

一种基于权重哈希化的深度人脸识别算法

曾 燕¹ 陈岳林¹ 蔡晓东²

(桂林电子科技大学机电工程学院 广西 桂林 541004)¹

(桂林电子科技大学信息与通信学院 广西 桂林 541004)²

摘 要 针对采用融合深度哈希的卷积神经网络进行人脸识别时可能存在准确率下降及内存占用率偏高的问题,提出了基于权重哈希化的深度人脸识别算法。首先,提出一种基于高低维特征维度拼接的全卷积深度哈希网络,用以保证融合深度哈希后网络模型的识别准确率;然后,提出一种基于权重哈希化的模型压缩方法,将浮点型权重量化为哈希编码来进行模型存储,用以减少模型的内存占用率。实验表明,该方法在基于 VGG 框架进行改进时,可将 VGG 原网络的识别总效率提高 68%,将准确率提高 1.67%且使模型尺寸压缩了 91.2%;该方法扩展到 Sphreface 框架时,在准确率略有提升的情况下将识别效率提高了 61%,将模型压缩了 42.24%。因此所提方法可提高识别准确率和效率,并减少内存占用率,同时还可扩展应用于其他网络。

关键词 深度哈希,维度拼接,模型压缩,全卷积网络

中图法分类号 TP183 文献标识码 A DOI 10.11896/j.issn.1002-137X.2019.06.041

Deep Face Recognition Algorithm Based on Weighted Hashing

ZENG Yan¹ CHEN Yue-lin¹ CAI Xiao-dong²

(School of Mechanical and Electrical Engineering,Guilin University of Electronic Technology,Guilin,Guangxi 541004,China)¹

(School of Information and Communication,Guilin University of Electronic Technology,Guilin,Guangxi 541004,China)²

Abstract In order to solve the problem that the accuracy rate may decrease and the memory occupancy rate may still be high when the convolution neural network with fused depth hash is used for face recognition,this paper proposed a deep face recognition algorithm based on weighted hashing. Firstly,a fully convolutional neural network of deep hash based on dimension splicing with high and low dimensional features is proposed to improve recognition accuracy. Secondly,a model compression method with floating-point weights quantized into hash coding is proposed to reduce memory occupancy rate of the model. The experimental results show that the proposed method improves efficiency by 68%,improves the Rank-1 accuracy by 1.67%,and the model size is compressed by 91.2% when it is improved based on VGG framework. In addition,it improves efficiency by 61% when the Rank-1 accuracy is slightly improved,and the model size is reduced by 42.24% when it is improved based on Sphreface framework. The results indicate that the proposed method can improve the recognition accuracy and efficiency,and reduce the memory usage. It also can be applied for other frameworks.

Keywords Deep hash,Dimension splicing,Model compression,Full convolution network (FCN)

1 引言

近年来,基于深度学习的人脸识别技术不断被工程化应用到各个领域,而在实际应用场景中,人脸识别工程的内存占用率以及识别效率是必须要考虑的问题。

基于深度学习的人脸识别技术采用卷积神经网络进行特征提取,再根据特征比对完成人脸识别任务。人脸特征提取网络包括 DeepFace^[1],Web-Scale Training (WST) Fusion^[2],

Visual Geometry Group (VGG)^[3]等大型网络结构,以及 Centerloss^[4],Sphreface^[5],ArcFace^[6]等基于 Resnet-20^[7]层的小型网络结构。大型网络结构应用于实际工程项目时存在很大困难,于是小型网络结构应运而生,但是对于很多实际应用的设备,这些小型网络结构也很难满足其高效率、低内存占用率的要求。针对此问题,文献[8]将深度哈希融入上述基础网络构架,能够在保证识别准确率的前提下有效提高识别效率并得到更小的网络模型,但通过观察网络训练过程可知,融合深

到稿日期:2018-05-31 返修日期:2018-10-27 本文受 2016 年广西重点研发计划(桂科 AB16380264),2018 年新疆重点研发专项(2018B03022)资助。

曾 燕(1992—),女,硕士生,主要研究方向为图像处理、人脸识别;陈岳林(1971—),男,教授,硕士生导师,主要研究方向为图像处理、模式识别;蔡晓东(1971—),男,教授,硕士生导师,CCF 会员,主要研究方向为图像处理、模式识别,E-mail:caixiaodong@guet.edu.cn (通信作者)。

度哈希后的网络收敛较困难,需要耗费更长的时间进行训练,如在文献[8]的激活函数对比实验中,融合深度哈希后对 VGG 网络进行微调时,需要迭代 50 000 次左右才能将损失值降低到 3;而在未融合深度哈希的情况下对 VGG 网络进行微调时,只需要迭代 10 000 次左右即可将网络的损失值降低到 3。

本文提出一种基于高低维特征维度拼接的全卷积深度哈希网络,将高维特征与近似哈希编码进行维度拼接并同时作为分类层的输入,使得网络收敛速度更快,可以大幅节约训练时间成本并能进一步提高识别准确率;与此同时,本文提出了一种基于权重哈希化的模型压缩方法,该方法将浮点型权重量化为哈希编码来进行模型存储,减少了网络模型的尺寸,从而减少了内存占用率。本文方法能够使融合深度哈希的人脸识别网络训练效率不变的情况下进一步保证识别效率,得到更低的内存占用率。

2 基于权重哈希化的特征提取方法

2.1 基于高低维特征拼接的人脸特征提取网络

在人脸特征提取网络中融合深度哈希的一般做法是:在特征提取层后加入隐藏层进行近似哈希编码的提取,再用量化误差损失与分类损失分别学习近似哈希编码和高维特征。相比于原网络,这种做法会导致网络收敛较慢,同时可能导致一些可调性较弱的网络模型的精度下降。文献[8]用基于全局平均池化层的全卷积神经网络来进一步提升效率并加快网络的训练速度,但是该方法仍存在上述两个问题,因此本文在文献[8]的基础上进行改进。网络构架如图 1 中模块 A 所示,具体设计思路如下。

(1)将分类层的输入层设计为高低维特征拼接层。本文采用全局与局部池化特征层和哈希层的维度拼接层作为网络分类层的输入层,使网络的分类损失分别同时作用于全局与局部池化特征层和哈希层。文献[8]采用哈希层作为分类层的输入,以得到具有分类性能的近似哈希编码,但是由于分类损失经过哈希层后才能对全局与局部池化特征层进行特征分

类,这个反向传播过程会弱化高维特征的分类能力,使得网络收敛性能降低,从而导致在二次识别过程中模型的精度下降,因此本文采用高低维特征拼接层作为分类层的输入。高低维特征的特征值在进行维度拼接后并未发生改变,因此本文将拼接后的特征作为分类层的输入,使得网络的分类损失层可以同时作用于高维特征层和哈希层,这种方式可以使网络的训练过程更接近原网络,从而提高其收敛性能,加快网络训练速度,并能得到具有强分类能力的高维特征和近似哈希编码,进而提高模型的识别准确率。

(2)根据图 1 中的模块 B 设计卷积层 2 的通道数。不同网络的可调性不同,性能较差的网络的可调性较高,改变其网络结构可大幅提高原网络的性能,而性能较好的网络的可调性较低,即使细微的改变也可能导致网络性能降低,因此本文根据网络性能来设计卷积层 2 的通道数。通常模块 B 中的原全连接特征层的维度较高,且为卷积层 2 和下采样层 2 之间通道数的 2 倍以上,网络的可调性较高;而原全连接特征层的维度较低且与卷积层 2 和下采样层 2 之间通道数相等时,网络的可调性较低。由于全局与局部池化特征层融合了全局与局部池化层,得到的特征维度为二者通道数之和,即为下采样层 2 通道数的 2 倍,针对可调性较高的网络,原全连接特征层维度很高,包含很多冗余信息,即使模块 A 中的特征层维度没有原全连接特征层的维度高,但其包含的有效信息足以达到与原网络模型相同的精度,因此本文设计卷积层 2 的通道数保持原网络设置不变;针对可调性较低的网络,原全连接特征层维度很低,而模块 A 中得到的特征层维度为原特征层维度的 2 倍,这在一定程度上会增加特征的冗余信息,导致训练过程中的有效信息受到干扰,从而使得更高维度的池化特征层的识别准确率降低,因此本文将卷积层 2 的通道数减半,以使模块 A 中的全局和局部池化特征层与模块 B 中的原全连接特征层的维度保持一致,通过微调原网络模型,用模块 B 中的网络参数优化模块 A 中的参数,以提高网络的泛化能力,从而进一步保证网络模型的识别准确率。

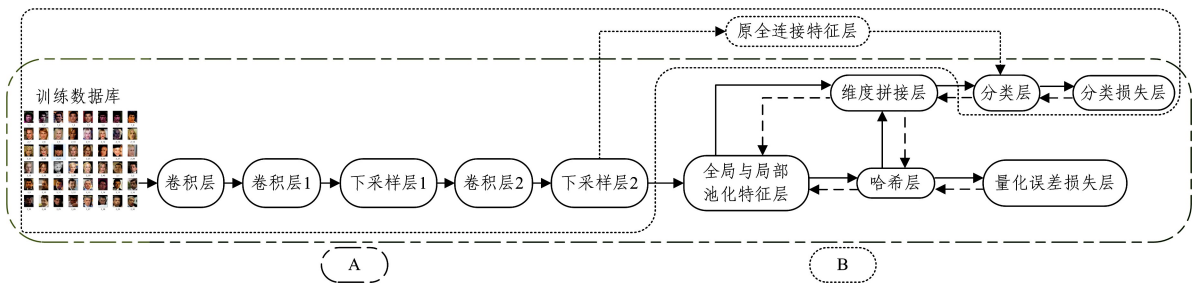


图 1 基于高低维特征拼接的人脸特征提取网络构架

Fig. 1 Face feature extraction network structure based on high-low-dimensional feature mosaic

2.2 基于权重哈希化的模型压缩方法

大型卷积神经网络在融合深度哈希后,其模型得到了大幅度压缩,但仍然不小,而很多小型卷积神经网络在融合哈希后,模型压缩幅度并不大,这两种网络的模型很有可能无法满足实际应用需求,因此为了进一步压缩融合深度哈希后的网络模型,本文提出了一种基于权重哈希化的模型压缩方法。

卷积神经网络的模型保存的权重是单精度浮点数,每个

浮点数所占内存为 4 个字节,即 32 位,而一个整型所占内存也为 4 个字节,因此本文将浮点型权重量化并整合为二值哈希编码来进行模型存储。TWN (Ternary Weight Networks)[9]将权重量化为三值编码来进行前向计算,本文借鉴文献[9],对量化后的三值权重进行重新编码,使其以二值哈希码作为模型的存储方式。浮点型权重量化为三值权重的阈值取值为:

$$\Delta = \frac{0.7}{n} \sum_{i=1}^n |W_i| \tag{1}$$

其中, n 为权重个数; W_i 为原浮点型权重。根据阈值定义量化函数如下:

$$W'_i = \begin{cases} 1, & W_i > \Delta \\ 0, & |W_i| \leq \Delta \\ -1, & W_i < -\Delta \end{cases} \tag{2}$$

其中, W'_i 为量化后的三值权重。三值权重哈希化的缩放因子定义为:

$$\alpha_\Delta^* = \frac{1}{|I_\Delta|} \sum_{i \in D} |W_i| \tag{3}$$

其中, $|I_\Delta|$ 为 $|W_i| > \Delta$ 的元素的个数。三值权重与二值哈希编码的转换规则如下:

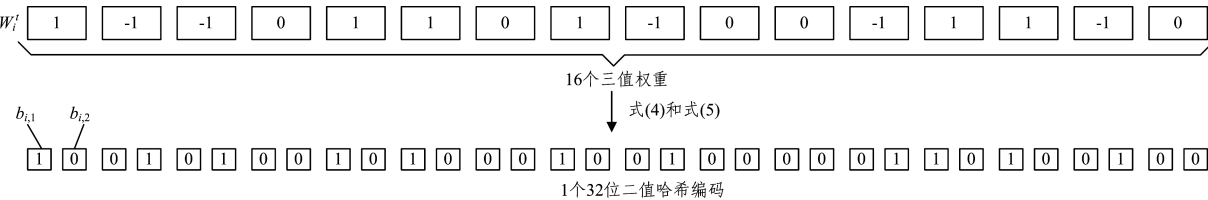


图 2 三值权重重编码规则简图

Fig. 2 Recoding rule diagram of ternary weight

3 实验结果与分析

本文实验平台的配置为 Intel i3 四核 3.4 GHz 处理器、8 GB 内存、GTX980Ti 显卡以及 Ubuntu 14.04 操作系统,并使用基于 C++ 编程语言的 Caffe 工具包。

3.1 实验数据库

3.1.1 训练数据库

本文采用 Institute of Automation, Chinese Academy of Sciences WebFace (CASIA-WebFace)^[10] 公开的数据库对网络进行训练,筛选掉错误的和重复的图片之后,最终的样本库包含 10 575 人,共 453 439 张图片。

3.1.2 测试数据库

本文使用公开人脸库 Labeled Faces in the Wild (LFW)^[11] 数据库进行测试。LFW 数据库含有 5 479 个人,共 13 233 张人脸图片。本文的测试内容是基于 LFW 数据库的人脸识别任务,根据文献[12]选取 LFW 闭集作为测试库来进行 Rank-1 准确率以及识别效率的测试。闭集 gallery 库选取 596 个多于两张的人脸样本,每个人选取一张人脸图像,再选取 3 653 个仅只有一张图片的人脸样本,共 4 249 张图片;probe 库选取了与 gallery 库对应的 596 个人脸剩下的样本,共 3 143 张。

3.1.3 数据预处理

本文对所有的训练样本和测试样本都做了相同的预处理。首先,使用 Multi-task Cascaded Convolutional Networks (MTCNN)^[13] 算法对每个样本做人脸检测和 5 个关键点定位(两只眼睛、一个鼻子和两个嘴角);然后根据定位出的 5 个关键点的位置做相似变换;最后将所有人脸裁剪成 112×96 大小的 Red Green Blue (RGB) 图片。

3.2 实验参数设置

本文实验的参数设置与文献[8]中基本一致,训练策略做

$$b_{i,1} = \begin{cases} 1, & W'_i > 0.5\alpha_\Delta^* \\ 0, & W'_i \leq 0.5\alpha_\Delta^* \end{cases} \tag{4}$$

$$b_{i,2} = \begin{cases} 1, & W'_i > -0.5\alpha_\Delta^* \\ 0, & W'_i \leq -0.5\alpha_\Delta^* \end{cases} \tag{5}$$

其中, W'_i 为文献[9]中量化后的三值权重; α_Δ^* 为文献[9]中的缩放因子。

三值权重重编码规则如图 2 所示,根据式(4)和式(5)将每个三值权重转换为一个 2 位的二值编码,再将 16 个三值权重按顺序整合为一个 32 位的二值编码作为最后的模型存储方式,即将 16 个 32 位的浮点型数经过量化后再重新编码成一个 32 位的整型数据,进而大幅压缩模型尺寸。

如下调整:学习率迭代方式先采用 poly 并设置参数 power 为 1,当训练产生过拟合时,提前停止训练;再将学习率降低为原来的 10%,权重衰减提高 1 倍进行训练,若产生过拟合,则提前停止训练;进一步将学习率降低为原来的 50%、权重衰减提高 1 倍进行训练。不断重复此过程,直到网络完全收敛为止。在测试时,不再提取测试图片水平翻转后的特征,只需要提取一次原图的特征进行测试即可。

3.2.1 基于高低维特征拼接的网络收敛性对比实验

该实验以 VGG 为基础框架,分别训练文献[8]中的网络(VGG-P)和根据 2.1 节思想进行改进的网络(VGG-O₁)。由于 VGG 的可调性较高,因此 VGG-O₁ 网络的卷积层 2 保持原通道数目不变。截取两个网络训练的部分结果绘制曲线图,结果如图 3 所示。

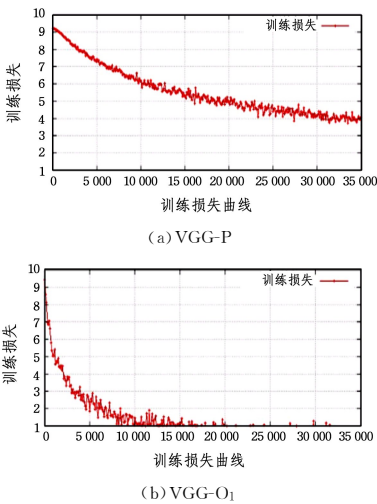


图 3 VGG-P 和 VGG-O₁ 的网络训练曲线图

Fig. 3 Network training curve of VGG-P and VGG-O₁

图 3(a)和图 3(b)分别为 VGG-P 和 VGG-O₁ 网络的训练

损失曲线,纵坐标为网络训练的损失值,横坐标为网络训练迭代次数。从图中可以看出,微调 VGG 时,VGG-O₁ 的损失下降速度远快于 VGG-P,由此可知,高低维特征进行维度拼接后,网络的收敛性更好,同时训练时间成本也节约了92.86%。

3.2.2 基于权重哈希化的模型压缩对比实验

该实验以 VGG-O₁ 为研究对象,根据 2.2 节的思想对 VGG-O₁ 进行再训练,得到 VGG-O₂,实验结果如图 4 所示。相比于 VGG-O₁,VGG-O₂ 模型压缩了约 42%,由此可知,网络权重哈希化再存储的方式可以有效地对模型尺寸进行压缩。

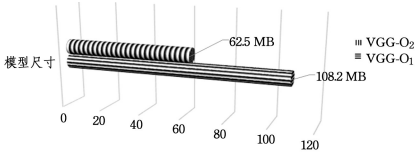


图 4 模型尺寸对比图

Fig. 4 Contrast diagram of model size

3.2.3 LFW 闭集测试结果及分析

识别效率分为 3 个指标,即总效率 S、特征提取效率 S₁ 和相似度比对效率 S₂,三者关系如下:

$$S=S_1+S_2 \tag{6}$$

其中,S 为测试数据库所有图片从输入模型到输出识别结果所耗费的时长;S₁ 为图片从输入模型到提取出所需特征所耗费的时长;S₂ 为提取出的特征进行相似度计算并输出识别结果所耗费的时长。本文实验均以单张图片的识别效率为比较对象,计算公式如下:

$$S_1'=\frac{S_1}{7392} \tag{7}$$

$$S_2'=\frac{S_2}{3143} \tag{8}$$

其中,7392 为 LFW 闭集测试库中的图片总数量;3143 为 LFW 闭集测试库中 gallery 库的图片数量。

该实验以 VGG 为基础框架,结合 2.1 节和 2.2 节的思想对公开的 VGG 人脸模型进行微调,并对训练出的模型进行测试,实验结果如表 1 所列。

表 1 本文方法与其他方法在 LFW 闭集上的对比结果
Table 1 Comparison of proposed method and other methods on closed-set of LFW

方法	Rank-1 准确率/%	总耗时/ s	单层特征 提取耗时/ (ms/张)	相似度 计算耗时/ (ms/张)	模型尺寸/ MB
DeepFace ^[1]	64.90	—	—	—	—
WTS Fusion ^[2]	82.50	—	—	—	—
VGG ^[3]	83.69	1014.74	16.83	284.18	710.3
VGG-P ^[8]	84.95	326.99	10.24	56.04	64.9
本文方法	85.36	326.52	10.07	55.63	62.5

其中 DeepFace 和 WTS Fusion 是对应文献上的结果。从表中可以看出,本文方法在准确率上优于其他 4 种方法,与 VGG-P 相比,在效率相当的情况下,本文方法的能够得到更高的准确率及更小的内存占用率;与基础框架 VGG 相比,本

文方法总识别效率提高约 68%,模型压缩了约 91.2%,同时识别准确率提高了 1.67%。从实验结果可知,本文方法可以有效地提高网络的泛化能力,使其得到的模型具有更快的识别效率和更高的准确率,同时减少内存占用率。

3.3 扩展实验

为了验证本文方法是否具有延展性,将本文方法移植到 Sphreface^[13] 网络中进行扩展实验。该实验以 Sphreface 网络为基础框架,根据图 1 中的模块 A 对网络重新设计,新设计的网络为 Hash-Sphreface,按照 3.2 节设置实验参数并进行训练和测试。

3.3.1 卷积层 2 通道数目对比实验

为确定 Hash-Sphreface 网络的卷积层 2 的通道数目,该实验分别将卷积层 2 的通道数目设置为不变和原数目的 1/2,1/3,1/4,3/4 进行网络训练并测试,实验结果如表 2 所列。可以看出,当卷积层 2 通道数目设置为原数目的 1/2 时,识别准确率最高,为 95.58%,而当原通道数不变时准确率比通道数设置为其 1/2 时低约 5%;通道数分别设置为原数目的 1/3,1/4,3/4 时,准确率均比通道数设置为 1/2 时有大幅度降低,且在训练过程中发现通道数分别设置为 1/3,1/4 时网络收敛更困难。因此本文将卷积层 2 的通道数目设置为原数目的一半。

表 2 卷积层 2 不同通道数目在 LFW 闭集上的测试结果
Table 2 Test results that number of different channels of convolution layer 2 on closed-set of LFW

现卷积层 2 通道数目 与该层原通道数目的比值	Rank-1 准确率/%
1	90.32
1/2	95.58
1/3	83.66
1/4	81.51
3/4	87.93

3.3.2 扩展实验在 LFW 闭集的测试结果及分析

根据 3.3.1 节的实验,将卷积层 2 通道的数目设为原数目的一半,并对 Hash-Sphreface 网络进行训练和准确率及效率的测试,其与 Sphreface 网络的对比结果如表 3 所列。

表 3 扩展实验在 LFW 闭集上的对比结果
Table 3 Comparison of extended experiment on closed-set of LFW

方法	Rank-1 准确率/%	总耗时 /s	单层特征 提取耗时/ (ms/张)	相似度 计算耗时/ (ms/张)	模型尺寸/ MB
Sphreface ^[5]	95.20	320.842	10.21	75.56	112.3
Hash-Sphreface	95.58	125.419	8.48	59.90	53.9

从表 3 中可以看出,Hash-Sphreface 识别准确率比 Sphreface 有略微提升,将识别总耗时从 320.842 s 降低到 125.419 s,因此整体识别效率提升了约 61%;模型尺寸从 112.3 MB 压缩到 53.9 MB,因此模型尺寸压缩了 50%。从实验可知,本文方法具有延展性,因为本文方法是基于现有网络进行改进的,并未对卷积层部分做过多的改进,只针对特征提取层及分类层做了重新设计,所以该方法基于现有网络进行

改进时并不会对卷积层部分造成太大的影响;同时本文基于 VGG 和 Sphreface 两个框架做了实验验证,本文方法可有效应用于两个网络,由此可知,本文方法亦可有效适用于 Alex-net,ResNet,Face-ResNet 等网络。

结束语 本文提出了一种基于权重哈希化的深度人脸识别算法。首先提出了一种基于高低维特征维度拼接的全卷积深度哈希网络,在已经融合有深度哈希的全卷积神经网络中,将融合全局与局部池化的高维特征和近似哈希编码进行维度拼接并作为分类层的输入;然后针对可调性不同的网络对特征层前面的卷积层通道数进行重新设计,改进后的网络在以 VGG 为基础框架的实验中,可获得比 VGG-P 更高的识别准确率,而其需要的训练时间成本仅为 VGG-P 的 1/10;同时提出了一种基于权重哈希化的模型压缩方法,将浮点型权重进行三值量化后再将 16 个三值权重重编码为一个 32 位的哈希编码作为模型的存储格式,在全卷积深度哈希网络中,其内存占用率为原网络的 1/2。在 LFW 库 1:N 人脸识别任务上的两个实验结果证明,本文方法可以提取更有效的特征,达到识别准确率与效率同步提升且内存占用率大幅降低的目的。该方法具有延展性,可适用于其他网络框架,满足实际场景的使用。

参 考 文 献

[1] 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. Columbus:IEEE Computer Society,2014:1701-1708.

[2] TAIGMAN Y,YANG M,RANZATO M,et al. Web-scale training for face identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston:IEEE,2015:2746-2754.

[3] PARKHI O M,VEDALDI A,ZISSERMAN A. Deep Face Recognition[C]// British Machine Vision Conference. Swansea:

BMVC,2015:41. 1-41. 12.

[4] WEN Y,ZHANG K,LI Z,et al. A Discriminative Feature Learning Approach for Deep Face Recognition[M]// Computer Vision-ECCV 2016. Amsterdam: Springer International Publishing,2016:499-515.

[5] LIU W,WEN Y,YU Z,et al. SphereFace: Deep Hypersphere Embedding for Face Recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition. Hawaii:IEEE Computer Society,2017:6738-6746.

[6] DENG J,GUO J,ZAFEIRIOU S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition [OL]. <http://arxiv.org/abs/1801.07698>.

[7] HE K,ZHANG X,REN S,et al. Deep Residual Learning for Image Recognition[C]// Computer Vision and Pattern Recognition. Las Vegas:IEEE,2016:770-778.

[8] 曾燕,陈岳林,蔡晓东. 结合全局与局部池化的深度哈希人脸识别算法 [OL]. <http://www.cnki.com.cn/Article/CJFDTotalsXDKD20180320000.htm>.

[9] LI F,ZHANG B,LIU B. Ternary Weight Networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[10] YI D,LEI Z,LIAO S,et al. Learning Face Representation from Scratch[C]//IEEE Conference on Computer Vision and Pattern Recognition. Columbus:IEEE Computer Society,2014.

[11] HUANG G B,RAMESH M,BERG T,et al. Labeled faces in the wild:A database for studying face recognition in unconstrained environments:07-49 [R]. Amherst: Technical Report,2007.

[12] BEAT-ROWDEN L,HAN H,KLARE B F,et al. Unconstrained Face Recognition: Identifying a Person of Interest From a Media Collection[J]. IEEE Transactions on Information Forensics & Security,2014,9(12):2144-2157.

[13] ZHANG K,ZHANG Z,LI Z,et al. Joint Face Detection and Alignment Using Multi-task Cascaded Convolutional Networks [J]. IEEE Signal Processing Letters,2016,23(10):1499-1503.