

基于余弦距离损失函数的人脸表情识别算法

吴慧华, 苏寒松, 刘高华*, 李燊, 苏晓

天津大学电气自动化与信息工程学院, 天津 300072

摘要 为解决人脸表情识别任务中存在的类内表情差异性大、类间表情相似度高的问题, 基于传统的 Softmax 损失函数和 Island 损失函数, 提出一种新的基于余弦距离损失函数来指导深度卷积神经网络的学习。该方法不仅可以减小特征空间中类内特征的差异, 而且可以增大类间特征分布, 从而提升特征判别效果。经过大量的实验和分析, 该算法在 RAF-DB 人脸表情数据集上的准确率达到 83.196%, 效果优于 Softmax 损失函数和 Island 损失函数, 所提算法在人脸表情识别任务中具有较高的优越性。

关键词 机器视觉; 深度学习; 表情识别; 卷积神经网络; 损失函数

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP56.241502

Facial Expression Recognition Algorithm Based on Cosine Distance Loss Function

Wu Huihua, Su Hansong, Liu Gaohua*, Li Shen, Su Xiao

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

Abstract This study proposes a new cosine distance loss function based on the traditional Softmax loss function and Island loss function to guide the learning of deep convolution neural networks and solve the problem of large difference in intra-class expressions and high similarity in inter-class expressions in the facial expression recognition tasks. The proposed method not only reduces the difference of intra-class features in the feature space, but also increases the distribution of inter-class features, thereby improving the effect of feature discrimination. After conducting several experiments and analyses, the accuracy of the facial expression recognition algorithm is observed to be 83.196% based on the RAF-DB facial expression dataset, and the effect is better than those obtained using the Softmax loss function and the Island loss function. Furthermore, the proposed algorithm is highly superior with respect to the facial expression recognition tasks.

Key words machine vision; deep learning; expression recognition; convolutional neural network; loss function

OCIS codes 150.1135; 100.2960; 100.3008

1 引言

人脸面部表情是表达人类情绪状态和意图的最强大最自然的信号之一, 在机器人、驾驶员疲劳驾驶检测及人机交互系统中都有着广泛应用^[1-2]。Ekman 等^[3]在 1971 年定义了 6 种基本表情, 即生气、厌恶、害怕、开心、难过和惊讶, 后期加入中性表情。真实生活环境中存在头部姿势变化、光照变化、遮挡以及细微的面部外观变化等, 会导致类内特征

差异大、类间特征相似度高, 故面部表情识别的准确率依然较低, 面部表情识别任务仍然面临着巨大挑战^[4]。

近几年来, 人工智能领域有了令人瞩目的发展。一方面 AI 芯片在处理能力和架构方面有一定的提升, 在算力上为并行计算提供了保障; 另一方面, 人工神经网络出现了各种新颖的结构和算法, 为各行业提供了更高效的智能化解决方案, 从而使得不少领域的研究开始转向深度学习方法, 并且获得了

收稿日期: 2019-04-25; 修回日期: 2019-05-23; 录用日期: 2019-06-05

基金项目: 广州市科技计划项目(201802020008)

* E-mail: suppig@126.com

非常高的识别准确率^[5-9]。当前,越来越多的研究者采用深度学习技术来处理面部表情识别在真实生活环境中的挑战,特别是卷积神经网络技术,其中多层卷积神经网络可以学习到输入对象的深层特征,在面部表情识别方面取得了很好的成果^[10-13]。

传统的卷积神经网络使用 Softmax 损失函数来优化类间特征的差异,但忽略了类内特征存在的差异性。为解决这个问题,许多新的损失函数被提出。Island 损失函数受人脸识别任务中的 Center 损失函数的启发^[14],在对特征与相应类的距离增加惩罚项的基础上,增大类间距离,这不仅压缩了聚类簇,而且增大了聚类中心之间的距离^[15]。人脸识别任务中提出的 AM-Softmax 损失函数是在 Softmax 损失函数的基础上引入余弦余量项 m ,最大化所学习到的特征在特征空间中的决策边缘,使得人脸特征具有更大的类间距离和更小的类内距离^[16]。考虑到 Island 损失函数依然是结合 Softmax 损失函数优化网络结构进行训练的,本文结合 Island 损失函数和 AM-Softmax 损失函数提出基于余弦距离

的损失函数,该函数可以监督卷积神经网络学习强区分度的人脸表情特征。

2 基本原理

2.1 人脸表情识别方法

基于深度学习的人脸表情识别算法分为 3 部分,人脸预处理、表情特征提取和表情分类。具有足够多的标记训练集、尽可能多的种族变化和環境变化对于人脸表情识别算法的设计至关重要。目前已经公开了多种人脸表情数据集用于人脸表情识别研究^[17-21]。对于特征提取,本文设计的网络结构是 Xception 结构的精简模式 mini-Xception^[22],结构如图 1 所示。每个卷积层(Conv)后面都加入批量正则化(BN)及线性整流激活单元(ReLU),这可以缓解由数据集数目不足造成的过拟合问题。在网络结构最后连接两个全连接层(FC),根据第 1 个 FC 层输出计算 Island 损失函数,根据第 2 个 FC 层输出计算 AM-Softmax 损失函数,二者联合损失对网络训练进行监督,验证了本文基于余弦距离损失函数的有效性。

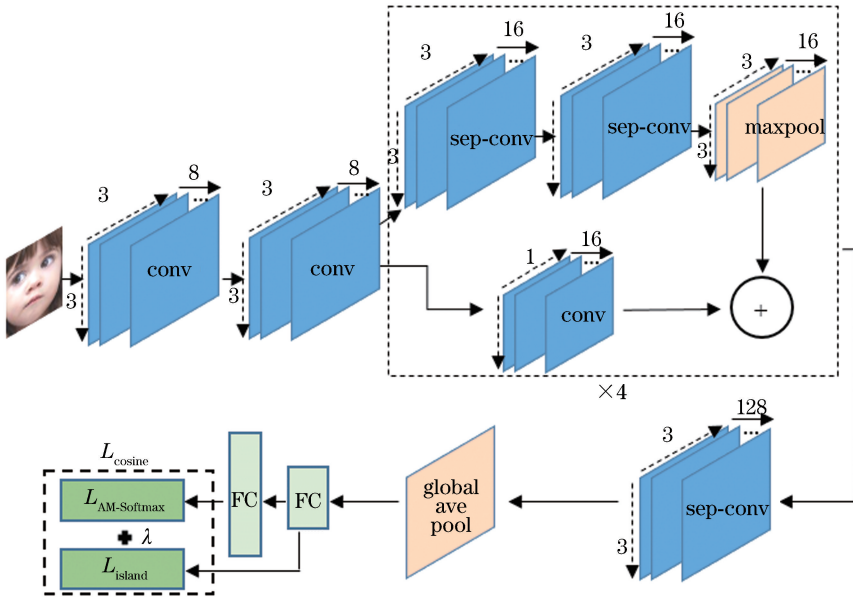


图 1 mini-Xception 网络结构图

Fig. 1 Schematic of mini-Xception network structure

2.2 传统损失函数

为更好地理解本文基于余弦距离的损失函数,简要介绍 Softmax 损失函数和 Island 损失函数。Softmax 损失函数表示为 L_S , 计算公式为

$$L_S = - \sum_{i=1}^n \log \frac{\exp(\mathbf{W}_{y_i}^T \mathbf{x}_i + b_{y_i})}{\sum_{j=1}^l \exp(\mathbf{W}_j^T \mathbf{x}_j + b_j)}$$

$$- \sum_{i=1}^n \log \frac{\exp[\|\mathbf{W}_{y_i}\| \|\mathbf{x}_i\| \cos \theta_{y_i} + b_{y_i}]}{\sum_{j=1}^l \exp[\|\mathbf{W}_j\| \|\mathbf{x}_j\| \cos \theta_j + b_j]}$$

(1)

式中: n 表示一个批次的样本数量; l 表示类别数量; \mathbf{x} 表示最后一个全连接层的输入; \mathbf{x}_i 表示第 i 个样本的特征向量; \mathbf{W}_j 表示第 j 个类别全连接层输出

的权重向量; θ_j 表示权重向量 \mathbf{W}_j 和特征向量 \mathbf{x}_j 之间的夹角; y_i 表示第 i 个样本的类别标签; $\mathbf{W}_{y_i}^T \mathbf{x}_i$ 表示第 i 个样本的目标评定。

Island 损失函数是基于 Center 损失函数的改进函数,在网络模型训练过程中,Center 损失函数可以减小类内特征之间的差异,Island 损失函数可以增大类间特征之间的距离。Center 损失函数表示为 L_C ,计算公式为

$$L_C = \frac{1}{2} \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{c}_{y_i}\|^2, \quad (2)$$

式中: \mathbf{c}_{y_i} 表示与 y_i 具有相同类别标签的所有样本的中心。Island 损失函数表示为 L_1 ,其定义为 Center 损失与聚类中心之间欧氏距离之和,计算公式为

$$L_1 = L_C + \lambda_1 \sum_{\substack{\mathbf{c}_k \in N \\ \mathbf{c}_j \in N \\ \mathbf{c}_k \neq \mathbf{c}_j}} \sum_{\substack{\mathbf{c}_k \in N \\ \mathbf{c}_j \in N \\ \mathbf{c}_k \neq \mathbf{c}_j}} \left(\frac{\mathbf{c}_k \cdot \mathbf{c}_j}{\|\mathbf{c}_k\|_2 \|\mathbf{c}_j\|_2} + 1 \right), \quad (3)$$

式中: N 表示表情标签的集合; \mathbf{c}_k 和 \mathbf{c}_j 分别表示具有 L_2 范数的 $\|\mathbf{c}_k\|_2$ 和 $\|\mathbf{c}_j\|_2$ 的第 k 个和第 j 个

表情类别中心。(3)式中,前项约束类内差异性,缩小样本与其对应中心之间的距离;后项惩罚表情之间的相似性,增大种类间的距离。 λ_1 是调节因子,用于平衡两项比例。最小化 Island 损失函数,可使类内差异减小,类间差异增大。网络模型的损失函数可以表示为

$$L = L_S + \lambda L_1, \quad (4)$$

式中: λ 为超参数,用于调节两项损失的比例。

Softmax 损失函数、Center 损失函数和 Island 损失函数的分类效果如图 2 所示^[15]。传统卷积神经网络利用 Softmax 损失进行优化,对错误分类样本进行惩罚,使得不同类别的分布彼此远离。如图 2(a)所示,由于同类表情存在较大差别导致每个群集内特征分散,又由于不同类表情间存在高度相似性导致群集间有较大重叠部分。Center 损失可以使人脸表情识别方法学习到的类内特征变化较小,如图 2(b)所示,样本被拉向相应特征中心,但仍然存在特征簇的相互重叠问题。Island 损失可以通过增大不同表情之间的差异,来增强所学习的深层特征的辨识力,如图 2(c)所示,这不仅会压缩群集,还会将两个群集分开,使特征簇彼此分离。

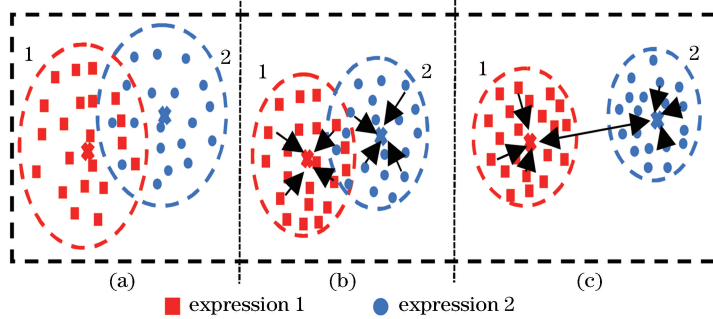


图 2 三种损失函数学习到的深度特征图示。(a) Softmax 损失;(b) Center 损失;(c) Island 损失

Fig. 2 Depth feature diagrams learned by three loss functions. (a) Softmax loss; (b) Center loss; (c) Island loss

2.3 基于余弦距离的损失函数

根据对 Island 损失函数的分析,Softmax 损失函数在很大程度上影响着 Island 损失函数的分类准确率,而基于余弦距离损失函数的基本思想就是通过改善 Softmax 损失函数来提高分类准确率。

人脸识别任务中的 AM-Softmax 损失函数就是在 Softmax 损失函数的基础上改进而来,它可以使人脸特征具有更大的类间距和更小的类内距。AM-Softmax 损失函数除了将网络偏置参数 b 置为 0,网络权重参数 $\|\mathbf{W}\|$ 置为 1,还把样本特征向量 $\|\mathbf{x}\|$ 进行了归一化,最终计算公式为

$$L_{AM-Softmax} = - \sum_{i=1}^n \log \frac{\exp[s \cdot (\mathbf{W}_{y_i}^T \mathbf{x}_i - m)]}{\exp[s \cdot (\mathbf{W}_{y_i}^T \mathbf{x}_i - m)] + \sum_{j=1, j \neq y_i}^l \exp(s \mathbf{W}_j^T \mathbf{x}_i)} - \sum_{i=1}^n \log \frac{\exp[s \cdot (\cos \theta_{y_i} - m)]}{\exp[s \cdot (\cos \theta_{y_i} - m)] + \sum_{j=1, j \neq y_i}^l \exp(s \cdot \cos \theta_j)}, \quad (5)$$

式中： m 为决策余量，能将网络权重 \mathbf{W}_j 和特征向量 \mathbf{x}_i 的余弦距离变为 $\cos \theta - m$ ，通过 m 来调节特征之间的距离； s 为缩放因子，通常取大于 1 的整数。AM-Softmax 损失函数可以减小对应标签项的概率，增大损失，从而对同一类的聚合更有帮助。传统 Softmax 损失函数和 AM-Softmax 损失函数的分类

效果示意图如图 3 所示^[23]，后者在类间施加了一个固定的决策余量 m ^[16]。

本文基于余弦距离损失函数的思想就是把 AM-Softmax 损失函数加入到 Island 损失函数中，可以得到基于余弦距离损失函数的计算公式为

$$\begin{aligned}
 L_{\text{cosine}} = L_{\text{AM-Softmax}} + \lambda L_1 = & - \sum_{i=1}^n \log \frac{\exp[s \cdot (\cos \theta_{y_i} - m)]}{\exp[s \cdot (\cos \theta_{y_i} - m)] + \sum_{j=1, j \neq y_i}^l \exp(s \cdot \cos \theta_j)} + \\
 & \lambda \left[L_C + \lambda_1 \sum_{\substack{c_j \in N \\ c_k \neq c_j}} \sum_{c_k \in N} \left(\frac{\mathbf{c}_k \cdot \mathbf{c}_j}{\|\mathbf{c}_k\|_2 \|\mathbf{c}_j\|_2} + 1 \right) \right] = \\
 & - \sum_{i=1}^n \log \frac{\exp[s \cdot (\cos \theta_{y_i} - m)]}{\exp[s \cdot (\cos \theta_{y_i} - m)] + \sum_{j=1, j \neq y_i}^l \exp(s \cdot \cos \theta_j)} + \\
 & \lambda \left[\frac{1}{2} \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{c}_{y_i}\|^2 + \lambda_1 \sum_{\substack{c_j \in N \\ c_k \neq c_j}} \sum_{c_k \in N} \left(\frac{\mathbf{c}_k \cdot \mathbf{c}_j}{\|\mathbf{c}_k\|_2 \|\mathbf{c}_j\|_2} + 1 \right) \right], \tag{6}
 \end{aligned}$$

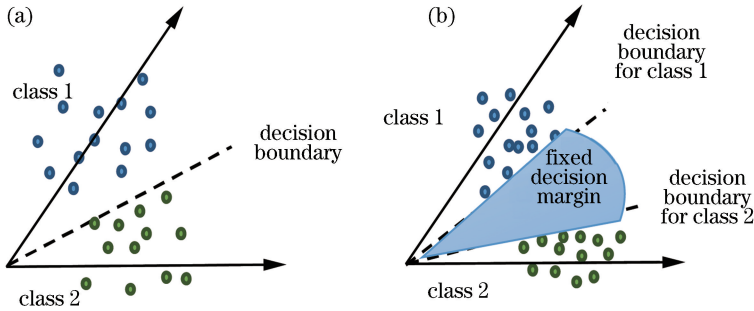


图 3 传统 Softmax 损失函数和 AM-Softmax 损失函数的对比示意图

Fig. 3 Comparison of traditional Softmax loss function and AM-Softmax loss function

式中：超参数 λ 用于调节两项损失的比例。第一项通过 m 来约束特征间的距离，可以更好地聚合同类；第二项约束样本与其中心之间的距离，可以再一次缓解由于类内差异性所造成的影响，同时拉大不同类别中心之间的距离，可以缓解由于类间相似性所造成的影响。实验中设置 $\lambda = 0.01, \lambda_1 = 10$ 。本文基于余弦距离的损失函数是以 AM-Softmax 损失函数为前提，即以权重向量模值置 1，偏置归零，特征向量归一化 ($\|\mathbf{W}_j\| = 1, b = 0, \|\mathbf{x}_i\| = 1$) 为前提条件。图 4 为 Island 损失函数和本文基于余弦距离损失函数的分类边界示意图。 m 的取值大小控制着分类边界的大小，在二分类里，类 1 的分类边界由 P_0 变为了 P_1 。相比 Island 损失函数，本文基于余弦距离损失函数的分类边界不仅与权重向量 \mathbf{W}_1

和 \mathbf{W}_2 有关，还与添加的决策余量 m 相关。需要注意的是，图中标记的距离并不代表实际距离，实际距离是角度的余弦距离。图中使用角度作为距离只是为获得更好的视觉效果。

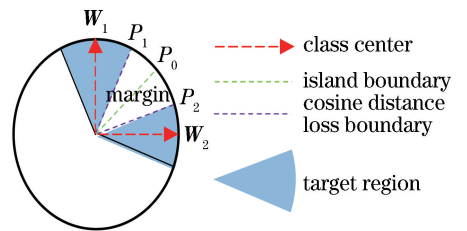


图 4 Island 损失函数和基于余弦距离损失函数的分类边界示意图

Fig. 4 Schematic of classification boundaries of Island loss function and loss function based on cosine distance

3 实验仿真

3.1 实验设置

3.1.1 预处理

为排除数据集中人脸尺寸、角度等不一致对表情识别造成的影响,本文对数据集中的所有图像都进行人脸对齐。人脸关键点检测采用人脸检测 Dlib 库^[24],基于三个关键点对准面部区域:两个眼睛和嘴巴中心。然后将对齐后的面部图像大小都调整为 100×100 ,将三个颜色通道的像素进行归一化处理,即把 $[0, 255]$ 范围内的像素值归一化到 $[0, 1]$ 范围内。

受限于人脸表情数据集的数据量,采用数据增强技术来扩充训练数据^[25],将图像块随机旋转 -20° 到 20° 、以 50% 的概率对图像进行随机水平翻转,在水平和垂直方向上对图像进行 10% 范围内的随机偏移,对图像进行 10% 范围内的随机缩放。

3.1.2 人脸表情数据集

RAF-DB 是一个真实世界的面部表情数据集^[21],包含从互联网上下载的 29672 个高度多样化的面部表情图像。通过手动注释和可靠估计,为表情样本提供了 7 种基本表情标签和 11 种复合表情标签。选用基本表情数据集来评估基于余弦距离损失函数的有效性。

3.1.3 卷积神经网络框架及训练配置

使用 Ubuntu16.04 系统下的 Keras 框架来实现基于余弦距离损失函数的 mini-Xception 卷积神经网络结构。本实验使用 GPU(显卡型号为 NVIDIA GeForce GTX 1080Ti)进行训练,迭代 1 万次。采用自适应矩估计(Adam)的训练策略,其中,批处理大小为 128,学习率初始设为 0.01,当测试损失 val_loss 值不再下降时,学习率乘以衰减因子 0.2。

3.2 实验结果

对基于余弦距离损失函数中的两个超参数 s 和 m 进行多次试验,表 1 所示为 s 取 10, m 取不同值时人脸表情识别的准确率, $m = 0.35$ 时,准确率最高;表 2 所示为 m 取 0.35, s 取不同值时人脸表情识别的准确率,可以看到 $s = 20$ 时,准确率最高。由此得出 $m = 0.35, s = 20$ 时人脸表情识别的准确率最高。混淆矩阵如图 5 所示,图片对角线位置对应每种表情的识别准确率,可以看到 $fear$ 和 $disgust$ 的准确率很低,这是由训练数据集中二者数目较少造成的,12271 张训练图像,标签为 $fear$ 的图像只有 281 张,占 2.29%,标签为 $disgust$ 的图像只有 717 张,占 5.84%。

表 1 $s = 10, m$ 取不同取值时的面部表情识别的准确率

Table 1 Accuracy of facial expression recognition when $s = 10$ and m takes different values

m	Accuracy / %
0.25	80.771
0.30	80.893
0.35	81.578
0.40	79.985
0.45	79.341
0.50	78.432

表 2 $m = 0.35, s$ 取不同取值时的面部表情识别的准确率

Table 2 Accuracy of facial expression recognition when $m = 0.35$ and s takes different values

s	Accuracy / %
5	80.443
10	81.578
15	82.811
20	83.196
25	81.239
30	80.025

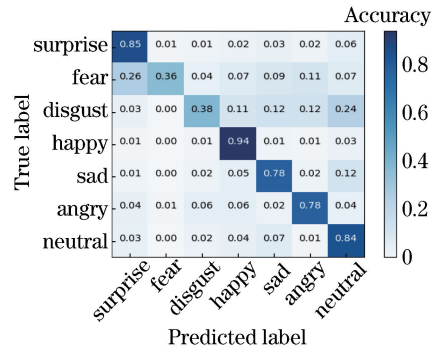


图 5 基于余弦距离损失函数的网络模型实验结果的混淆矩阵

Fig. 5 Confusion matrix of experimental results of network model based on cosine distance loss function

将本文基于余弦距离损失函数与人脸表情识别经典算法进行准确率的对比,对比结果如表 3 所示。可以看到基于余弦距离的损失函数相比传统 Softmax 损失函数准确率提高了 4.471%,相比目前最新的 Island 损失函数准确率提高了 1.383%。与其他人脸表情识别算法相比,本文方法具有很大优势,比 DLP-CNN 算法^[26]提高了 2.299%,比 GAN-Inpainting 算法^[27]提高了 1.322%。此外,本文只采用了结构模型较简单的 mini-Xception 网络模型,人脸表情识别准确率就取得显著的提升。实验结果表明,本文基于余弦距离的损失函数在表情识别任务中具有良好的性能,可以学习到具有更强区分度的表情特征。

表3 不同损失函数下面部表情识别的准确率

Table 3 Accuracy of facial expression recognition under different loss functions

Method	Accuracy / %
DLP-CNN	80.897
GAN-Inpainting	81.874
Softmax loss	78.725
Island loss	81.813
Cosine distance loss	83.196

4 结 论

提出一种基于余弦距离的损失函数,实验采用 mini-Xception 神经网络模型,模型结构精简,参数量较传统神经网络模型大大减小。采用基于余弦距离的损失函数对网络模型进行优化训练,可以学习到具有强区分度的表情特征,达到最小化类内距离、最大化类间距离的效果。本文的 mini-Xception 神经网络模型在加入基于余弦距离的损失函数以后,人脸表情识别准确率得到提升。实验结果证明,本文损失函数在表情识别任务中具有巨大的优势。后续研究将主要致力于针对数目较少的类别训练集进一步提升准确率;及对损失函数进行进一步改进,以更好地指导神经网络训练,使得人脸表情识别准确率得到进一步提升。

参 考 文 献

- [1] Tian Y I, Kanade T, Cohn J F. Recognizing action units for facial expression analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(2): 97-115.
- [2] Darwin C, Prodger P. The expression of the emotions in man and animals [M]. USA: Oxford University Press, 1998.
- [3] Ekman P, Friesen W V. Constants across cultures in the face and emotion [J]. Journal of Personality and Social Psychology, 1971, 17(2): 124-129.
- [4] Valstar M F, Mehu M, Jiang B H, *et al.* Meta-analysis of the first facial expression recognition challenge [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part b (Cybernetics), 2012, 42(4): 966-979.
- [5] Long X, Su H S, Liu G H, *et al.* A face recognition algorithm based on angular distance loss function and convolutional neural network [J]. Laser & Optoelectronics Progress, 2018, 55(12): 121505.
龙鑫, 苏寒松, 刘国华, 等. 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 [J]. 激光与光

电子学进展, 2018, 55(12): 121505.

- [6] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C] // Advances in Neural Information Processing Systems 25 (NIPS 2012), December 3-6, 2012, Lake Tahoe, Nevada, United States. Canada: NIPS, 2012.
- [7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J/OL]. (2015-04-10) [2019-04-24]. <https://arxiv.org/abs/1409.1556>.
- [8] Szegedy C, Liu W, Jia Y Q, *et al.* Going deeper with convolutions [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 15523970.
- [9] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [10] Kim B K, Lee H, Roh J, *et al.* Hierarchical committee of deep CNNs with exponentially-weighted decision fusion for static facial expression recognition [C] // Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15, November 9-13, 2015, Seattle, Washington, USA. New York: ACM, 2015: 427-434.
- [11] Yu Z D, Zhang C. Image based static facial expression recognition with multiple deep network learning [C] // Proceedings of the 2015 ACM on International Conference on Multimodal Interaction-ICMI'15, November 9-13, 2015, Seattle, Washington, USA. New York: ACM, 2015: 435-442.
- [12] Ng H W, Nguyen V D, Vonikakis V, *et al.* Deep learning for emotion recognition on small datasets using transfer learning [C] // Proceedings of the 2015 ACM on International Conference on Multimodal Interaction-ICMI'15, November 9-13, 2015, Seattle, Washington, USA. New York: ACM, 2015: 443-449.
- [13] Yao A B, Cai D Q, Hu P, *et al.* HoloNet: towards robust emotion recognition in the wild [C] // Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016, November 12-16, 2016, Tokyo, Japan. New York:

- ACM, 2016: 472-478.
- [14] Wen Y D, Zhang K P, Li Z F, *et al.* A discriminative feature learning approach for deep face recognition[M] // Leibe B, Matas J, Sebe N, *et al.* Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9911: 499-515.
- [15] Cai J, Meng Z B, Khan A S, *et al.* Island loss for learning discriminative features in facial expression recognition [C] // 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), May 15-19, 2018, Xi'an, China. New York: IEEE, 2018: 302-309.
- [16] Wang F, Cheng J, Liu W Y, *et al.* Additive margin softmax for face verification [J]. IEEE Signal Processing Letters, 2018, 25(7): 926-930.
- [17] Kanade T, Cohn J F, Tian Y L. Comprehensive database for facial expression analysis [C] // Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), March 28-30, 2000, Grenoble, France. New York: IEEE, 2000: 6577271.
- [18] Lucey P, Cohn J F, Kanade T, *et al.* The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression [C] // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE, 2010: 94-101.
- [19] Zhao G Y, Huang X H, Taini M, *et al.* Facial expression recognition from near-infrared videos[J]. Image and Vision Computing, 2011, 29(9): 607-619.
- [20] Dhall A, Ramana Murthy O V, Goecke R, *et al.* Video and image based emotion recognition challenges in the wild [C] // Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15, November 9-13, 2015, Seattle, Washington, USA. New York: ACM, 2015: 423-426.
- [21] Li S, Deng W H, Du J P. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2584-2593.
- [22] Chollet F. Xception: deep learning with depthwise separable convolutions [C] // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, Hawaii, USA. New York: IEEE, 2017: 1251-1258.
- [23] Deng J K, Zhou Y X, Zafeiriou S. Marginal loss for deep face recognition [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2006-2014.
- [24] Saragih J M, Lucey S, Cohn J F. Face alignment through subspace constrained mean-shifts [C] // 2009 IEEE 12th International Conference on Computer Vision, September 29-October 2, 2009, Kyoto, Japan. New York: IEEE, 2009: 1034-1041.
- [25] Simard P Y, Steinkraus D, Platt J C. Best practices for convolutional neural networks applied to visual document analysis [C] // Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings., August 6-6, 2003, Edinburgh, UK. New York: IEEE, 2003.
- [26] Zhao L M, Li X, Zhuang Y T, *et al.* Deeply-learned part-aligned representations for person re-identification [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 3239-3248.
- [27] Isola P, Zhu J Y, Zhou T H, *et al.* Image-to-image translation with conditional adversarial networks [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 5967-5976.