



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020
04
VOL.25

ISSN1006-8961
CN11-3758/TB



海岛岸线自动分割 P768



第25卷第4期（总第288期）

2020年4月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 顾行发
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@radi.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

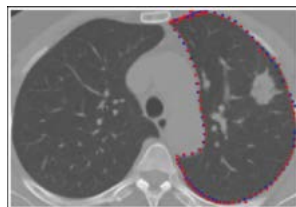
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



自适应在线判别外观学习的
分层关联多目标跟踪
(第0708页)



低秩和主动形状模型的肺区
分割(第0759页)



结合曲率滤波的HTM算法去
除遥感影像云雾(第0791页)

综述

深度学习目标检测方法综述

赵永强, 饶元, 董世鹏, 张君毅 0629

图像处理和编码

结合矩阵分解与差分隐私的人脸图像发布

张啸剑, 付聪聪, 孟小峰 0655

图像分析和识别

多尺度深度特征融合的变化检测

樊玮, 周末, 黄睿 0669

面向类内差距表情的深度学习识别

陈亮, 吴攀, 刘韵婷 0679

带后缀印刷体维吾尔文映射关系检索

伊克萨尼·普尔凯提, 阿布都萨拉木·达吾提, 艾斯卡尔·艾木都拉 0688

结合混沌鸟群算法的阴极铜板表面缺陷检测

王卓, 张长胜, 李伟, 钱俊兵, 唐都作, 蔡兵, 常以涛 0697

自适应在线判别外观学习的分层关联多目标跟踪

方岚, 于凤芹 0708

图像理解和计算机视觉

多方向Laplacian能量和与tetrolet变换的图像融合

沈瑜, 陈小鹏, 杨倩 0721

双重对偶生成对抗网络的跨年龄素描—照片转换

吴柳玮, 孙锐, 阚俊松, 高隽 0732

医学图像处理

残差密集相对平均CGAN的脑部图像配准

王丽芳, 张程程, 秦品乐, 蔺素珍, 高媛, 窦杰亮 0745

低秩和主动形状模型的肺区分割

孙申申, 田丹, 吴微, 康雁, 赵宏 0759

遥感图像处理

改进Deeplab网络的遥感影像海岸线分割

王振华, 钟元芾, 何婉雯, 曲念毅, 徐利智, 张文苹, 刘智翔 0768

卷积神经网络提取风云影像土壤湿度

韩颖娟, 周顺武, 王文文, 张学艺, 张承明, 栗佩琪 0779

结合曲率滤波的HTM算法去除遥感影像云雾

杨珂珂, 贾渊, 沈川 0791

ChinaMM 2019会议专栏

利用稳健非负矩阵分解实现无监督高光谱解混

宋晓瑞, 吴玲达, 孟祥利 0801

关键点深度特征驱动人脸表情识别

王善敏, 帅惠, 刘青山 0813

融合区域和边缘特征的水平集水下图像分割

孙杨, 陈哲, 王慧斌, 张振, 沈洁 0824

深度时空能量特征表示下的人体行为识别

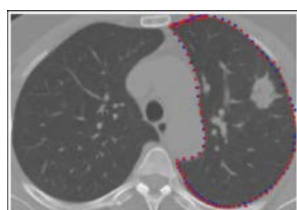
巢新, 侯振杰, 李兴, 梁久桢, 宦娟, 刘浩昱 0836

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Multi-object tracking based on adaptive online discriminative appearance learning and hierarchical association(P0708)



Lung segmentation by active shape model approach based on low rank theory(P0759)



Haze and cloud removal from remote sensing image using HTM algorithm based on curvature filtering(P0791)

Review

Survey on deep learning object detection

Zhao Yongqiang, Rao Yuan, Dong Shipeng, Zhang Junyi 0629

Image Processing and Coding

Private facial image publication through matrix decomposition

Zhang Xiaojian, Fu Congcong, Meng Xiaofeng 0655

Image Analysis and Recognition

Multiscale deep features fusion for change detection

Fan Wei, Zhou Mo, Huang Rui 0669

Depth learning recognition method for intra-class gap expression

Chen Liang, Wu Pan, Liu Yunting 0679

Mapping relationship retrieval of Uyghur-printed suffix word

Eksan Firkat, Abdusalam Dawut, Askar Hamdulla 0688

Cathodic copper plate surface defect detection based on bird swarm algorithm with chaotic theory

Wang Zhuo, Zhang Changsheng, Li Wei, Qian Junbing, Tang Duzuo, Cai Bing, Chang Yitao 0697

Multi-object tracking based on adaptive online discriminative appearance learning and hierarchical association

Fang Lan, Yu Fengqin 0708

Image Understanding and Computer Vision

Image fusion of multidirectional sum modified Laplacian and tetrolet transform

Shen Yu, Chen Xiaopeng, Yang Qian 0721

Double dual generative adversarial networks for cross-age sketch-to-photo translation

Wu Liuwei, Sun Rui, Kan Junsong, Gao Jun 0732

Medical Image Processing

Image registration method with residual dense relativistic average CGAN

Wang Lifang, Zhang Chengcheng, Qin Pinle, Lin Suzhen, Gao Yuan, Dou Jieliang 0745

Lung segmentation by active shape model approach based on low rank theory

Sun Shenshen, Tian Dan, Wu Wei, Kang Yan, Zhao Hong 0759

Remote Sensing Image Processing

Island shoreline segmentation in remote sensing image based on improved Deeplab network

Wang Zhenhua, Zhong Yuanfu, He Wanwen, Qu Nianyi, Xu Lizhi, Zhang Wenping, Liu Zhixiang 0768

Extracting soil moisture from Fengyun satellite images using a convolutional neural network

Han Yingjuan, Zhou Shunwu, Wang Wenwen, Zhang Xueyi, Zhang Chengming, Li Peiqi 0779

Haze and cloud removal from remote sensing image using HTM algorithm based on curvature filtering

Yang Keke, Jia Yuan, Shen Chuan 0791

Column of ChinaMM 2019

Unsupervised hyperspectral unmixing based on robust non-negative matrix factorization

Song Xiaorui, Wu Lingda, Meng Xiangli 0801

Facial expression recognition based on deep facial landmark features

Wang Shanmin, Shuai Hui, Liu Qingshan 0813

Level set method combining region and edge features for segmenting underwater images

Sun Yang, Chen Zhe, Wang Huibin, Zhang Zhen, Shen Jie 0824

Action recognition under depth spatial-temporal energy feature representation

Chao Xin, Hou Zhenjie, Li Xing, Liang Jiuzhen, Huan Juan, Liu Haoyu 0836

中图法分类号: TP183 文献标识码: A 文章编号: 1006-8961(2020)04-0679-09

论文引用格式: Chen L, Wu P and Liu Y T. 2020. Depth learning recognition method for intra-class gap expression. Journal of Image and Graphics, 25(04):0679-0687 [陈亮, 吴攀, 刘韵婷. 2020. 面向类内差距表情的深度学习识别. 中国图象图形学报, 25(04):0679-0687] [DOI:10.11834/jig.190235]

面向类内差距表情的深度学习识别

陈亮, 吴攀, 刘韵婷

沈阳理工大学, 沈阳 110159

摘要: 目的 为解决真实环境中由类内差距引起的面部表情识别率低及室内外复杂环境对类内差距较大的面部表情识别难度大等问题, 提出一种利用生成对抗网络 (generative adversarial network, GAN) 识别面部表情的方法。方法 在 GAN 生成对抗的思想下, 构建一种 IC-GAN (intra-class gap GAN) 网络结构, 使用卷积组建编码器、解码器对自制混合表情图像进行更深层次的特征提取, 使用基于动量的 Adam (adaptive moment estimation) 优化算法进行网络权重更新, 重点针对真实环境面部表情识别过程中的类内差距较大的表情进行识别, 使其更好地适应类内差异较大的任务。结果 基于 Pytorch 环境, 在自制的面部表情数据集上进行训练, 在面部表情验证集上进行测试, 并与深度置信网络 (deep belief network, DBN) 和 GoogLeNet 网络进行对比实验, 最终 IC-GAN 网络的识别结果比 DBN 网络和 GoogLeNet 网络分别提高 11% 和 8.3%。结论 实验验证了 IC-GAN 在类内差距较大的面部表情识别中的精度, 降低了面部表情在类内差距较大情况下的误识率, 提高了系统鲁棒性, 为面部表情的生成工作打下了坚实的基础。

关键词: 深度学习; 生成对抗网络; IC-GAN (intra-class gap GAN); 面部表情识别

Depth learning recognition method for intra-class gap expression

Chen Liang, Wu Pan, Liu Yunting

Shenyang Ligong University, Shenyang 110159, China

Abstract: **Objective** China has a large population of 1.3 billion in 2005, accounting for 19% of the world's population. This number is equivalent to the population of Europe or Africa added to the populations of Australia, North America, and Central America. It is one of the few populous countries in the world, and its huge population size has brought many problems. With the rapid development of the economy, the number of people working outside of their homes is increasing, the population is moving frequently, and the safety of the floating population is even difficult to control. The huge mobile population provides the city's infrastructure and public services tremendous pressure. Thus, to conduct a comprehensive check on the area of adult traffic, which is time-consuming and labor-intensive, is difficult for security and related staff. Particularly, complex environmental safety problems, such as subways, railway stations, and airports, are becoming increasingly serious. Unstable events occur frequently, security situation is receiving much attention, and urban management and service systems are seriously lagging behind. These conditions need to be improved, especially after the September 11 incident

收稿日期: 2019-06-06; 修回日期: 2019-09-17; 预印本日期: 2019-09-24

基金项目: 国家重点研发项目 (2017YFC0821001, 2017YFC0821004-5); 辽宁省自然科学基金项目 (20170540788); 辽宁省教育厅基本科研项目 (LG201707)

Supported by: National Key Research and Development Program of China (2017YFC0821001, 2017YFC0821004-5); Natural Science Foundation of Liaoning Province, China (20170540788)

in the United States. The situation has aroused widespread concern in the international community. Meanwhile, expression is the most intuitive way for humans to express emotions. In addition to language communication, expressions are extremely effective means of communication. People usually express their inner feelings through specific expressions. Expression can be used to judge other person's thoughts. For expressions used to express information, psychologist Mehrabian summed up a formula: emotional expression = 7% of words + 38% of sound + 55% of facial expressions. Expression is one of the most important features of human emotion recognition. Expression is the emotional state expressed by facial muscle changes. Through the facial expression of the person's face, to evaluate abnormal psychological state, speculate on extreme emotions, and observe the facial expressions of pedestrians in the subway, railway station, and airport to further judge the psychology of the person is possible. We provide technical support to determine who is suspicious and prevent certain criminal activities in a timely manner. Strengthening urban surveillance and identifying the facial expressions of criminals are especially important. Expression plays an important role in human emotion cognition. However, factors affecting facial expression recognition in safety screening are extremely large, and the large intra-class gap seriously influences the accuracy of facial expression recognition. The problem of large gaps in facial expression recognition in a real environment is solved by identifying suspected molecules to be monitored should be identified, and the security personnel should prepare in advance to accurately identify them. Facial expressions are also particularly important for preventing security problems. The era of large data has arrived. Meanwhile, with the advancement of computer hardware, deep learning continues to develop. The traditional facial expression recognition method cannot meet the needs of the development of the times, and a new algorithm based on deep learning facial expression recognition is coming soon. Learning methods are widely used in facial expression recognition. Although facial recognition intelligent recognition technology has a long history of research, a large number of research methods have been proposed. However, due to the large facial expression gap, the expression is complex, and the influencing factors are many. The current intelligent recognition effect of facial expression results is not ideal. Considering the deep learning because of its powerful expressive ability, this study introduces the model structure of traditional neural network and carries out corresponding experiments and analysis in the context of real-life facial expression recognition and proposes real-world facial expression recognition research based on deep learning. In the next period, real-world facial expression recognition will make considerable progress. This work further studies the realistic facial expression recognition based on deep learning.

Method This study constructs a new IC-GAN (intra-class gap GAN (generative adversarial network)) recognition network model, providing good adaptability to the facial expression recognition task with large gap within the class. The network consists of a convolutional layer, a fully connected layer, an active layer, a Bath-Norm layer, and a Softmax layer, in which a convolutional assembly encoder and a decoder are used to perform deep feature extraction on facial expression images, and download and parse from the network. The video self-made mixed facial expression data set is based on the real environment, the image is expanded, and the facial expression data are normalized. The complexity of the facial expression features with large differences within the class also increased the network training and network recognition. The momentum-based Adam is used to update the network weight, adjust the network parameters, and optimize the network structure based on this factor. In this study, the facial expression category data are trained based on the Pytorch platform in deep learning and tested on the verification set of the self-made mixed facial expression data set.

Result When the input image is 256×256 pixels, the IC-GAN network model can reduce the false positive rate of the expression in large difference in the class, image blur, and facial expression incompleteness, and improve the system robustness. Compared with deep belief network (DBN) deep trust network and GoogLeNet network, the recognition result of IC-GAN network is 11% higher than that of DBN network and 8.3% higher than that of GoogLeNet network.

Conclusion The IC-GAN accuracy in facial expression recognition with large gaps in the class is verified by experiments. This condition reduces the misunderstanding rate of facial expressions in large intra-class differences, improves the system robustness, and lays down the solid foundation for facial expression generation.

Key words: deep learning; generative adversarial network (GAN); intra-class gap GAN (IC-GAN); facial expression recognition

0 引言

我国庞大的流动人口对城市基础设施和公共服务构成巨大的压力,恶性事件时有发生,安防形势备受关注,城市管理与服务体系亟待完善。美国911事件后,安全形势引起了国际社会的广泛关注(崔成,2017),加强城市监控和对不法分子的面部表情识别变得尤为重要。表情在人类情感认知中起着重要作用(Yu和Zhang,2015),是人类情绪识别最重要的特征之一。通过对人的面部情感表达的识别,可以判断出异常心理状态(徐琳琳等,2017)、推测极端情绪(姚雪,2010)。观测复杂环境出现的行人的面部表情,能够为进一步判断人的心理提供技术支撑,大致判断出哪些人比较可疑,可以及时阻止某些犯罪活动。传统的面部表情识别主要是基于模板匹配(肖瑞琪,2017)和神经网络(唐守军和李鹏程,2014)的面部表情识别方法,而且传统的面部表情在特征选择过程需要人为干预,依靠人工对特征提取算法进行精细设计,缺乏足够的计算能力,训练难度大、准确率低且易丢失原有的表情信息。

目前,深度学习模型通过卷积神经网络(convolutional neural network, CNN)对数据进行深层次的特征提取,已经在计算机视觉尤其是图像识别领域取得了巨大成功(曹金梦等,2018)。Zhang等人(2003)使用深度置信网络(deep belief network, DBN),通过BP(back propagation)神经网络实现对面部表情的识别,取得了可观的识别结果;超密集网络(ultra dense network, UDN)(Liu等,2015;Simonyan和Zisserman,2014),通过CNN与玻尔兹曼机对面部动作进行特征提取,使用支持向量机(support vector machine, SVM)作为网络末端分类器,对表情进行分类。Szegedy等人(2015)提出了深层网络结构GoogLeNet,网络结构进一步加深,将大尺度卷积分解成多个小尺度卷积,计算量大大减少,Top-5错误率降至3.57%,相比之前的网络取得了非常可观的结果。GoogLeNet的创新主要在于Inception,且在Inception v1的基础上,相继提出了Inception v2, Inception v3, Inception v4(Szegedy等,2015,2016), Top-5错误率逐步降低,取得令人满意的识别效果, Inception v4在Network in Network结构中识别效果更是尤为突出,将错误率进一步降至3.08%。

面部表情识别已有大量研究,但对真实环境下由类内差距引起的识别效果不理想等问题的研究相对较少。针对上述问题,本文利用生成对抗网络(generative adversarial network, GAN)(Bowles等,2018),融合Softmax分类器构建一种新的IC-GAN(intral-class gap GAN)网络,在面部表情识别中建立对抗关系,提取面部表情特征,实现对类内差距较大的面部表情图像进行识别,降低面部表情识别过程中因受类内差距影响导致识别效率低和误识的问题。

1 面部表情识别模型的构建

1.1 IC-GAN模型

本文构建的面部表情识别网络结构IC-GAN如图1所示。IC-GAN输入真实数据 x ,通过编码器将 x 编码为潜在向量 z ,解码器作为生成网络将 z 重构为 $\hat{x}$,判别器将原始输入图像 x 与重构图像 $\hat{x}$ 进行对比,同时在生成对抗网络的基础上加入全连接层与Softmax分类器,在本文构建的类内差距较大的面部表情数据集上进行识别,训练生成器 G 和判别器 D 两个模型来进行网络优化。IC-GAN包括如下3个部分:

1)生成器。由编码器 $G_{E(x)}$ 和解码器 $G_{D(z)}$ 构成,编码器将输入图像 x 通过卷积层进行特征提取, x 映射得到面部表情矢量矩阵 z ,其中生成器的解码部分 $G_{D(z)}$ 采用深度卷积生成对抗网络(deep convolutional GAN, DCGAN)结构(Suárez等,2017),使用反转置卷积层、ReLU、BatchNorm和Tanh激活层将图像的矢量 z 重构为 $\hat{x}$,随后使用卷积,批量归一化和ReLU对重构图像 $\hat{x}$ 进行编码,其中,BatchNorm层对上一层的输出进行归一化操作,ReLU和LeakyReLU激活函数对卷积进行非线性操作。

2)判别器。对原始图像 x 判为真,重构图像 $\hat{x}$ 判为假,判别器通过不断优化重构图像 $\hat{x}$ 与原始图像 x 的差距,进行网络的迭代训练,直到达到最好效果,理想情况下重构图像与原始图像无异,判别器使用Softmax通过一个 $[0,1]$ 之间的数来表示数据来自真实数据的概率,Softmax在这里仅判断某类表情识别结果是真或假的概率,判断训练何时停止。其中1表示输入的数据来自真实数据 x ,0表示输入的数据是生成器生成的并尽可能服从真实数据分布

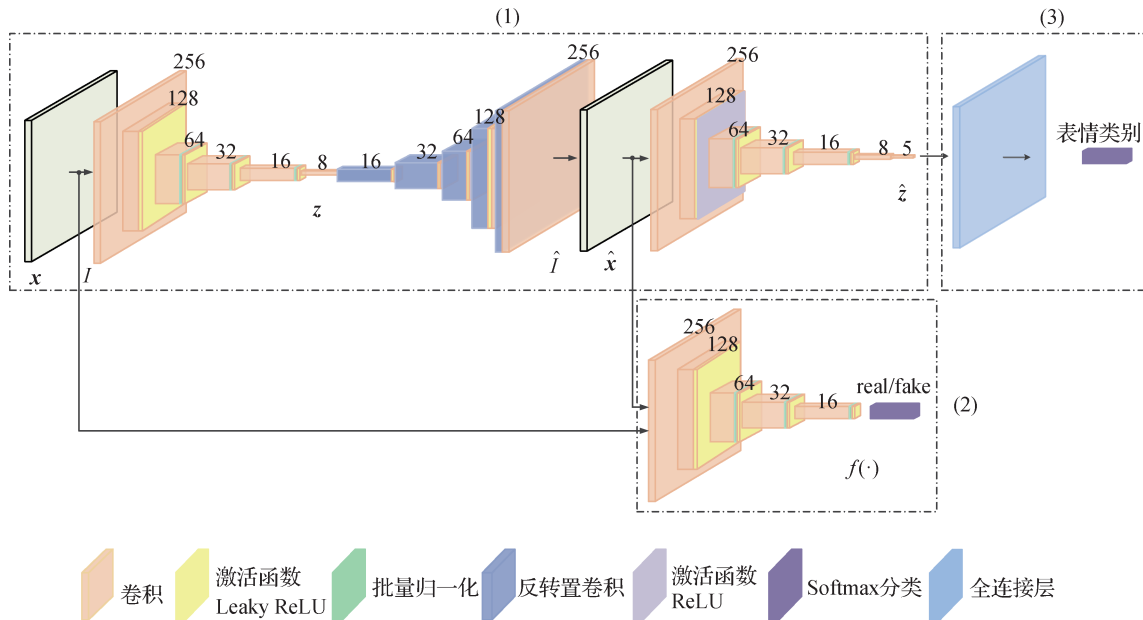


图1 IC-GAN 网络结构

Fig.1 IC-GAN network structure

的假数据 $\hat{x}$, 判别器的输入类型为 $(x_{\text{real}}, 1)$ 和 $(x_{\text{fake}}, 0)$, 输出为 real 或 fake。

3) 全连接层和 Softmax 分类器。通过 Softmax 分类器对面部表情进行识别, 以概率大小依次向下排列, 在最上方输出最大概率的表情, 输出层 output 是 5 个输出, 分别对应 5 个表情类别, 用来关联真实图像 x 和重构图像 $\hat{x}$ 的相关性, 加快生成器和判别器的迭代进程, 保证重构图像 $\hat{x}$ 的正确性。模型 f 学习正常类别数据分布, 并最小化输出异常分数 $A(x)$ 。

1.2 Softmax 分类器

作为网络的表情输出层, Softmax 将网络得到的多个值进行归一化处理, 输出一个 $[0, 1]$ 之间的值来表示面部表情类别的概率, Softmax 输出的表情置信度越高, 面部表情识别结果越准确。Softmax 采用较为简单且容易计算的交叉熵损失参与网络权重的更新, Softmax 的计算过程为

$$z_i = \sum_j \omega_{ij} x_{ij} + b \quad (1)$$

$$S_i = \frac{e^{z_i}}{\sum_k e^{z_k}} \quad (2)$$

$$y_i = \frac{e^{z_i}}{\sum_k e^{z_k}} \quad (3)$$

式中, z_i 表示网络的第 i 个输出, ω_{ij} 表示第 i 个神经

元的第 j 个权重, b 是偏置; S_i 表示第 i 个神经元的输出, y_i 表示 Softmax 的第 i 个输出值, k 表示向量维数。

2 网络目标优化

为优化生成器 G 的参数并学习训练数据的分布, 在生成器与判别器 D_I 之间执行最小—最大策略。利用目标函数 $L_I(G, D_I, I, P)$ 联合训练, D_I 试图最大化正确分类真实和重构图像的概率, G 则尽量骗过判别器, 生成器和判别器的目标函数可优化为

$$L_I(G, D_I, I, P) = E_{I \sim p_{\text{data}}(I)} [\log D_I(I)] + E_{I \sim p_{\text{data}}(I)} [\log (1 - D_I(G(I|p)))] \quad (4)$$

式中, p_{data} 表示真实数据的分布, $I \sim p_{\text{data}}(I)$ 表示 I 服从真实数据分布, $E(\cdot)$ 表示数学期望。

2.1 损失函数定义

根据网络结构和实验特点, 将网络损失分为以下 4 个部分。

1) 重构误差损失。在像素层面上减小原始图像和重构图像的差距, 具体为

$$L_{\text{con}} = E_{x \sim p_X} \|x - G(x)\|_1 \quad (5)$$

式中, p_X 表示数据分布。

2) 特征匹配误差损失。本文使用特征匹配方法 (Chen 和 Hu, 2019), 以减少训练的不稳定性, 在

图像特征层面进行优化,具体为

$$L_{\text{adv}} = E_{x \sim p_X} \|f(x) - f(G(x))\|_2 \quad (6)$$

式中, $f(\cdot)$ 代表判别器模型变换。

3) 潜在向量 z 与重构潜在向量 $\hat{z}$ 的面部表情信息的编码损失。 z 与 $\hat{z}$ 之间的损失函数为

$$L_T = \|h(z) - h(\hat{z})\|_2^2 \quad (7)$$

式中, $h(\cdot)$ 代表编码变换。式(7)保证了 z 和 $\hat{z}$ 的内容相关性,可以防止网络解码过程中与图像无关信息的干扰。

4) Softmax 层的交叉熵损失。面部表情的真实结果 y 与表情的识别结果 $\hat{y}$ 间的损失函数为

$$L_S = \|k(y) - k(\hat{y})\|_2^2 \quad (8)$$

式中, $k(\cdot)$ 代表 Softmax 的交叉熵损失过程, $k(y)$ 代表真实结果, $k(\hat{y})$ 代表识别结果。式(8)可以降低网络损失,最终使得面部表情识别的结果更为精确。

综上,整体网络损失函数为

$$L = \omega_{\text{adv}} L_{\text{adv}} + \omega_{\text{con}} L_{\text{con}} + \omega_T L_T + \omega_S L_S \quad (9)$$

式中, ω_{adv} , ω_{con} , ω_T , ω_S 是调节损失的参数。

2.2 参数优化算法

基于梯度的优化算法有随机梯度下降(stochastic gradient descent, SGD)、AdaGrad(Cherry 等, 1998)、RMSProp(Hadgu 等, 2015)和 Adam(adaptive moment estimation)(Kingma 和 Ba, 2014)等多种。Adam 是动量梯度下降 Momentum 和 RMSprop 的结合,在 IMDB(the internet movie database)表情分类数据集上应用优化 logistic 回归算法,适用于卷积神经网络,其在偏差修正上体现出当梯度稀疏时比其他算法更加快速和优秀,在解决深度学习问题上更具有高效性。因此本文选用 Adam 优化方法更新网络的权重,使网络的损失值最小,根据经验选择产生最佳结果,算法具体步骤如下:

1) 对参数 v_{dw} 和 s_{dw} 初始化,利用动量梯度下降算法进行加权计算。具体为

$$v_{dw}^{[l]} = \beta_1 v_{dw}^{[l]} + (1 - \beta_1) \frac{\partial J}{\partial W^{[l]}} \quad (10)$$

$$s_{dw}^{[l]} = \beta_2 s_{dw}^{[l]} + (1 - \beta_2) \frac{\partial J^2}{\partial W^{[l]}} \quad (11)$$

式中, l 是隐层数, β_1 为动量梯度下降算法的超参数, β_2 为 RMSprop 算法的加权超参数, W 为权重, J 为损失函数。

2) 对加权后的梯度进行偏差纠正,具体为

$$v_{dw}^{\text{cor}} = \frac{v_{dw}^l}{1 - (\beta_1)^t} \quad (12)$$

$$s_{dw}^{\text{cor}[l]} = \frac{s_{dw}^{[l]}}{1 - (\beta_1)^t} \quad (13)$$

式中, t 为迭代次数,初始化时间为 0。

3) 采用 RMSprop 算法对梯度平方进行更新,同时纠正矩估计偏差。

4) 对算法梯度值权重更新,更新模型的参数,具体为

$$W^{[l]} = W^{[l]} - \alpha \frac{v_{dw}^{\text{cor}[l]}}{\sqrt{s_{dw}^{\text{cor}[l]} + \epsilon}} \quad (14)$$

式中, α 为步长, ϵ 为小常数,防止分母为 0。

3 实验

实验配置为 Inter Core i7-7700 CPU, 主频 3.6 GHz, GeForce GTX 1070Ti 显卡, 16 GB 内存, 操作系统 Ubuntu16.04, 平台 PyTorch (v0.4.0)。

3.1 实验数据

实验为真实环境中的面部表情识别。为更贴近真实环境,评价本文识别算法的有效性,构建了一个类内差距较大的混合数据集,以 Multi-PIE 和 JAFFE (the Japanese female facial expression database) 表情数据集为基础,从网络下载面部表情图像并进行样本扩充,自制所需的面部表情数据集,选择不同国家、不同年龄、不同职业的人群的 5 种面部表情进行实验,包括憎恶(abomination)、快乐(happy)、中立(neutral)、焦虑(anxious)、惊讶和恐惧(surprise and fear),增加了类内差距较大的面部表情特征的复杂度,样本总数为 4 452 幅,分为 5 类标签,如表 1 所示。

表 1 实验数据

Table 1 Experimental data

标签	数量/幅
憎恶	849
快乐	917
中立	1 052
焦虑	785
惊讶和恐惧	852

为解决样本数量少的问题,使用 OpenCV 对部

分不清晰表情图像进行预处理,通过旋转图像角度、调整色调、改变饱和度和曝光度等方法对样本集进行扩充,将模糊和受光线影响较严重的图像进行锐化和增强处理,然后将图像调整为 256×256 像素,以期完成面部表情识别实验。为验证实验的准确性,将数据集的 80% 作为训练集,20% 作为测试集。

3.2 网络训练

根据网络特点和研究者的经验设置网络参数,其中学习率为 0.000 2, $\beta_1 = 0.5$, $\beta_2 = 0.999$, epoch 分别为 100、200、300、400, batch 为 64。实验中利用以上参数对 IC-GAN 网络和传统的 DBN、GoogLeNet (Simonyan 和 Zisserman, 2014; Hinton, 2011) 网络进行训练,并结合本文的面部表情图像数据,应用 IC-GAN 模型进行实验验证。图 2 为 IC-GAN 模型中 epoch 最大设为 400 时的网络训练集 Loss 曲线图。

由图 2 可知,IC-GAN 模型在训练时达到了较好的训练效果, Loss 曲线缓慢下降,当经过 400 次迭代运算后,训练集 Loss 值下降到了 0.05 以下,网络达到了很好的识别效果。

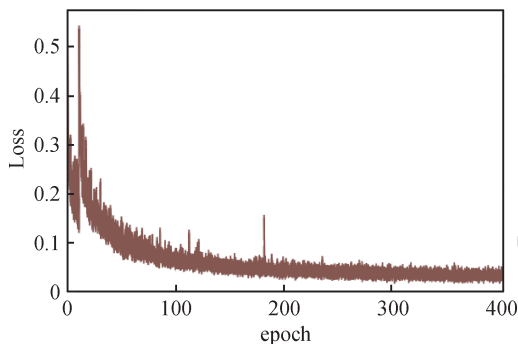


图 2 网络训练 Loss 曲线图

Fig. 2 Network training Loss curve

在本文预留的测试数据集上对训练好的 IC-GAN 模型进行测试,对于测试样本,某类表情的异常分数 $A(\hat{x})$ 定义为

$$A(\hat{x}) = \|G_E(x) - E(G(\hat{x}))\|_1 \quad (15)$$

式中, E 表示数学期望。

通过计算测试集 $\hat{D}$ 内的单个测试样本 x 的 $A(\hat{x})$ 来评估整体的性能,产生一组异常分数 $S = \{s_i: A(\hat{x}_i), \hat{x}_i \in \hat{D}\}$ 。应用特征缩放在 $[0, 1]$ 的概率范围内的异常分数为

$$s'_i = \frac{s_i - \min(S)}{\max(S) - \min(S)} \quad (16)$$

根据分类器对表情样本分别进行预测,设判别为正常表情的概率为 p_0 , 判别为不正常表情样本的概率为 p_1 , 则 $p_0 > p_1$ 的概率即为 AUC (area under curve), AUC 表示曲线 ROC (receiver operating characteristic) 下的面积,值的范围在 0.1 ~ 1 之间,最大分数对应的标签的 rank 记为 n , 次之记为 $n - 1$, M 代表正类分数为最小的值, N 为不正常表情样本, 则

$$A_{UC} = \frac{\sum_{s_i \in p-c} rank(s_i) - \frac{M \times (M + 1)}{2}}{M \times N} \quad (17)$$

式中, $p-c$ 代表对分数取平均值。

本文通过 AUC 值对比判断网络性能, AUC 值越大则分类越准确。图 3 是 epoch 为 400 时 3 个不同网络在数据集上的结果,其中 x 轴是网络的 AUC 值, y 轴是 5 种表情类别。通过图 3 可以看出, IC-GAN 在各个表情上的 AUC 值明显比 DBN 和 GoogLeNet 网络高, 其中对于类内差距较大的“惊讶和恐惧”类表情, BP 神经网络、DBN 网络和 IC-GAN 网络的 AUC 值都较低, IC-GAN 的 AUC 值明显高于其他两个网络, 因此 IC-GAN 网络在表情识别中达到了最佳的性能, 能更好地适应面部表情的识别任务。

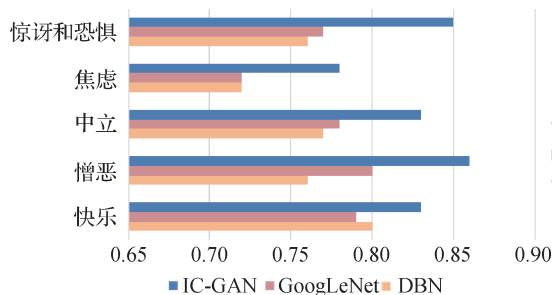


图 3 不同网络的训练 AUC 值

Fig. 3 Training AUC values for different networks

图 4 是 IC-GAN 输入潜在向量 z 中的元素个数分别为 32、100、256、512、1 024 时的 AUC 值, 其中横轴刻度上的 1、2、3、4、5 分别代表“快乐”、“憎恶”、“中立”、“焦虑”、“惊讶和恐惧”。由图 4 可知, 当输入潜在向量 z 的大小为 100 时, IC-GAN 网络的 AUC 值最高, 网络模型的性能最佳, 但是对于“焦虑”等负面类别表情, 训练集少, 且面部表情变化特征不明显, 易与中性表情特征相混淆, 模型的学习能力仍有欠缺, 所以识别准确率比较低。

图 5 是随着 z 值的变化, IC-GAN 网络的 AUC 值。由图 5 可知, z 中元素个数为 100 时, 网络的

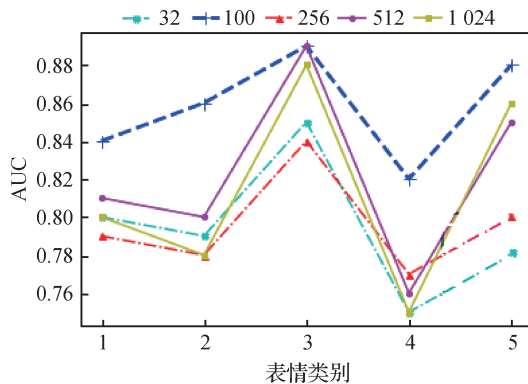


图 4 基于潜在向量 z 的不同大小模型的整体性能
Fig. 4 Overall performance of models based on different sizes of potential vectors z

AUC 值最高。实验证明 z 在 $0 \sim 100$ 时,网络的识别准确率逐渐增高,但当 z 值高于 100 时,网络的识别准确率反而呈现下降趋势,因此过多的潜在向量输入容易导致网络发生过拟合,引起识别率下降。

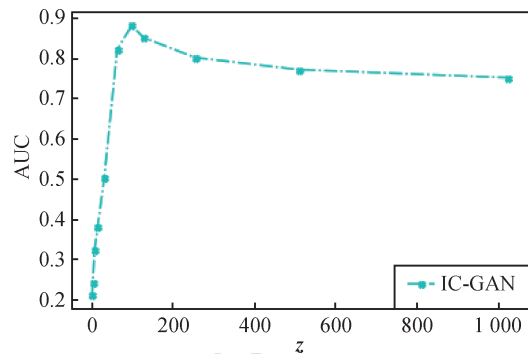


图 5 IC-GAN 网络在不同 z 值时的 AUC 值
Fig. 5 AUC value of IC-GAN network at different z

3.3 对比实验

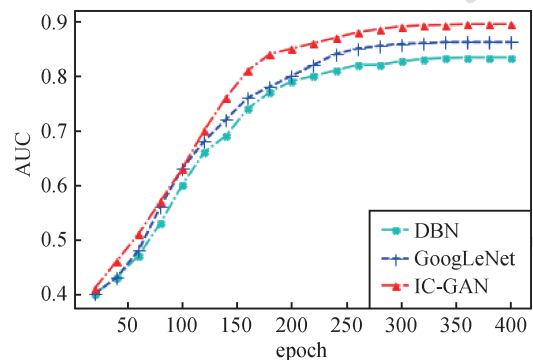
将 IC-GAN 模型与传统的 DBN 和 GoogLeNet 模型进行对比,验证本文构建的 IC-GAN 模型的性能,3 种网络的平均耗时对比如表 2 所示。由表 2 可知,本文构建的 IC-GAN 网络的平均耗时最少。

表 2 3 种网络的平均耗时对比
Table 2 Comparison of average time-consuming among three networks

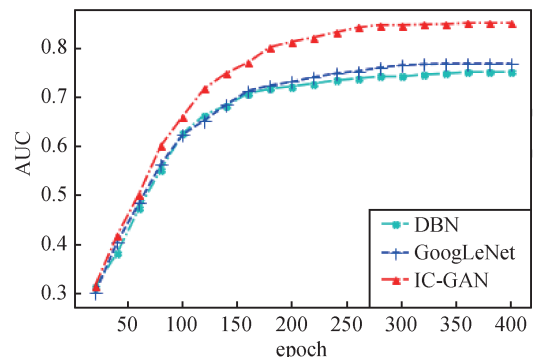
网络模型	平均耗时 (s/步)
DBN	0.077
GoogLeNet	0.039
IC-GAN	0.025

面部表情识别问题中的难点是存在类内差距、图像模糊和面部表情拍摄不全的情况,且识别过程中易受这些条件的影响,导致表情识别效果差,识别错误率高,而且面部表情识别任务与 ImageNet 数据集不同,不需要大型、极深的网络模型。对此本文主要在原始 GAN 网络的基础上,加入了面部识别中必须的条件,限制了原始图像 x 与重构图像 $\hat{x}$ 之间的相关性差距,构建了新型的 IC-GAN 网络,并使用 IC-GAN 网络模型与 DBN、GoogLeNet 模型在面部表情验证集上进行对比实验,分析并对比 IC-GAN 网络与 DBN、GoogLeNet 在面部表情识别的性能。面部表情识别验证集包括 700 幅图像,在网络输入尺寸为 256×256 像素时,IC-GAN 网络与 DBN、GoogLeNet 在面部表情识别验证集上的识别准确率如图 5 所示。

图 6 是 DBN、GoogLeNet 和 IC-GAN 网络在 Multi-PIE 类内差距较小的数据集和混合数据集上 5 种表情类别的识别结果。可以看出,在实验参数相同的情况下,不同网络对清晰且类内差距较小的



(a) 类内差距较小



(b) 类内差距较大

图 6 不同网络对面面部表情识别的准确率
Fig. 6 Accuracy of discriminating facial expression by different networks ((a) when intraclass gaps are smaller; (b) when intraclass gaps are larger)

面部表情图像的识别准确率不相上下,但是对于类内差距较大的面部表情,IC-GAN 网络对面表情识别的准确率远高于 DBN 和 GoogLeNet 对表情识别的准确率,表明 IC-GAN 网络的识别结果更加精准。

4 结 论

本文以 DCGAN 网络为基础,构建了用于面部表情识别的 IC-GAN 网络结构。将 IC-GAN 网络结构与深度学习中常见的识别网络 DBN、GoogLeNet 进行实验对比,结果表明,本文构建的面部表情识别模型在类内差距较大的面部表情识别任务中效果最好。在识别率方面,IC-GAN 网络识别准确率最高,比 DBN 和 GoogLeNet 网络分别高 11% 和 8.3%。在误识方面,IC-GAN 网络的误识率最低,识别精度能够满足实际应用的需要。但 IC-GAN 还是存在对部分负面表情以及受拍摄角度影响导致识别效果不好的情况。未来将在这方面进一步展开工作,继续进行以图像识别网络为基础的面部表情识别算法研究,提高网络对负面表情的识别率,尤其是对“焦虑”表情类别的识别率。

参考文献 (References)

- Bowles C, Chen L, Guerrero R, Bentley P, Gunn R, Hammers A, Dickie D A, Hernández M V, Wardlaw J and Rueckert D. 2018. GAN augmentation; augmenting training data using generative adversarial networks [EB/OL]. (2018-11-27) [2019-08-29]. <https://arxiv.org/pdf/1810.10863v1.pdf>
- Cao J M, Ni R R and Yang B. 2018. Binary-channel convolutional neural network for facial expression recognition. *Journal of Nanjing Normal University (Engineering Technology Edition)*, 18(3): 1-9 (曹金梦, 倪蓉蓉, 杨彪. 2018. 面向面部表情识别的双通道卷积神经网络. *南京师范大学学报(工程技术版)*, 18(3): 1-9) [DOI: 10.3969/j.issn.1672-1292.2018.03.001]
- Chen Y Z and Hu H F. 2019. An improved method for semantic image inpainting with GANs: progressive inpainting. *Neural Processing Letters*, 49(3): 1355-1367 [DOI: 10.1007/s11063-018-9877-6]
- Cherry J M, Adler C, Ball C, Chervitz S A, Dwight S S, Hester E T, Jia Y K, Juvik G, Roe T Y, Schroeder M, Weng S and Botstein D. 1998. SGD: saccharomyces genome database. *Nucleic Acids Research*, 26(1): 73-79 [DOI: 10.1093/nar/26.1.73]
- Cui C. 2017. Design and Research on Face Recognition System Based on Deep Convolutional Neural Network. Harbin: Harbin Institute of Technology (崔成. 2017. 基于深度卷积神经网络人脸识别系统设计与研究. 哈尔滨: 哈尔滨工业大学)
- Hadgu A T, Nigam A and Diaz-Aviles E. 2015. Large-scale learning with AdaGrad on spark//Proceedings of 2015 IEEE International Conference on Big Data. Santa Clara, United States; IEEE: 2828-2830 [DOI: 10.1109/BigData.2015.7364091]
- Hinton G. 2011. Deep belief nets//Sammut C, Webb G I, eds. *Encyclopedia of Machine Learning*. Boston: Springer: 1527-1554 [DOI: 10.1007/978-0-387-30164-8_208]
- Ioffe S and Szegedy C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-05-02) [2019-08-29]. <https://arxiv.org/pdf/1502.03167.pdf>
- Kingma D P and Ba J. 2014. Adam: a method for stochastic optimization [EB/OL]. (2014-01-30) [2019-08-29]. <https://arxiv.org/pdf/1412.6980.pdf>
- Liu M Y, Li S X, Shan S G and Chen X L. 2015. AU-inspired deep networks for facial expression feature learning. *Neurocomputing*, 159: 126-136 [DOI: 10.1016/j.neucom.2015.02.011]
- Simonyan K and Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2014-09-04) [2019-08-29]. <https://arxiv.org/pdf/1409.1556.pdf>
- Suárez P L, Sappa A D and Vintimilla B X. 2017. Infrared image colorization based on a triplet DCGAN architecture//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, HI, USA; IEEE: 212-217 [DOI: 10.1109/CVPRW.2017.32]
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA; IEEE: 1-9 [DOI: 10.1109/CVPR.2015.7298594]
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z. 2016. Rethinking the inception architecture for computer vision//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA; IEEE: 2818-2826 [DOI: 10.1109/CVPR.2016.308]
- Tang S J and Li P C. 2014. Image feature extraction in face recognition and research on matching technology. *Computer CD-ROM Software and Application*, 17(14): 133-133, 135 (唐守军, 李鹏程. 2014. 人脸识别中图像特征提取与匹配技术研究. *计算机光盘软件与应用*, 17(14): 133-133, 135)
- Xiao R Q. 2017. The Influence of Facial Expression and Body Gestures on Peak Emotion Perception. Shanghai: East China Normal University (肖瑞琪. 2017. 极端情绪下面部表情与身体姿势对表情识别的影响. 上海: 华东师范大学)
- Xu L L, Zhang S M and Zhao J L. 2017. Summary of facial expression recognition methods based on image. *Journal of Computer Applications*, 37(12): 3509-3516 (徐琳琳, 张树美, 赵俊莉. 2017. 基于图像的面部表情识别方法综述. *计算机应用*, 37(12): 3509-

3516) [DOI: 10.11772/j.issn.1001-9081.2017.12.3509]

Yao X. 2010. The Affecting Factors of Facial Expression Recognition: Expression Intensity and the Presentation. Jilin: Jilin University (姚雪. 2010. 面部表情识别的影响因素: 表情强度和呈现方式. 吉林: 吉林大学)

Yu Z D and Zhang C. 2015. Image based static facial expression recognition with multiple deep network learning//Proceedings of 2015 ACM on International Conference on Multimodal Interaction. Seattle: ACM: 435-442 [DOI: 10.1145/2818346.2830595]

Zhang Y M, Diao Q, Huang S and Hu W. 2003. DBN based multi-stream models for speech//Proceedings of 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing. Hong Kong, China; IEEE: 1520-6149 [DOI: 10.1109/I-CASSP.2003.1198911]

作者简介



陈亮, 1979年生, 男, 教授, 主要研究方向为人工智能技术、大数据分析和嵌入式仪表。

E-mail: kongkuchen@126.com



刘韵婷, 通信作者, 女, 副教授, 主要研究方向为无线传感器网络、人工智能技术和数据分析。

E-mail: liuyunting0224@163.com

吴攀, 女, 硕士研究生, 主要研究方向为人工智能技术和智能检测与信息处理。

E-mail: 1509481705@qq.com