一组卷积层。这些分支能够并行学习不同的特征表示,然后将它们进行融合,以获得最终的特征表示。这种并行学习与融合的方式可以增强网络的非线性能力,提高模型对复杂数据的建模能力。图1给出了残差模块结构。

### 1.2 构建具有注意力的知识蒸馏模型

选取ResNeXt模型作为骨干特征提取网络,教师网络和学生网络均选用ResNeXt-100来构建整体的知识蒸馏框架。在每一个残差结构中引入CBAM模块<sup>[17]</sup>,构建具有注意力机制的Res\_CBAM结构,如图2所示。

注意力是引导特征聚焦到重要区域进行识别的一种简单有效的方法。设  $f_i = H_i(x)$  是来自 CNN 第  $i$  层的中间层特征输出, 其中  $H_i$  表示特征提取函数,  $x$  为输入数据。 $f_i$  的注意力映射可以表示为  $A_i(f_i)$ , 其中  $A_i$  是注意力模块。中间层特征通过注意力模块得到对应的注意力图并与其相乘, 然后传递到下一层。在训练过程中, 通过不断更新注意力模块中可训练的参数以提高模型性能, 从而得到更精确的注意力图。注意力机制可以表示为

$$f'_i = A_i^c(f_i) \otimes f_i \quad (1)$$

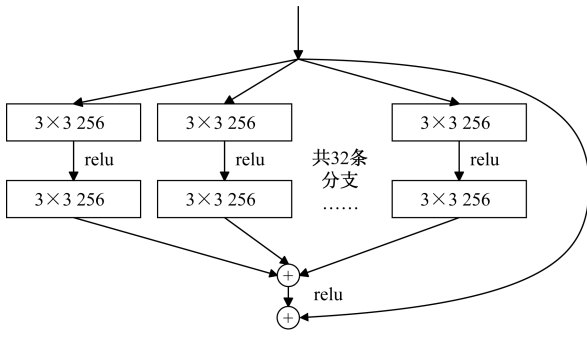
$$f''_i = A_i^s(f'_i) \otimes f'_i \quad (2)$$

式中:  $A_i^c$  和  $A_i^s$  分别为通道注意力模块和空间注意力模块;  $f'_i$  和  $f''_i$  分别为通过通道注意力模块  $A_i^c$  和通过空间注意力模块  $A_i^s$  作用后的特征;  $\otimes$  为特征相乘。

$$A^c(f) = \sigma \{ \text{FC}[\text{AvgPool}(f)] + \text{FC}[\text{MaxPool}(f)] \} \quad (3)$$

$$A^s(f) = \sigma \{ f^{7 \times 7}[\text{AvgPool}(f); \text{MaxPool}(f)] \} \quad (4)$$

式中:  $\sigma(\cdot)$  为 sigmoid 函数;  $\text{FC}(\cdot)$  和  $f^{7 \times 7}(\cdot)$  分别为全连接层和卷积核为  $7 \times 7$  的卷积层。



relu 为激活函数  
图1 残差模块

Fig. 1 Residual block

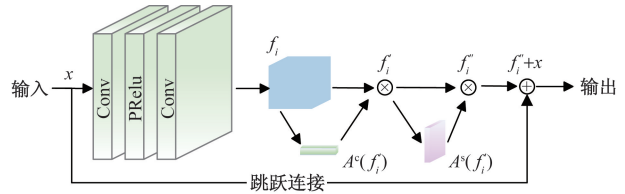


图2 Res\_CBAM 模块分支结构

Fig. 2 Res\_CBAM module branch structure

### 1.3 标签知识蒸馏

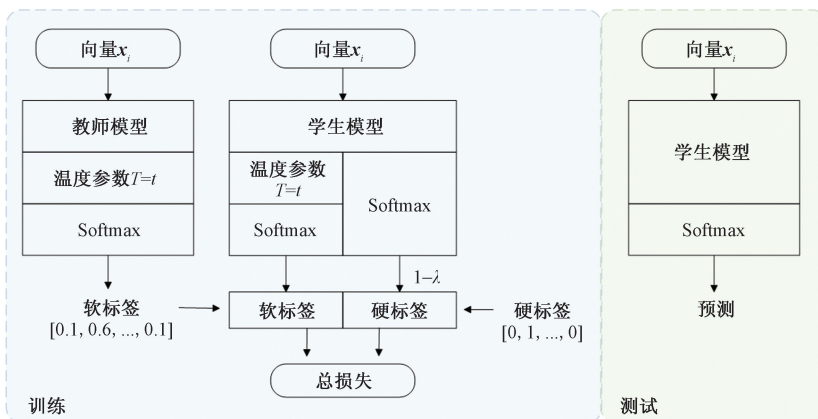
标签知识 (label knowledge) 是指通过给定的标签或类别来传递知识。在监督学习任务中, 教师模型通过将正确的标签与相应的输入数据关联, 将这种知识传递给学生模型。标签知识蒸馏首先生成“软标签”, 即使用教师模型对训练数据进行推理, 得到每个样本的输出概率分布, 这些概率分布被称为软标签。相比于硬标签 (真实标签) 软标签包含更多的信息, 因此能够提供更丰富的知识。然后使用软标签来训练学生模型。标签知识蒸馏过程如图3所示。在训练过程中, 通常会引入一个温度参数  $T$  来平滑教师模型的输出概率分布, 公式为

$$q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \quad (5)$$

式(5)中:  $q_i$  为每个类别输出的概率;  $z_j$  为每个类别输出的 logit, 其中  $j$  的取值范围是所有类别;  $T$  为温度, 当  $T = 1$  时, 就是标准的 Softmax 公式。

$T$  越大, Softmax 输出的概率分布越平滑, 分布的熵越高, 负标签中的信息就会相对丰富, 从而使模型在训练过程中更加关注负标签, 而这种信息也对区分不同类有着关键作用。标签知识蒸馏总的损失函数  $L_{\text{logit\_distill}}$  表示为

$$L_{\text{logit\_distill}} = L_{\text{hard}} + L_{\text{soft}} \quad (6)$$



$\lambda$  为权重参数

$1-\lambda$  为权重参数

图3 标签知识蒸馏过程

Fig. 3 Label knowledge distillation process

式(6)中:  $L_{\text{hard}}$  和  $L_{\text{soft}}$  分别为硬标签损失和软标签损失。

#### 1.4 注意力相似性蒸馏

输出层的标签知识所提供的信息较为有限,中间层知识能够有效弥补输出层知识信息单一和不够丰富的不足,为模型提供更加丰富的特征知识。中间层知识的形式多种,各种形式的知识提供的信息各有优缺点。本文研究中选取注意力图作为中间层知识,通过将高质量人脸图像的注意力图传递给低质量人脸图像,从而提高低质量图像的认识性能。

$$\begin{aligned} \rho_i &= 1 - \langle A_{T,i}(f_{T,i}), A_{S,i}(f_{S,i}) \rangle \\ &= 1 - \frac{A_{T,i}(f_{T,i}) \cdot A_{S,i}(f_{S,i})}{\|A_{T,i}(f_{T,i})\|_2 \|A_{S,i}(f_{S,i})\|_2} \end{aligned} \quad (7)$$

式(7)中:  $\rho_i$  为教师网络和学生网络第  $i$  层注意力图之间的余弦距离;  $\langle \cdot, \cdot \rangle$  为余弦相似度;  $\|\cdot\|_2$  为 L2 标准化;  $A_i(f_i)$  为第  $i$  层特征的注意力图;  $T$  和  $S$  分别为教师网络和学生网络;  $A_{T,i}(f_{T,i})$  和  $A_{S,i}(f_{S,i})$  分别为根据第  $i$  层教师和学生网络特征估计所得的注意力图。

使得异质图像注意力图之间的余弦距离相近可以提高它们之间的相似性。注意力相似性蒸馏损失表示为

$$L_{\text{attention\_distill}} = \sum_{i=1}^N \frac{\rho_i^s + \rho_i^c}{2} \quad (8)$$

式(8)中:  $N$  为用于蒸馏的层数。

对教师网络和学生网络之间的通道和空间注意力图的余弦距离进行平均,并在主干网络各层

( $i = 1, 2, \dots, N$ ) 上求和。

#### 1.5 算法总的损失函数

Arcface 损失是一种用于人脸识别任务的损失函数,它在解决人脸识别中的类内差异小、类间差异大的问题上表现出色。Arcface 损失函数是基于传统的 Softmax 损失函数进行修改得到的,传统的 Softmax 损失可以表示为

$$L_{\text{Softmax}} = -\frac{1}{N} \sum_{i=1}^N \ln \frac{\exp\{\mathbf{W}_{y_i}^T \mathbf{x}_i + \mathbf{b}_{y_i}\}}{\sum_{j=1}^n \exp\{\mathbf{W}_j^T \mathbf{x}_i + \mathbf{b}_j\}} \quad (9)$$

式(9)中:  $\mathbf{x}_i \in \mathbf{R}^d$  为属于第  $y_i$  类的第  $i$  个样本的嵌入(embedded)特征;  $N$  和  $n$  分别为批次大小和类别数;  $\mathbf{W}_j \in \mathbf{R}^d$  为最后一个全连接层权重  $W \in \mathbf{R}^{d \times n}$  的第  $j$  列的偏置项;  $\mathbf{b}_j$  为偏置;  $\mathbf{W}_{y_i}$  和  $\mathbf{b}_{y_i}$  分别为第  $y$  类第  $i$  个样本的权值和偏置。

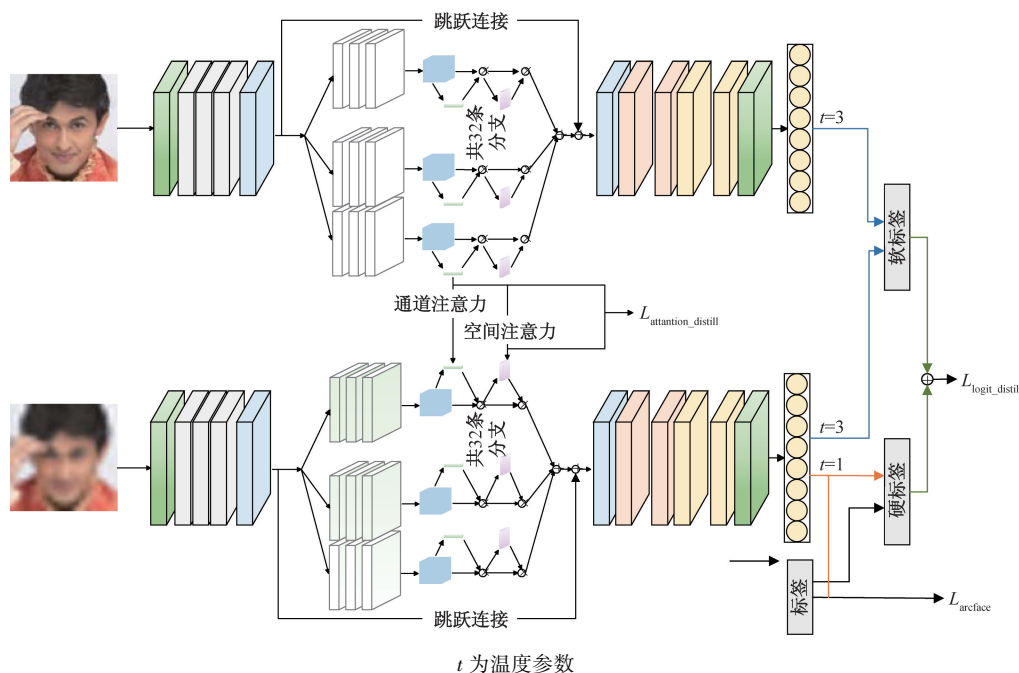
为了简单起见,把偏置  $\mathbf{b}_j$  设置为 0,那么  $\mathbf{W}_{y_i}^T \mathbf{x}_i$  可表示为

$$\mathbf{W}_{y_i}^T \mathbf{x}_i = \|\mathbf{W}_j\| \|\mathbf{x}_i\| \cos \theta_j \quad (10)$$

式(10)中:  $\theta_j$  为  $\mathbf{W}_j$  和  $\mathbf{x}_i$  之间的夹角。

Arcface 损失通过 L2 正则化使  $\|\mathbf{W}_j\| = 1$ ,  $\|\mathbf{x}_i\| = 1$  来最大化类间差异性和最小化类内差异性,引入了加性角度间隔  $m$  以避免收敛困难。Arcface 损失函数可表示为

$$L_{\text{Arcface}} = -\frac{1}{N} \sum_{i=1}^N \ln \frac{\exp\{s[\cos(\theta_{y_i} + m)]\}}{\exp\{s[\cos(\theta_{y_i} + m)]\} + \sum_{i=1, j \neq y_i}^n e^{s[\cos \theta_j]}} \quad (11)$$



$t$  为温度参数

图4 本文算法整体网络结构

Fig. 4 The overall network structure of the algorithm in this paper

式(11)中:  $m$  为  $x_i$  和  $W_{y_i}$  之间的加性角边缘惩罚;  $s$  为缩放因子。

本文算法总的损失为识别损失、标签知识蒸馏损失和注意力相似性蒸馏损失的相结合。本文算法使用 Arcface 作为网络的识别损失,因此,标签知识蒸馏中网络最后一层使用 Arcface 输出的概率分布。总的损失函数表示为

$$L_{\text{Total}} = L_{\text{arcface}} + L_{\text{logit\_distill}} + L_{\text{attention\_distill}} \quad (12)$$

式(12)中:  $L_{\text{arcface}}$  为识别损失;  $L_{\text{logit\_distill}}$  为标签知识蒸馏损失;  $L_{\text{attention\_distill}}$  为注意力相似性蒸馏损失。

## 1.6 整体网络结构

基于知识蒸馏的低质量人脸图像识别模型如图4所示,本文研究中采用的总体蒸馏方法由两部分组成:一是基于 logit 的标签蒸馏,二是基于注意力的中间层蒸馏。教师网络和学生网络均选择 ResNeXt-100,其基本单元模块是在引入双通道注意力模块后的残差块。网络的总层数为 100,每个阶段对应的卷积层数目分别为 3、13、30 和 3。

从图4可看出,整体网络是分为两分支:第一分支是教师网络,用于训练高质量人脸图像以提取丰富特征知识;第二分支是学生网络,用于训练低质量人脸图像。训练过程中,学生网络参考教师网络,将教师网络的预测结果作为标签,通过计算学生网络与教师网络预测结果之间的差异来计算标签蒸馏损失函数。此外,通过将高质量人脸图像的注意力图传递给低质量人脸图像,提供完整的特征信息,从而弥补低质量图像的缺陷,提高识别性能。

## 2 实验结果与分析

### 2.1 实验数据集

本文采用公开人脸数据集 CASIA-WebFace<sup>[18]</sup>、AgeDB-30<sup>[19]</sup> 和 CPLFW<sup>[20]</sup> 进行模型训练和测试。CASIA-WebFace 是一个用于人脸识别的数据集,由中国科学院自动化研究所创建。该数据集包含 10 575 个身份的 494 414 张人脸图像,其中每个身份有 2~804 张图像。AgeDB-30 由香港中文大学创建,包含 30 个身份的 16 488 张人脸图像,其中每个身份有 548 张图像。CPLFW 是基于野外标记人脸 LFW 数据集标注的跨姿态数据集,包含 5 749 个身份的 13 233 张人脸图像。由于训练网络需要高低质量人脸图像对,采用双三次插值算法对高质量图像以比例因子 2、4、8 进行下采样,并加入高斯模糊生成逼真的低质量人脸图像。本文研究中使用 CASIA-WebFace 数据集进行训练, AgeDB-30 与 CPLFW 用于测试。CPLFW 数据集样本如图5所示。



图5 CPLFW 数据集样本

Fig. 5 Example of CPLFW dataset

### 2.2 评价指标

准确率(accuracy, Acc)是人脸识别中最常用的评价指标。简而言之,准确率是指在给定的测试数据集,模型正确分类的样本数与总样本数之比。其计算公式为

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (13)$$

式(13)中: TP 为相似度大于给定阈值且真实标签为同一个人的图像对数; FP 为相似度大于给定阈值且真实标签不是同一个人的图像对数; FN 为相似度小于给定阈值且真实标签为同一个人的图像对数; TN 为相似度小于给定阈值且真实标签不是同一个人的图像对数。

### 2.3 实验环境与设置

本次实验所采用的硬件处理器为 Intel(R) Xeon(R) Gold 6330 CPU@2.00 GHz,内存为 80 G, GPU 为 RTX3090,显存为 24 G。实验虚拟环境为 Ubuntu18.04,深度学习框架为 Pytorch = 1.8.1。初始学习率为 0.01,每经过 5 轮训练,学习率下降至原学习率的 0.1 倍,共训练 20 轮。

### 2.4 消融实验

为了验证本文算法提出的知识蒸馏对整体模型的贡献度,对采用不同模式训练的模型性能在 AgeDB-30 和 CPLFW 数据集上进行了消融实验。首先,不使用任何知识蒸馏的模型作为基础模型。其次,分别单独使用不同的知识蒸馏方法进行比较和分析,在基础模型上使用标签知识蒸馏将教师网络的输出特征知识传递给学生网络;在基础模型上通过注意力相似性知识蒸馏将教师网络的中间层知识传递给学生网络。最终,结合标签知识蒸馏和注意力相似性蒸馏,将高质量人脸图像的丰富信息传

递给低质量图像,从而提高低质量图像的认识准确率。实验结果如表 1、表 2 及图 6、图 7 所示。

从表 1 和表 2 可以看出,采用标签知识蒸馏方法的模型在不同质量图像下的认识性能,相比无蒸馏方法的模型在 AgeDB-30 和 CPLFW 测试集上分别提升了 1.93%、10.61%、22.05% 和 2.11%、9.55%、24.75%。这表明标签知识蒸馏通过将教师网络输出层的有效信息传递至学生网络,从而提高了模型的准确率,初步验证了知识蒸馏对提升模型性能的有效性。使用注意力相似性知识蒸馏方法的模型认识性能在不同质量图像下,相比无蒸馏方法的模型分别提升了 1.97%、10.92%、1.24% 和 2.38%、10.33%、27.50%。此外,与标签知识蒸馏方法相比,注意力相似性知识蒸馏方法的模型性能分别提高了 0.04%、0.11%、1.24% 和 0.27%、0.78%、2.75%。这进一步说明中间层知识蒸馏方法较输出层知识蒸馏方法更加有效。最后,结合两种知识蒸馏方法的模型认识性能比单独使用注意力相似性知识蒸馏方法的模型分别提高了 0.28%、0.61% 和 1.35% 和 0.42%、0.25%、0.35%。从而证明了结合两种知识蒸馏方法可使的模型认识准确率得到一定程度的提高。

由图 6 和图 7 可以直观地看出,未经过任何蒸馏的模型随着图像质量的降低,模型的认识准确率逐渐下降。而本文提出的方法表明,经过蒸馏的模型在面对不同质量的图像时,具有更高的认识准确率。不同的蒸馏方式对模型性能的贡献度也有所不同,相比于标签知识蒸馏,注意力相似性蒸馏能够从教师模型中将有利于认识的中间层特征传递给学生模型。当结合两种知识蒸馏方法同时对教师

表 1 在 AgeDB-30 测试集上消融实验结果  
Table 1 Results of ablation experiments on the AgeDB-30 test set

方法	准确率/%		
	×2	×4	×8
BasicNet	91.98	79.34	55.28
BasicNet + logit_KD	93.91	89.95	77.33
BasicNet + attention_KD	93.95	90.06	78.57
BasicNet + logit_KD + attention_KD	94.23	90.67	79.92

表 2 在 CPLFW 测试集上消融实验结果  
Table 2 Results of ablation experiments on the CPLFW test set

方法	准确率/%		
	×2	×4	×8
BasicNet	88.07	78.50	55.45
BasicNet + logit_KD	90.18	88.05	80.20
BasicNet + attention_KD	90.45	88.83	82.95
BasicNet + logit_KD + attention_KD	90.87	89.08	83.30

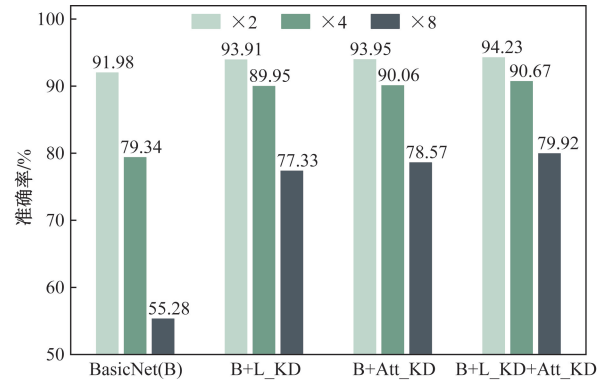


图 6 在 AgeDB-30 测试集不同质量图像上消融实验结果  
Fig. 6 Results of ablation experiments on human images of different quality in the AgeDB-30 test set

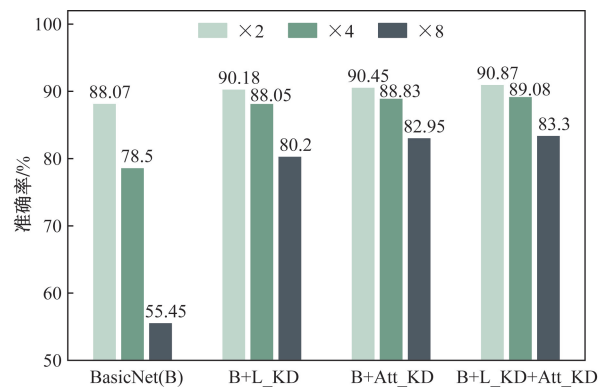


图 7 在 CPLFW 测试集不同质量图像上消融实验结果  
Fig. 7 Results of ablation experiments on human images of different quality in the CPLFW test set

模型的输出 logit 和中间层的注意力图进行蒸馏时,其认识准确率往往高于任何单独采用一种知识蒸馏的方法。因此可以看出,结合多种知识蒸馏的方法效果更好。通过对比基础模型认识准确率和最终模型的认识准确率可以发现,本文算法在面对较低质量的人脸图像时,仍能保持良好的认识准确率,表明使用知识蒸馏可以有效提升模型性能,并增强模型的鲁棒性。

## 2.5 对比实验

为了验证本文算法的性能,目前主流的人脸识别算法 FaceNet<sup>[5]</sup>、ArcFace<sup>[6]</sup>、CosFace<sup>[21]</sup>、CenterFace<sup>[22]</sup>、AT<sup>[23]</sup> 和 A-SKD<sup>[24]</sup> 与本文算法在 AgeDB-30 和 CPLFW 测试集中针对不同质量图像进行性能对比测试。对比实验结果如表 3、表 4 和图 8、图 9 所示。

从表 3 和表 4 可知,在 AgeDB-30 和 CPLFW 测试集的不同质量图像上,本文算法的人脸识别率显著优于 FaceNet 和 ArcFace。在 AgeDB-30 测试集上,与 CenterFace 和 CosFace 相比,本文算法在不同质量图像上的认识率分别提高了 6.41%、8.08%、

15.36%、23.83% 和 0.22%、2.07%、9.73%、21.26%。前几种算法由于主要针对高质量人脸图

表3 在 AgeDB-30 测试集上本文算法与其他算法的识别率比较

Table 3 Comparison of the recognition rate of the algorithm proposed in this paper with other algorithms on the AgeDB-30 test set

方法	准确率/%			
	×1	×2	×4	×8
FaceNet	71.32	57.93	56.23	50.05
ArcFace	72.50	72.50	66.71	55.42
CenterFace	88.07	86.15	75.30	55.90
CosFace	94.26	92.06	80.93	58.46
AT	87.03	85.78	79.29	75.95
A-SKD	93.73	93.26	88.38	78.65
本文方法	94.48	94.23	90.66	79.72

表4 在 CPLFW 测试集上本文算法与其他算法的识别率比较

Table 4 Comparison of the recognition rate of the proposed algorithm in this paper with other algorithms on CPLFW test set

方法	准确率/%			
	×1	×2	×4	×8
FaceNet	70.81	56.54	54.24	56.24
ArcFace	56.61	54.43	53.47	53.93
CenterFace	84.85	83.20	73.58	55.60
CosFace	89.53	88.75	80.61	61.67
A-SKD	88.81	88.83	87.70	82.91
本文方法	90.87	89.87	89.08	83.30

像识别,因此在图像质量最好的情况下取得了较高的识别准确率,但随着图像质量的降低,算法的性能急剧下降。这表明,面对现实环境中获得的低质量人脸图像,通用人脸识别算法并不具备高鲁棒性,不适用于低质量人脸识别问题。AT 和 A-SKD 算法与本文算法相似,同样将知识蒸馏方法引入分类任务中,前者是将单通道注意力图传递给了学生网络,后者则是将双通道注意力图传递给了学生网络。与本文算法不同的是,两者都仅用了一种知识蒸馏的方法。与 AT 和 A-SKD 相比,本文算法的识别率分别提高了 7.45%、8.45%、11.37%、3.77% 和 0.75%、0.97%、2.28%、1.07%。在 CPLFW 测试集上,与 CenterFace 和 CosFace 相比,本文算法在该数据集不同质量图像上分别提高了 6.02%、6.67%、15.5%、27.7 和 1.34%、1.12%、8.47%、21.63%。与 A-SKD 相比,本文算法提高了 2.06%、1.04%、1.38%、0.39%。可以看出,相比通用的人脸识别算法以及只采用一种知识蒸馏的算法,本文提出的结合两种知识蒸馏的算法性能更好。

由图 8 和图 9 可见,在 AgeDB-30 和 CPLFW 测试集上,通用人脸识别算法 FaceNet 和 ArcFace 的人脸识别率在不同图像质量下波动较大,而 CenterFace 和 CosFace 算法的识别率随着图像质量的降低逐步下降,尤其是在图像尺寸较小时。由此可见,通用人脸识别算法面对低质量人脸图像时,算法稳定性不足。AT 和 A-SKD 算法同样使用了知识蒸馏方法,其识别率明显高于前几种方法,表明了知识

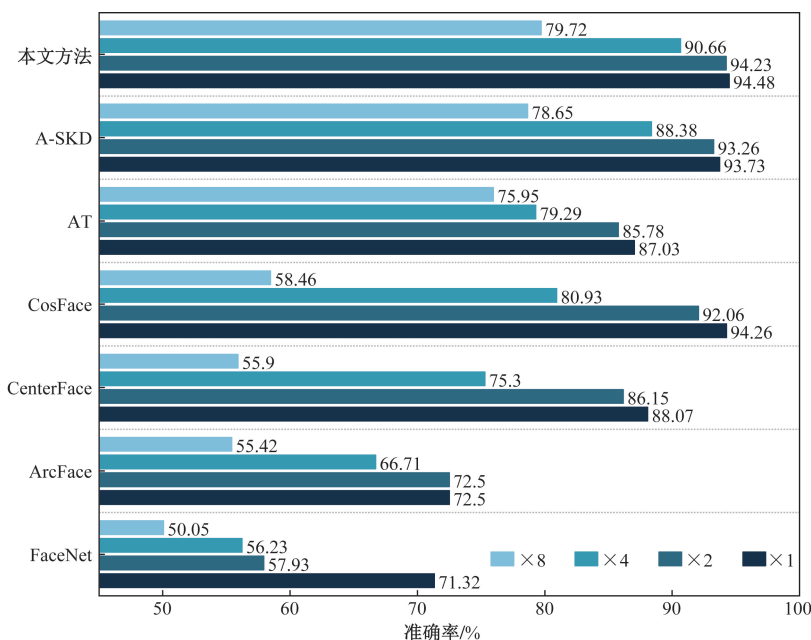


图8 在 AgeDB-30 测试集不同质量图像上本文算法与其他算法的识别率比较

Fig. 8 Comparison of recognition rates of the proposed algorithm of this paper with other algorithms on different quality images of AgeDB-30 test set

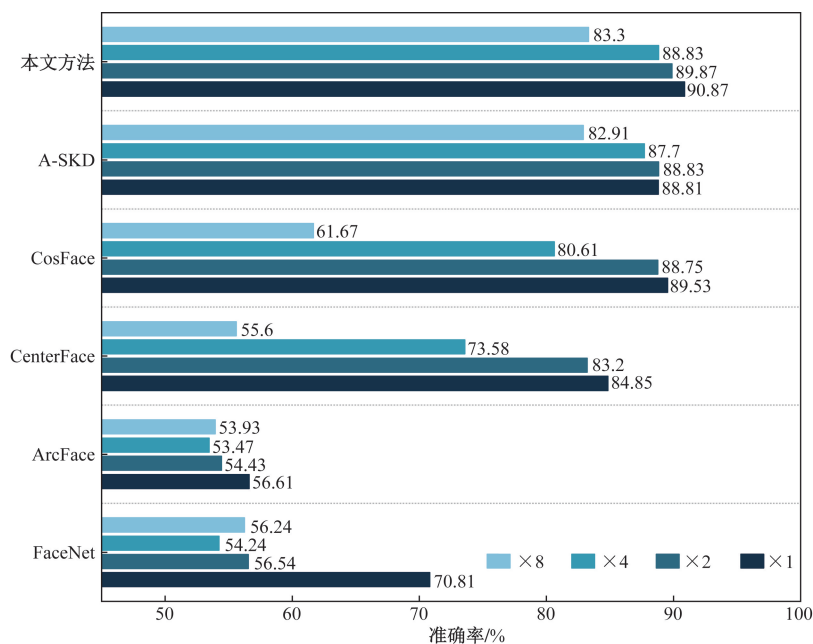


图9 在CPLFW测试集不同质量图像上本文算法与其他算法的识别率比较

Fig. 9 Comparison of recognition rates of the proposed algorithm of this paper with other algorithms on different quality images of CPLFW test set

蒸馏方法的有效性。相比之下,本文研究提出的结合两种知识蒸馏的方法在识别准确率上相较于 AT 和 A-SKD 算法也有明显提升。通过对比可以发现,本文算法在两种测试集上表现出较好的性能,所提出的网络更具有鲁棒性和泛化性。

### 3 结论

提出了一种基于知识蒸馏的低质量人脸图像识别算法。首先,将 ResNeXt 网络作为骨干特征提取网络,并引入 CBAM 模块以构建具有注意力机制的知识蒸馏框架,然后,将教师网络最后的输出作为标签知识,将注意力图作为中间层,通过结合两种知识蒸馏方法,保证从教师网络中可传递的特征知识多样性,最后,将标签知识蒸馏损失、注意力图蒸馏损失和识别损失的加权平均融合作为网络的总损失函数,保证了学生网络模型具有更强的特征表示能力。实验结果表明所提算法相比通用人脸识别算法更具鲁棒性和泛化性。

#### 参 考 文 献

[1] 张飞. 基于特征脸的 Adaboost 检测算法与识别应用研究[J]. 大理大学学报, 2022, 7(6): 18-21.  
Zhang Fei. Research on Adaboost detection algorithm and recognition application based on eigenface[J]. Journal of Dalian University, 2022, 7(6): 18-21.

[2] 崔希, 刘首明. 基于朴素贝叶斯算法的智能电厂监控夜间视频人脸识别方法[J]. 电子测试, 2021(24): 588-595.  
Cui Xi, Liu Shouming. Face recognition method for nighttime video

of intelligent power plant surveillance based on simple Bayes algorithm[J]. Electronic Test, 2021(24): 588-595.

[3] 张敦凤, 高宁化, 王姮, 等. 基于分块 LBP 融合特征和 SVM 的人脸识别算法[J]. 传感器与微系统, 2019, 38(5): 154-156, 160.  
Zhang Dunfeng, Gao Ninghua, Wang Heng, et al. Face recognition algorithm based on chunked LBP fusion features and SVMs[J]. Transducer and Microsystem Technologies, 2019, 38(5): 154-156, 160.

[4] 蔡晓云, 尹贺峰. 基于低秩投影与稀疏表示的人脸识别算法[J]. 科学技术与工程, 2019, 19(17): 244-249.  
Cai Xiaoyun, Yi Hefeng. Face recognition algorithm based on low rank projection and sparse representation[J]. Science Technology and Engineering, 2019, 19(17): 244-249.

[5] Schroff F, Kalenichenko D, Philbin J. Facenet: a unified embedding for face recognition and clustering[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015: 815-823.

[6] Deng J, Guo J, Xue N, et al. Arcface: additive angular margin loss for deep face recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 4690-4699.

[7] Ma C, Jiang Z, Rao Y, et al. Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation[C]// Proceeding of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 5569-5578.

[8] Liu S, Xiong C, Shi X, et al. Progressive face super-resolution with cascaded recurrent convolutional network[J]. Neurocomputing, 2021, 449: 357-367.

[9] Lu Z, Jiang X, Kot A. Deep coupled ResNet for low-resolution face recognition[J]. IEEE Signal Processing Letters, 2018, 25(4): 526-530.

- [10] Zangeneh E, Rahmati M, Mohsenzadeh Y. Low resolution face recognition using a two-branch deep convolutional neural network architecture [J]. *Expert Systems with Applications*, 2020, 139: 112854.
- [11] 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 review of research on knowledge distillation in deep learning [J]. *Chinese Journal of Computers*, 2022, 45(8): 1638-1673.
- [12] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network [J]. *arXiv preprint arXiv: 1503.02531*, 2015.
- [13] 孟宪法, 刘方, 李广, 等. 卷积神经网络压缩中的知识蒸馏技术综述 [J]. *计算机科学与探索*, 2021, 15(10): 1812-1829.  
Meng Xianfa, Liu Fang, Li Guang, et al. A review of knowledge distillation techniques in convolutional neural network compression [J]. *Journal of Frontiers of Computer Science & Technology*, 2021, 15(10): 1812-1829.
- [14] Xu G, Liu Z, Li X, et al. Knowledge distillation meets self-supervision [C]// *Proceedings of the European Conference on Computer Vision (ECCV)*. Cham: Springer International Publishing, 2020: 588-604.
- [15] Massoli F V, Amato G, Falchi F. Cross-resolution learning for face recognition [J]. *Image and Vision Computing*, 2020, 99: 103927.
- [16] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 492-1500.
- [17] Woo S, Park J, Lee J Y, et al. Cbam: convolutional block attention module [C]// *Proceedings of the European Conference on Computer Vision (EC-CV)*. Munich: Springer International Publishing, 2018: 3-19.
- [18] Yi D, Lei Z, Liao S, et al. Learning face representation from scratch [J]. *arXiv preprint arXiv: 14-11.7923*, 2014.
- [19] Moschoglou S, Papaioannou A, Sagonas C, et al. Agedb: the first manually collected, in-the-wild age database [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Honolulu: IEEE, 2017: 51-59.
- [20] Zheng T, Deng W. Cross-poselfw: a database for studying cross-pose face recognition in unconstrained environments [J]. *Beijing University of Posts and Telecommunications, Technical Report*, 2018, 5(7): 1-6.
- [21] Wang H, Wang Y, Zhou Z, et al. Cosface: large margin cosine loss for deep face recognition [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake: IEEE, 2018: 5265-5274.
- [22] Xu Y, Yan W, Yang G, et al. CenterFace: joint face detection and alignment using face as point [J]. *Scientific Programming*, 2020, 2020: 1-8.
- [23] Zagoruyko S, Komodakis N. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer [J]. *arXiv preprint arXiv: 1612.03928*, 2016.
- [24] Shin S, Lee J, Lee J, et al. Teaching where to look: attention similarity knowledge distillation for low resolution face recognition [C]// *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022: 631-664.