文章编号:1672－9331(2017)02－0085－07

基于 KL 距离的卷积神经网络人脸特征提取模型

罗 可,周安众

(长沙理工大学 计算机与通信工程学院,湖南 长沙 410004)

摘 要:为了克服欧式距离的度量方法在人脸特征表达上的不足,提出了一种基于 KL 距离的卷积神经网络人脸特征提取模型。通过卷积神经网络将输入样本转换为一个概率分布,利用 KL 距离度量不同样本之间概率分布的差异,并定义了一个代价函数对此距离进行优化,最后使用反向传播算法修改卷积神经网络的参数,使网络对人脸特征有更强的区分能力。将提取的特征向量通过神经网络分类器进行人脸验证,在 YouTube 等人脸库上进行了测试。试验结果表明,该方法不仅能提高正确率,而且还具有更好的泛化性能。

关键词:人脸识别;人脸验证;特征提取;*KL* 距离;度量学习;卷积神经网络

中图分类号:TP301 **文献标识码**:A

Face feature extraction model of convolutional neural network based on KL divergence

LUO Ke, ZHOU An-zhong

(School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410004, China)

Abstract: In order to overcome the shortcomings of Euclidean distance measurement in face feature expression, a neural network face feature extraction model based on KL divergence is proposed. The convolution neural network is used to transform the input sample into a probability distribution. The distance between different samples is measured by the KL divergence, and a cost function is defined to optimize the distance. The back propagation algorithm is used to modify the parameters of convolution neural network, the network has a stronger ability to distinguish between facial features. The extracted face feature vector is transformed into neural network classifier to performs face validation with YouTube face database. The experimental results show that the method can not only improve the error rate but also improve the generalization performance.

Key words: face recognition; face verification; feature extraction; KL divergence; metric learning; convolutional neural network

人脸识别作为一项重要的生物识别技术,在安全领域有着重要的实际用处,是模式识别和机器视觉领域的研究热点,而人脸特征提取则是整个人脸识别环节中重要的一环。由于人脸样本采集环境通常是在不可控的自然环境下进行的,人脸样本常常含有诸如光照、姿态、遮挡、表情等变

收稿日期:2017－04－20

基金项目:国家自然科学基金资助项目(11671125,71371065)

作者简介:罗 可(1961－),男,湖南长沙人,长沙理工大学教授,博士,主要从事数据挖掘和计算机应用等方面的研究。

化,如何在含有较大干扰的样本中提取鲁棒的特征向量,便成了当前许多研究工作者所关心的问题。受线性判别分析(Linear Discriminant Analysis,简称LDA)[1]的启发,文献[2-4]将人脸图像映射到某一特征向量上,使用距离来度量向量之间的相似度,让同类人脸向量的距离彼此接近,而不同类人脸向量的距离相互远离,取得了较好的效果。这种基于距离来度量向量之间相似性的方法在人脸特征提取中得到了广泛的应用。但是,由于之前浅层特征学习模型[5]不能很好地对人脸高维、非线性特点进行提取,因此难以通过距离度量的方式表达类间和类内差异。

近几年来,随着卷积神经网络(Convolutional Neural Network,简称CNN)的兴起,为人脸特征提取提供了强大的模型表达能力,也使得基于距离度量的方法在人脸特征提取中再次得到重视。文献[6]根据欧式距离判断两个人脸特征向量的相似度,在样本数量较多但样本类别较少的情况下,具有不错的效果。在此基础上,一系列基于度量学习的CNN人脸特征提取模型的研究取得了一定的进展。文献[7]在文献[6]的基础上提出了三元组算法,利用谷歌的超大数据集将三元组算法应用于卷积神经网络的人脸识别,取得了最好的结果。但此方法对正负样本的选择有很大的依赖,直接影响结果收敛的稳定性。文献[8]在CNN的代价函数中同时使用交叉熵和距离度量,每次只比较两个样本,减少了对样本的依赖,但是两种信号的度量方式不一致,需要从不同的层中提取不同尺度的信号。

考虑到CNN在人脸特征提取上的优势以及基于欧式距离的度量学习方法对人脸特征的区分性不够,作者提出一种基于KL距离(Kullback-Leibler Divergence)[9]度量方法的CNN模型。人脸图像作为输入单元以成对的形式通过CNN,在网络的最后一层利用Softmax[10]转换为概率分布,直接判断两个样本概率分布的差异,并将此差异作为需要优化的目标值。首先设计一个CNN提取人脸特征向量,然后使用神经网络分类器进行人脸验证。CNN使用了3种不同的代价函数,分别为交叉熵代价函数、基于欧式距离的代价函数和基于KL距离的代价函数,利用YouTube等人脸库对该方法进行验证,以期在一定程度上证明该方法比基于欧式距离的度量方法有更好的效果。

1 相关基础

1.1 度量学习方法

度量学习是机器学习和计算机视觉领域的一个重要的研究方向。被广泛应用到分类、聚类、检索和身份认证等问题中,并取得了很好的效果。其目的是学习一种合适的距离度量方式,使类内距离更小,类间距离更大,从而达到更好的分类性能。基于度量学习的CNN通过优化某个代价函数,得到一个能有效反映样本空间关系的度量方式。传统的距离度量学习将每一个数据样本表示成高维空间中的一个点,然后在这些点上通过欧式距离度量样本相似性。图1所示为基于欧式距离的度量算法[7]。

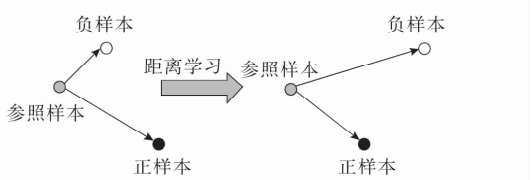

图1 基于距离度量的三元组算法

Fig. 1 Triplet method based on distance metric

根据图1所示,正样本与参照样本的距离要小于负样本与参照样本的距离,且要相差1个阈值。三元组的距离关系式为:

$$\| x_i^a - x_i^p \|_2^2 + \alpha < \| x_i^a - x_i^n \|_2^2,$$
$$\forall (x_i^a, x_i^p, x_i^n) \in T。 \quad (1)$$

式中:α是预先定义的阈值;x_i^a为参照样本;x_i^p与x_i^a为同类样本;x_i^n与x_i^a为异类样本。

为了减小类内距离,增大类间距离,代价函数定义为:

$$\sum_{i=1}^{N} [\| f(x_i^a) - f(x_i^p) \|_2^2 - \| f(x_i^a) - f(x_i^n) \|_2^2 + \alpha]。 \quad (2)$$

以式(2)为代价函数时,需要从训练样本中选择所有的三元组作为网络的输入,其中,一个为参照样本,一个为与参照样本同类的正样本,一个为与参照样本不同类的负样本。

1.2 基于度量学习的CNN人脸特征提取框架

CNN是深度学习框架的一种,以其局部权值共享的特殊结构在图像处理方面有着独特的优越

性,其结构更接近于实际的生物神经网络,权值共享降低了网络的复杂性,特别是多维输入向量的图像可以直接输入网络这一特点,避免了特征提取和分类过程中数据重建的复杂度。

基于 CNN 的人脸特征提取框架将人脸图像通过非线性映射,从低层次的局部特征转换为高层次的全局抽象特征。在 CNN 结构的全连接层(FC 层)提取人脸特征向量,最后用提取到的特征向量进行人脸识别或人脸验证等任务。图 2 显示了整个算法框架。

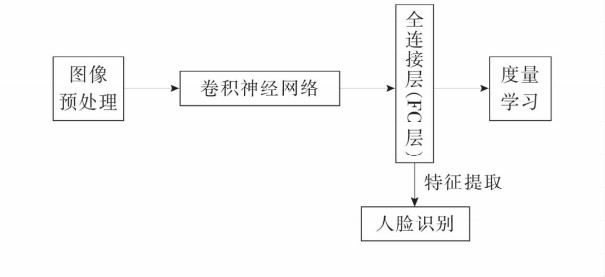

图 2　人脸特征提取框架

Fig. 2　Face feature extraction framework

CNN 通常在最后一层使用 Softmax 激活函数结合交叉熵代价函数进行一个有监督的分类,用于将输入样本分类到样本标记的类别,优化样本与标记类别的距离。随着样本越来越多,样本之间具有更丰富的多样性,类间与类内方差对结果的影响也越明显,于是用于人脸识别的 CNN 的代价函数大多采用了度量学习方法。利用各种变化下的人脸图像训练神经网络产生初始的网络权重,然后利用反向传播算法[11]调整网络权重。使用距离度量人脸图像之间的相似度,目的是减小类内距离及增大类间距离,这样可以使训练后的模型对样本有更好的鉴别能力。

2　基于 KL 距离的度量学习方法

通过度量学习方法与 CNN 的结合,让模型有了更强的区分性。然而,由于 CNN 本身的稀疏性,提取的人脸特征向量是高度冗余的,人脸特征本身又是非线性的,仅仅用欧式距离表达人脸特征向量的差异存在明显的不足,需要一种更合适的度量方式,更好地表达类内和类间差异。

针对以上欧式距离度量方法存在的缺点,考虑到 Softmax 函数输出为一个概率分布,作者提出一种基于 KL 距离的度量学习方法。文献[12]将 KL 距离引入贝叶斯分类器的参数学习中,并取得了更高的分类正确率。KL 距离是一种用于描述 2 个概率分布之间差异性的工具[13],也称为 KL 散度,其具有如下形式:

$$KL(p,q)=\sum_{x\in X}p(x)\times\log\frac{p(x)}{q(x)}。\qquad(3)$$

式中:p 和 q 是概率空间 X 下的两个概率分布;$KL(p,q)$称为分布 p 关于分布 q 的 KL 散度;$KL(q,p)$ 则为分布 q 关于分布 p 的 KL 散度。

由式(3)易知,$KL(p,q)\neq KL(q,p)$,即 KL 散度不满足对称性。对于形如 $KL(p,q)$的 KL 散度而言,称分布 p 为真实分布,分布 q 为分布 p 的近似分布。$KL(p,q)$的取值越大,表明真实分布 p 与近似分布 q 相异度越大,反之则越小。由于 Softmax 激活函数的输出为一个概率分布函数,可以用 KL 距离描述 2 个样本概率分布的差异;又因为是有监督的学习,可以将样本的标记值作为参照样本的概率分布(对于多分类,只有标记项的值为 1,其他项都为 0),这样每次只需选择 2 个样本,如果这两个样本是同类则相互靠近,不同类则相互远离,同时还要减小与参照样本的距离(如图 3 所示)。

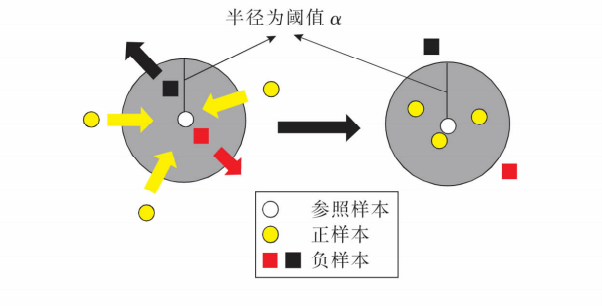

图 3　同类样本互相靠近,异类样本互相远离

Fig. 3　Similar samples close to each other, heterogeneous samples away from each other

训练完成后,提取 FC 层的值作为人脸的特征向量,用于人脸识别、人脸验证等任务。由式(3)可知,KL 距离为非负度量,当且仅当 $p=q$ 时为零,它根据发生概率的加权比来评估 2 个分布的差异性。这里将 KL 距离引入 CNN 的代价函数进行参数优化,每次选取 2 个样本作为 CNN 的输入,在最后一层利用 Softmax 输出 2 个样本的概

率分布 p_i 和 p_j。于是,代价函数为:

$$KL(\hat{p}_i,p_i)+KL(\hat{p}_j,p_j)+\lambda\cdot \mathrm{Div}(p_i,p_j)。 \tag{4}$$

式中:第一项和第二项分别表示2个样本实际输出分布 p_i 和 p_j 在期望分布 $\hat{p}_i$,$\hat{p}_j$ 上的逼近程度;最后一项 $\mathrm{Div}(p_i,p_j)$ 展开为:

$$\mathrm{Div}(p_i,p_j)=\begin{cases}KL(p_i,p_j)+KL(p_j,p_i), & y_{ij}=1\\ \max\{0,m-[KL(p_i,p_j)+KL(p_j,p_i)]\}, & y_{ij}=-1\end{cases}。 \tag{5}$$

式(5)描述了2个类别之间的分离程度。因为KL距离不具有对称性,使用式(5)作为 p_i 和 p_j 之间的KL距离时,又称为对称KL距离[14],λ 是权值系数。优化上述目标函数,不仅要使模型能够尽可能地拟合期望分布,而且需要表达出类别之间的鉴别性。值得注意的是,对代价函数关于输入 x 求梯度时,在式(4)第一项和第二项中,期望概率分布 $\hat{p}_i$,$\hat{p}_j$ 中只有1项输出为1,其他项输出都为0,代入KL距离,式(3)实际上得到的就是交叉熵代价函数:

$$KL(\hat{p}_i,p_i)=-\sum_{i=1}^{n}-\hat{p}_i\log p_i=-\log p_t。 \tag{6}$$

式中:p_t 为Softmax在标记类上的概率值,所以式(4)也可以看作是以交叉熵为识别信号,以 $\mathrm{Div}(p_i,p_j)$ 为验证信号的代价函数。

3 模型的构建

本研究使用CNN结构基于LeNet[15]改进,输出为500个类别,使用KL距离度量学习方法,交叉熵代价函数作为识别信号,KL距离作为验证信号度量2个人脸的相似性。CNN结构如图4所示,激活函数使用Relu[16]函数,图4中为了表达更简洁去掉了Relu层显示。为了对比不同的距离度量方法带来的效果差异,另外训练了2个网络,代价函数分别为交叉熵函数和欧式距离度量函数,总共训练3个网络。

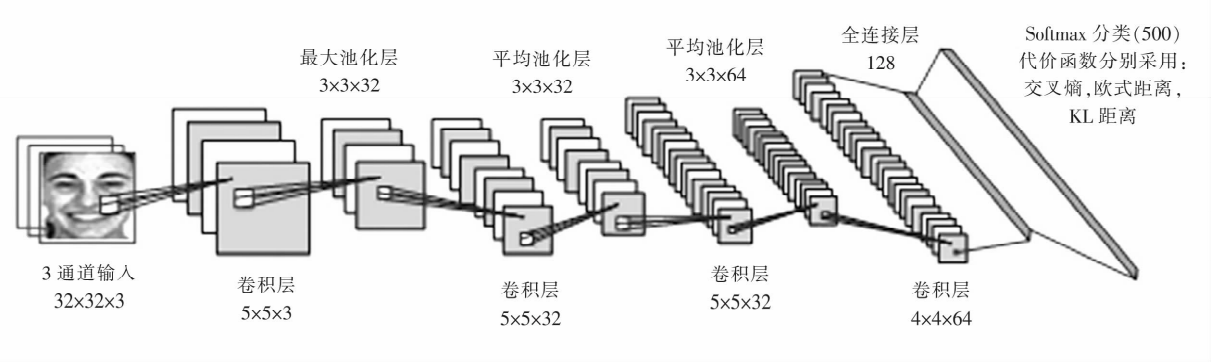

图4 进行人脸特征提取的CNN结构

Fig. 4 The CNN structure for face feature extraction

如图4所示,网络共4层卷积层,第一、三和第五层卷积层后面有池化层。池化层接在卷积层后面用来降低卷积层输出的特征向量维度,同时防止过拟合。池化层的相临区域一般没有重叠,不重叠池化忽视了邻近像素对特征的影响,会造成网络精度的下降,很难迅速收敛。为了获得更多的局部细节,这里采用带重叠的池化层,在一定程度上还可以减轻过拟合。值得注意的是,重叠结构在提高精度的同时,还可能引入噪声,如果对重叠的池化层依旧使用最大池化,那么有更大可能把噪声放大,而使用平均池化[16]就能把引入的噪声减小。于是,在第三、五卷积层后使用了平均池化层,起到降噪的作用。

网络全连接层的128维输出值作为人脸特征向量,可以作为后续人脸验证任务。通过CNN提取了人脸特征向量后,利用二分类的神经网络进行人脸验证。图5展示了训练的三层神经网络结构[17],用来对2个人脸向量进行判断是否为同一个人。输入层同时输入2个人脸特征向量,大小为128×2;输出层为1个节点,大于0代表是同一个人,小于0代表不是同一个人;隐藏层和输出层都使用Tanh激活函数。各层类型和节点数如图5所示。

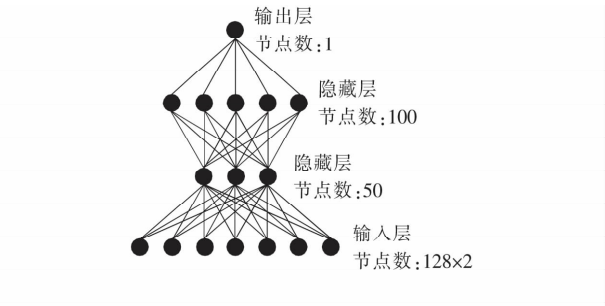

图 5　进行人脸验证的神经网络结构

Fig. 5　The neural network structure for face verification

4　试验结果与分析

4.1　模型训练

模型训练使用了 YouTube[18] 的人脸数据库，YouTube 库包含从 1 595 个人的 3 425 段视频中截取的 60 多万张图片，因为数据量太大，所以从中选择了 500 个人，共 6 万张图片，平均每个人 120 张图片，其中，5.5 万张为训练数据，5 000 张为验证数据。再从 YouTube 数据集中重新选择 100 个人的 5 000 张图片用作测试数据，训练集、验证集和测试集之间没有交集。每张人脸图片都是已对齐的。由于在 YouTube 库中，人脸在图片中所占的比率较小，所以对图片边缘进行了适当裁剪，使人脸所占比重更大，再缩放到 32×32 的大小作为输入。为防止过拟合，模型的 FC 层使用了 Dropout[16]。

图 6 展示了部分 YouTube 人脸库的图片。从图 6 可以看到，YouTube 的人脸图片是在非限制条件下拍摄的，光照、表情、姿态和背景的变化对模型更富有挑战性。

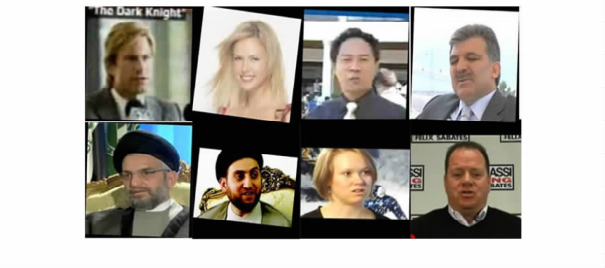

图 6　部分 YouTube 人脸库的图片

Fig. 6　Part of the picture of the YouTube face database

图 7 是经过训练后 3 种网络的 Top1 分类错误率的收敛情况。

图 7 中，使用交叉熵时错误率收敛得最快，Top1 错误率也最低；使用欧式距离和 KL 距离后，错误率有所升高，大约在 30%～40%之间。因为欧式距离对样本敏感，曲线出现抖动的情况，而 KL 距离的曲线整体呈现平稳的下降。在使用度量学习方法后，分类的错误率上升是因为特征向量中包含了更多的冗余信息，后面会看到，虽然不利于分类，但应用于人脸验证任务时泛化性能会有明显的提升。

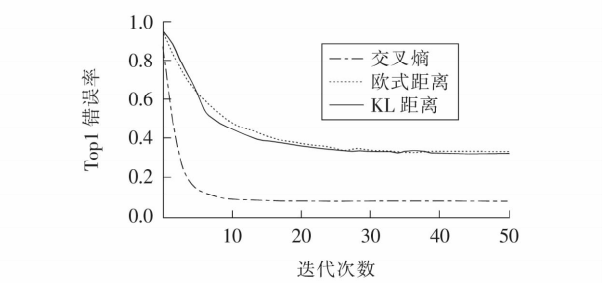

图 7　不同代价函数下验证样本 Top1 错误率

Fig. 7　The Top1 error rate of verify sample under different cost functions

CNN 训练完后，随机选择 2.5 万个训练样本再次输入 CNN，在 FC 层提取它们的特征向量，用来训练神经网络分类器。训练过程如图 8 所示。经过了 380 次迭代，模型达到理想的收敛状态。

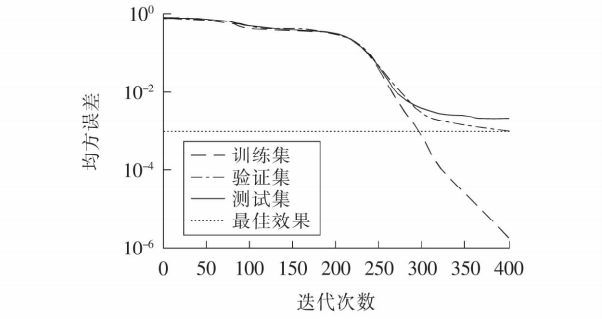

图 8　用于人脸验证的神经网络收敛情况

Fig. 8　The convergence of neural networks for face verification

4.2　人脸验证

将测试集输入 CNN 提取到人脸特征后，再输入以上训练过的神经网络分类器进行人脸验证。为了对泛化性能做更多的对比，另外又挑选了 ORL 人脸库和 AR 人脸库进行同样方式的验证。图 9 展示了 ORL 人脸库与 AR 人脸库的部分图片。ORL 人脸样本姿态和表性的变化相比 AR 库来说都比较小，且 AR 库有大量的遮挡因素（眼镜、围巾），识别会有很大的难度。表 1 对比了不同方法及不同人脸库上的结果。

(a) 部分ORL人脸库 (b) 部分AR人脸库

图9 ORL与AR人脸库图片展示

Fig. 9 The display of ORL and AR face database

表1 不同方法在3种人脸库上的检测结果

Table 1 The test results on three face databases with different methods %

方法	YouTube	ORL	AR
Softmax分类	93.50	87.90	65.69
欧式距离	93.35	90.21	76.03
KL距离(本研究方法)	94.20	91.55	76.50

在YouTube和ORL人脸库上,本研究方法都达到了最好,在AR人脸库上的正确率有明显的下降。这是因为AR库的人脸除了有较大的表情、光照变化外,还有一定数量的人脸遮挡(如图9所示),使识别难度加大。但使用了度量算法尤其是KL距离度量后,正确率比仅用Softmax分类提高了10.81%,可见本研究方法对泛化性能有明显的提升。

4.3 平衡识别信号与验证信号

KL距离代价函数里的λ用来控制验证信号所占的比重,选择适当的λ可以对网络的识别信号即用于分类的交叉熵函数和验证信号即用于优化类内、类间距离的KL距离度量函数进行权衡。增大λ验证信号占主导作用,减小λ识别信号占主导作用。为了研究λ对识别率的影响,调节了λ在不同取值下的人脸检测结果。如图10所示,在0.1附近的取值获得了最高的正确率。

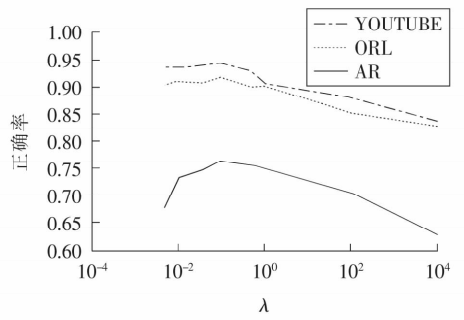

图10 不同λ取值下在3种人脸库上的检测结果

Fig. 10 The results of the test on the three face database with different λ

5 结论

作者将距离度量学习与CNN模型相结合,提出了基于KL距离的人脸识别模型。使用此模型在YouTube,ORL和AR人脸库上进行特征提取,并对提取的特征向量进行了人脸验证,得出如下结论。

1) CNN模型最后一层Softmax激活函数的输出可以看作是样本在概率空间中的一个分布,通过KL距离度量概率分布的差异,达到了减小类内距离、增加类间距离的作用,使模型具有更强的区分能力。

2) 利用神经网络分类器对提取的特征向量进行多次人脸验证,对比欧式距离度量与KL距离度量的试验结果可知,本研究算法对模型的识别正确率和泛化能力都有明显的提升,证明其在人脸识别应用中具有较好的竞争力。

3) 权重系数λ起到平衡识别信号与验证信号比重的作用。减小λ使模型有更好的分类能力,增大λ使模型对未知类别有更强的区分能力。通过对不同λ取值的多次试验,结果显示,λ在取值为0.1时,2种信号达到最优的组合。

〔参考文献〕

[1] Lu G F, Zou J, Wang Y. Incremental complete LDA for face recognition[J]. Pattern Recognition, 2012, 45(7): 2510-2521.

[2] Fernandes S, Bala J. Performanceanalysis of PCA-based and LDA-based algorithms for face recognition[J]. International Journal of Signal Processing Systems, 2013, 1(1): 1-6.

[3] Wu P, Hoi S C H, Zhao P, et al. Online multi-modal distance metric learning with application to image retrieval[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(2): 454-467.

[4] Lu J, Wang G, Deng W, et al. Reconstruction-based metric learning for unconstrained face verification [J]. Information Forensics and Security IEEE Transactions on, 2015, 10(1): 79-89.

[5] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification[J].

Journal of Machine Learning Research,2009,10(2):207-244.

[6] Hoffer E,Ailon N.Deep metric learning using triplet network[A].International Workshop on Similarity-Based Pattern Recognition[C].Dubai:Springer International Publishing,2015:84-92.

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

[8] Sun Y,Chen Y,Wang X,et al.Deep learning face representation by joint identification-verification[A].Advances in neural information processing systems[C].Montreal:NIPS,2014:1988-1996.

[9] Li X Y,Zhang Z G,Zhang Y S,et al.KL-divergence based feature selection algorithm with the separate-class strategy[J].Computer Science,2012,39(12):224-227.

[10] Liu W,Wen Y,Yu Z,et al.Large-margin softmax loss for convolutional neural networks[A].Proceedings of the 33rd International Conference on Machine Learning[C].New York:ICML,2016:507-516.

[11] Fernando T G I.Persons' personality traits recognition using machine learning algorithms and image processing techniques[J].Advances in Computer Science,2016,5(1):40-44.

[12] 冯奇,田凤占,黄厚宽.基于 KL 距离的 TAN 分类器判别性学习方法[J].模式识别与人工智能,2008,21(6):806-811.
FENG Qi,TIAN Feng-zhan,HUANG Hou-kuan.Discriminative learning of TAN classifier based on KL divergence[J].Pattern Recognition and Artificial Intelligence,2008,21(6):806-811.

[13] 沈媛媛,严严,王菡子.有监督的距离度量学习算法研究进展[J].自动化学报,2014,40(12):2673-2686.
SHEN Yuan-yuan,YAN Yan,WANG Han-zi.Recent advances on supervised distance metric learning algorithms[J].Acta Automatica Sinica,2014,40(12):2673-2686.

[14] 姚志均,刘俊涛,周瑜,等.基于对称 KL 距离的相似性度量方法[J].华中科技大学学报:自然科学版,2011,39(11):1-4.
YAO Zhi-jun,LIU Jun-tao,ZHOU Yu,et al.Similarity measure method using symmetric KL divergence[J].Journal of Huazhong University of Science and Technology:Natural Science Edition,2011,39(11):1-4.

[15] Lin S,Cai L,Lin X,et al.Masked face detection via a modified LeNet[J].Neurocomputing,2016(218):197-202.

[16] Krizhevsky A,Sutskever I,Hinton G E.Imagenet classification with deep convolutional neural networks[A].Advances in Neural Information Processing Systems[C].Lake Tahoe:NIPS,2012:1097-1105.

[17] 施徐敢,张石清,赵小明.融合深度信念网络和多层感知器的人脸表情识别[J].小型微型计算机系统,2015,36(7):1629-1632.
SHI Xu-gan,ZHANG Shi-qing,ZHAO Xiao-ming.Facial expression recognition by integrating deep belief networks with multi-layer perceptron[J].Journal of Chinese Computer Systems,2015,36(7):1629-1632.

[18] Wolf L,Hassner T,Maoz I.Face recognition in unconstrained videos with matched background similarity[A].Computer Vision and Pattern Recognition[C].Colorado:Springs,2011:529-534.