

引入感知模型的改进孪生卷积神经网络实现人脸识别算法研究

徐先峰,张丽,郎彬,夏振
(长安大学电子与控制工程学院,陕西西安710064)

摘要: 针对非限定性条件下人脸识别困难问题,设计了一种引入感知模型的改进孪生卷积神经网络结构(Inception Module Incorporated Siamese Convolutional Neural Networks, IMISCNN),在充分利用孪生结构有效减少外界干扰并避免过拟合等优点的基础上,为其增加感知模型实现更丰富特征的提取.为寻找最优学习率引入了循环学习率策略,加速模型收敛.在CASIA-webface和Extended Yale B标准人脸数据库上的仿真实验表明,所提IMISCNN算法提升了人脸识别精度.

关键词: 人脸识别;孪生卷积神经网络;感知模型;循环学习率

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112(2020)04-0643-05

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2020.04.004

Research on Inception Module Incorporated Siamese Convolutional Neural Networks to Realize Face Recognition

XU Xian-feng, ZHANG Li, LANG Bin, XIA Zhen

(College of Electronics and Control Engineering, Chang'an University, Xi'an, Shaanxi 710064, China)

Abstract: To solve the problem of face recognition under unlimited conditions, a simple network structure named Inception Module Incorporated Siamese Convolutional Neural Networks (IMISCNN) was designed, which was suitable for small-scale data sets. On the basis of making full use of the Siamese structure to effectively reduce external interference and avoid over-fitting, inception module was incorporated to the Siamese network to extract richer features. Furthermore, a cyclical learning rate strategy was adopted to accelerate the convergence of the model. Simulation results on the CASIA-webface and Extended Yale B standard face database showed that the recognition accuracy of IMISCNN was significantly improved compared with other face recognition algorithms.

Key words: face recognition; siamese convolutional neural networks; inception module; cyclical learning rate

1 引言

安防意识的提升促使人们对公共及个人安全需求不断攀升,如何准确快速的鉴定个人身份,保护信息安全,成为亟待解决的关键性社会问题.因此,多种生物特征识别技术应运而生^[1,2],人脸识别技术由于其方便、快捷、非侵入性等优点而备受关注,研究成果丰富^[3].

总结经典的人脸识别算法可以发现,主成分分析法(PCA)^[4]通过矩阵变换有效地减少了计算复杂度,但当存在遮挡等外界因素干扰时,无法获取数据真实子空间结构,识别效果变差.支持向量机(SVM)^[5]方法具有较强的泛化能力,且能很好地处理高维度数据,但其

无法适用于具有大规模训练样本的应用场合.随着深度学习技术的快速发展^[6],基于卷积神经网络的人脸识别算法^[7-10]被广泛提出.其中,Chan等人^[11]提出的PCANet将深度学习与传统的特征提取方法建立了联系,对光照变化和遮挡等表现出了很强的鲁棒性^[12].DeepFace^[13]引入了3D人脸模型对有姿态的人脸分片仿射对齐,提高了识别精度.Google的FaceNet^[14]直接学习从图像到欧式空间的编码方法.但DeepFace和FaceNet无论在训练还是使用时都需要相当巨大的计算量和数据集,在小规模数据集上并不适用.

在实际应用中,识别对象往往客观存在着光照、表情、姿态和遮挡等复杂干扰,导致人脸识别较为困难,为

收稿日期:2018-12-09;修回日期:2019-05-13;责任编辑:李勇锋

基金项目:国家自然科学基金(No. 61201407, No. 61473047);陕西省自然科学基金基础研究计划(No. 2016JQ5130);陕西省重点研发计划(No. 2019GY-002);长安大学中央高校基本科研业务费(No. 300102328202);西安市智慧高速公路信息融合与控制重点实验室(No. ZD13CG46)

为了更好地对外界干扰鲁棒,提高数据集的处理速度并解决由于数据集较少而导致的过拟合等问题^[15],本文在深度学习理论上,提出了一种引入感知模型^[16]的改进孪生卷积神经网络结构(Inception Module Incorporated Siamese Convolutional Neural Networks, IMISCNN),并对整个训练过程辅以循环学习率策略优化算法^[17],降低外界噪声干扰,提高网络学习能力,同时使得寻找最优学习率变得容易,提升网络性能,有效的实现了非限定性条件下的人脸识别.利用 CASIA-web-face^[18]和 Extended Yale B^[19] 标准人脸数据库对所提算法进行测试,结果表明所提 IMISCNN 网络结构与经典的基于 PCA、SVM 和基于传统的卷积神经网络的人脸识别算法相比,在识别精度上有大幅提升.

2 网络结构模型及学习率策略

本文所提 IMISCNN 由一对结构相同、权值共享的 CNN 组成的孪生卷积神经网络^[20]构成,与传统结构不同的是,为孪生卷积神经网络引入了 Inception 模型以提取更加丰富的特征,同时网络引入循环学习率^[17]优化策略来加快训练速度.

2.1 网络结构模型

IMISCNN 整体的网络构架如图 1 所示.

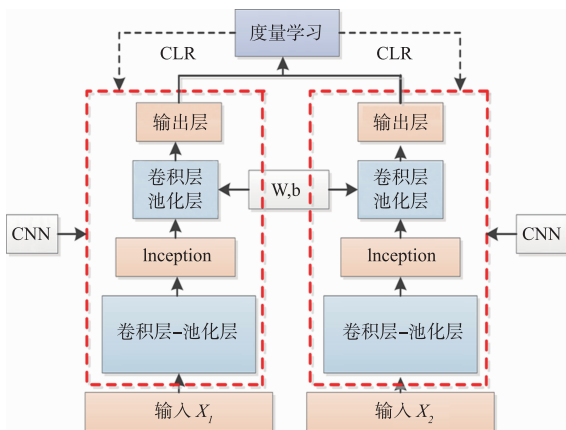


图1 整体网络架构

使用度量学习以衡量输入图像之间的相似性.首先,定义一对输入样本 $(\mathbf{x}_1, \mathbf{x}_2)$,当 \mathbf{x}_1 和 \mathbf{x}_2 属于同一个人时,标签 $y=1$,当 \mathbf{x}_1 和 \mathbf{x}_2 不属于同一人时,标签 $y=0$.将 \mathbf{x}_1 和 \mathbf{x}_2 输入孪生网络结构,寻找参数 \mathbf{W}, \mathbf{b} ,并利用可微的映射函数 $G_{\mathbf{w}}(\mathbf{x})$ 将输入映射到目标空间,在目标空间使用距离参数定义相似性,如式(1)所示.

$$E_{\mathbf{w}}(\mathbf{x}_1, \mathbf{x}_2) = \|G_{\mathbf{w}}(\mathbf{x}_1) - G_{\mathbf{w}}(\mathbf{x}_2)\| \quad (1)$$

给定正样本对 $(\mathbf{x}_1, \mathbf{x}_2)$ 和负样本对 $(\mathbf{x}_1, \mathbf{x}_2')$,期望正样本对的相似性度量 $E_{\mathbf{w}}$ 较小,负样本对的相似性度量 $E_{\mathbf{w}}$ 较大,因此模型需要满足条件:任意正数 $m > 0$,

$$E_{\mathbf{w}}(\mathbf{x}_1, \mathbf{x}_2) + m < E_{\mathbf{w}}(\mathbf{x}_1, \mathbf{x}_2') \quad (2)$$

m 为设定阈值.为了能够对孪生神经网络进行训练,定义可微分的代价函数如式(3)所示:

$$L(\mathbf{W}, (y, \mathbf{x}_1, \mathbf{x}_2)) = (1-y)L_c(E_{\mathbf{w}}(\mathbf{x}_1, \mathbf{x}_2)^i) + yL_l(E_{\mathbf{w}}(\mathbf{x}_1, \mathbf{x}_2)^i) \quad (3)$$

其中, L_c 和 L_l 分别为只计算相同类别对图片和只计算不同类别对图片的损失函数, $(y, \mathbf{x}_1, \mathbf{x}_2)^i$ 表示第 i 个样本.进一步定义损失函数如式(4)所示:

$$L(\mathbf{W}, (y, \mathbf{x}_1, \mathbf{x}_2)) = \frac{1}{2N} \cdot \sum_{n=1}^N (yE_{\mathbf{w}}^2 + (1-y)\max(m - E_{\mathbf{w}}, 0)^2) \quad (4)$$

N 为训练的样本数.

当 $y=0$,损失为:

$$L(\mathbf{W}, (y, \mathbf{x}_1, \mathbf{x}_2)) = \frac{1}{2N} \sum_{n=1}^N \max(m - E_{\mathbf{w}}, 0)^2 \quad (5)$$

即原本不相似的样本,如果距离度量小于阈值,损失减小,则 $E_{\mathbf{w}}$ 增大,扩大图像距离.

当 $y=1$,损失为:

$$L(\mathbf{W}, (y, \mathbf{x}_1, \mathbf{x}_2)) = \frac{1}{2N} \sum_{n=1}^N E_{\mathbf{w}}^2 \quad (6)$$

即原本相似的样本,损失减小,则 $E_{\mathbf{w}}$ 减小,图像距离进一步缩小.

显然,度量学习能够减少复杂干扰造成的相同身份人脸特征向量之间的差异,并增大不同身份人脸特征向量间的差异,从大量的训练人脸特征向量中,学习该特征空间的更鲁棒、更具可区分的距离度量,将原本在原始空间中分辨困难的数据进行维度规约,降低干扰影响,提升识别精度.

2.2 循环学习率策略

循环学习率^[17](Cyclical Learning Rate, CLR)是一种有效的学习率调整策略,其具体优化过程分为两步:学习率循环范围的确定和在此范围内学习率的具体调整.

在正式训练模型之前需要对样本进行一次预先训练以确定学习率的循环变化范围^[17].根据经验给定一个非常小的学习率初始值,在训练过程的每个批次中使学习率在初基础上以非始学习率的常低的速率线性增大,绘制损失值关于学习率的变化曲线,损失值下降最快的区域所对应的学习率范围为最优学习率的循环变化范围,其范围的上下界分别称为最大边界(max_lr)和最小边界(min_lr)^[17].

在已确定的学习率范围内进行相应调整策略优化.如图2所示,使学习率在已确定的边界内随迭代次数(iterations)的增加周期性变化,选择2个步长(s)为一个循环周期.学习率(Learning Rate, LR)从最小边界开始,在设定的第一个迭代步长中由初始值逐渐上升至最大边界,在下一个迭代步长中学习率由最大边界

下降至最小边界,依次循环往复,完成训练.

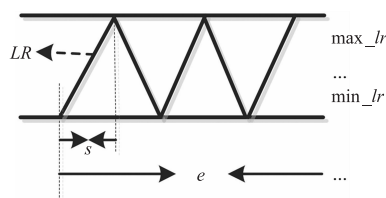


图2 学习率策略

训练时,每经过一个批(batch)学习率更新一次,其更新的具体值由式(7)计算.

$$LR = \min_lr + (\max_lr - \min_lr) \cdot f(0, (1 - x)) \quad (7)$$

其中, f 函数表示在两个变量中取较大值, $x = f_1(\frac{b}{s} - 2$

$\cdot c + 1)$, $c = f_2(1 + \frac{b}{2 \cdot s})$, f_1 为取绝对值函数, f_2 为对浮点型数据向下取整函数, s 为步长,表示半个循环周期,一般设定为 $(2 \sim 10) \cdot epoch$,其中 $epoch$ 的大小等于使用训练集中的全部样本除以批尺寸($batch_size$), b 为变量,在 $[0, 2s]$ 范围内变化,表示此次训练为第 b 个 batch,在训练迭代过程中通过 LR 值的计算对学习率更新从而提高识别准确率,加速模型收敛. IMISCNN 算法学习率具体参数选择将在后文给出.

3 仿真分析

3.1 实验数据集

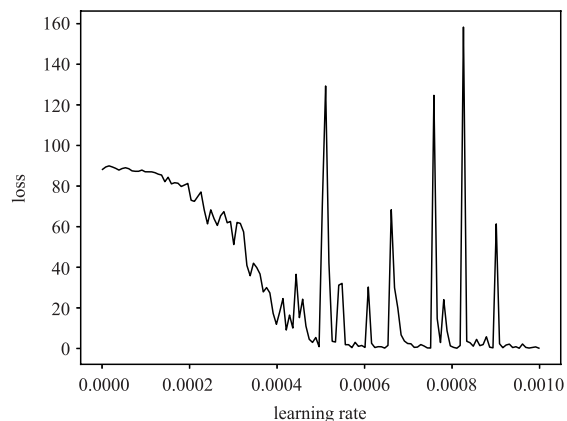
CASIA-webface 数据集:由于本文方法针对小规模数据集,同时为验证其对混合干扰有效,选择两组数据集以对比结果:(1)随机选取 335 个不同人物,每人不超过 30 张图像,每张图像只含有轻微光照干扰,不含实物遮挡及姿势遮挡等其他干扰.(2)同样随机选取 335 个不同人物,每人不超过 30 张图像,其中随机包含实物遮挡及姿态变化等干扰但不包括极端遮挡.为方便表示,分别记数据集(1)为数据集 CASIA-A,数据集(2)为数据集 CASIA-B.

3.2 实验过程与结果分析

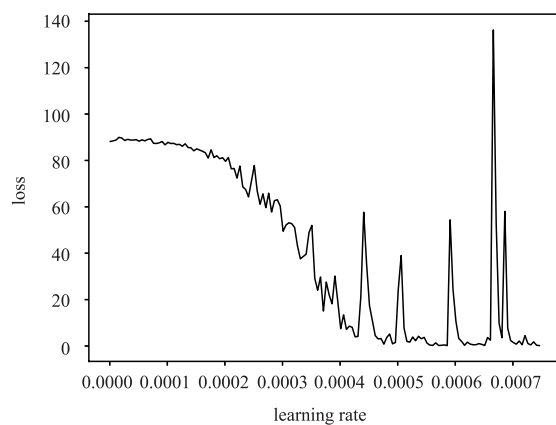
本文根据不同的结构进行对比试验确定网络卷积层的层数为 5 层时,网络结构识别率最高;当卷积层数不变,Inception 个数为 1 时网络结构识别率最高,损失收敛最低,因此确定网络整体结构为 5 层卷积层、5 层池化层、1 层 Inception 层以及 1 层全连接层.为方便后文讨论,分别将不添加、添加 Inception 模型的网络结构记为 Conv_和 Conv_IN.

确定学习率范围.按照经验设置学习率初始值为 $1e-6$,在训练过程的每个批次中使学习率以 $1e-5$ 的增长速率线性增大,分别记录 CASIA-B 数据集下 Conv_结构和 Conv_IN 结构的学习率与损失值并绘制其关系

图,如图 3 所示.



(a) Conv_结构下损失随学习率变化曲线



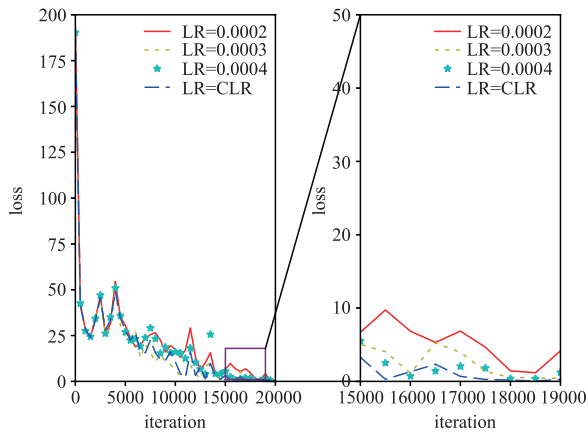
(b) Conv_IN结构下损失随学习率变化曲线

图3

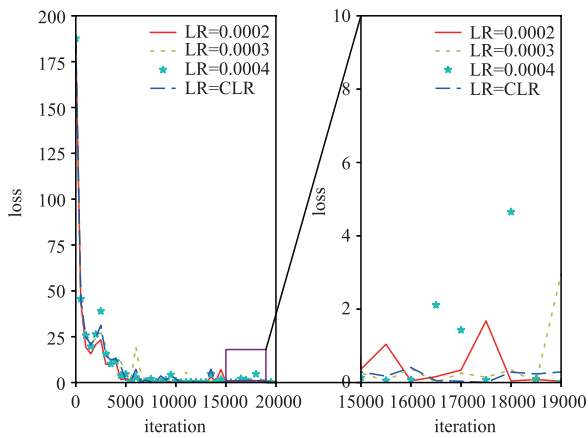
分析图 3 可知,当学习率线性增大时损失值会随之减少,一旦学习率增大到一定程度损失值反而会上升,因此学习率周期性变化的范围选择为相应损失值快速下降的区间^[17],对于 CASIA-B 数据集,Conv_结构和 Conv_IN 结构均选择学习率下界 \min_lr 为 0.0002,学习率上界 \max_lr 为 0.0004.除此之外,还需要设置参数步长 s ,文献[18]认为步长一般为 $(2 \sim 10) \cdot iterations$,文献[19]通过实验表明,stepsize 设成 2 倍或者 10 倍的 iterations,两者结果并没有太大的不同.因此在本文中选取 2 和 10 的中间数 6,设置为 6 倍的 iterations.

CASIA-B 数据集下的损失收敛如图 4 所示.

对于 CASIA-B 数据集,在大约前 5000 次的迭代中,固定学习率与循环学习率策略对损失收敛情况表现大致相同,而在 5000 次迭代之后,循环学习率策略相较于固定学习率在迭代次数相同的情况下具有更低的损失值(见局部放大图),且训练样本的损失降到相同值时,Conv_IN 结构只需要训练 5000 次左右,而 Conv_结构需要训练 15000 次,即循环学习率策略使模型收敛到更低的值,且加快了模型的收敛速度.同时,比较相同迭代次数 10000



(a) 不同学习率下Conv_结构损失收敛及局部放大



(b) 不同学习率下Conv_IN结构损失收敛及局部放大

图4

次下不同学习率的识别准确率,如表1所示。

表1 数据集 CASIA-B 下相同迭代次数下识别精度比较

LR	test accuracy/% (Conv_)	test accuracy/% (Conv_IN)
0.0002	97.32	97.85
0.0003	97.41	98.03
0.0004	97.27	97.56
CLR	97.65	98.89

表1表明,当迭代次数相同时,循环学习率策略明显有更高的识别准确率,即当达到同一准确率时,循环学习率策略显然需要更少的迭代次数。本文验证了寻找最优学习率方法的有效性,经过较少的全局循环就可以为模型找到最优的学习率,不再需要人工进行多次尝试,且对比固定学习率,循环学习率使网络具有更好的收敛性。

在 CASIA-B 人脸数据集下经过 20000 次迭代后,最终的识别结果如表2所示。

由表2可知,在 CNN 其余结构不变,仅添加 Inception 结构的情况下,在 CASIA-webface 标准人脸数据库

上识别率均有明显的提升,其中,Conv_IN 结构使用循环学习率策略识别准确率达到到了 99.25%,表现出优异的识别效果。

表2 数据集 CASIA-B 下不同学习率识别结果

neural network structure	recognition rate/%			
	LR = 0.0002	LR = 0.0003	LR = 0.0004	LR = CLR
Conv_	99.02	98.72	99.18	99.13
Conv_IN	99.11	99.01	99.14	99.25

3.3 IMISCNN 与经典算法比较结果分析

为了进一步表明 IMISCNN 算法的泛化性能,将上述基于 CASIA-webface 数据集确定的相同网络结构参数直接应用于 Extended Yale B 标准人脸数据库,并与其他经典 PCA、PCA_SVM、CNN、LBP + DBN^[21] 和 NN + PCANet^[12] 算法进行比较。最终实验结果如表3所示。

表3 不同算法识别准确率比较

methods	recognition rate /%		
	CASIA-A	CASIA-B	Extended Yale B
PCA	60.79	51.13	67.25
PCA_SVM	72.36	68.23	75.92
CNN	84.55	80.68	80.09
LBP + DBN	92.36	90.86	96.06
NN + PCANet	99.22	99.16	99.36
IMISCNN	99.36	99.39	99.21

分析表3可知,PCA 算法并不是通过训练来提取特征,因此本文算法相较于 PCA 算法在识别正确率上有较大提升;SVM 的核函数及其参数需人为选取,引起算法低效;CNN 算法数据集规模较小,识别率较低;LBP + DBN 算法,能够克服 DBN 不能学习到人脸图像局部结构特征的缺点^[21],但当图像缺乏纹理特征时,其识别性能减弱;NN + PCANet 具有高精度的识别准确率,但其并不擅长处理缺乏纹理变化的强光遮挡。

IMISCNN 算法在三大数据集所达到的高精度识别率,充分验证了其所具有的强泛化能力。

4 结束语

本文提出了将感知模型和孪生神经网络相结合的架构,并引入循环学习率策略,使得网络收敛速度加快,降低了相同识别率所需要的迭代次数。同时,所提网络结构简单,适用于非限定性条件下的小规模数据集的人脸识别,在 CASIA-webface 和 Extended Yale B 标准人脸数据库进行了训练与测试,能够达到较高的准确率,是一种性能优异的人脸识别算法。

参考文献

- [1] Sundararajan K, Woodard D L. Deep learning for biometrics: A survey [J]. ACM Computing Surveys (CSUR), 2018, 51(3): 65.
- [2] Nguyen K, Fookes C, Sridharan S, et al. Super-resolution for biometrics: A comprehensive survey [J]. Pattern Recognition, 2018, 78: 23 – 42.
- [3] Takalkar M, Xu M, Wu Q, et al. A survey: facial micro-expression recognition [J]. Multimedia Tools and Applications, 2018, 77(15): 19301 – 19325.
- [4] Sharma R, Patterh M S. A new hybrid approach using PCA for pose invariant face recognition [J]. Wireless Personal Communications, 2015, 85(3): 1561 – 1571.
- [5] Tao Q Q, Zhan S, Li X H, et al. Robust face detection using local CNN and SVM based on kernel combination [J]. Neurocomputing, 2016, 211: 98 – 105.
- [6] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436 – 444.
- [7] Chen G, Shao Y, Tang C, et al. Deep transformation learning for face recognition in the unconstrained scene [J]. Machine Vision and Applications, 2018, 29(3): 513 – 523.
- [8] Atoum Y, Liu Y, Jourabloo A, et al. Face anti-spoofing using patch and depth-based cnns [A]. 2017 IEEE International Joint Conference on Biometrics (IJCB) [C]. IEEE, 2017. 319 – 328.
- [9] Bao J, Chen D, Wen F, et al. Towards open-set identity preserving face synthesis [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. IEEE, 2018. 6713 – 6722.
- [10] 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 Qian-yu, JIANG Jian-guo, QI Mei-bin. Face recognition algorithm based on improved deep networks [J]. Acta Electronica Sinica, 2017, 45(3): 619 – 625. (in Chinese)
- [11] Chan T H, et al. PCANet: A simple deep learning baseline for image classification? [J]. IEEE Transactions on Image Processing, 2015, 24(12): 5017 – 5032.
- [12] 李小薪, 梁荣华. 有遮挡人脸识别综述: 从子空间回归到深度学习 [J]. 计算机学报, 2018(1): 177 – 207.
LI Xiao-xin, LIANG Rong-hua. A review for face recognition with occlusion: from subspace regression to deep learning [J]. Chinese Journal of Computers, 2018(1): 177 – 207. (in Chinese)
- [13] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. IEEE, 2014. 1701 – 1708.
- [14] 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]. IEEE, 2015. 815 – 823.
- [15] 吴长虹, 苏剑波, 陈叶飞. 抗年龄干扰的人脸识别 [J]. 电子学报, 2018, 46(7): 1593 – 1600.
WU Chang-hong, SU Jian-bo, CHEN Ye-fei. Age invariant face recognition [J]. Acta Electronica Sinica, 2018, 46(7): 1593 – 1600. (in Chinese)
- [16] Szegedy C, Liu W, et al. Going deeper with convolutions [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. IEEE, 2015. 1 – 9.
- [17] Smith L N. Cyclical learning rates for training neural networks [J]. Computer Science, 2015: 464 – 472.
- [18] Yi D, Lei Z, Liao S, et al. Learning face representation from scratch [J]. arXiv preprint arXiv:1411.7923, 2014.
- [19] Lee K C, Ho J, Kriegman D J. Acquiring linear subspaces for face recognition under variable lighting [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, (5): 684 – 698.
- [20] Bromley J, Bentz J W, Bottou L, et al. Signature verification using a “SIAMESE” time delay neural network [A]. International Conference on Neural Information Processing Systems [C]. Morgan Kaufmann Publishers Inc, 1993. 737 – 744.
- [21] 梁淑芬, 刘银华, 李立琛. 基于 LBP 和深度学习的非限制条件下人脸识别算法 [J]. 通信学报, 2014, 35(6): 154 – 160.
LIANG Shu-fen, LIU Yin-hua, LI Li-chen. Face recognition under unconstrained based on LBP and deep learning [J]. Journal on Communications, 2014, 35(6): 154 – 160. (in Chinese)

作者简介



徐先峰 男, 1982 年 3 月出生, 山东宁阳人。博士, 副教授, 硕士生导师, IEEE 会员。2004 年和 2010 年分别在哈尔滨工程大学和西安电子科技大学获工学学士和工学博士学位, 2018 年为美国加州大学洛杉矶分校 (UCLA) 访问学者。主要研究方向: 信号处理、深度学习算法及应用、机械系统状态监测与故障诊断、智能交通信息挖掘、智能电网等。

E-mail: xuxianfeng1982@163.com



张丽 女, 1993 年 7 月出生于陕西。长安大学电子与控制工程学院在读研究生, 研究方向为人脸识别、图像处理。