

# 一种基于角度距离损失函数和卷积神经网络的人脸识别算法

龙鑫, 苏寒松, 刘高华\*, 陈震宇

天津大学电气自动化与信息工程学院, 天津 300072

**摘要** 在开集协议下设计了一种基于角度距离损失函数和密集连接卷积神经网络的人脸识别算法, 以实现深度人脸识别。所设计的网络结构使用基于角度距离的损失函数, 让人脸特征的分度度更高, 符合特征的理想分类标准。同时, 所提出的神经网络结构采用先进的密集连接模块, 在很大程度上减少了传统网络结构的参数冗余。经过大量的分析和实验, 该算法在 LFW 数据集上的人脸识别准确率达到 99.45%, 在 MegaFace 数据集上的人脸确认任务和人脸验证任务中的人脸识别准确率分别为 72.534% 和 85.348%, 因此所提算法在人脸识别任务中具有较高的优越性。

**关键词** 机器视觉; 人脸识别; 卷积神经网络; 深度学习; 角度损失函数; 密集连接

**中图分类号** TP391.4 **文献标识码** A

**doi:** 10.3788/LOP55.121505

## A Face Recognition Algorithm Based on Angular Distance Loss Function and Convolutional Neural Network

Long Xin, Su Hansong, Liu Gaohua\*, Chen Zhenyu

*School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China*

**Abstract** A face recognition algorithm based on the angular distance loss function and densely connected convolutional neural network is proposed under the open-set protocol to achieve deep face recognition. The loss function based on angle distance is adopted in the proposed network structure, which makes the facial features more distinguishable and meets the ideal criteria of feature classification. At the same time, the advanced dense connection module is adopted in the proposed neural network structure, which greatly reduces the parameter redundancy of the traditional network structure. After extensive analysis and repeated experiments, the face recognition accuracy reaches 99.45% on the LFW dataset, and the recognition accuracy rates of face identification task and face verification task on MegaFace dataset are 72.534% and 85.34%, respectively. The superiority of the proposed algorithm is confirmed in the face recognition domain.

**Key words** machine vision; face recognition; convolutional neural network; deep learning; angular loss function; dense connection

**OCIS codes** 150.1135; 100.2000; 100.3008

## 1 引言

近几年,卷积神经网络(CNN)在人脸识别等领域受到越来越多关注,这主要是因为新网络结构<sup>[1-4]</sup>和各种强区分度损失函数<sup>[5-7]</sup>的不断优化,人脸识别技术达到了较高水平。

目前,人脸识别算法的改进主要集中在人脸预

处理(包括人脸检测和人脸对齐)、人脸特征提取(主要指设计人工神经网络结构)和人脸特征分类损失函数的设计。人脸识别任务一般分成人脸确认和人脸验证<sup>[8-9]</sup>两大类,前者将人脸分类到特定身份,后者则确定两张人脸图片对是否属于同一个身份。

文献[10]和[11]的算法使用传统的 softmax 损失函数学习人脸特征,但 softmax 损失函数只能学

收稿日期: 2018-05-25; 修回日期: 2018-06-21; 录用日期: 2018-07-12

基金项目: 国家自然科学基金(61471260)

\* E-mail: suppig@126.com

习可分的特征,学习到的人脸特征区分度不高,人脸识别效果较差。为了解决这个问题,一些结合 softmax 损失函数和 contrastive 损失函数<sup>[6,12]</sup>或者结合 center 损失函数<sup>[7]</sup>的算法被相继提出,用于增强特征的区分度。文献[5]的算法采用 triplet 损失函数来监督人脸特征的学习,获得了目前较好的人脸识别效果。文献[7]提出的 center 损失函数缩小了特征的类内距离,但并没有增加特征的类间距离。contrastive 损失函数<sup>[13]</sup>和 triplet 损失函数<sup>[5]</sup>的缺点是不能约束独立的人脸样本,这两个算法都需要提前进行人脸对或人脸三元组的提取,这个预处理步骤既耗时又让算法性能变得不稳定,算法性能依赖于人脸预处理效果的好坏。

目前,大部分基于深度学习的算法都选择欧氏距离来进行特征学习,但欧氏距离不适用于学习区分度较强的人脸特征。文献[6]、[7]、[12]提出的算法结合了基于欧氏距离的损失函数和 softmax 损失函数,构造了一种联合监督函数。但在一些场景下,基于欧氏距离的损失函数和 softmax 损失函数相互排斥。本文在传统 softmax 损失函数的基础上加入了角度距离的思想,进而提出一种基于角度距离的损失函数,可以监督 CNN 学习强区分度的人脸特征。

CNN 是图像识别领域最主要的机器学习方法。由于计算机硬件和神经网络结构的限制,CNN 直到最近几年才开始广泛应用于图像识别领域。最初的 LeNet5<sup>[14]</sup>仅由 5 层神经网络组成,VGG<sup>[15]</sup>模型也只有 19 层,但从 2016 年开始,高速公路网络<sup>[4]</sup>和残差网络(ResNet)<sup>[1]</sup>被相继提出,这两个网络的深度均突破 100 层的限制,ResNet 甚至还开始了对 1000 层网络的探索。真正的深度学习神经网络时代已经到来。

但随着神经网络深度的不断增加,各种问题也开始浮出水面,输入信号和梯度在经过多个中间层的传递后,信号强度变得越来越弱,甚至消失。为了解决这个问题,深度残差网络<sup>[1]</sup>和高速公路网络<sup>[4]</sup>通过恒等连接方式将信号从一层直接传递到下一层。随机深度网络<sup>[16]</sup>在模型训练期间,随机丢掉某些层来简化残差网络的结构,以减少网络参数,避免过拟合。虽然这些方法的网络结构和训练过程都有所不同,但它们都有一个共同的特征,即都建立了从前面层到后面层的短连接路径。

本文提出的 CNN 结构采用密集连接的方式,加入了密集连接模块,并在公开数据集 CASIA<sup>[17]</sup>

上训练的人脸特征,在两个测试数据集(LFW、MegaFace)上取得了很好的效果。为了确保网络中各层之间信息传递的最大化,直接让所有层相互连接,并加入了层与层之间的较短连接,以前馈的方式让每一层都从前面的所有层获得附加输入,并将本层的信息传递到后续所有层,减少了传统 CNN 模型的参数,提高了识别精度,虽然网络有较深的深度,但并不会出现梯度消失和过拟合的现象。

## 2 基本原理

### 2.1 开集人脸识别

按照人脸识别的测试协议,人脸识别可以分为闭集人脸识别和开集人脸识别<sup>[18]</sup>,这两种测试协议的具体实现流程如图 1 所示。在闭集人脸识别中,所有的测试身份都在训练集中提前定义,这类问题就是把待测试的人脸图片分类到特定的身份类别中[图 1(a)],可将其看作一个简单的多分类问题。在开集人脸识别中,测试集中的身份和训练集中的身份相互排斥,这种情况更接近于实际应用场景,同时挑战的难度也更大,这就需要把人脸映射到一个区分度更高的特征空间,人脸确认就是人脸图像和身份候选集人脸图像之间的特征识别[图 1(b)]。因此,开集人脸识别任务本质上是特征间距离学习,关键在于学习强区分度的特征。

在理想状态下,开集人脸识别的特征应该符合最大类内特征距离小于最小类间特征距离这个理想特征分类标准,如图 1(b)的等价任务所示。然而,由于人脸类内特征间距普遍较大、人脸类间特征相似度又较高,人脸特征很难符合理想特征分类标准。

### 2.2 传统的损失函数

在传统的距离损失函数中,距离可以通过学习矩阵  $\mathbf{A}$  表示为<sup>[19-22]</sup>  $\|\mathbf{x}_1 - \mathbf{x}_2\|_{\mathbf{A}} = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^{\top} \mathbf{A} (\mathbf{x}_1 - \mathbf{x}_2)}$ ,其中  $\mathbf{x}_1, \mathbf{x}_2$  分别为两个特征向量。目前,大多数深度学习的距离损失函数<sup>[5-7,11,23-26]</sup>都采用欧氏距离作为距离度量标准,且性能良好的人脸识别算法几乎都对传统损失函数进行改进,使学习到的特征有更高的区分度,提升开集人脸识别任务的准确率。在二分类问题中,由 softmax 损失函数得到的两个类别的后验概率分别为

$$p_1 = \frac{\exp(\mathbf{W}_1^{\top} \mathbf{x} + b_1)}{\exp(\mathbf{W}_1^{\top} \mathbf{x} + b_1) + \exp(\mathbf{W}_2^{\top} \mathbf{x} + b_2)}, \quad (1)$$

$$p_2 = \frac{\exp(\mathbf{W}_2^{\top} \mathbf{x} + b_2)}{\exp(\mathbf{W}_1^{\top} \mathbf{x} + b_1) + \exp(\mathbf{W}_2^{\top} \mathbf{x} + b_2)}, \quad (2)$$

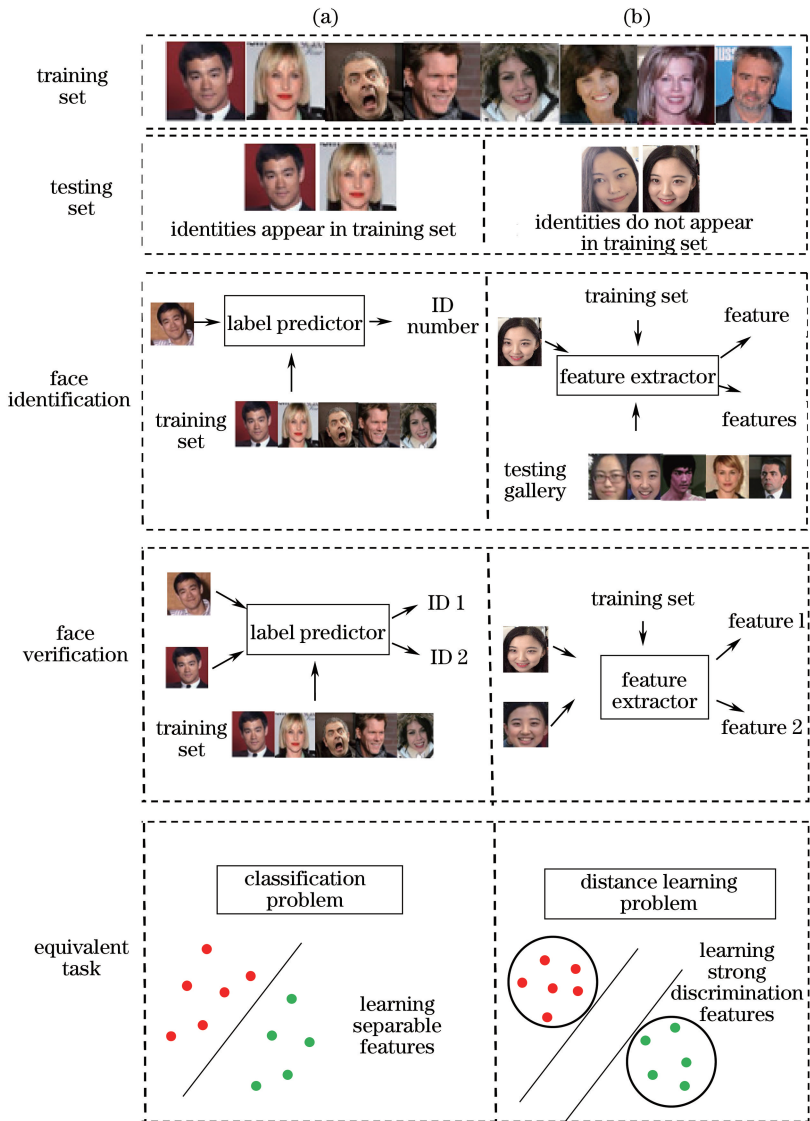


图1 人脸识别测试协议的比较。(a)闭集人脸识别;(b)开集人脸识别

Fig. 1 Comparison of test protocol of face recognition. (a) Closed-set face recognition; (b) open-set face recognition

式中:向量  $\mathbf{x}$  为学习到的特征向量;  $\mathbf{W}_i$  和  $b_i$  ( $i=1, 2$ ) 分别为第  $i$  个类别全连接层输出的权重向量和偏置。若  $p_1 > p_2$ , 则把预测的标签分到类别 1 中; 反之, 则把预测的标签分到类别 2 中。通过比较  $p_1$  和  $p_2$  的大小, 可以得到在二分类问题中, 分类结果取决于  $\mathbf{W}_1^T \mathbf{x} + b_1$  和  $\mathbf{W}_2^T \mathbf{x} + b_2$  的大小。  $\mathbf{W}_i^T \mathbf{x} + b_i$  也可以改写为  $\|\mathbf{W}_i^T\| \|\mathbf{x}\| \cos \theta_i + b_i$ , 其中  $\theta_i$  是权重向量  $\mathbf{W}_i$  和特征向量  $\mathbf{x}$  之间的夹角。如果令  $\|\mathbf{W}_i\| = 1$  且  $b_i = 0$ , 即权重向量的模值置 1, 偏置归零, 也就是对权重和偏置进行标准化处理。经过标准化处理后, (1) 式和 (2) 式中的后验概率分别变为

$$p_1 = \frac{\exp(\|\mathbf{x}\| \cos \theta_1)}{\exp(\|\mathbf{x}\| \cos \theta_1) + \exp(\|\mathbf{x}\| \cos \theta_2)}, \quad (3)$$

$$p_2 = \frac{\exp(\|\mathbf{x}\| \cos \theta_2)}{\exp(\|\mathbf{x}\| \cos \theta_1) + \exp(\|\mathbf{x}\| \cos \theta_2)}. \quad (4)$$

本文定义分类边界为  $p_1 - p_2 = 0$ 。由上述分析可以得到, 二分类问题的分类边界为  $(\mathbf{W}_1 - \mathbf{W}_2) \cdot \mathbf{x} + b_1 - b_2 = 0$ 。定义进行规范化处理后的 softmax 损失函数为改进的 softmax 损失函数, 改进的 softmax 损失函数的分类边界为  $\|\mathbf{x}\| \cos \theta_1 - \|\mathbf{x}\| \cos \theta_2 = 0$ 。当  $\|\mathbf{x}\| \neq 0$  时, 分类边界变为  $\cos \theta_1 - \cos \theta_2 = 0$ , 即分类边界完全取决于  $\theta_1$  和  $\theta_2$ 。

虽然上述分析是基于二分类问题讨论的, 但这很容易推广到多分类问题中。使用改进的 softmax 损失函数, 把标签分到类别  $i$  中, 只需满足  $\cos \theta_i > \cos \theta_{\text{else}}$ , 即  $\theta_i < \theta_{\text{else}}$ , 其中  $\theta_{\text{else}}$  为其他类别特征向量和权重向量之间的夹角。这就让权重向量  $\mathbf{W}_i$  和特征向量  $\mathbf{x}$  之间的夹角  $\theta_i$  成为一个可靠的分类标准。

传统的 softmax 损失函数和改进的 softmax 损失函数的分类边界如图 2 所示。由图 2 可以看到, 无论是传统的 softmax 损失函数还是改进的 softmax 损失函数, 其分类边界都是两个权重向量

间的角平分线, 并没有提高特征间的区分度。因此, 采用基于角度距离的损失函数来学习强区分度的特征, 以增加类间特征距离, 压缩类内特征距离, 从而提高特征间的区分度。

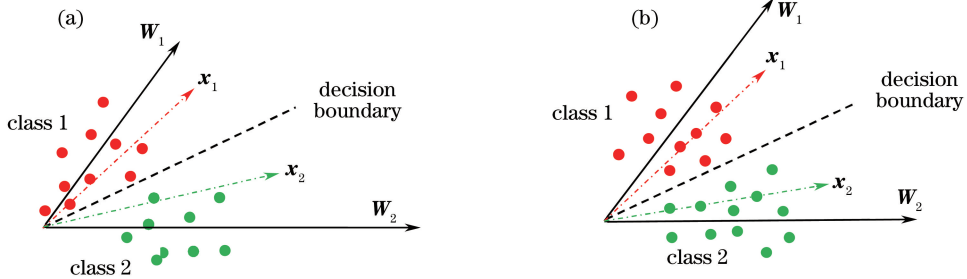


图 2 softmax 损失函数的比较。(a)传统的 softmax 损失函数;(b)改进的 softmax 损失函数

Fig. 2 Comparison of softmax loss function. (a) Traditional softmax loss function; (b) improved softmax loss function

### 2.3 基于角度距离的损失函数

根据前文对 softmax 损失函数的分析, 分类边界在很大程度上影响特征的分布, 而基于角度距离的损失函数的最基本思想就是通过控制分类边界来产生类间的角度距离。假设已知类别 1 中的一个特征向量为  $\mathbf{x}$ ,  $\theta_i$  为权重向量  $\mathbf{W}_i$  和特征向量  $\mathbf{x}$  之间的夹角。改进的 softmax 损失函数的分类边界为  $\cos \theta_1 - \cos \theta_2 = 0$ 。若要让特征向量  $\mathbf{x}$  被正确分到类别 1, 需要满足  $\cos \theta_1 > \cos \theta_2$ , 即只需满足  $\theta_1 < \theta_2$ 。若要让分类条件变得更加严格, 产生类别间的角度距离, 增加类间距离, 缩小类内距离, 需满足  $\theta_1$  远小于  $\theta_2$ 。

本文选择一种更为严格的分类条件:  $\cos \theta_1 > \cos[(1-\omega)\theta_2]$ , 可写为  $\cos[\theta_1/(1-\omega)] > \cos \theta_2$ ,

其中  $\omega$  是在区间  $[0, 1)$  上的实数, 引入参数  $\omega$  可以控制角度距离的大小。  $(1-\omega)$  在  $(0, 1]$  上取值, 因此新的分类边界须满足  $\theta_1$  远小于  $\theta_2$ 。由上述分析可知, 类别 1 的决策边界是  $\cos \theta_1 = \cos[(1-\omega)\theta_2]$ , 只需令  $\cos \theta_1 > \cos[(1-\omega)\theta_2]$ , 即满足  $\theta_1 < (1-\omega)\theta_2$  时, 特征向量  $\mathbf{x}$  才能被正确分到类别 1。同理, 满足  $\theta_2 < (1-\omega)\theta_1$  时, 特征向量  $\mathbf{x}$  分才能被到类别 2, 此时类别 2 的决策边界为  $\cos \theta_2 = \cos(\theta_1 - \omega\theta_1)$ 。上述的分类边界会产生大小为  $\omega\theta_1^2/(2-\omega)$  的角度距离, 其中  $\theta_1$  是向量  $\mathbf{W}_1$  和  $\mathbf{W}_2$  间的夹角。

把基于角度距离的损失函数的思路加入到改进的 softmax 损失函数中, 可以得到基于角度距离损失函数的表达式:

$$L_{\text{angular}} = -\frac{1}{N} \sum_{i=1}^N \log_2 \frac{\exp\left[\|\mathbf{x}_i\| \cos\left(\frac{1}{1-\omega}\theta_{y_i}\right)\right]}{\exp\left[\|\mathbf{x}_i\| \cos\left(\frac{1}{1-\omega}\theta_{y_i}\right)\right] + \sum_{j \neq y_i, j=1}^K \exp(\|\mathbf{x}_j\| \cos \theta_j)}, \quad (5)$$

式中:  $j$  的取值范围为  $[1, K]$ ,  $K$  为类别的数量;  $N$  为训练样本的数量;  $\mathbf{x}_i$  为第  $i$  个训练样本的特征向量;  $\theta_j$  为权重向量  $\mathbf{W}_j$  和特征向量  $\mathbf{x}_i$  之间的夹角;  $\theta_{y_i}$  为权重向量  $\mathbf{W}_{y_i}$  和特征向量  $\mathbf{x}_i$  之间的夹角。当  $\omega = 0$  时, (5) 式即为改进的 softmax 损失函数。

所提出的基于角度距离的损失函数对于不同的类别有不同的分类边界, 可以产生不同类别间的角度距离。图 3 所示为二分类任务中基于角度距离损失函数、改进的 softmax 损失函数和本文所提损失函数的分类边界。本文提出的基于角度距离的损失

函数的分类边界不仅是权重向量  $\mathbf{W}_1$  和  $\mathbf{W}_2$  间的角平分线, 还产生了角度距离。3 种损失函数分类边界条件的对比如表 1 所示。加入超参数  $\omega$  的目的是为了学习不同身份类别间的角度距离, 角度距离随着参数  $\omega$  的增大而增大。本文提出的基于角度距离的损失函数是以改进的 softmax 损失函数为前提, 即以权重向量的模值置 1, 偏置归零 ( $\|\mathbf{W}_i\| = 1, b_i = 0$ ) 为前提条件。类别的预测结果仅依赖于样本特征向量  $\mathbf{x}$  和权重向量  $\mathbf{W}_i$  之间的角度, 样本  $\mathbf{x}$  将被分到角度最小的身份类别中去。



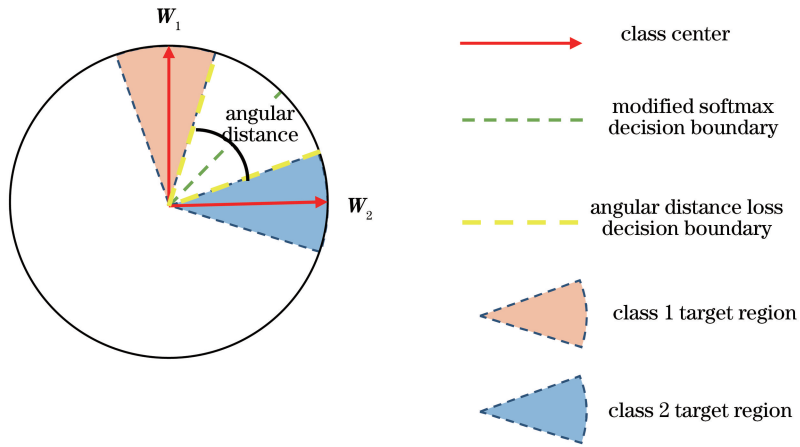


图 3 本文提出的基于角度距离损失函数的示意图

Fig. 3 Schematic of the proposed angular distance loss function

表 1 三种损失函数的分类边界对比

Table 1 Comparison of classification boundaries of loss functions

Loss function	Decision boundary
Original softmax loss	$(W_1 - W_2) \cdot x + b_1 - b_2 = 0$
Modified softmax loss	$\ x\  (\cos \theta_1 - \cos \theta_2) = 0$
Angular distance loss	$\ x\  \{ \cos \theta_1 - \cos[(1-\omega)\theta_2] \} = 0$ for class 1 $\ x\  \{ \cos[(1-\omega)\theta_1] - \cos \theta_2 \} = 0$ for class 2

### 2.4 密集连接 CNN 结构

所设计的 CNN 结构使用了目前精度高、参数少的密集连接模块<sup>[27]</sup>。密集连接模块的结构具体如图 4 中的密集连接块 1 所示,其中图 4 中的  $X_i$  和  $H_i$  的分别表示输入特征图矩阵和非线性变换操作。密集连接模块为了确保网络中各层之间信息传递的有效性,直接将所有层相互连接,每一层都从前面的所有层获得附加输入,并将本层的特征传递到后续

所有层。密集连接模块通过对特征的重复利用,搭建了一种易于训练、参数效率高的精简模型。该模型改善了整个网络的信息传递效率和梯度传递效率,减少了当训练集规模较小时的过拟合现象。与现有的其他模型相比,密集连接模型需要更少的参数,并且具有较高的准确性。本文设计的密集连接 CNN 结构主要由三个部分组成:密集连接模块、过渡模块、分类器模块。

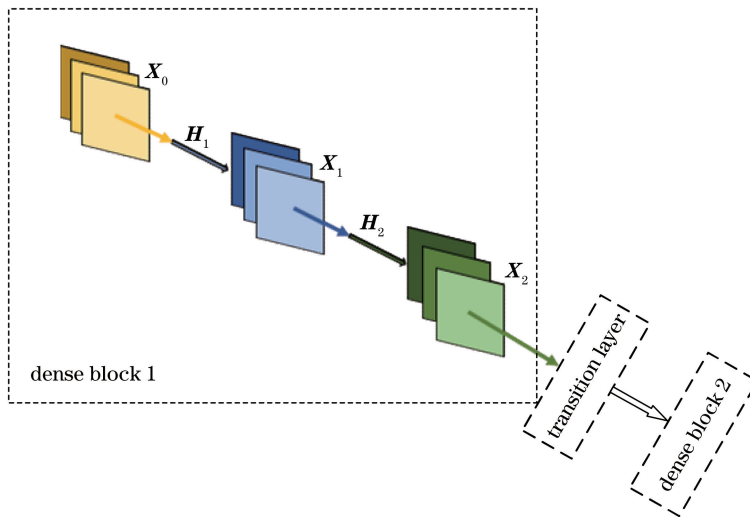


图 4 密集连接网络的结构

Fig. 4 Structure of densely connected networks

### 2.4.1 密集连接模块

假设输入图像用  $x_0$  表示,密集连接模块共有  $L$  层,每一层都进行一次非线性变换  $H_l(\cdot)$ ,其中  $l$  是层数的序号。非线性变换  $H_l(\cdot)$  包括批量归一化 (BN)<sup>[14]</sup>、激活函数<sup>[6]</sup>、卷积、池化<sup>[19]</sup>等复合非线性运算。将  $x_l$  定义为第  $l$  层的输出。

密集连接模块的结构如图 4 的密集连接块 1 所示。密集连接模块的层间变换可表示为

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]), \quad (6)$$

式中:  $[x_0, x_1, \dots, x_{l-1}]$  为从输入层、第 1 层, ..., 第  $l-1$  层网络所有特征图的连接。  $H_l(\cdot)$  包含六个连续非线性变换,依次为: BN、线性整流函数 (ReLU)、 $1 \times 1$  卷积、BN、ReLU 和  $3 \times 3$  卷积。

### 2.4.2 过渡模块

过渡模块是两个密集连接模块之间的 CNN 结构,过渡层的作用是为了改变不同密集连接模块特征图的尺寸,对特征图进行下采样。过渡模块包含 3 个连续的神经网络层,依次为 BN 层,  $1 \times 1$  卷积层和一个  $2 \times 2$  平均池化层。

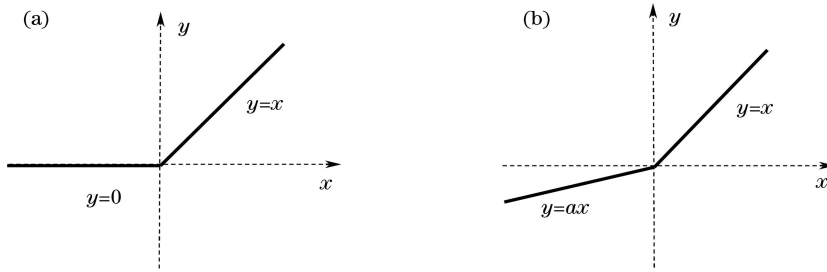


图 5 激活函数的对比。(a) ReLU; (b) PReLU

Fig. 5 Comparison of activation functions. (a) ReLU; (b) PReLU

设计了 4 种深度、2 种宽度的 8 个密集连接 CNN 结构,分别为 42 层网络、54 层网络、78 层网络和 122 层网络,每种网络分别设计了宽度为 16 和宽度为 32 的两组。依照层数依次由浅到深的规律设计了 4 种深度的密集连接 CNN 结构,目的是为了验证本文提出的网络结构随着网络深度的增加对分

表 2 密集连接结构的具体配置

Table 2 Specific configuration of the dense connection structure

Layer	Output size	DenseFace-42	DenseFace-54	DenseFace-78	DenseFace-122
Dense block 1	$56 \times 56$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 6$
Dense block 2	$28 \times 28$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 5$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 12$
Dense block 3	$14 \times 14$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 5$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 24$
Dense block 4	$7 \times 7$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 16$

### 2.4.3 分类器模块

分类器模块是对神经网络结构提取到的特征通过全连接层和最后的损失函数层进行分类,本文选择的分类器是基于角度距离损失函数的分类器。该分类器可以让前面的神经网络结构学习到具有强区分度的人脸特征,以符合最大类内特征距离小于最小类间特征距离这个理想的特征分类标准。

### 2.4.4 整体网络结构

所设计的密集连接 CNN 结构,对传统的密集连接模块的激活函数作了改进,将原来密集块和过渡层中所有的 ReLU 激活函数都替换为 PReLU (parametric rectified linear unit) 激活函数。ReLU 和 PReLU 激活函数的对比如图 5 所示。PReLU 激活函数是在原来的 ReLU 激活函数的基础上增加了参数修正<sup>[28]</sup>。原来的 ReLU 激活函数随着训练次数的推进,可能会出现权重无法更新的情况。PReLU 激活函数加入了一个学习参数  $a_i$ ,可在一定程度上避免梯度为零的情况,让训练结果更容易收敛。

类准确率的影响。这 4 种深度网络结构密集连接结构配置如表 2 所示,4 种网络的整体结构如图 6 所示。每种网络结构输入的人脸图像尺寸均为  $112 \text{ pixel} \times 112 \text{ pixel}$ ,含有 4 个密集连接模块,每个密集连接块有不同的层数和相同的网络宽度,都具有较小的参数量。

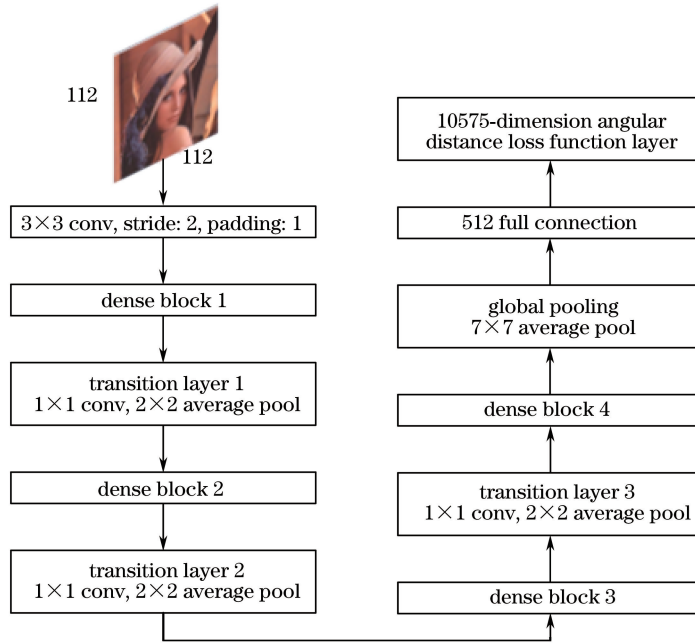


图6 网络整体结构

Fig. 6 Integral structure of network

### 3 分析与讨论

#### 3.1 超参数 $\omega$ 的影响

基于角度距离的损失函数,在传统的 softmax 损失函数的基础上,加强了分类边界的约束条件,并引入了超参数  $\omega$  来控制角度距离的大小。

分类边界条件如表 1 所示,也可以表示为  $\cos \theta_1 = \cos(1-\omega)\theta_2$ ,角度距离的大小完全取决于超参数  $\omega$  的选择。角度距离  $\omega\theta_1^2/(2-\omega)$  随着参数  $\omega$  的增大而增大,故一定存在一个最小的  $\omega$ ,让所提的损失函数符合理想的特征分类标准。将  $\omega_{\min}$  定义为超参数  $\omega$  符合理想特征分类标准的最小值。

在二分类场景下,特征的最大类内角度距离为  $(1-\omega)\theta_1^2/\omega + (1-\omega)\theta_1^2/(2-\omega)$ ,最小类间角度距离为  $\omega\theta_1^2/(2-\omega)$ ,若要符合理想的特征分类标准,则应满足:

$$\frac{1-\omega}{\omega}\theta_1^2 + \frac{1-\omega}{2-\omega}\theta_1^2 \leq \frac{\omega}{2-\omega}\theta_1^2, \quad (7)$$

式中:  $\theta_1$  是向量  $\mathbf{W}_1$  和  $\mathbf{W}_2$  间的夹角。求解(7)式可以得到  $\omega_{\min} \geq \sqrt{3}-1 \approx 0.73$ 。在二分类问题中,为了符合理想的特征分类标准, $\omega$  的最小值为  $\sqrt{3}-1$ 。

在多分类场景下,假设  $k$  个类别的权重向量  $\mathbf{W}_i$  均匀地分布在单位圆上,那么向量  $\mathbf{W}_i$  和  $\mathbf{W}_{i+1}$  间的夹角为  $\theta_i^{i+1} = 2\pi/k$ ,第  $i$  个类别的最大类内角度距离为  $(1-\omega)\theta_i^{i+1}/(2-\omega) + (1-\omega)\theta_{i-1}^i/(2-\omega)$ ,最

小类间角度距离为  $\min\{\omega\theta_i^{i+1}/(2-\omega), \omega\theta_{i-1}^i/(2-\omega)\}$ ,且  $\theta_i^{i+1} = \theta_{i-1}^i$ 。若要符合理想的特征分类标准,则应满足:

$$\frac{1-\omega}{2-\omega}\theta_i^{i+1} + \frac{1-\omega}{2-\omega}\theta_{i-1}^i \leq \min\left\{\frac{\omega}{2-\omega}\theta_i^{i+1}, \frac{\omega}{2-\omega}\theta_{i-1}^i\right\}. \quad (8)$$

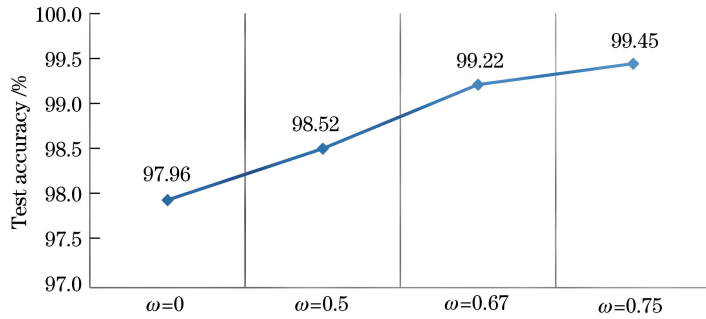
求解(8)式可以得到  $\omega_{\min} \geq 2/3 \approx 0.67$ 。

在后续试验中,如果没有特殊说明,无论在二分类场景还是多分类场景下,均把超参数  $\omega$  设置为 0.75。

为了验证上述结论,在 LFW 数据集上验证超参数  $\omega$  对人脸识别准确率的影响。实验结果如图 7 所示,该结果是基于本文提出的宽度为 32、层数为 122 的密集连接 CNN 模型得到。可以观察到随着超参数  $\omega$  的增加,在 LFW 数据集上的人脸识别准确率也不断提高,当  $\omega$  取 0.75 时,其准确率比无角度距离的损失函数提高了 1.49%,这就验证了上述结论的正确性。

#### 3.2 密集连接 CNN 结构的影响

所设计的 4 种深度(42 层、54 层、78 层和 122 层)的密集连接 CNN 模型中,每个模型都有 4 个密集连接模块,并且每个模块都输出相同个数的特征图,这 4 种模型的每个密集块的宽度分别设置为 16 和 32。4 种网络模型的详细配置如表 2 所示。密集连接的 CNN 模型比传统的 CNN 模型拥有更少的

图 7 不同超参数  $\omega$  的人脸识别准确率Fig. 7 Face recognition accuracy versus hyperparameter  $\omega$ 

参数,模型更加紧凑,整个网络的特征可以重复利用,参数效率也更高<sup>[29]</sup>,表 3 比较了目前几种 CNN 模型的参数量。

由表 3 可得,密集连接模块比传统 CNN 结构的参数效率更高,LeNet、AlexNet、VGGNet 网络结构的层数都较少,不超过 20 层,但参数量都很大,尤其是 AlexNet 和 VGGNet,其参数量都超过了  $6 \times 10^7$ ,而 GoogleNet 虽然只有 22 层,参数量却和 42 层的密集连接网络的参数量相当,目前效果较好的

152 层 ResNet 的参数量甚至超过  $6 \times 10^7$ 。本文设计的密集连接模块深度为 122 层,当宽度为 32 时,CNN 模型的参数量为  $1.28 \times 10^7$ ;当宽度为 16 时,CNN 模型的参数量更少,只有  $7.4 \times 10^6$ 。表 3 计算的所有密集连接网络结构的参数量都是在加上最后 10575 维全连接层的  $5.4 \times 10^6$  参数量后计算出的结果。若将本文的网络模型用在类别较少的其他数据集上,参数量会更少。因此,密集连接的 CNN 模型比传统的 CNN 模型拥有更少的参数量。

表 3 几种 CNN 模型参数量的比较

Table 3 Comparison of parameter quantities of several convolutional neural network models

Net structure	Input size /pixel	Depth /layer	Parameter / $10^6$
LeNet	$32 \times 32 \times 1$	5	0.062
AlexNet	$227 \times 227 \times 3$	8	62.4
VGGNet	$224 \times 224 \times 3$	16	138.4
GoogleNet	$224 \times 224 \times 3$	22	5.3
ResNet	$224 \times 224 \times 3$	152	61.3
DenseFace (width: 32)	$112 \times 112 \times 3$	42	6.7
		54	7.3
		78	8.9
		122	12.8
DenseFace (width: 16)	$112 \times 112 \times 3$	42	5.78
		54	5.9
		78	6.37
		122	7.4

基于角度距离的损失函数设计了密集连接 CNN 模型,用不同层数的密集连接 CNN 来训练基于角度距离的损失函数和传统的 softmax 损失函数。实验是在 LFW 数据集上进行测试,实验结果如图 8 所示。

基于角度距离的损失函数监督的网络模型比传统 softmax 损失函数监督的网络模型在 LFW 数据集上的准确率更高,准确率提升了 1.67%~1.93%,验证了基于角度距离损失函数在人脸识别任务中比 softmax 损失函数的性能更好,更适合于开集人脸

识别任务。随着网络层数的增加,人脸识别准确率也不断提高,由 98.24% 提升到 99.45%,提高了 1.21%,验证了本文的密集连接 CNN 结构随着网络层数的增加,并没有出现过拟合现象,网络的精度不断提高,网络很容易优化。

设计了两组密集连接 CNN 结构,一组的宽度为 16,另一组的宽度为 32,图 9 显示了 4 种不同深度的网络在宽度分别为 16 和 32 时在 LFW 数据集上的测试准确率。网络的宽度越宽,参数量越大,测试准确率也越高。当 54 层网络模型的宽度由 16 增



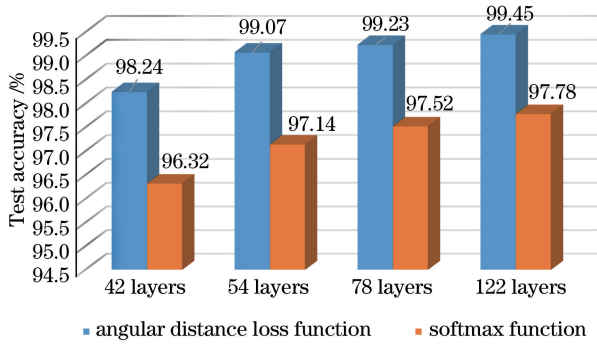


图 8 不同层数和不同损失函数的网络结构在 LFW 数据集上的测试准确率

Fig. 8 Test accuracy of LFW dataset for network structures with different layer numbers and different loss functions

加到 32 时,准确率提高幅度最大,提高了 1.91%;122 层网络测试准确率提高了 0.36%。随着网络宽度和深度的增加,测试准确率都有所提高。

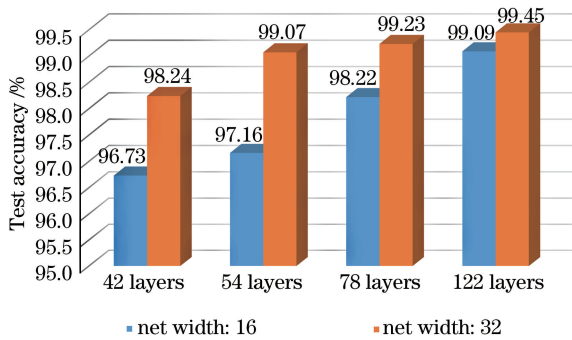


图 9 不同层数和宽度的网络结构在 LFW 数据集上的测试准确率

Fig. 9 Test accuracy of LFW dataset for network structures with different layer numbers and widths

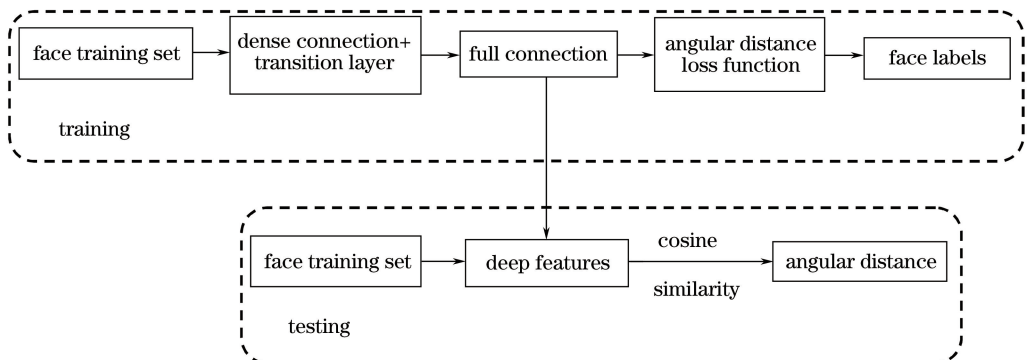


图 10 本文实验实施流程

Fig. 10 Proposed implementation process

#### 4.1.3 训练数据集

使用一个公开的西方人脸数据库 CASIA-WebFace<sup>[17]</sup>来训练本文的密集连接 CNN 模型。这个数据集共有 494414 张人脸照片,分别属于 10575

## 4 实验

### 4.1 实验设置

#### 4.1.1 人脸预处理

使用标准的人脸识别预处理流程,所有图像中的人脸关键点检测均使用 MTCNN 算法<sup>[30]</sup>。使用相似性变换实现图像的人脸对齐和剪裁。将检测到的人脸都剪裁成 112 pixel×112 pixel,每个 RGB 颜色通道的像素都进行规范化处理,即原来 3 个通道在 $[0,255]$ 范围内的像素值,每个像素点的像素值都减去 127.5 然后除以 128,规范化到 $[0,1)$ 区间上的像素值。

#### 4.1.2 CNN 框架及训练配置

使用 Ubuntu16.04 系统下的 Caffe<sup>[31]</sup>框架来实现基于角度距离的损失函数和密集连接的 CNN 结构,具体实施流程如图 10 所示,主要包括对网络结构的训练、人脸特征的提取和人脸身份的识别。使用密集连接模块,分别设计了 42、54、78、122 层的宽度分别为 16 和 32 的 CNN 结构作为本文的网络结构。为了公平起见,所有损失函数效果的对比都使用相同的 CNN 结构。不同 CNN 结构的配置如表 2 所示。本实验均用一个 GPU(显卡型号为 NVIDIA Quadro P4000)批次为 32 进行训练。学习率一开始设置为 0.1,到 48 万次和 96 万次迭代时,学习率依次除以 10,迭代到 144 万次时结束训练。训练时,权重衰减设置为 0.0005,权重动量设置为 0.9,对训练数据仅采用随机镜像对称作为数据增产<sup>[32]</sup>。

个不同身份。训练数据的规模只有  $0.49 \times 10^6$ ,是相对较小的数据集。DeepFace<sup>[11]</sup>的数据规模达到  $4 \times 10^6$ ,VGGFace<sup>[33]</sup>的数据规模达到  $2 \times 10^6$ ,FaceNet<sup>[5]</sup>的数据规模甚至达到了  $2 \times 10^8$ 。本文后

续实验中的所有网络模型均使用 CASIA-WebFace 数据集进行训练。

#### 4.1.4 测试数据集

##### 1) LFW 数据集

LFW 数据集<sup>[8]</sup>包含 13233 张人脸图片,分别来自 5749 个不同身份。这个数据集包含的人脸图片存在大量的姿势、表情和光照变化。按照 LFW 数据集的测试协议<sup>[8]</sup>,对 LFW 数据集的 6000 张人脸对进行人脸验证。

##### 2) MegaFace 数据集

MegaFace 数据集<sup>[9]</sup>是一个公开的人脸测试数据集,引入了百万数量级的干扰项,用来评估人脸识别算法的性能。MegaFace 数据集具有多个测试场

景,包括人脸确认、人脸验证、大数据量和小数据量的测试协议。如果训练集的规模小于  $0.5 \times 10^6$ ,那么该数据集可视为一个小数据集。本文是在小训练集数据量协议下对本文提出的算法进行评价。

#### 4.1.5 测试流程

从全连接层的输出提取人脸特征。在后文所有的实验中,测试集中的人脸特征均是由原始人脸特征和水平翻转人脸特征所共同决定。最后的测试准确率是由两张人脸特征间的余弦距离计算得到,得分最高的类别将作为人脸识别的身份确认。

## 4.2 实验结果

### 4.2.1 LFW 数据集的实验结果

LFW 数据集的实验结果如表 4 所示。

表 4 不同损失函数、人脸识别算法的测试准确率

Table 4 Test accuracy of different loss functions or face recognition algorithms

Method	Dataset	Data amount / $10^6$	Accuracy / %
DeepFace	LFW	4	97.33
FaceNet	LFW	200	<b>99.67</b>
Deep FR	LFW	2.6	98.85
DeepID2+	LFW	0.3	98.74
Center Face	LFW	0.7	99.31
Softmax loss	CAISA-WebFace	0.49	97.78
Triplet loss	CAISA-WebFace	0.49	98.65
Center loss	CAISA-WebFace	0.49	99.02
L-softmax loss	CAISA-WebFace	0.49	99.15
Angular distance loss	CAISA-WebFace	0.49	<b>99.45</b>

由表 4 可知,与其他损失函数相比,基于角度距离的损失函数学到的特征具有区分度更强,识别准确率更高等特点,这证明了基于角度距离的损失函数在人脸识别任务具有优越性。在 CAISA-WebFace 数据集上进行训练,采用基于角度距离的损失函数进行监督,在 LFW 数据集上的测试准确率为 99.45%,高于其他损失函数监督得到的网络模型。基于角度距离的损失函数在 LFW 数据集上的测试准确率比 softmax 损失函数、triplet 损失函数、center 损失函数和 L-softmax 损失函数提高了 0.3%~1.67%。

提出的人脸识别算法与其他人脸识别算法进行对比也有很强的优势,比同在小数据集上训练的 DeepID2+ 和 Center Face 算法的准确率分别提高了 0.71% 和 0.14%,比用大数据集训练的 DeepFace 算法和 Deep FR 算法的准确率提高了 0.6%~2.12%,效果仅次于用超大规模数据集训练的 FaceNet 算法,该算法训练集图片数量为  $2 \times 10^8$ ,是 CAISA-WebFace 数据集图片数量的 400 倍。因

此,本文提出的基于角度距离损失函数和密集连接 CNN 的人脸识别算法具有很好的人脸识别性能。

### 4.2.2 MegaFace 数据集的实验

MegaFace 数据集的实验结果如表 5 所示。本文提出的基于角度距离损失函数和密集连接 CNN 的人脸识别算法在 MegaFace 数据集的小规模测试协议下取得了很有优势的准确率,人脸确认准确率达到 72.534%,人脸验证准确率达到 85.348%。在小数据集的人脸确认任务中比第二名的 Deepsense-Small 算法的准确率高出 1.551%,人脸验证任务中比第二名的 Deepsense-Small 算法的准确率高出 2.497%。与在大数据集上训练的模型相比,本文提出的算法性能仍然有很大竞争力,其人脸确认和人脸验证的准确率仅比大数据集上第一名的算法低 2%,而比一般的大数据集测试协议下的 FaceNet 算法高出 2.038%,人脸验证准确率只降低了 1.125%。此外,本文只采用了密集连接的 CNN 模型,由基于角度距离的损失函数进行监督,就取得了良好的性能,已经超过了大多数目前性能领先的算法。同时,本文

还对比了不同损失函数的性能,每种进行比较的损失函数均使用相同的 CNN 网络模型,本文提出的基于角度距离的损失函数的人脸识别准确率都有明显的性能优势。其人脸确认和人脸验证准确率分别比原

始的 softmax 损失函数提高了 17.906% 和 19.616%; 其人脸确认准确率比 triplet 损失函数、center 损失函数和 L-softmax 损失函数高出 5.499%~7.836%, 人脸验证准确率高出 5.163%~7.318%。

表 5 不同损失函数、人脸识别算法在 MegaFace 数据集上的测试准确率

Table 5 Test accuracy of different loss functions or face recognition algorithms on the MegaFace dataset

Method	Test protocol	Accuracy / %	
		Face identification	Face verification
FaceNet	large	70.496	86.473
Deepsense	large	<b>74.798</b>	<b>87.764</b>
Deepsense	small	70.983	82.851
Softmax loss	small	54.628	65.732
Triplet loss	small	64.698	78.030
Center loss	small	65.334	80.106
L-softmax loss	small	67.035	80.185
Angular softmax loss	small	<b>72.534</b>	<b>85.348</b>

上述实验结果可以证明本文所提出的基于角度距离损失函数和密集连接 CNN 的人脸识别算法在开集人脸识别任务中具有良好的性能,具有类间角度距离的损失函数可以学习到更有区分度的人脸特征,在人脸识别任务中性能可以获得显著提升。

## 5 结 论

提出了一种基于角度距离损失函数和密集连接 CNN 的人脸识别算法,并比较了 8 种不同深度和宽度的密集连接 CNN 模型,模型结构简单、参数复杂度低,可显著减少传统 CNN 模型的参数量。同时采用基于角度距离的损失函数对网络模型进行监督,学习具有角度距离并且区分度强的人脸特征,使得人脸特征满足特征的理想分类条件——最大类内角度距离小于最小类间角度距离。本文的密集连接 CNN 模型在加入基于角度距离的损失函数后,人脸识别准确率大幅提高。本文提出的人脸识别算法在 CAISA-WebFace 数据集上进行训练,然后在 LFW 和 MegaFace 人脸识别数据集上进行测试。实验结果验证了本文提出的算法在开集人脸识别任务中的优越性和巨大潜力。后续研究将主要致力于损失函数的进一步改进,使得网络更容易收敛,以降低训练的迭代次数和训练时长。

## 参 考 文 献

[1] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[2] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]//International Conference on Neural Information Processing Systems, 2012, 60(2): 1097-1105.

[3] Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.

[4] Srivastava R K, Greff K, Schmidhuber J. Training very deep networks [C] // Conference and Workshop on Neural Information Processing System, 2015.

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

[6] Sun Y, Chen Y, Wang X, *et al.* Deep learning face representation by joint identification-verification[C]// Conference and Workshop on Neural Information Processing System, 2014.

[7] Wen Y, Zhang K, Li Z, *et al.* A discriminative feature learning approach for deep face recognition[C]//Springer International Publishing, 2016, 47(9): 499-515.

[8] Huang G B, Learned-Miller E. Labeled faces in the wild: updates and new reporting procedures [R]. USA: Massachusetts Amherst, 2014: 3-14.

[9] Kemelmacher-Shlizerman I, Seitz S M, Miller D, *et al.* The megaface benchmark: 1 million faces for recognition at scale [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4873-4882.

[10] Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes [C] // IEEE Conference on Computer Vision and Pattern

- Recognition, 2014: 1891-1898.
- [11] Taigman Y, Yang M, Ranzato M, *et al.* Deepface: closing the gap to human-level performance in face verification [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1701-1708.
- [12] Sun Y, Wang X, Tang X. Sparsifying neural network connections for face recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4856-4864.
- [13] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2006: 1735-1742.
- [14] LeCun Y, Bottou L, Bengio Y, *et al.* Gradient based learning applied to document recognition [C] // Intelligent Signal Processing of the IEEE, 1998, 86 (11): 2278-2324.
- [15] Russakovsky O, Deng J, Su H, *et al.* Imagenet large scale visual recognition challenge [C] // International Journal of Computer Vision, 2015, 115 (3): 211-252.
- [16] Huang G, Sun Y, Liu Z, *et al.* Deep networks with stochastic depth [C] // European Conference on Computer Vision, 2016: 646-661.
- [17] Yi D, Lei Z, Liao S, *et al.* Learning face representation from scratch [J]. arXiv, 2014: 1411.7923.
- [18] Liu W Y, Wen Y D, Yu Z D, *et al.* SphereFace: deep hypersphere embedding for face recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6738-6746.
- [19] Xing E P, Ng A Y, Jordan M I, *et al.* Distance metric learning with application to clustering with side information [C] // International Conference on Neural Information Processing Systems, 2002, 15: 521-528.
- [20] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification [J]. Journal of Machine Learning Research, 2009, 10(1): 207-244.
- [21] Köstinger M, Hirzer M, Wohlhart P, *et al.* Large scale metric learning from equivalence constraints [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2288-2295.
- [22] Ying Y, Li P. Distance metric learning with eigenvalue optimization [J]. Journal of Machine Learning Research, 2012, 13(1): 1-26.
- [23] Hu J, Lu J, Tan Y P. Discriminative deep metric learning for face verification in the wild [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2014, 27(10): 1875-1882.
- [24] Lu J, Wang G, Deng W, *et al.* Multi-manifold deep metric learning for image set classification [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1137-1145.
- [25] Song H O, Xiang Y, Jegelka S, *et al.* Deep metric learning via lifted structured feature embedding [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4004-4012.
- [26] W Liu, Y Wen, Z Yu, and M Yang. Large-margin softmax loss for convolutional neural networks [C] // International Conference on International Conference on Machine Learning, 2016: 507-516.
- [27] Ming Z H, Chazalon J, Luqman M M, *et al.* Simple triplet loss based on intra/inter-class metric learning for face verification [C] // IEEE International Conference on Computer Vision Workshops, 2018: 1656-1664.
- [28] Zhang R Y, Wang Q Q, Lu Y. Combination of ResNet and center loss based metric learning for handwritten Chinese character recognition [C] // 14th IAPR International Conference on Document Analysis and Recognition, 2017: 25-29.
- [29] Huang G, Liu Z, Maaten L, *et al.* Densely connected convolutional networks [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2261-2269.
- [30] Zhang K P, Zhang Z P, Li Z F, *et al.* Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.
- [31] Jia Y Q, Shelhamer E, Donahue J, *et al.* Caffe: convolutional architecture for fast feature embedding [J]. arXiv, 2014: 1408.5093.
- [32] Choudhury S K, Padhy R P, Sa P K. Faster R-CNN with DenseNet for scale aware pedestrian detection vis-à-vis hard negative suppression [C] // IEEE 27th International Workshop on Machine Learning for Signal Processing, 2018: 1-6.
- [33] Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition [C] // British Machine Vision Conference, 2015.