计算机科学与探索
Journal of Frontiers of Computer Science and Technology
1673-9418/2023/17(08)-1893-11
doi: 10.3778/j.issn.1673-9418.2209111

开放科学(OSID)

融合注意力机制和课程式学习的人脸识别方法

王海勇[1,2],潘海涛[1,2+],刘贵楠[1]
1. 南京邮电大学 计算机学院,南京 210023
2. 南京邮电大学 智慧校园研究中心,南京 210023
+ 通信作者 E-mail: 1220045123@njupt.edu.cn

摘　要:针对当前人脸识别算法中提取的人脸特征可区分性不强、难易样本区分度不够的问题,提出一种融合注意力机制和课程式学习的人脸识别算法(ECACFace)。该算法提出一种高效的空间通道注意力模块(ESCA)并将其融入特征提取网络的基本模块中,使用高效的通道注意力模块(ECA)获取通道关注度并在ECA之后加入空间注意力模块,在关注图像通道信息的基础之上进一步获取空间关注度,从而得到信息更加丰富的人脸特征向量用于人脸分类。同时在训练时引入基于课程式学习的损失函数,做到在训练过程中区分难易样本,并在前期着重训练简单样本,后期着重训练困难样本,实现有区分度的样本学习。在CASIA-WebFace数据集上训练基于轻量级网络和浅层网络的ECACFace,与原始网络相比有超过1.5个百分点的精度提升。在百万规模的MS1MV2数据集上训练基于深层网络的ECACFace,在CPLFW数据集上比ArcFace精度提升了1.14个百分点,实验结果表明,融合ESCA模块和基于课程式学习的损失函数能够进一步提升人脸识别性能。

关键词:人脸识别;特征提取;课程式学习;注意力机制
文献标志码:A　　**中图分类号:**TP181

Face Recognition Method Based on Attention Mechanism and Curriculum Learning

WANG Haiyong[1,2], PAN Haitao[1,2+], LIU Guinan[1]
1. College of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
2. Smart Campus Research Center, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract: Aiming at the problems that the facial features extracted from current face recognition algorithms are not distinguishable and the discrimination of difficult and easy samples is not enough, a face recognition algorithm combining attention mechanism and curriculum learning is proposed, which is called efficient cooperative attention and curriculum face (ECACFace). The algorithm proposes an efficient spatial channel attention module (ESCA) and integrates it into the basic module of the feature extraction network. The efficient channel attention module (ECA) is used to obtain the channel attention, and the spatial attention module is added after the ECA. On the basis of paying attention to the image channel information, the spatial attention is further obtained, and the face feature vector with richer information is obtained for face classification. At the same time, the loss function based on curriculum learning is introduced to distinguish the difficult and easy samples in the training process. The simple samples are trained in the early stage and the difficult samples are trained in the later stage to realize the discriminative sample

基金项目:国家自然科学基金(61872190);赛尔网络下一代互联网技术创新项目(NGII20190612);江苏省博士后科研资助计划项目(2020Z058)。
This work was supported by the National Natural Science Foundation of China (61872190), the Cel Network Next Generation Internet Technology Innovation Project (NGII20190612), and the Postdoctoral Research Funding Program of Jiangsu Province (2020Z058).
收稿日期:2022-09-30　　**修回日期:**2022-11-21

learning. Training ECACFace based on lightweight network and shallow network on CASIA-WebFace dataset and it has an accuracy improvement of more than 1.5 percentage points compared with the original network. ECACFace based on deep network is trained on MS1MV2 dataset which has millions of data, and the accuracy tested on CPLFW dataset is increased by 1.14 percentage points compared with ArcFace. Experimental results show that the cooperation of ESCA module and the loss function based on curriculum learning can further improve the performance of face recognition.

Key words: face recognition; feature extraction; curriculum learning; attention mechanism

近年来人脸识别技术得到迅速发展并取得众多研究成果,人脸识别已成为计算机视觉领域的重要研究方向。随着深度学习的迅速发展,人脸识别技术得到了质的提升,如今的人脸识别已经成为一项重要的生物识别技术。目前人脸识别技术已经非常成熟并且普遍应用于实际生活,如安检系统、门禁系统、刷脸支付等。

基于深度学习的人脸识别算法研究[1]取得了重大进展,卷积神经网络强大的特征提取能力和较少的参数量使其逐渐取代了传统的人工神经网络以及手工提取人脸特征的方法。随着硬件资源的发展,基于卷积神经网络的人脸识别方法成为了主流。从AlexNet[2]、VGGNet[3],再到GoogleNet[4]、ResNet[5]、DenseNet[6],利用卷积神经网络进行图像分类已经取得显著效果。人脸识别问题也可以看成一个多分类问题,其要求降低类内散度而增大类间散度。通过卷积神经网络可以得到更具代表性的人脸表示,然而普通的卷积操作只会对图像不同通道的特征简单求和,默认不同通道特征对于后续任务的重要性相同,这对于通道数较多的特征图并不适用,可能会引入过多无用的特征。受人类视觉机制的启发,注意力机制被引入卷积神经网络,让网络学习更重要的特征,这对于提取人脸图像中的高质量信息有很大的作用,常用来辅助卷积神经网络提取图像的关键信息。最近有效的通道注意模块(efficient channel attention,ECA)[7]被应用于视觉任务,在图像分类和目标检测等计算机视觉任务上显现出先进的性能。人脸图像的某些区域对人脸分类起着关键作用,但ECA模块只关注图像的通道信息,忽视了空间信息。针对以上问题提出一种高效的空间通道注意模块(efficient spatial channel attention,ESCA),该模块在原始ECA模块的基础之上加入空间注意力模块,着重关注对人脸分类起重要作用的人脸区域。

近些年来对于人脸识别模型的优化主要集中在损失函数上,旨在缩短类内距离,扩大类间距离。ArcFace[8]模型的提出使深度人脸识别取得了巨大进步,其使用角余量在特征空间表示的超球面上划分人脸特征,进一步提升了人脸识别性能,但是ArcFace忽视了训练样本的难易程度对特征空间分布的影响,实际上在训练过程中不同难度的训练图像会产生不同程度的影响。MV-Arc-Softmax[9]提出对不同难度的样本给予不同的训练关注度,强调困难样本对于模型训练的重要性,但忽视了简单样本对于早期训练的重要性。针对以上问题,引入基于课程式学习的损失函数CurricularFace[10],提出模型的训练应该由易到难,模型先由简单样本中学习经验,再在后期通过学习困难样本进一步优化特征分布。综上所述,本文提出了一种融合注意力机制和课程式学习的人脸识别方法。主要贡献如下:

(1)提出一种高效的空间通道注意模块(ESCA),通过融入特征提取网络进一步提高网络的特征提取能力。

(2)引入CurricularFace中基于课程式学习的损失函数,在模型训练期间动态调整不同难度样本的权重分配,做到在训练前期着重训练简单样本,后期着重训练困难样本。

(3)提出一种融合注意力机制和课程式学习的人脸识别算法(efficient cooperative attention and curriculum face,ECACFace),并在五个主流的人脸测试数据集上进行测试。在轻量级、浅层和大型人脸识别网络中分别融合注意力模块ESCA和基于课程式学习的损失函数,实验证明本文提出的模型优于其他主流的人脸识别模型。

1 相关工作

1.1 注意力机制

卷积神经网络是在卷积运算的基础上建立起来的,它通过融合局部感受野中的空间和通道信息来提取信息特征,而使用注意力机制增强空间编码可以提高网络的表现力。将注意力转移到图像中最重

要的区域而忽略不相关的部分的方法被称为注意力机制。SE（squeeze-and-excitation）[11]模块通过显式建模通道之间的相互依赖性，自适应地重新校准通道特性响应。SK（selective kernel）[12]模块使用多个不同大小的卷积核，获取不同感受野的特征图，通过对不同特征图赋予不同的权值来获取不同感受野的重要性。BAM（bottleneck attention module）[13]和CBAM（convolutional block attention module）[14]选择同时关注通道信息和空间信息，BAM采用并行的方式，在获取空间注意图时使用空洞卷积增大感受野，而CBAM采用串行的方式，先获取通道注意图，再获取空间注意图。SGE（spatial group-wise enhance）[15]模块对特征图进行通道分组，通过注意力掩码对不同分组的特征进行语义增强，突出正确区域并抑制可能的噪声点。以上注意力模块大都通过复杂的卷积或者全连接的方式获取每个通道的特征，在提高性能的同时，显著增加了参数量，这对于大型网络来说是很消耗性能的。最近ECA模块克服了注意力机制设计中性能和复杂性的矛盾，通过避免维数约简提高性能，同时使用局部跨通道交互方式降低模块复杂度。

罗思诗等人[16]将ECA模块应用在人脸表情识别，增强部分面部区域的特征，这些区域通常对表情判别有很大影响。张宏鸣等人[17]在ECA模块中加入空间注意力模块并应用于MobileFaceNet[18]进行羊脸识别，提高了识别性能，但其空间注意力模块本质上还是基于全局平均池化和全局最大池化的操作，关注的还是单个通道特征图的全局信息。本文方法对ECA模块做进一步的改进，提出一种高效的空间通道注意力模块ESCA。针对ECA模块无法获取图像的空间信息的问题，加入空间注意力模块，使用3×3的池化核进行平均池化和最大池化操作提取图像的空间信息，在图像像素域动态分配权重，更多地关注对分类起主要作用的人脸区域。

1.2 损失函数

人脸识别是在闭集上训练，开集上测试，因此只能在有限的数据集上学习能分开未知类特征的算法模型。该算法模型的核心就是定义一个损失函数来进行度量学习，近些年来损失函数的设计已经成为人脸识别领域的一个重要研究课题。Google提出FaceNet[19]模型，将模型映射到欧几里德空间并使用三元组损失Triplet Loss训练来增大类间距，然而先构建三元组再基于三元组进行计算增加了计算复杂度。权重矩阵 $\boldsymbol{W}$ 与特征向量 $\boldsymbol{X}$ 的外积可以表示为 $\|\boldsymbol{W}\|\cdot\|\boldsymbol{X}\|\cos\theta$，其中 θ 为权重矩阵与特征向量的夹角。但 $\boldsymbol{W}$ 和 $\boldsymbol{X}$ 较大的参数量会导致模型训练难度大。随后SphereFace[20]模型被提出，使用A-Softmax Loss训练模型，提出权重矩阵归一化的概念，进一步降低了训练难度，但特征向量的存在导致模型训练难度依然较大。CosFace[21]模型提出将特征向量归一化，并将人脸特征空间映射到半径为 s 的超球面上，将优化重点集中在特征向量与权重向量之间的夹角，使用余弦相似度划分决策边界。2019年，ArcFace模型被提出并广泛用于深度人脸识别，使用角余量代替余弦余量，使决策边界的划分更明确并具有更好的可解释性。2020年，MV-Arc-Softmax被提出，开始强调错误分类的样本对于模型训练的重要性，在训练过程中增加对错误分类的样本权重，进行更有区分度的特征学习，但整个过程只关注困难样本对于训练的影响，忽视了简单样本对于训练的重要性。随后基于课程式学习的人脸识别算法CurricularFace被提出，该模型通过巧妙的损失函数设计，在不增加训练难度的前提下充分利用训练样本，在训练前期着重训练简单样本，训练后期着重训练困难样本。为了在有限的训练数据集下合理利用训练样本，本文算法引入基于课程式学习的损失函数，通过将该损失函数与注意力机制相结合进一步提升人脸识别性能。

2 本文算法

本文算法在人脸特征提取网络的基础之上，将有效的空间通道注意模块ESCA加入到特征提取网络中并引入基于课程式学习的损失函数，进一步提高了人脸识别性能。方法主要分为四个步骤：人脸检测、人脸对齐、人脸表示和人脸分类。训练过程如图1所示，首先采用MTCNN（multi-task convolutional neural network）[22]来对图像进行人脸检测和对齐，将人脸图像裁剪到112×112，裁剪后的图像送入改进的特征提取网络进行人脸表示，将输出的人脸特征向量和人脸标签一起输入课程式损失函数获取正负余弦相似度，并将其输入Softmax交叉熵损失计算分类损失。使用SGD进行反向传播更新网络参数、权重矩阵 $\boldsymbol{W}$ 和特征向量与类中心的夹角 θ。其中ESCA嵌入在每个bottleneck中。

2.1 特征提取网络

特征提取网络用于提取人脸图像的特征，提取

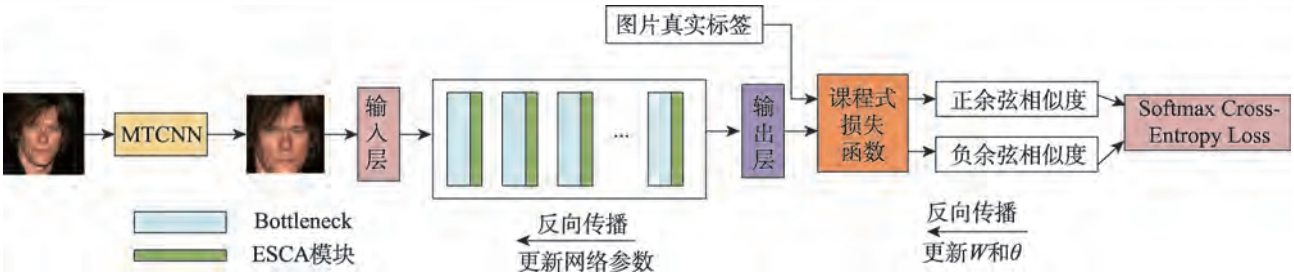

图1 本文算法整体框架

Fig.1 Overall framework of proposed algorithm

出的人脸特征向量将直接用于后续分类，因此特征提取网络的结构会间接影响最后的人脸识别性能。卷积神经网络强大的特征提取能力使得其在图像分类、目标检测等计算机视觉领域有着广泛的应用。为了体现本文方法的先进性能，本文使用不同种类的特征提取网络来提取人脸特征，将高效的空间通道注意力模块分别嵌入不同特征提取网络的bottleneck，主要使用三种网络模型：轻量级网络MobileFaceNet、浅层网络IR_50、深层网络IR_101。嵌入方式如图2所示，其中IR_50和IR_101的嵌入方式相同，三种网络模型的bottleneck数量如表1所示，bottleneck数量逐渐增多，其中三种网络模型对应的嵌入ESCA模块的bottleneck结构如图2所示。这三种网络结构均使

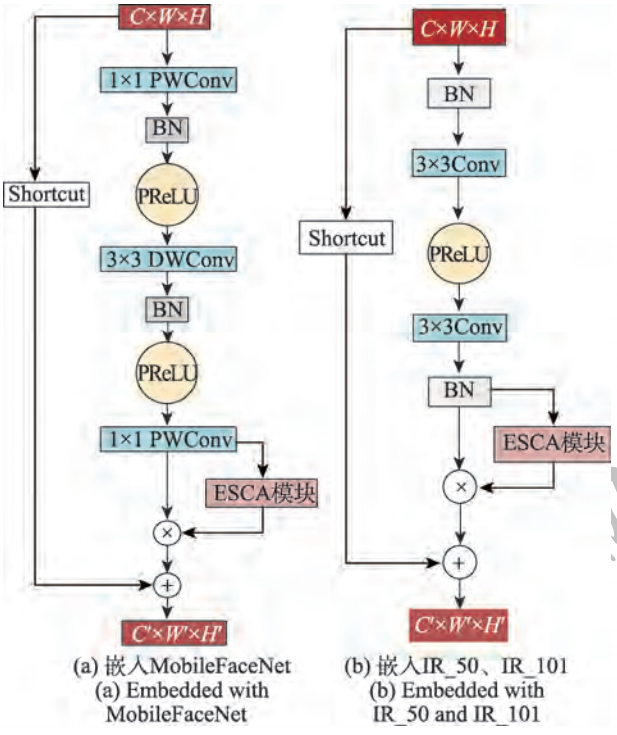

图2 嵌入注意力模块的bottleneck结构

Fig.2 Structure of bottleneck embedded attention module

表1 不同网络模型结构细节

Table 1 Structural details of different network models

网络模型	bottleneck数量	bottleneck结构
MobileFaceNet	15	图2(a)
IR_50	24	图2(b)
IR_101	49	图2(b)

用到残差单元，缓解因网络层数的增加导致的性能退化问题，Shortcut通过卷积或池化操作将输入特征图的通道数映射到与输出特征图相同。通过将ESCA模块融入特征提取网络，可以使网络更多地关注于提取人脸图像中的关键信息，进一步提升特征提取网络的特征提取能力。

2.2 高效的空间通道注意力模块ESCA

近年来，注意力机制在改善深度卷积神经网络的性能方面显示出巨大的潜力，但大多数现有方法致力于设计更为复杂的注意力模块，这会增加人脸识别模型的参数量和复杂度。ECA模块针对复杂度较大的问题做出了修改，降低了模型的参数量和复杂度并在视觉任务中保持相当的性能。原始的ECA模块如图3(a)所示，输入的特征图首先经过全局平均池化，获取各个通道特征图的全局信息，再通过自适应的一维卷积获取相邻通道的相关性，最后通过激活函数输出不同通道的权值，再作用于原始特征图。但是全局平均池化将每个通道的特征图取平均值，忽视了单个通道特征图的空间信息。人脸识别是一个对人脸图像分类的过程，人的眼睛、鼻子、嘴巴等一些重要的空间信息对分类起到关键作用。

针对ECA模块存在的问题，提出一种高效的空间通道注意力模块ESCA，模块结构如图3(b)所示。ESCA模块主要包括两部分：空间注意力模块和通道注意力模块。假设ESCA模块的输入为 $\boldsymbol{x}\in\mathbf{R}^{W\times H\times C}$，$W$、$H$、$C$ 分别为图像宽度、高度和通道数，ESCA的过程如式(1)所示：

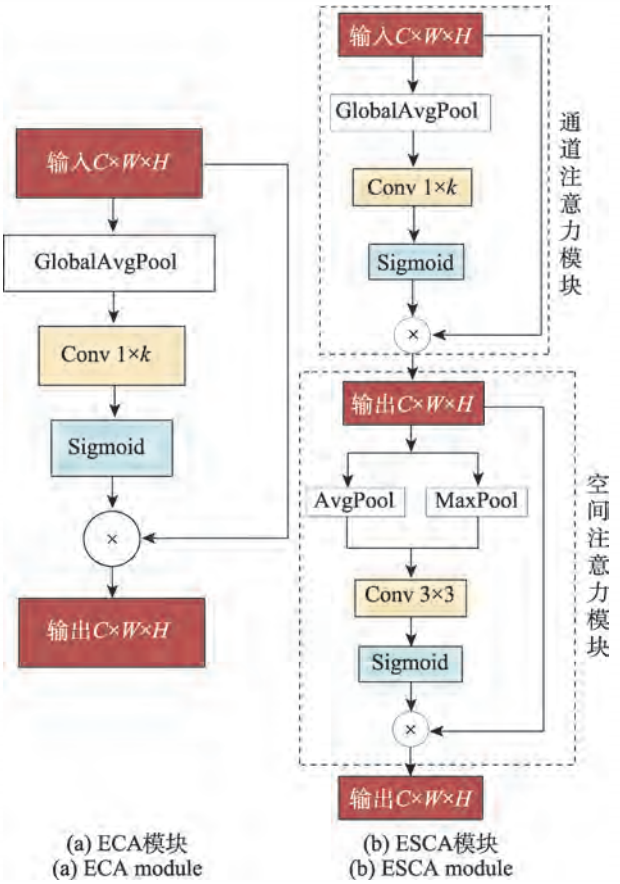

图3 注意力模块结构

Fig.3 Structure of attention module

$$x_{out}=CA(x_{in})\otimes SA(CA(x_{in})) \tag{1}$$

其中，CA为通道注意力模块，SA为空间注意力模块。CA采用原始的ECA模块，经过该模块后输出的特征图进入SA，进一步获取人脸图像的空间关注度，即每个像素对应的权值，并将该权值与通过CA输出的特征图对应像素相乘，最终得到ESCA模块的输出。空间注意力模块SA的过程如式(2)所示：

$$SA(x)=\sigma(f2(f1(AP(x),MP(x)))) \tag{2}$$

其中，AP和MP分别表示平均池化和最大池化操作，池化核大小选择为3×3，保持特征图的大小。$f1$表示通道连接操作，在通道维度连接经过平均池化和最大池化的特征图。$f2$表示一个3×3卷积，进行降维，σ表示Sigmoid激活函数，用于获取最终的空间像素权值。空间注意力模块在图像的像素空间为重要的像素分配大的权值，从而增强图像的空间信息，空间注意力模块的引入在提取特征时更多地关注人脸关键部位信息，进而提高人脸特征的可区分性。

2.3 基于课程式学习的损失函数

基于课程式学习的损失函数采用课程学习的思想，在训练前期强调简单样本的重要性，后期强调困难样本的重要性。后期对困难样本的学习会进一步优化从简单样本学习到的特征，基于课程式学习的损失函数与MV-Arc-Softmax相似，MV-Arc-Softmax中负余弦相似度的定义过度强调训练过程中的困难样本而忽视简单样本的重要性。假设第i个样本属于第y_i类，人脸类别数为n，θ_j为样本特征与第j类人脸类中心的夹角，则课程式损失的定义如式(3)所示：

$$L=-\ln\frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)}+\sum_{j=1,j\neq y_i}^{n}e^{sN(t^{(k)},\cos\theta_j)}} \tag{3}$$

其中，$\cos(\theta_{y_i}+m)$为正余弦相似度，用来衡量样本特征与其真实类中心的相似度，$N(t^{(k)},\cos\theta_j)$为负余弦相似度，用来衡量样本特征与非真实类中心的相似度，由式(3)可以看出该损失函数的核心就是计算每个样本特征对应的正余弦相似度和与其他非真实类中心的负余弦相似度。由样本标签可以直接计算出每个样本特征对应的正余弦相似度，负余弦相似度的计算如式(4)所示：

$$N(t,\cos\theta_j)=\begin{cases}\cos\theta_j,\cos(\theta_{y_i}+m)\geqslant\cos\theta_j\\ \cos\theta_j(t+\cos\theta_j),\cos(\theta_{y_i}+m)<\cos\theta_j\end{cases} \tag{4}$$

式中，当$\cos(\theta_{y_i}+m)<\cos\theta_j$时，即一幅人脸图像特征向量与某个非真实人脸类中心的夹角小于该特征向量与真实类中心的夹角加上角余量m时，此时该图像很有可能会被误分类，认为该图像为困难样本，改变困难样本的负余弦相似度为$\cos\theta_j(t+\cos\theta_j)$，否则该样本为简单样本。简单样本的负余弦相似度为$\cos\theta_j$，其中参数$t$是动态变化的，不需要手动设置，初期$t$接近于0，则$t+\cos\theta_j$小于1，即简单样本的负余弦相似度要大于困难样本的负余弦相似度，因此前期简单样本的交叉熵损失较大，重点学习优化简单样本，而后期$t+\cos\theta_j$大于1时，困难样本的负余弦相似度超过简单样本，困难样本交叉熵损失较大，开始学习优化困难样本，从而达到课程式学习的目的。文献[10]中发现正余弦相似度的平均值是评估t的一个很好的指标，但是基于小批量统计的方法通常面临一个问题：当在一个小批量中对许多极端数据进行采样时，统计数据可能会非常嘈杂，估计也会不稳定，指数移动平均(exponential moving average，EMA)[23]是解决这一问题的常用解决方案。通过EMA动态计算t值可以避免手动调参，t由式(5)自适应变化：

$$t^{(k)}=\alpha r^{(k)}+(1-\alpha)t^{(k-1)} \tag{5}$$

式中，$r^{(k)}$ 是第 k 个批次的正余弦相似度的平均值，α 为一个动量参数，参考文献[10]取 0.99。$r^{(k)}$ 的计算方式如式(6)所示：

$$r^{(k)}=\frac{1}{N^{(k)}}\sum_{i=1}^{N^{(k)}}\cos(\theta_{y_i}+m) \tag{6}$$

式(6)即计算一个批次中正余弦相似度的平均值，其中 $N^{(k)}$ 为第 k 个批次的图片数量，$\cos(\theta_{y_i}+m)$ 为正余弦相似度。

图 4 为训练集中样本的一些示例图，每一行代表同一身份，从左到右样本难度逐渐增加。可以看出简单样本大都是清晰无遮挡的人脸图像，而困难样本可能是有部分遮挡、光照或大姿态的人脸图像，这类图像很有可能会被误分类。基于课程式学习的损失函数可以在训练初期使类中心的优化更多依靠清晰的简单样本，在训练后期更多依靠复杂多变的困难样本优化以增强人脸识别对困难样本的适应性。

图 4 训练样本示例图

Fig.4 Example of training samples

3 实验结果与分析

3.1 实验环境与设置

实验设备使用 ubuntu 服务器，配置 4 个 NVIDIA Tesla P40 GPU，采用 Pytorch 深度学习框架进行网络模型训练。考虑到 GPU 通信时间，未使用分布式训练策略，参数设置遵循 CurricularFace，用 SGD 算法训练，动量为 0.9，权重衰减为 5E−4。使用 CASIA-WebFace 训练集训练，分别在 28、38、46 个 epoch 将学习率除以 10，模型在第 50 个 epoch 完成收敛，使用 MS1MV2 训练集训练，分别在第 10、18、22 个 epoch 将学习率除以 10，模型在第 24 个 epoch 完成收敛，其中超球面半径 s 为 64，余量 $m=0.5$。

3.2 数据预处理

使用 MTCNN 检测人脸并获取 5 个人脸关键点，再利用 5 个人脸关键点进行相似性变换并裁剪到大小为 112×112 的人脸图像。使用精炼的 CASIA-WebFace 和 MS1MV2 数据集作为训练数据。在几个主流的基准上测试人脸识别性能，主要包含 LFW、CFP_FP、CPLFW、AgeDB 和 CALFW。训练集和测试集对应的人脸身份数量以及图像数量如表 2 所示。CASIA-WebFace 包含了 10 575 个身份的 455 594 张人脸图像，MS1MV2 包含 85 000 个人脸身份的约 580 万张图片。LFW 包含了 5 749 人的 13 233 张不同姿态、表情的人脸图片，图片均来自生活中的自然场景；CFP_FP 由 500 个身份的 7 000 张图片组成，每个身份都有 10 张正面图像和 4 张侧面图像；CPLFW 和 CALFW 是基于 LFW 的跨姿态人脸数据集和跨年龄人脸数据集；AgeDB 包含 6 000 对共 440 个身份的 12 240 张图片，这些图片包含不同姿态、表情、年龄和性别。

表 2 实验相关数据集

Table 2 Datasets related to experiment

数据集	类型	身份数量	图片数量
CASIA-WebFace	训练集	10 575	455 594
MS1MV2	训练集	85 402	5 800 000
LFW	测试集	5 749	13 233
CFP_FP	测试集	500	7 000
CPLFW	测试集	5 749	13 233
AgeDB	测试集	440	12 240
CALFW	测试集	5 749	13 233

3.3 数据分析

3.3.1 注意力机制性能分析

在 IR_101 网络中分别加入多种主流的注意力模块，并使用基于课程式学习的损失函数训练模型，在 CASIA-WebFace 数据集上进行训练，验证性能如表 3 所示。由表 3 可以看出，相较于加入其他注意力模块，加入 BAM 和 ECA 对于人脸识别性能的提升较为明显，但加入 BAM 模块的网络训练过程中平均每个批次的训练耗时较长，导致网络训练耗时较长。加入 ECA 模块的 IR_101 在各个测试集上与原始 IR_101 相比均有所提升，与加入 SE 模块的网络相比，在 AgeDB 数据集上的精度提升了 0.75 个百分点，并且训练耗时较短，说明对于不同姿态、年龄、表情和性别的人脸图像，ECA 模块提取关键信息的能力更强，

表3 加入不同注意力模块识别性能对比

Table 3 Comparison of recognition performance with different attention modules

注意力模块	识别精度/%			训练耗时/(ms/batch)
	LFW	CFP_FP	AgeDB	
无	99.50	94.70	93.30	168.5
SE	99.33	95.25	93.20	187.0
SK	99.18	94.48	93.58	222.1
CBAM	99.24	95.21	93.81	254.9
BAM	99.40	95.30	94.00	243.4
SGE	99.25	94.98	93.00	196.0
ECA	99.41	95.31	93.95	**185.2**
ESCA	**99.50**	**95.40**	**94.50**	203.0

并且ECA模块在提升特征提取能力的同时具有较小的复杂度。本文提出的ESCA模块在ECA模块基础上加入了空间注意力模块。由表3可以看出，加入ESCA模块相比加入ECA模块有进一步的精度提升，在AgeDB数据集上相比原始网络有超过1个百分点的精度提升，ESCA模块中加入的空间注意力模块可以进一步获取特征图的空间关注度，突出对图像分类有决定性作用的空间特征，使网络提取到的特征向量可区分性更强，进而提高人脸识别模型的识别性能，同时ESCA模块保留了ECA模块中的局部跨通道交互方式，在获取通道关注度的同时保持较低的复杂度。加入ESCA模块相比加入CBAM和BAM模块依然有性能的提升，这是因为CBAM和BAM采用全连接的方式捕获所有通道之间的相关性，这是耗时且不必要的，ESCA中的通道注意力模块采用局部跨通道交互方式，只关联相邻的部分通道，在提升网络特征提取能力的同时减低模块复杂度。

在MS1MV2上重新训练原始的CurricularFace和ECACFace，两者均使用基于课程式学习的损失函数，两次训练经过24次epoch后完成收敛，准确率和迭代次数曲线如图5所示。由图5可以看出，前期ESCA模块的作用较小，到了后期趋于平稳，测试准确率均在99%以上，接近饱和。训练中期的准确率和迭代次数曲线如图6所示。由图6可以看出，训练中期ECACFace的识别准确率高于CurricularFace，并且都是向准确率高的方向波动，说明训练中期ESCA模块的加入使提取到的人脸特征更有可能被正确识别，这是因为ESCA模块突出了图像的关键特征信息，包括关键的通道特征信息和空间特征信息，使得关键信息在经过多层网络后不会丢失，而这些关键信息可以很大程度地决定是否是同一张人脸。训练前中期的损失和迭代次数曲线如图7所示。由图7可以看出，训练中期ECACFace比CurricularFace的损失降低得更快，即人脸识别模型的优化速度更快，因此加入ESCA模块可以加快人脸识别模型的优化速度，同时ECACFace训练时的损失向着小的方向波

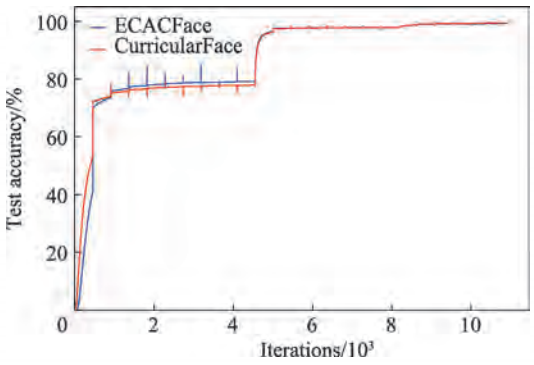

图5 准确率和迭代次数曲线

Fig.5 Curves of accuracy and number of iterations

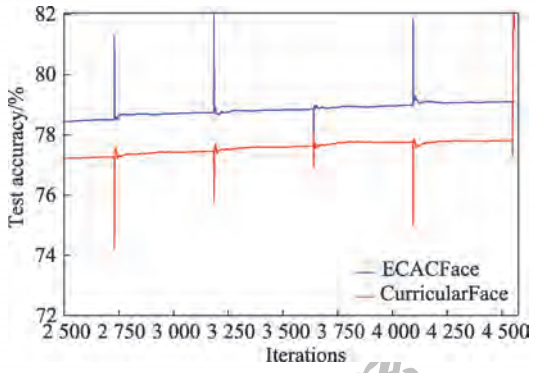

图6 2 500~4 500轮迭代的准确率和迭代次数曲线

Fig.6 Accuracy and iteration times curves of 2500 to 4500 iterations

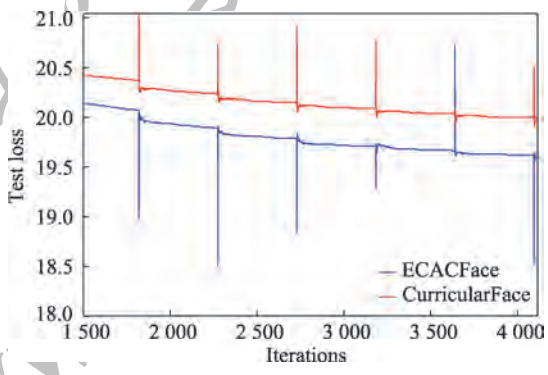

图7 1 500~4 000轮迭代的分类损失和迭代次数曲线

Fig.7 Classification loss and iteration number curves of 1500 to 4000 iterations

动。这是因为在训练中期,ECA模块将特征图像中的关键信息加强,使得大部分关键信息最后被用于分类,大部分保留关键信息的图像被正确识别,因此分类损失降低得较快。由图8可以看出,到了10 600次迭代时采用ECACFace训练的模型已经基本收敛,而采用CurricularFace训练的模型还在优化过程中,因此加入ESCA模块可以使模型收敛得更快。同时图8中ECACFace的最终分类损失较大,但能获得更好的精度,这是因为ECACFace训练不会过度拟合训练数据,对测试数据有一定的泛化能力。

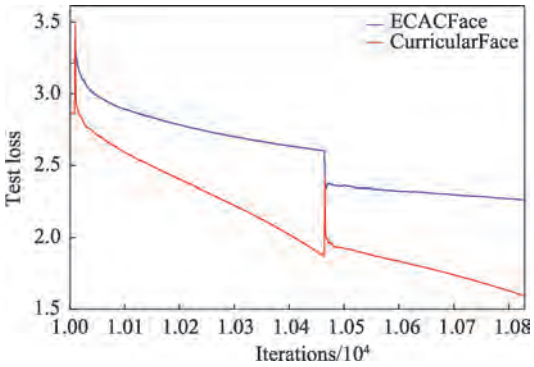

图8 10 000~10 800轮迭代的分类损失和迭代次数曲线

Fig.8 Classification loss and iteration number curves of 10000 to 10800 iterations

综上所述,ESCA模块的加入对人脸识别模型的性能提升最大,充分说明ESCA模块的有效性。

3.3.2 ECACFace性能分析

本文提出的人脸识别方法ECACFace将高效的空间通道注意力机制ESCA模块加入人脸识别模型的特征提取网络中,并且采用基于课程式学习的损失函数训练模型。为了进一步分析式(5)中参数α对损失函数性能的影响,设计一组对比实验,使用ECACFace在CASIA-WebFace数据集上训练模型,在基于课程式学习的损失函数中将α依次设置为0.80、0.85、0.90、0.99,在5个测试集上测试,结果如表4所示。平均精度为5个测试集上识别精度的平均值。由式(5)可以看出t的取值与上一个训练批次t的取值以及当前批次的正余弦相似度的平均值有关,α的取值越大,说明t的取值越依赖于当前批次正余弦相似度的平均值,由表4可以看出α的取值越大,识别效果越好,进一步说明正余弦相似度的平均值是衡量t的一个重要指标。

使用ArcFace的损失函数和基于课程式学习的损失函数进行对比实验,网络模型使用IR_101,均未使用注意力机制,在CPLFW数据集上进行测试,在测试结果中选取了两对测试图片,如图9所示。同一行表示同一身份的两张图像,这两对测试图片在ArcFace的损失函数训练的模型下验证为不同人脸,在基于课程式学习的损失函数训练的模型下验证为相同人脸。由图片可以看出,使用基于课程式学习的损失函数训练的模型对于光照和姿态变化的人脸具有一定的适应性,这是因为基于课程式学习的损失函数对训练过程中的难易样本进行有区分度的学习,并且在前期着重使用简单样本优化类中心,后期着重使用类似图9的困难样本优化类中心,使得训练出来的特征提取网络提取的人脸特征对同一身份的人脸变化具有一定的鲁棒性。

表4 不同α取值的对比实验

Table 4 Comparison experiment of value α

α	平均精度/%
0.80	87.7
0.85	90.1
0.90	95.3
0.99	**97.2**

图9 测试数据示例

Fig.9 Example of test data

为了体现本文算法对于不同网络的适用性,选择使用轻量级网络模型MobileFaceNet以及浅层网络IR_50作为基础网络进行实验。在不同网络模型的bottleneck中嵌入ESCA模块,并且使用基于课程式学习的损失函数训练,实验结果如表5、表6所示。

表5和表6中ECACFace_Mob以及ECACFace_IR50分别表示使用MobileFaceNet和IR_50作为特征提取网

表5 使用MobileFaceNet网络的识别性能对比

Table 5 Comparison of recognition performance using MobileFaceNet network 单位：%

网络模型	LFW	CFP_FP	AgeDB	CPLFW	CALFW
MobileFaceNet	98.50	89.94	89.36	83.90	91.09
ECACFace_Mob	**98.76**	**91.67**	**90.36**	**85.08**	**91.33**

表6 使用IR_50网络的识别性能对比

Table 6 Comparison of recognition performance using IR_50 network 单位：%

网络模型	LFW	CFP_FP	AgeDB	CPLFW	CALFW
IR_50	98.78	92.25	90.23	85.35	91.75
ECACFace_IR50	**99.80**	**93.51**	**91.96**	**86.90**	**92.33**

络的ECACFace。由表5可以看出，ECACFace同样适用于轻量级人脸识别网络模型，在CFP_FP和CPLFW测试集上有超过1个百分点的精度提升。MobileFaceNet使用深度可分离卷积，简化了卷积运算，但无法突出关键信息的重要性，并且使用的ArcFace损失函数不能在训练过程中区分难易样本，无法合理利用训练样本，而在ECACFace中ESCA模块通过获取图像的通道关注度和空间关注度突出人脸图像的关键信息，同时基于课程式学习的损失函数充分利用了训练样本，在训练期间划分难易样本并动态分配权值，使训练前期着重训练简单样本，后期着重训练困难样本。由表6可以看出，ECACFace同样适用于浅层网络，并且对于浅层网络效果更明显，在LFW、CFP_FP、AgeDB、CPLFW测试集上精度分别提升了1.02、1.26、1.73以及1.55个百分点，ESCA模块有效提升了浅层网络的特征提取能力，在一定程度上弥补了网络层数较少的缺点，同时基于课程式学习的损失函数合理利用了训练样本，进一步提升了人脸识别模型性能。

为了进一步验证本文方法的有效性，使用IR_101作为特征提取网络，该网络与ArcFace和CurricularFace所使用的网络结构相同，基于该网络重新训练ECACFace，训练集均使用MS1MV2，实验结果如表7所示。由表7可以看出，ECACFace相比其他算法在5个测试集上精度均有不同程度的提升，其中在CFP_FP、CPLFW和CALFW数据集上比ArcFace精度分别提升了0.14个百分点、1.14个百分点和0.88个百分点，在AgeDB数据集上比MV-Arc-Softmax精度提升了0.42个百分点，在CPLFW数据集上比CurricularFace精度提升了0.82个百分点。说明融合了注意力模块ESCA和基于课程式学习的损失函数的人脸识别算法可以进一步提升深度人脸识别模型的性能。相比CurricularFace，ECACFace将ESCA模块融入特征提取网络中，在提取特征的时候重点关注一些对人脸分类有较大作用的区域进行提取，而忽视一些背景区域，进一步提升网络的特征提取能力。同时相较于ArcFace与MV-Arc-Softmax，ECACFace使用的基于课程式学习的损失函数能够充分利用训练样本，在训练过程中划分难易样本并实时调整训练策略，在前期着重训练简单样本，后期着重训练困难样本。最近的MagFace[24]通过设计损失函数控制难易样本的特征分布，本文方法与MagFace相比在LFW、CPLFW、AgeDB、CALFW数据集上均有提升，说明比起控制难易样本在训练过程中的特征分布，控制难易样本在训练过程中的训练方式更有效。

表7 不同人脸识别算法性能对比

Table 7 Performance comparison of different face recognition algorithms 单位：%

方法	LFW	CFP_FP	CPLFW	AgeDB	CALFW
ArcFace[8]	99.70	98.27	92.08	98.15	95.45
MV-Arc-Softmax[9]	99.80	98.28	92.83	97.95	96.10
CurricularFace[10]	99.80	98.10	92.40	98.00	96.10
MagFace[24]	99.80	**98.46**	92.87	98.17	96.15
ECACFace	**99.80**	98.41	**93.22**	**98.37**	**96.33**

综上所述，融合注意力模块ESCA和基于课程式学习的损失函数的人脸识别方法ECACFace适用于不同的人脸特征提取网络，并且可以进一步提升人脸识别模型的识别性能。

4 结束语

本文所提出的ECACFace在特征提取网络中引入ESCA模块并结合课程式损失函数，通过获取图像的通道关注度以及空间关注度来强调图像的关键信息，提升特征提取网络的特征提取能力。基于课程式学习的损失函数在训练阶段充分利用训练样本，采用先易后难的思想训练模型。实验结果证明，该算法可以提升人脸识别模型的识别性能。对于注意力机制和人脸识别感兴趣的研究人员可以进一步尝试在网络的不同位置融入不同的注意力模块来探索对人脸识别性能的影响。

虽然本文算法在人脸识别方面取得了一定的优势，但ECACFace的人脸特征提取网络模型较大，对

于模型迁移的开销较大,难以用于嵌入式开发,对于如何在保持识别精度的同时降低模型复杂度方面还需要进一步研究。

参考文献:

[1] MASI I, WU Y, HASSNER T, et al. Deep face recognition: a survey[C]//Proceedings of the 31st SIBGRAPI Conference on Graphics, Patterns and Images, Parana, Oct 29-Nov 1, 2018. Piscataway: IEEE, 2018: 471-478.

[2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[3] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//Proceedings of the 3rd International Conference on Learning Representations, San Diego, May 7-9, 2015: 1-17.

[4] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, Jun 7, 2015. Washington: IEEE Computer Society, 2015: 1-9.

[5] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 26-Jul 1, 2016. Washington: IEEE Computer Society, 2016: 770-778.

[6] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 4700-4708.

[7] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition, Washington, Jun 14-19, 2020. Washington: IEEE Computer Society, 2020: 11531-11539.

[8] DENG J, GUO J, XUE N, et al. ArcFace: additive angular margin loss for deep face recognition[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 4690-4699.

[9] WANG X, ZHANG S, WANG S, et al. Mis-classified vector guided softmax loss for face recognition[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence, the 32nd Innovative Applications of Artificial Intelligence Conference, the 10th AAAI Symposium on Educational Advances in Artificial Intelligence, New York, Feb 7-12, 2020. Menlo Park: AAAI, 2020: 12241-12248.

[10] HUANG Y, WANG Y, TAI Y, et al. CurricularFace: adaptive curriculum learning loss for deep face recognition[C]//Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition, Washington, Jun 14-19, 2020. Washington: IEEE Computer Society, 2020: 5901-5910.

[11] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 7132-7141.

[12] LI X, WANG W, HU X, et al. Selective kernel networks[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Washington: IEEE Computer Society, 2019: 510-519.

[13] PARK J, WOO S, LEE J Y, et al. BAM: bottleneck attention module[C]//Proceedings of the 29th British Machine Vision Conference, Newcastle, Sep 3-6, 2018. Britain: BMVA, 2019: 1-14.

[14] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//LNCS 11211: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Cham: Springer, 2018: 3-19.

[15] LI X, HU X, YANG J. Spatial group-wise enhance: improving semantic feature learning in convolutional networks [J]. arXiv:1905.09646, 2019.

[16] 罗思诗, 李茂军, 陈满. 多尺度融合注意力机制的人脸表情识别网络[J]. 计算机工程与应用, 2023, 59(1): 199-206.
LUO S S, LI M J, CHEN M. Multi-scale integrated attention mechanism for facial expression recognition network [J]. Computer Engineering and Applications, 2023, 59(1): 199-206.

[17] 张宏鸣, 周利香, 李永恒, 等. 基于改进MobileFaceNet的羊脸识别方法[J]. 农业机械学报, 2022, 53(5): 267-274.
ZHANG H M, ZHOU L X, LI Y H, et al. Sheep face recognition method based on improved MobileFaceNet[J]. Transactions of the Chinese Society of Agricultural Machinery, 2022, 53(5): 267-274.

[18] CHEN S, LIU Y, GAO X, et al. MobileFaceNets: efficient CNNs for accurate real-time face verification on mobile devices[C]//LNCS 10996: Proceedings of the 2018 Chinese Conference on Biometric Recognition, Urumqi, Aug 11-12, 2018. Cham: Springer, 2018: 428-438.

[19] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]//Proceedings of the 2015 IEEE Conference on Computer

Vision and Pattern Recognition, Boston, Jun 7-12, 2015. Washington: IEEE Computer Society, 2015: 815-823.

[20] LIU W, WEN Y, YU Z, et al. SphereFace: deep hypersphere embedding for face recognition[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 212-220.

[21] WANG H, WANG Y, ZHOU Z, et al. CosFace: large margin cosine loss for deep face recognition[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 5265-5274.

[22] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

[23] LI B, LIU Y, WANG X. Gradient harmonized single-stage detector[C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence, the 31st Innovative Applications of Artificial Intelligence Conference, the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, Jan 27-Feb 1, 2019. Menlo Park: AAAI, 2019: 8577-8584.

[24] MENG Q, ZHAO S, HUANG Z, et al. MagFace: a universal representation for face recognition and quality assessment [C]//Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition, Jun 19-25, 2021. Washington: IEEE Computer Society, 2021: 14225-14234.

王海勇(1979—)，男，江苏连云港人，博士，副研究员，CCF 会员，主要研究方向为计算机网络安全、计算机视觉等。

WANG Haiyong, born in 1979, Ph.D., associate researcher, member of CCF. His research interests include computer network and security, computer vision, etc.

潘海涛(1998—)，男，江苏盐城人，硕士，主要研究方向为深度学习、人脸识别等。

PAN Haitao, born in 1998, M.S. His research interests include deep learning, face recognition, etc.

刘贵楠(1997—)，男，江苏盐城人，硕士，主要研究方向为深度学习、目标检测等。

LIU Guinan, born in 1997, M.S. His research interests include deep learning, object detection, etc.