

文章编号:1007-2780(2019)01-0110-08

角度空间三元组损失微调的人脸识别

任克强*, 胡 慧

(江西理工大学 信息工程学院, 江西 赣州 341000)

摘要:针对角度 Softmax 损失强约束存在的问题,提出一种用角度空间三元组损失对角度 Softmax 损失预训练模型进行微调的算法。算法首先对原来的卷积神经网络结构进行改进,将 1×1 卷积核与池化层加在不同残差块间,以进行选择更有效的特征。然后用角度空间下的三元组损失对预训练模型进行微调,以降低困难样本的强约束条件。最后在测试时,分别提取原始人脸图像特征和水平翻转的人脸图像特征,对两个特征相加作为最终的人脸特征表达,以丰富人脸特征信息,提高识别率。实验结果表明,在 LFW 和 YTF 人脸数据集分别取得了 99.25% 和 94.52% 的识别率,在大规模人脸身份识别中,本文提出的方法在仅用单模型和比较小的训练集就能有效地提高人脸识别率。

关键词:人脸识别; 卷积神经网络; 角度 Softmax 损失; 三元组损失; 特征相加

中图分类号: TP391 **文献标识码:** A **doi:** 10.3788/YJYXS20193401.0110

Face recognition of triple loss fine-tuning in angular space

REN Ke-qiang*, HU Hui

(School of Information Engineering, Jiangxi University of Science and Technology,
Ganzhou 341000, China)

Abstract: Aiming at the problem that angular Softmax loss exists in strong constraint, an algorithm to fine-tune the angular Softmax loss pre-training model by using triple loss in angle space was proposed in this paper. Firstly, the original structure of convolution neural network was improved and the convolution kernels of 1×1 and pooling layers were added between different residual blocks to select more effective features. Secondly, the pre-training model was fine-tuned by triple loss in angular space to reduce the strong constraint condition of difficult samples. Finally, in the test, the features of original and the horizontal flipped face images were extracted respectively, and the two features were added as the final facial feature expression, so as to enrich the face feature information and improve the recognition rate. The experimental results show that the recognition rates of the LFW and YTF face data sets are 99.25% and 94.52%, respectively. The method proposed in this paper can effectively improve the face recognition rate in large-scale face identification by using only a single model and a relatively small training set.

收稿日期:2018-07-18; 修订日期:2018-10-19.

基金项目:国家自然科学基金(No.61501210)

Supported by National Natural Science Foundation of China(No.61501210)

* 通信联系人, E-mail: jxrenkeqiang@163.com

Key words: face recognition; convolutional neural network; angular Softmax loss; triple loss; feature addition

1 引言

近年来,卷积神经网络的不断进步极大地提升了各种计算机视觉技术的性能,人脸识别作为最广泛的计算机视觉技术之一迅速发展,主要应用于军事,金融,电子商务,安全防务等领域^[1-2]。人脸识别早期的研究是基于手工设计的特征,这种设计比较困难,需要很强的先验知识,很难设计出有效的人脸特征,随着深度卷积神经网络的出现,深度卷积神经网络有着天然的自动提取特征的能力,已成为人脸识别的主流方向^[3]。

目前有关深度卷积神经网络人脸识别算法的研究大多围绕“基于损失函数的人脸识别”展开,实现方法大致可以分为基于分类损失函数、基于度量学习损失函数或两者的结合。用于人脸识别的分类损失函数大多是基于 Softmax 损失函数。文献[4]使用 3D 人脸模型对齐,不共享卷积核的神经网络设计,大规模的训练数据来表达人脸特征。文献[5]提取人脸的不同区域,形成多个网络结构进行融合,每个网络由 10 000 个不同类身份的 Softmax 损失函数作为分类器。文献[6]使用卷积、池化、多层稀疏自动编码器构建网络特征提取器,用 Softmax 回归模型对提取的特征进行分类。然而,Softmax 分类损失函数只优化类内变化,而忽略了类间变化,不具有判别性。度量学习损失函数是使同一身份的人脸特征相互靠近,不同身份的人脸特征相互远离。文献[7]提出了对比损失(Contrastive Loss)函数,产生一种有效的人脸特征表达来使不同类特征相互分离,相同类特征变得紧凑。文献[8]提出了三元组损失(Triple Loss)函数,学习一种从人脸图像到欧式空间的映射,直接对应人脸相似度的衡量。然而,度量学习损失函数不能约束每一个样本,因此要求精心设计的样本对或三元组挖掘过程,而这两者非常耗时和对性能比较敏感。文献[9]和文献[10]分别提出了中心损失(Center Loss)函数和

对比损失函数与 Softmax 损失函数结合来增强特征的判别能力,但 Softmax 损失和度量损失结合需设计多个损失函数,算法过于复杂。文献[11]提出角度 Softmax 损失函数,仅在 Softmax 损失函数上使用一个角度间距,迫使每一类的分类边界更靠近每一类的权重向量,使类间最小距离大于类内最大距离,这种强约束使同类进一步靠近,不同类进一步远离,进一步提高识别率,然而一些困难的样本很难满足这种条件。

针对困难样本难以满足角度 Softmax 损失函数的强约束条件,本文提出一种角度空间的三元组损失微调的人脸识别算法,并对卷积神经网络结构进行改进和使用特征相加作为人脸表达,以提高算法的识别率。

2 相关工作

2.1 网络模型

人工神经网络用于人脸识别取得了较好的性能^[12],但面对大规模人脸数据集,比较流行的人脸识别方法普遍采用卷积神经网络结构。文献[13]采用了 VGGNet 网络结构用于人脸识别,VGGNet 是由牛津大学的视觉几何组提出的一种卷积神经网络,是 ILSVRC-2014 中定位任务第一名和分类任务第二名,该网络比 AlexNet 网络具有更深层的网络结构,在该网络中,所有卷积层都使用了大小为 3×3 的卷积核,而且 VGGNet 对其他数据集有很好的泛化能力。但随着网络的加深,梯度弥散或梯度爆炸成为训练深层网络的障碍,导致无法收敛,有一些方法可以弥补,如归一初始化,各层输入归一化。然而,虽然收敛了,但网络开始退化,即增加网络层数导致更大的误差,Resnet 通过在一个浅层网络上增加残差块,可以让网络随深度增加而不退化,使得深层网络有较好的收敛能力和防止深层网络出现过拟合,因此在 VGGNet 基础上添加了多个残差块结构,构建人脸残差网络用于人脸识别,网络结构具体如表 1 所示。

表 1 人脸残差网络结构
Tab.1 Face residual network structure

层类型	Conv1.x	Conv2.x	Conv3.x	Conv4.x	FC
层结构	$[3 \times 3, 64] \times 1, S2$ $\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 1$	$[3 \times 3, 128] \times 1, S2$ $\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$[3 \times 3, 256] \times 1, S2$ $\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 4$	$[3 \times 3, 512] \times 1, S2$ $\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 1$	—
输出大小	56×48	28×24	14×12	7×6	512

在表 1 中, Conv1.x, Conv2.x, Conv3.x, Conv4.x 表示卷积单元, 里面含有多个卷积层和残差单元。例如: $[3 \times 3, 64] \times 1, S2$ 表示一个级联的卷积层, 其中有 64 个大小为 3×3 的滤波器, S2 表示步长为 2, $\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 1$, 表示一个残差单元, 为了简洁, 每个卷积之后省略了参数修正线性单元, 输出大小是不同尺寸的特征图, FC 是全连接层, 输出大小为 512 维的人脸特征。

2.2 角度 Softmax 损失函数算法

卷积神经网络广泛使用的分类损失函数 Softmax 表示如下:

$$L_1 = -\frac{1}{m} \sum_{i=1}^m \log \frac{\exp(W_{y_i}^T x_i + b_{y_i})}{\sum_{j=1}^n \exp(W_j^T x_i + b_j)} = -\frac{1}{m} \sum_{i=1}^m \log \frac{\exp(\|W_{y_i}\| \|x_i\| \cos(\theta_{y_i}) + b_{y_i})}{\sum_{j=1}^n \exp(\|W_j\| \|x_i\| \cos(\theta_j) + b_j)}, \quad (1)$$

式中: $x_i \in R^d$ 表示第 i 个样本的特征向量, 属于第 y_i 类, d 表示特征向量的维数, 在最后的全连接层中, $w_j \in R^d$ 代表权重矩阵 $W \in R^{d \times n}$ 的第 j 列, $b \in R^n$ 代表偏置项, m 表示批量大小, n 表示类别数量, θ_j 表示第 j 类权重向量与第 j 个样本的特征向量之间的角度。

在角度 Softmax 损失中, 为了简便, 首先把偏置项设置为 0, 然后归一化权重向量 ($\|W\| = 1$), 使得预测只取决于特征向量和权重向量之间的角度, $\cos(\theta_{y_i})$ 推广为 $\phi(\theta_{y_i})$, 得到式 (2) 形式。

$$L_2 = -\frac{1}{m} \sum_{i=1}^m \log \frac{\exp(\|x_i\| \phi(\theta_{y_i}))}{\exp(\|x_i\| \phi(\theta_{y_i})) + \sum_{j=1, j \neq y_i}^n \exp(\|x_i\| \cos(\theta_j))}, \quad (2)$$

式中:

$$\phi(\theta_{y_i}) = \frac{(-1)^k \cos(m_1 \theta_{y_i}) - 2k + \lambda \cos(\theta_{y_i})}{1 + \lambda},$$

$$\theta_{y_i} \in \left(\frac{k\pi}{m_1}, \frac{(k+1)\pi}{m_1} \right), k \in [0, m_1 - 1], m_1 \geq 1$$

1 为整数, 控制角度间距的大小, λ 为平衡因子。

为了满足最大类内距离小于最小类间距离, 得出 $m_1 = 4$ 。

2.3 人脸的特征表达

在训练时, 对训练集中的每一张人脸样本进行翻转以增强样本数据集, 提高人脸识别算法的鲁棒性和防止过拟合。在测试阶段, 为了与训练阶段增加的人脸样本对应, 使最终提取到的人脸特征更具有表达能力, 提取原始人脸的特征和水平翻转的特征进行级联, 作为最终的人脸特征表达。

3 本文算法

角度 Softmax 损失函数相对于分类损失函数和度量学习损失函数而言, 复杂性降低且其性能有较大提高, 但由于施加的角度间距, 使每一类的特征更靠近该类的特征向量, 强制压缩同一类人脸样本的特征空间变小, 导致对一些困难样本很难满足这种强约束条件。针对这些问题, 本文提出了角度空间下三元组损失微调的人脸识别算法。首先对人脸残差网络进行改进, 利用改进的网络进行训练, 得到一个预训练模型, 然后利用角度空间下的三元组损失微调预训练模型, 最后, 通过特征相加作为最终的人脸表达。

3.1 人脸残差网络结构的改进

角度 Softmax 损失中所采用的人脸残差网

络,逐层提取特征的过程中没有对特征加以筛选,产生一些影响特征判别的信息和信息的冗余。卷积神经网络里的卷积都是多通道的特征图和多通道的卷积核之间的操作,如果使用 1×1 的卷积核,这个操作实现的是多个特征图的线性组合,可以实现特征图在通道上的变化。并且 1×1 的卷

积核也能潜在地在卷积层间做特征选择,因而在残差块间加入了大小为 1×1 ,步长为 1 的卷积核 ($1 \times 1/1$)。为了增强模型的鲁棒性,减少参数和防止一定的过拟合能力,加入了大小为 2×2 ,步长为 2 的池化层 ($2 \times 2/2$),改进的人脸残差网络如表 2 所示。

表 2 网络改进模型

Tab.2 Improved face residual network model

层类型	Conv1.x	Conv2.x	Conv3.x	Conv4.x	FC
层结构	$[3 \times 3, 64] \times 1, S2$ $2 \times 2/2$ $\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 1$	$1 \times 1/1$ $[3 \times 3, 128] \times 1, S2$ $2 \times 2/2$ $\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$1 \times 1/1$ $[3 \times 3, 256] \times 1, S2$ $2 \times 2/2$ $\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 4$	$1 \times 1/1$ $[3 \times 3, 512] \times 1, S2$ $2 \times 2/2$ $\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 1$	—
输出大小	55×47	27×23	13×11	6×5	512

3.2 角度空间中的三元组损失微调模型

在 2.2 节,从角度 Softmax 损失中,得出角度间距 $m_1 = 4$,在没加 m_1 之前,也就是式(1),假设在一个两分类情况下,能够得出两类的决策边界为 $\|x\| \cos \theta_1 - \|x\| \cos \theta_2 = 0$ ($\theta_1, \theta_2 \in [0, \pi]$),其中 $\|x\| \neq 0$,即 $\theta_1 = \theta_2$,当 $m_1 = 4$ 时,对于第一类的决策边界为 $\cos(4\theta_1) = \cos(\theta_2)$,即 $\theta_2 = 4\theta_1$,同理,第二类决策边界为 $\cos(4\theta_2) = \cos(\theta_1)$,即 $\theta_1 = 4\theta_2$,相比于 $\theta_1 = \theta_2$ 这种约束变得更强了,一些困难的样本很难满足这种条件,而三元组损失是对比较困难的样本进行优化,具体的优化过程如图 1 所示。

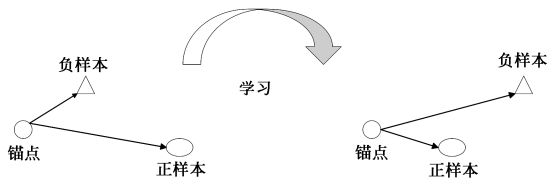


图 1 欧式空间下的三元组损失优化过程

Fig. 1 Process of triple loss optimization in European space

锚点和正样本属于同一身份,三元组损失最小化它们之间的距离,而锚点和负样本属于不同身份,三元组损失最大化它们之间的距离,三元组

损失表示如下:

$$\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+, \quad (3)$$

式中: x_i^a 表示锚点, x_i^p 表示正样本, x_i^n 表示负样本, α 表示正样本和负样本对间的间距。

式(3)是在欧式空间下的损失,而角度 Softmax 损失是在角度空间下的,这就造成了不同度量空间下的优化的不一致,这样对样本间的相似度衡量造成了干扰,为了与角度的 Softmax 损失相匹配,把欧式空间中的三元组损失转变为角度空间下的三元组损失,三元组损失在二维约束角度空间下的优化过程如图 2 所示。

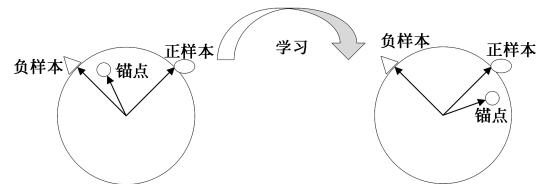


图 2 角度空间下的三元组损失优化过程

Fig. 2 Process of triple loss optimization in angular space

在基于角度 Softmax 损失的角度空间中,在两分类情况下,每一类的权重向量均进行归一化 ($\|W\| = 1$),假设正负样本代表这两类的权重

向量,故图 2 中的正样本和负样本都在单位圆上,而锚点在原来的角度空间没有进行归一化,因此这里也没有进行归一化,角度三元组损失微调如式(4)所示。

$$L = -\frac{1}{m} \sum_{i=1}^m \log \frac{\exp(\|x_i\| \cos(\theta_{y_i}))}{\sum_{j=1}^n \exp(\|x_i\| \cos(\theta_j))} \quad (4)$$

在微调的过程中,正负样本代表这两类的权重不更新,即:

$$\frac{\partial L}{\partial w_j} = 0, \quad (5)$$

式中: w_j 表示每一类的权重。

在角度空间中,用角度度量样本间距离的大小,由于三元组损失中的困难三元组的选择比较麻烦,从头开始训练比较难收敛,在这里只选择了较容易的样本,在角度空间实现了一个相对简单的三元组损失进行微调,该二维情况下很容易推广到多维超球面上。

3.3 人脸特征提取

特征提取是人脸表达的重要方式,提取到的特征代表人脸信息,不同身份的人脸特征应该有不同信息,而相同身份的人脸特征应该有相同的信息。在测试阶段,角度 Softmax 损失算法分别提取原始人脸图像的特征和水平翻转的特征,然后进行级联作为最终的人脸表达。每次提取一张人脸图像的特征是 512 维,进行级联之后得到 1 024 维的图像。在进行人脸特征比对的过程中,如此高的特征维数虽然在一定程度上丰富了特征信息,提高了特征表达的鲁棒性,但是面对如此大规模的人脸图像,特征提取和特征比对计算是一个非常耗时的过程。为了保持提取到的人脸特征信息丰富和鲁棒性不变的特点,用角度空间三元组损失微调改进的人脸残差网络模型进行特

征提取,分别提取原始的人脸图像特征和水平翻转的人脸图像特征,然后进行特征相加作为最终的人脸表达,具体的人脸特征提取过程如图 3 所示。

4 实验结果与分析

传统人脸识别的测试都是在同一个数据集上进行训练和测试,而本文实验采用公开人脸识别训练集 CASIA-WebFace 进行训练, LFW 和 YTF 作为测试集进行实验,得出的识别率更有说服力。本文使用特征向量的余弦距离作为相似性得分来评估本文算法的性能。实验中先对人脸数据集进行归一化处理并设置网络的训练参数,然后在 LFW 和 YTF 上评估相关算法以及改进后的人脸残差网络的性能。

4.1 实验数据预处理和网络参数设置

本文首先使用 MTCNN 对整个人脸训练集(CASIA-WebFace)和测试集(LFW, YTF)进行人脸和面部关键点检测,然后,采用相似性变换对面部的五个点(左右眼睛,鼻子,和左右嘴角)进行对齐。最后,将获得的人脸图片进行裁剪,裁剪的大小为 112×96 。在 RGB 图像中,每个像素 $[0, 255]$ 通过减去 127.5 然后除以 128 进行归一化。

为了降低硬件资源的消耗和节省训练的时间,网络结构是在角度 Softmax 损失的 20 层网络结构上改进后的人脸残差网络,并且使用了一个小的训练集(CASIA-WebFac)来训练模型。CASIA-WebFace 数据集包括 10 575 个不同身份的人,49 万张不同的人脸图像,为了增强数据集,对所有的人脸训练图像进行水平翻转,为保证本文算法测试的有效性,在实验中,移除了训练集中与测试集有相同身份的人脸图像。本文使用深度学习框架 caffe 来实现修改的残差网络和角度空间下的三元组损失并进行训练,使用随机梯度下降法进行优化。权重衰减设置为 0.0005,动量设置为 0.9,学习率最初设置为 0.1,当训练迭代次数达到 16 000, 24 000, 28 000 时,分别设为 0.01, 0.001, 0.000 1,最大迭代次数为 28 000 已基本收敛,完成训练。

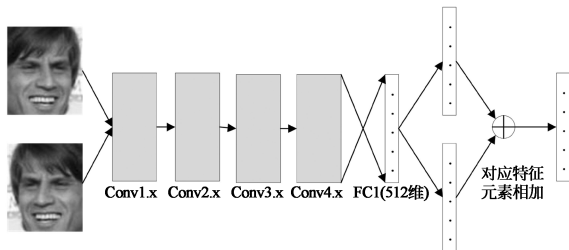


图 3 人脸特征提取过程

Fig.3 Process of face feature extraction

4.2 人脸数据集 LFW 上评估

4.2.1 在 LFW 标准协议上评估

LFW 数据集包含来自 5 749 个不同身份, 13 233 张人脸图像。该数据集在姿势, 表情, 光照等方面有较大的变化。在无约束带有标签的外部数据的标准协议上评估, LFW 是一个标准的人脸验证测试集, 包含 6 000 个人脸对, 人脸图像被划分为十折, 每一折包含不同的身份和 600 个人脸对。表 3 是在 LFW 标准协议上的人脸识别算法性能比较。

表 3 标准协议的人脸识别率

Tab.3 Face recognition rate of standard protocol

方法	模型	数据	识别率 / %
DeepFace ^[4]	3	4 M	97.35
FaceNet ^[8]	1	200 M	99.65
DeepFR ^[13]	1	2.6 M	98.95
DeepID2+ ^[14]	25	300 K	99.47
CenterFace ^[9]	1	WebFace	99.05
DeepID2 ^[10]	1	WebFace	98.78
A-Softmax ^[12]	1	WebFace	99.10
本文算法	1	WebFace	99.25

由表 3 可知, 本文算法的识别率比仅用单模型和 WebFace 较小的训练集上的算法要高, 这是由于本文使用了角度 Softmax 损失, 压缩了每一类的特征空间, 使得相同类更加靠近, 不同类更加远离。比角度 Softmax 算法 (A-Softmax) 有所提高, 是由于本文使用了角度空间下的三元组损失对角度 Softmax 进行微调。甚至比 DeepFace, DeepFR 算法使用的大型数据集和多模型融合都有较大的提升。比 FaceNet, DeepID2+ 算法识别率要低, 是由于 FaceNet 使用了比较困难的样本挖掘算法, DeepID2+ 使用了众多模型的融合。

4.2.2 在 BLUFR 协议上评估

BLUFR 是一个大规模无约束的人脸识别基准, 完全利用 13 233 个 LFW 人脸图像作为新的评估协议, 该协议包含人脸验证和开集人脸识别。在 BLUFR 协议上有十折实验, 每折实验包含 156 915 个真实匹配和 46 960 863 个错误匹配用于

性能评估。对于 LFW 数据集来说, BLUFR 协议更具有挑战性和普遍性, 表 4 是 LFW BLUFR 上的人脸识别算法性能比较。

表 4 BLUFR 协议的人脸识别率

Tab.4 Face recognition rate of BLUFR protocol

方法	VR@FAR=0.1%/ %	DIR@FAR=1%/ %
LightCNN ^[15]	98.71	90.42
CenterFace	93.35	67.86
NormFace ^[16]	95.83	77.18
A-Softmax	96.02	78.58
本文算法	97.28	80.03

在表 4 中 VR@FAR=0.1% 表示误识率为 0.1% 情况下的验证率, DIR@FAR=1% 表示误识率为 1% 的情况下的识别率。由表 4 可知, 在较低的误识率情况下, 本文算法比其他大部分算法的人脸验证率和识别率要高, 但是比 LightCNN 要低, 这是由于 LightCNN 使用数量庞大的训练集, 并且使用了 29 层的残差网络, 而本文只使用了改进的 20 层残差网络。整体验证率和识别率比 LFW 标准协议识别率要低, 这是由于 BLUFR 协议是大规模人脸身份的识别, 利用了 LFW 所有的人脸图像, 且在较低的误识率情况下, 识别更加困难和更有挑战性。

4.3 人脸数据集 YTF 上评估

YTF 是为研究视频中无约束人脸识别问题而设计的一种人脸视频数据集, 该数据集包含来自 1 595 个不同人的 3 425 个视频, 短片时长最短为 48 帧, 最长为 6 070 帧, 平均长度为 181.3 帧。在 YTF 上使用 5 000 个视频对进行测试, 表 5 是 YTF 数据集上的人脸识别算法性能比较。

由表 5 可知, 对于视频中的人脸识别, 在仅用单模型和比较小的训练集数据集上, 就能在数据集 YTF 上取得 94.52% 的识别率, 比同样大小数据集和单模型的算法要好, 甚至比一些大型数据集和多模型融合的相关算法都有较大的提升。但整体识别率比 LFW 数据集上要低, 这是由于 YTF 数据集上的人脸样本比 LFW 上的人脸样本更加模糊, 变化更大, 识别更加困难。

表 5 YTF 数据集上的人脸识别率

Tab.5 Face recognition rate on YTF dataset

方法	模型	数据	识别率/%
DeepFace	3	4 M	91.4
FaceNet	1	200 M	95.1
DeepFR	1	2.6 M	97.3
DeepID2+	25	300 K	93.2
CenterFace	1	WebFace	94.4
DeepID2	1	WebFace	93.5
A-Softmax	1	WebFace	94.26
本文算法	1	WebFace	94.52

4.4 网络结构性能的评估

不同的网络结构所提取到的特征也是不同的,而不同的特征具有不同的表达能力。本文也对网络结构进行研究,对原来的人脸残差网络结构进行改进,改进的人脸残差网络结构和原来的人脸残差网络结构在 LFW 和 YTF 人脸数据集上进行实验,网络结构性能如表 6 所示。

由表 6 可知,在各类人脸数据集上,本文的网络结构性能比原人脸残差网络结构性能略好,这是由于本文在不同的残差块间加入了 1×1 的卷积核与池化层,选择更有效的特征。

表 6 不同网络结构性能比较

Tab.6 Performance comparison of different network structures

(%)

网络结构	LFW	YTF	BLUFR	BLUFR
			VR@FAR=0.1%	DIR@FAR=1%
人脸残差网络	99.10	94.2	95.96	77.50
改进的人脸残差网络	99.18	94.4	96.68	78.98

5 结 论

角度 Softmax 损失强制使得相同类的特征空间更小,导致比较困难的人脸样本很难满足这种强约束条件。为了减轻这种强约束,但又不降低识别率,本文提出了一种角度空间中三元组损失微调的人脸识别算法。在原来的人脸残差网络结构不同卷积单元的残差块间加入了 1×1 卷积

核和池化层,增强了鲁棒性和捕获具有极强表达能力的深层特征,提高了识别率,并用角度空间下的三元组损失微调,降低了困难样本的强约束性。在测试阶段,分别把原始的人脸图像和水平翻转的人脸图像输送到微调网络中提取相应的特征,再把对应的特征相加,进一步提高了识别率。实验结果表明,本文算法在公开数据集 LFW 和 YTF 上取得了 99.25% 和 94.52% 的识别率,证明了本文算法在大规模人脸身份识别的有效性。

参 考 文 献:

- [1] 卢洋,王世刚,赵文婷,等.基于人脸姿态估计的虚拟眼镜试戴技术[J].中国光学,2015,8(4):582-588.
LUY, WANG S G, ZHAO W T, et al. Technology of virtual eyeglasses try-on system based on face pose estimation [J]. *Chinese Optics*, 2015, 8(4): 582-588. (in Chinese)
- [2] 郑明秋,杨帆.改进非负矩阵分解的神经网络人脸识别[J].液晶与显示,2017,32(3):213-218.
ZHENG M Q, YANG F. Face recognition based on mproved NME and neurd network [J]. *Chinese Journal of Liquid Crystals and Displays*, 2017, 32(3): 213-218. (in Chinese)
- [3] 严严,陈日伟,王茜子.基于深度学习的人脸分析研究进展[J].厦门大学学报(自然科学版),2017,56(1):13-24.
YAN Y, CHEN R W, WANG H Z. Resent advcmces on deep-leaming based face analysis [J]. *Journal of Xiamen University (Natural Science)*, 2017, 56(1): 13-24. (in Chinese)
- [4] TAIGMAN Y, YANG M, RANZATO M, et al. Deepface: closing the gap to human-level performance in face ver-

- ification [C] // *IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA; IEEE, 2014: 1701-1708.
- [5] SUN Y, WANG X, TANG X. Deep learning face representation from predicting 10,000 classes [C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA; IEEE Computer Society, 2014: 1891-1898.
- [6] 李倩玉, 蒋建国, 齐美彬. 基于改进深层网络的人脸识别算法[J]. 电子学报, 2017, 45(3): 619-625.
LI Q Y, JIANG J G, QI M B. Face recognition algorithm based on improved deep network [J]. *Acta Electronica Sinica*, 2017, 45(3): 619-625. (in Chinese)
- [7] CHOPRA S, HADSELL R, LECUN Y. Learning a similarity metric discriminatively, with application to face verification [C]// *IEEE Conference on Computer Vision and Pattern Recognition*. New York, USA; IEEE, 2005: 539-546.
- [8] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: a unified embedding for face recognition and clustering [C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA; IEEE, 2015: 815-823.
- [9] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition [C]//*Euro-pean Conference on Computer Vision*. Amsterdam: Springer International Publishing, 2016: 499-515.
- [10] SUN Y, WANG X, TANG X, et al. Deep learning face representation by joint identification-verification [C]//*Conference and Workshop on Neural Information Processing Systems*. Montreal, Canada; MIT Press, 2014: 1988-1996.
- [11] LIU W, WEN Y, YU Z, et al. Sphereface: deep hypersphere embedding for face recognition [C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA; IEEE, 2017: 6738-6746.
- [12] 沈凌云, 郎百和, 朱明. 一种基于人工神经网络的人脸识别方法[J]. 液晶与显示, 2011, 26(6): 836-840.
SHEN L Y, LANG B H, ZHU M. Face recognition method based on artificial neural network [J]. *Chinese Journal of Liquid Crystals and Displays*, 2011, 26(6): 836-840. (in Chinese)
- [13] PARKHI O M, VEDALDI A, ZISSERMAN A. Deep face recognition [C]// *British Machine Vision Conference*. Swansea, UK; British Machine Association, 2015: 41-42.
- [14] SUN Y, WANG X, TANG X. Deeply learned face representations are sparse, selective, and robust [C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA; IEEE, 2015: 2892-2900.
- [15] WU X, HE R, SUN Z, et al. A light CNN for deep face representation with noisy labels [J]. *IEEE Transactions on Information Forensics and Security*, 2018, 13(11): 2884-2896.
- [16] WANG F, XIANG X, CHENG J, et al. Normface: L2 hypersphere embedding for face verification [C]//*Proceedings of the 2017 ACM on Multimedia Conference*. Mountain View, USA; ACM New York, 2017: 1041-1049.

作者简介:



任克强(1959—),男,湖南汨罗人,硕士,教授,1993年于华中理工大学获得硕士学位,主要从事图像与视频处理、信息隐藏方面的研究。E-mail:jxrenkeqiang@163.com