计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2020,29(2):157−162 [doi: 10.15888/j.cnki.csa.007312]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

# 基于多 Inception 结构的卷积神经网络人脸识别算法①

李　楠[1], 蔡坚勇[1,2,3,4], 李　科[1], 程　玉[1], 张明伟[1]

[1](福建师范大学 光电与信息工程学院, 福州 350007)
[2](福建师范大学 医学光电科学与技术教育部重点实验室, 福州 350007)
[3](福建师范大学 福建省光子技术重点实验室, 福州 350007)
[4](福建师范大学 福建省光电传感应用工程技术研究中心, 福州 350007)
通讯作者: 蔡坚勇, E-mail: cjy@fjnu.edu.cn

**摘　要**: 人脸识别是视觉识别的一个重要领域, 由于人脸识别尺度变化范围大, 光照、姿态变化剧烈以及遮挡问题, 导致该类非限制条件下的识别难度较大, 为了解决该类问题, 本文提出了一种基于 Tensorflow 平台的多 Inception 模型, 通过将多个 Inception 结构进行串联, 再通过分解卷积核的方式减少输入参数, 实现了多维度同时卷积再聚合, 提高了人脸识别的精度. 实验结果表明, 该方法在较少参数的条件下能提取出更具区分度的人脸特征, 与分类损失方法及融合了其他度量学习方式的方法相比, 提高了识别准确率, 减少了计算时间.

**关键词**: 人脸识别; Tensorflow; Inception; 卷积神经网络

引用格式: 李楠,蔡坚勇,李科,程玉,张明伟.基于多 Inception 结构的卷积神经网络人脸识别算法.计算机系统应用,2020,29(2):157–162. http://www.c-s-a.org.cn/1003-3254/7312.html

## Face Recognition Algorithms Based on Convolutional Neural Network with Multi-Inception Structure

LI Nan[1], CAI Jian-Yong[1,2,3,4], LI Ke[1], CHENG Yu[1], ZHANG Ming-Wei[1]

[1](College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350007, China)
[2](Key Laboratory of Optoelectronic Science and Technology for Medicine (Ministry of Education), Fujian Normal University, Fuzhou 350007, China)
[3](Fujian Provincial Key Laboratory of Photonics Technology, Fujian Normal University, Fuzhou 350007, China)
[4](Fujian Provincial Engineering Technology Research Center of Photoelectric Sensing Application, Fujian Normal University, Fuzhou 350007, China)

**Abstract**: Face recognition is an important field of visual recognition. Because of the large scale of variations in face recognition, namely drastic changes in illumination and pose, occlusion problems, and complex image background, it is difficult to recognize the face under such unrestricted conditions. In order to solve these problems, a multi-Inception model based on Tensorflow platform is proposed in this study. By combining multiple Inception knots, a multi-Inception-V3 model based on Tensorflow platform is proposed. The structure is connected in series, which realizes the convolution and re-aggregation of multiple dimensions at the same time, and improves the accuracy of face recognition. The experimental results show that the proposed method can extract more discriminant face features with fewer parameters. Compared with the classification loss method and the fusion of other metric learning methods, it improves the accuracy of face recognition under unconstrained conditions.

**Key words**: face recognition; Tensorflow; Inception; convolution neural network

① 收稿时间: 2019-06-18; 修改时间: 2019-07-12, 2019-09-03; 采用时间: 2019-09-08; csa 在线出版时间: 2020-01-16

随着卷积神经网络 (CNN) 在视觉识别领域的广泛应用, 人脸识别领域在计算时间, 识别准确率等方面取得了明显的进展. 相对于传统的模式识别算法, 基于卷积神经网络的人脸识别方法具有准确率更高, 输入参数更少, 细节识别能力更强等优势[1]. 大部分卷积神经网络都采用分类损失函数来衡量预测值和实际值的差距, 再通过训练过程完成图像的分类从而扩大不同类别图像的距离. Taigman 等利用 3 维人脸信息作为特征信息, 通过大量数据训练, 提升了算法的鲁棒性及精度[2]. Sun 等利用 DeepID 用于人脸识别, 通过将人脸不同部位进行分区, 分别进行特征提取, 再使用贝叶斯算法对特征进行复合运算, 最终得到人脸特征信息, 有效提升了识别准确度[3]. 但是上述算法都没有解决非限制条件下的识别问题, 即在不同环境下, 人脸识别率会明显降低. 因此在识别过程中如何增大类间距离的同时减少类内距离, 是人脸识别的重要课题. Peri 课题组通过加入一个验证损失的方法, 实现训练过程中损失函数的反馈, 利用训练过程中生成的正样本来减少类间距离, 但是该方法较为依赖样本, 对训练参数的设置要求较高, 在训练数据集有限的情况下容易出现过拟合[4]. Schroff 及其课题组提出了一种三元损失算法, 将训练数据统一为三元组元素, 每个三元组都包含正值、负值和样本锚点, 该方法可以有效减少类内距离[5].

上述方法虽能解决部分非限制性问题, 但在收敛速度上性能较差, 特别是当网络层数太多时, 会出现梯度弥散现象. 为了解决这类问题, 本文提出基于多 Inception 结构的卷积网络神经算法用于人脸识别, 通过改造传统的 SoftmaxLoss 方法, 结合 Softmax 和 TripletLoss 可以获得更大的类间距离和更小的类内距离. 实验证明本文提出的算法在增加网络深度和宽度的同时减少了参数个数, 在训练过程中能有效减少类内间距, 在同等条件下能获取更高的特征提取能力.

## 1 基于 Inception 结构的卷积神经网络

### 1.1 卷积神经网络

CNN (Convolutional Neural Network) 是一个多层次结构的神经网络[6], 通常由输入、特征提取层 (多层) 以及分类器组成, 每层都有多个二维独立神经元. CNN 网络通过逐层的特征提取来提升特征准确度, 最后将其输入到分类器中对结果进行分类. 卷积层是 CNN 的特征映射层, 具有局部连接和权值共享的特征. 这两种特征降低了模型的复杂度, 并使参数数量大幅减少. 下采样 (池化) 层是 CNN 的特征提取层, 它将输入中的连续范围作为池化区域, 并且只对重复的隐藏单元输出特征进行池化, 该操作使 CNN 具有平移不变性. 实际上每个用来求局部平均和二次提取的卷积层后都紧跟一个下采样层, 这种两次特征提取的结构使 CNN 在对输入样本进行识别时具有较高畸变容忍力. 网络的最后是分类器, 通常由 Softmax 方法实现, 该层将之前提取到的特征进行综合, 使图像特征信息由二维降为一维. 分类器层 (如 Softmax 层) 一般位于网络尾端, 对前面逐层变换和映射提取的特征进行回归分类等处理也可作输出层.

### 1.2 Inception 结构

主流的卷积神经网络在特征提取过程中主要采用加深网络层数来实现, 但是由此引入了过度拟合、梯度弥散和计算复杂度提升的问题. 因此, Szegedy 等提出了 Inception 结构用于解决该类问题[7]. 这种结构能够有效地减少网络的参数数量, 同时也能加深加宽网络, 增加网络的特征提取能力.

最初的 Inception 是所有卷积核都放到上层的输出来实现, 即 1×1, 3×3, 5×5 的卷积和 3×3 池融合在一起, 因此也造成 5×5 的卷积核计算复杂度太高, 特征图厚度过大. 随后在 Inception 的第一个稳定版本中, Szegedy 将 Inception 结构进行优化, 在 3×3 前, 5×5 前, max pooling 后都分别加上了 1×1 的卷积核从而降低特征图厚度的. 最后的模型如图 1 所示.

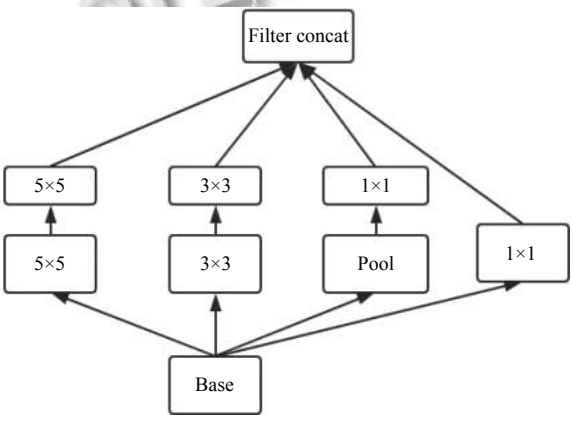


图 1 Inception V1 结构图

在接下来的 Inception V2 中, Google 团队为了进一步减少计算量并提升性能, 加入了 BN 层减少 Internal Covariate Shift, 将两个 5×5 的卷积分解成两个 3×3 的卷积进行叠加, 节省了 72% 左右的开销, 再将

3×3 的 conv 用 1×3 和 3×1 的卷积来代替, 在此基础上, Santurkar 等[8]认为 $n\times n$ 的卷积在理论上都可以由 $n\times 1$ 和 $1\times n$ 的卷积来进行替代, 从而节约 CPU 和内存损耗. 最终在训练参数较少的情况下提升了分类准确率. Inception V2 如图 2 所示.

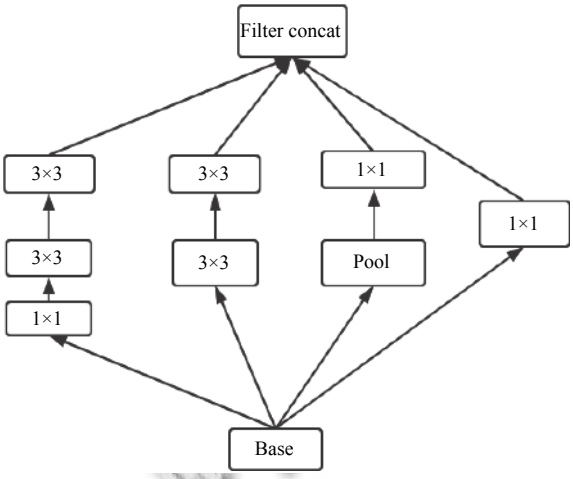


图 2 Inception V2 结构图

Inception V3 一个最重要的改进是分解 (factorization), 将 7×7 分解成两个一维的卷积 (1×7, 7×1), 3×3 也是一样 (1×3, 3×1), 这样的好处, 既可以加速计算 (多余的计算能力可以用来加深网络), 又可以将 1 个卷积拆成 2 个卷积, 使得网络深度进一步增加, 增加了网络的非线性.

### 1.3 归一化 (Batch Normalization, BN) 层

BN 层是 Ioffe S 及其团队在 2015 年提出[9], 该算法的核心功能是对网络的输入数据进行归一化. 其主要作用是将每一层神经网络的神经元的输入分布归一化到均值为 0 方差为 1 的标准正态分布, 从而让梯度变大, 避免产生梯度消失, 最终提升学习的收敛速度. 假设有数据 $x=\left\{x^{(1)},x^{(2)},\cdots,x^{(m)}\right\}$, 则使用下列公式进行归一化:

$$x''^{(k)}=\frac{x^{(k)}-E[x^{(k)}]}{\sqrt{\mathrm{var}[x^{(k)}]}} \tag{1}$$

其中, 分母为第 $i$ 维数据的标准差, 分子为 $i$ 维数据和均值的差.

## 2 多结构 Inception 算法

### 2.1 多 Inception 结构

本文提出的多 Inception 结构特征提取算法的核心思路是将网络层的输入输出尺寸进行修改, 并将滤波器的结构进行调整, 每个 Inception 结构都在对应的卷积层特征上提取不同尺度的特征图, 从而对同一目标不同尺寸的特征进行提取, 能有效将卷积特征层更好地结合, 从而在整体上提升人脸特征的精度. 本文构建 Inception 的核心思路是: ① 使用更小的核来减少网络参数, 例如在第一层将原本 7×7 的卷积核减少为两个 5×5 和 3×3 的卷积核, 从而减少参数. ② 利用 BN 层降低梯度消失的问题. ③ 利用 TripletLoss 和 Softmax 结合的方法来降低类内距离, 提高类间距离. ④ 采用瓶颈性结构, 充分利用深层次提高抽象能力, 同时节约计算. 本文提出的多 Inception 结构如下:

(1) 第 1 层为卷积层, 使用 5×5 的卷积核 (滑动步长 2, padding 为 3), 在 64 通道卷积后进行 ReLU 操作经过 3×3 的 max pooling(步长为 2), 输出为 ((112–3+1)/2)+1=56, 即输出大小为 56×56×64, 再进行 ReLU 操作.

(2) 第 2 层继续使用卷积层, 使用 3×3 的卷积核 (滑动步长为 1, padding 为 1), 192 通道, 输出大小转化为 56×56×192, 卷积后进行归一化操作, 经过 3×3 的 max pooling(步长为 2), 输出为 ((56–3+1)/2)+1=28, 即输出大小为 28×28×192, 再进行 ReLU 操作.

(3) 第 3 层分成 4 个部分, 采用不同尺度的卷积核来进行处理, 4 个卷积核分别为: 1) 64 个 1×1 的卷积核, 然后进行 ReLU 操作, 输出 28×28×64, 2) 96 个 3×3 的卷积核, 进行 ReLU 计算, 再进行 128 个 3×3 的卷积 (padding 为 1), 输出 28×28×128. 3) 16 个 5×5 的卷积核, 大小变为 28×28×16. 4) MaxPool 层, 使用 3×3 的核 (padding 为 1), 然后进行 32 个 1×1 的卷积, 大小变为 28×28×32. 最后将 4 个结果进行连接, 对这 4 部分输出结果的第三维进行并联, 即 64+128+32+32=256, 最终输出大小变为 28×28×256.

(4) 第 4 层有 4 部分, 分别是: ① 128 个 1×1 的卷积核, 进行 ReLU 操作, 输出 28×28×128. ② 128 个 1×1 的卷积核, 作为 3×3 卷积核之前的降维, 进行 ReLU 操作, 再进行 192 个 3×3 的卷积操作 (padding 为 1), 输出大小为 28×28×192. ③ 将 32 个 1×1 的卷积核作为 5×5 卷积核之前的降维, 进行 ReLU 操作后, 再进行 96 个 5×5 的卷积 (padding 为 1), 输出大小变为 28×28×96. ④ pool 层, 使用 3×3 的核 (padding 为 1) 进行 64 个 1×1 的卷积, 输出大小转换

为 28×28×64. 将 4 个结果进行连接, 对这 4 部分输出结果的第三维并联.

由于本文提出的结构在每个卷积层都加入 Inception, 使网络可以充分考虑每个卷积层的特征维度, 获取不同场景下 (即各种非限制条件) 的目标特征, 具有更强的鲁棒性, 同时 BN 层能进一步优化了参数, 能实现快速收敛. 本文提出的多 Inception 结构如图 3.

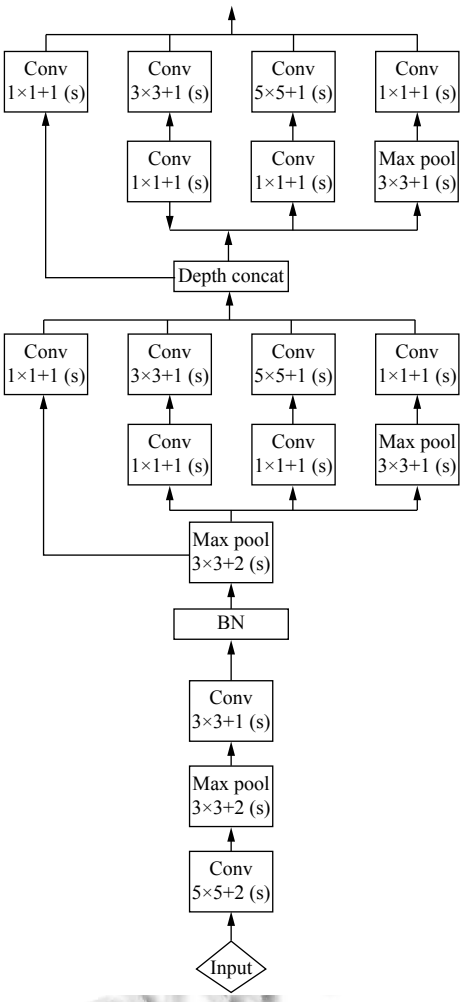


图 3 多 Inception 结构

### 2.2 损失函数的度量学习

大部分基于 CNN 的特征提取方法都会采用 SoftmaxLoss[10]作为训练网络的损失函数, 通过迭代过程中损失函数反馈的损失值来动态优化网络参数, 但是在人脸识别的任务中, 由于人脸表情的复杂性, 环境的多变性, 传统的 Softmax 函数只能增大类间距离, 而类内距离无法有效的减少, 因此非限制条件下的人脸识别效率很低. 而三元组损失函数 TripletLoss 在传统的基于正负样本对的基础上, 引入了 Anchor 作为第三个约束条件, 在减低同类样本的同时, 增大非同类样本的类间距离[11]. 假设为锚样本, 为正样本, 为负样本, 则可以将三元损失函数定义为:

$$Loss_{\text{TripletLoss}} = \frac{1}{N}\sum_{i=1}^{N}\max\{d(A_i,P_i)-d(A_i-N_i)+a\} \quad (2)$$

因此综合了 SoftmaxLoss 和 TripletLoss 的损失函数能有效的优化类间距离和类内距离的分类问题, 同时避免 TripletLoss 的收敛缓慢问题. 因此本实验采用的损失函数计算方法是将两者进行加权, 具体为:

$$L = Loss_{\text{SoftmaxLoss}} + \beta\, Loss_{\text{TripletLoss}} \quad (3)$$

其中, $\beta$ 是权重值, 用来平衡两个损失函数. 在本文提出的网络中, Softmax 层和 TripletLoss 层的使用方式如下:

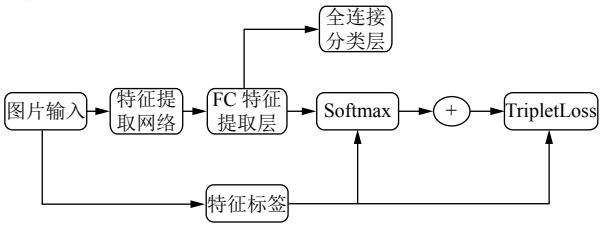


图 4 基于多 Inception 结构和融合 TripletLoss 和 SoftmaxLoss 的特征提取网络

## 3 实验方案及结果

本实验采用配置为 Intel 的 I7-8900K 4 核 3.7 G 处理器, 32 GB 内存, M40 显卡, Windows10 操作系统的 PC 电脑作为运行环境, 使用基于 Python 的 Google 大数据平台 TensorFlow.

### 3.1 数据样本

本文使用 LFW (Labeled Faces in the Wild)[12]人脸数据库, 该数据库是由美国马萨诸塞州立大学提供, 一共 13 000 张图片, 每张图片都被标识出对应的人的名字, 其中有 1680 人对应不只一张图像, 即大约 1680 个人包含两张以上人脸.

### 3.2 实验流程

首先将训练和测试用的样本进行预处理, 流程如下: (1) 用 Adaboost 算法[13]面部检测器将样本进行人脸检测和面部关键点定位 (双眼, 鼻子, 嘴角, 耳朵). (2) 根据 Adaboost 算法定位出的 6 个关键点位置进行数据剪裁, 得到统一的 112×96 的人脸图片. (3) 开始训练, 核心训练参数如下: 学习衰减率设置为 0.001, 训练批次为 50 次, 迭代次数为 10 万次. 在测试结果的比对

中, 将测试的人脸原图特征和水平翻转图提取的特征进行对比, 计算其相似度, 最后标识出人物名字. 最后识别效果图如图 5.


图 5 识别效果图

### 3.3 测试结果及分析

在 LFW 库中选取 8000 个人脸, 在非限制条件下对人脸进行检测的过程中, 在采用 Softmax 和 Triplet 结合的损失函数作为训练监督信号的情况下, 能取得 98.54% 的准确率. 在 ROC 的对比方面, 如图 6 所示, 本文将 DeepID, DeepFace, 传统 Inception 等 3 种算法的结果和本文提出算法进行比较. 以虚检率为横坐标, 检出率为纵坐标, 可以看到, 本文提出的算法在 ROC 上表现优于其他对比算法. 实验对比结果如表 1.

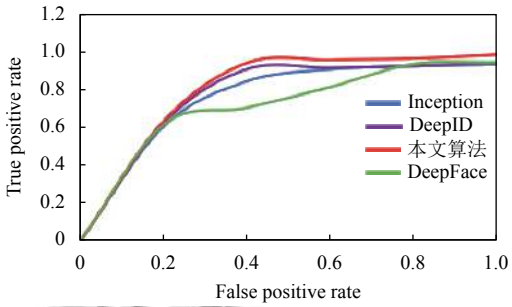


图 6 ROC 对比图

在最高识别准确率方面传统的基于 BP 神经网络的 PCA 算法[14]和 Fisher[15]判别分析的算法相比, 本文提出的算法具有更高的识别准确率. 其次, 与基于分类网络的算法如 DeepFace、DeepID 相比, 本文提出的算法的网络个数明显减少, 计算的复杂度大幅度下降. 同时, 与基于 cosine 距离的 Contrastive loss 的算法相比, 在保证相同识别率的条件下, 明显减少了人脸关键点. 和传统的 Inception 的对比中, 本文对卷积核进行了简化, 例如第三层用一个 3×3 和 1×1 的的卷积核代替了 5×5 的卷积核等等, 大量减少了网络参数.

表 1 人脸识别对比结果

| 算法 | 监督函数 | 网络个数 | 人脸关键点 | 识别准确率 (%) |
|---|---|---|---|---|
| PCA | No | 0 | 27 | 94.97 |
| Fisher | No | 0 | 10 | 94.02 |
| DeepID | Softmax | 50 | 4 | 94.21 |
| DeepFace | Softmax+ConLoss | 20 | 23 | 95.02 |
| Inception-v2 | No | 1 | 6 | 95.01 |
| 本文的多 Inception | Softmax+TripletLoss | 1 | 6 | 98.54 |

## 4 结论

本文提出了基于多 Inception 结构的神经网络人脸识别算法, 在进一步分解简化了卷积核之后采用 Softmax 和 TripletLoss 相结合的方式, 实现了加深和加宽网络的能力, 在增强网络性能的同时, 减少网络参数的数量, 在 LWF 库中进行实验, 实验证明本文提出的算法可以在减少输入参数的情况下提高识别率, 同时降低计算复杂度, 为人脸识别算法的研究提供有益参考.

## 参考文献

1 王守觉, 曲延锋, 李卫军, 等. 基于仿生模式识别与传统模式识别的人脸识别效果比较研究. 电子学报, 2004, 32(7): 1057–1061. [doi: 10.3321/j.issn:0372-2112.2004.07.001]

2 Taigman Y, Yang M, Ranzato M, *et al*. DeepFace: Closing the gap to human-level performance in face verification. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 1701–1708.

3 Sun Y, Wang XG, Tang XO. Deep learning face representation from predicting 10, 000 classes. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 1891–1898.

4 Peri SV, Dhall A. DisguiseNet: A contrastive approach for

disguised face verification in the wild. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, UT, USA. 2018. 25–31.

5 Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 815–823.

6 蔡娟, 蔡坚勇, 廖晓东, 等. 基于卷积神经网络的手势识别初探. 计算机系统应用, 2015, 24(4): 113–117. [doi: 10.3969/j.issn.1003-3254.2015.04.019]

7 Szegedy C, Liu W, Jia YQ, *et al*. Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 1–9.

8 Santurkar S, Tsipras D, Ilyas A, *et al*. How does batch normalization help optimization? Proceedings of the 32nd Conference on Neural Information Processing Systems. Montreal, Canada. 2018. 2483–2493.

9 Ioffe S, Szegedy S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on International Conference on Machine Learning. Lille, France. 2015. 448–456.

10 Martins AFT, Astudillo RF. From softmax to sparsemax: A sparse model of attention and multi-label classification. Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York, NY, USA. 2016. 1614–1623.

11 Bessette EE, Goodenough AK, Langouët S, *et al*. Screening for DNA adducts by data-dependent constant neutral loss-triple stage mass spectrometry with a linear quadrupole ion trap mass spectrometer. Analytical Chemistry, 2009, 81(2): 809–819. [doi: 10.1021/ac802096p]

12 Grujicic M, Arakere G, Pandurangan B, *et al*. Process modeling of Ti-6Al-4V Linear Friction Welding (LFW). Journal of Materials Engineering and Performance, 2012, 21(10): 2011–2023. [doi: 10.1007/s11665-011-0097-8]

13 Hastie T, Rosset S, Zhu J, *et al*. Multi-class adaboost. Statistics and its Interface, 2009, 2(3): 349–360. [doi: 10.4310/SII.2009.v2.n3.a8]

14 Perlibakas V. Distance measures for PCA-based face recognition. Pattern Recognition Letters, 2004, 25(6): 711–724. [doi: 10.1016/j.patrec.2004.01.011]

15 Shan SG, Cao B, Gao W, *et al*. Extended Fisherface for face recognition from a single example image per person. Proceedings of 2002 IEEE International Symposium on Circuits and Systems. Phoenix-Scottsdale, AZ, USA. 2002. II-81–84.