

基于光照归一化和 ResNet18 的面部表情识别方法

马铭骏, 崔倩芳, 李 晓

浙江树人大学信息科技学院 浙江杭州

【摘要】真实的人机交互场景下, 人脸面部图像会受到光照等因素影响, 从而降低面部表情识别准确率。针对该问题, 提出了一种基于光照归一化和 ResNet18 的优化模型。采用直方图均衡化和线性变换加权求和的光照归一化方法对原图进行亮度平衡, 并利用预训练的 ResNet18 网络提取面部特征; 使用 Softmax 函数对面部表情结果进行预测。实验结果表明, 该网络模型在 RAF-DB 和 FERPlus 上分别取得 87.03% 和 87.46% 识别准确率。

【关键词】人机交互; 面部表情识别; 光照归一化; ResNet18

【基金项目】2021 年国家级大学生创新创业训练计划项目 (202111842016)

Illumination Normalization and ResNet18 for Facial expression recognition

Mingjun Ma, Qianfang Cui, Xiao Li

School of Information Technology, Zhejiang Shuren University, Hangzhou, Zhejiang

【Abstract】In the real human-computer interaction scene, facial images of human faces are affected by factors such as lighting, which reduces the accuracy of facial expression recognition. In view of the problem, an optimization model based on improved illumination normalization and ResNet18 was proposed. The illumination normalization method of histogram equalization and linear transformation weighted summation is used to balance the brightness of the original image, and extract facial features using the pre-trained ResNet18 network; Softmax function was used to predict the facial expression results. The experimental results show that the network model achieves 87.03% and 87.46% recognition accuracy on RAF-DB and FERPlus datasets.

【Keywords】Human-computer interaction; Facial expression recognition; Illumination normalization; ResNet18

引言

人脸面部表情是人类表达情感状态和意图最有力、最自然和最普遍的信号之一^[1]。面部表情识别技术在社交机器人、医疗诊断、疲劳监测等人机交互领域中有非常广泛的应用^[2]。据第七次全国人口普查数据显示, 我国 60 岁及以上人口比重达到 18.70%^[3], 人口老龄化程度已高于世界平均水平, 帮助空巢老人获得情感慰藉是当前社会重点关注的问题。一些研究人员将研究重心倾注在情感交互机器人上, 其原因为情感机器人能够从文字、语音、

人脸面部特征等多方面理解人类情感状态并与之交互, 从而能够在空巢老人急需情感帮助的情况下给予他们贴心的互动交流。然而在真实的人机交互场景下, 空巢老人通常会进行一系列动态行为(转头、行走、拿取物品、开关灯等), 这可能会导致机器人通过摄像头捕获到的面部图像受到遮挡、姿态变化、光照等因素影响, 从而降低面部表情识别的准确率。

本文提出了一种基于光照归一化和 ResNet18 的优化模型(illumination normalization and ResNet18,

作者简介: 马铭骏 (2002-) 浙江台州, 男, 计算机科学与技术;
崔倩芳 (1987-) 陕西蒲城, 女, 博士, 讲师, 研究方向: 遥感大数据分析, 医学数据分析与挖掘;
李晓 (1983-) 男, 汉, 研究生, 讲师, 主要从事数据科学与大数据技术专业。

ILLR), 通过平衡图像中人脸面部的亮度, 以更准确地提取细节特征。首先, 采用改进的光照归一化算法(Illumination Normalization, ILL)对原始面部图像进行处理, 削弱光照对表情识别的影响。紧接着, 利用预训练的 ResNet18 网络来提取图像的面部特征。最后, 使用 Softmax 函数对学习结果进行表情预测。本文第一节介绍真实环境下的面部表情识别算法的发展以及本文所采用的算法模块的相关工作, 第二节阐述 ILLR 模型的核心思想, 第三节在两个公共数据集上评估模型的可行性。

1 相关工作

1.1 面部表情识别

人脸面部表情识别主要包括三个阶段^[2]: 人脸图像收集与检测、面部特征提取和表情识别。当前大多数学者致力于研究准确面部特征提取的方法。文献^[4]提出了一种改进的 LeNet-5 卷积神经网络, 将提取到的低层次特征和高层次特征相结合, 取得了较好的结果。文献^[5]提出了三种不同类型的神经网络, 以此用于面部表情特征的提取。Li 等人^[6]提出了一种带有注意力机制的卷积神经网络(Attention Convolution Neural Network, ACNN), 其可以感知人脸的遮挡区域, 并关注最具判别性的未遮挡区域。Wang 等人^[7]提出了自修复网络(Self-Cure Network, SCN), 能够有效地抑制真实环境下面部表情识别的不确定性, 其原理是利用自注意力机制对每个训练样本进行排序正则化加权, 并采用重标签机制纠正权重最低的样本标签。上述研究均忽略了光照对面部特征提取的影响。

1.2 图像的光照归一化

同一个人表现出相同的面部表情, 在不稳定的光源环境下, 识别结果可能会产生较大的偏差^[2]。为了削弱光照的影响, 文献^[8]提出了双向重光照(Bidirectional Relighting)算法, 以最小化纹理之间的光照差异, 该方法具有更少的约束条件, 有助于更广泛地适用于一般光照归一化。文献^[9]通过构建表征皮肤表面、光源和摄像头传感器之间相互作用的底层反射模型, 生成对光照变化具有鲁棒性的色度本征图像, 能够有效地消除光照影响下的各种阴影。相关研究表明, 应用直方图均衡化的光照归一化可以得到更优越的识别性能。但是采用单一的直方图均衡化会导致图像的局部对比度过度增强, 为此,

Kuo 等人^[10]提出一种结合直方图均衡化和线性映射的加权求和方法, 使图像的全局和局部对比度达到平衡。

2 ILLR 模型

2.1 模型概述

用于面部表情识别的 ILLR 模型的总体架构如图 1 所示, 定义一个数据集 $D = \{X_i, y_i\}$, 其中 X_i 表示大小为 $3 \times H \times W$ 的输入面部表情图像, y_i 表示 C 类表情标签。首先, 对输入图像 X_i 其进行光照归一化方法得到光照适中图像 X_{ILL} , 这将在 2.2 节中详细介绍。然后, 将 ResNet18 网络作为特征提取主干网, 并在 MS-Celeb-1M 人脸数据集上预训练权重, 以此来提取光照适中图 X_{ILL} 的大小为 $C_{out} \times \frac{H}{DS} \times \frac{W}{DS}$ 的特征映射 X_{ILL}^* , 其中 C_{out} 为输出的通道数, DS 为下采样的倍率。紧接着, 使用分类头得到表情的概率分布。最后, 应用 Softmax 函数预测每张人脸面部图像的表情识别结果。

2.2 光照归一化

真实环境下摄像头捕捉的面部表情图像通常都会受到不同光照的影响, 这可能会阻碍模型的训练。为得到光照适中的面部图像, 我们采用了直方图均衡化^[11]和线性变换(将最小和最大像素映射到区间 $[0,1]$)加权求和的归一化方法。给定大小为 $3 \times H \times W$ 的面部表情图像 X_i , 首先对其进行两个不同的灰度变换操作: 直方图均衡化和线性变换, 计算方法如公式(1)(2)所示:

$$X_{iHE} = \text{concat}(\text{HE}(R), \text{HE}(G), \text{HE}(B)) \quad (1)$$

$$X_{iHN} = \text{concat}(\text{HN}(R), \text{HN}(G), \text{HN}(B)) \quad (2)$$

式(1)(2)中, concat 表示连接操作, (R, G, B) 表示图像 X_i 的三色彩通道, HE 表示直方图均衡化, HN 表示线性变换。紧接着, 将 X_{iHE} 和 X_{iHN} 按合适的融合权重组合成光照适中图像 X_{ILL} , 具体可表示为:

$$X_{ILL} = (1-\delta) \times X_{iHE} + \delta \times X_{iHN} \quad (3)$$

式(3)中, δ 表示权重因子, 在本论文中设置 $\delta = 0.5$

3 实验

3.1 数据集描述

RAF-DB^[12]是真实环境下收集的面部表情数据集, 包含由 315 名训练有素的工作人员标注的 29672 张面部图像, 其中每张面部图像由 40 名独立的工作

人员进行标注。RAF-DB 包含基本表情集和复合表情集, 本实验只使用了基本表情集, 其包含 7 种基本情绪(中性、快乐、惊讶、悲伤、愤怒、厌恶、恐惧)。将基本表情数据集中的面部图像分组得到 12271 个训练样本和 3068 个验证样本。样本的总体精度被用于模型评估。图 2 为 RAD-DB 数据集示例图像。

FERPlus^[13]是对 ICML 2013 挑战赛中 FER2013 数据集的扩展。其中每张面部图像由 10 名工作人员进行标注, 得到比 FER2013 数据集质量更好的标签。FERPlus 包含 8 种情绪(除 7 种基础情绪外, 还包括了轻视)。由于该数据集采用多标签的标注方法, 因此使用多数投票模式对标签进行处理, 从而得到 25045 个训练样本和 3191 个验证样本。样本的总体精度被用于模型评估。图 3 为 FERPlus 数据集示例图像。

3.2 实验环境及参数

本文采用 Pytorch 搭建 ILLR 模型, 并在单张 NVIDIA Tesla P100 GPU 显卡上完成了所有的实验。

本实验使用 MTCNN 网络^[14]检测面部图像中的人脸并裁剪, 将裁剪后的图像大小调整为 224×224 。设置批量大小为 16, 进行训练时, 采用 0.5 的概率对样本进行随机水平翻转, 以达到数据增强的效果。我们采用 ResNet18 作为主干网络, 将其在 MS-Celeb-1M 人脸识别数据集上进行预训练, 并通过其最后一个池化层提取所需的面部特征。在所有的数据集上, 学习率初始化为 0.001, 使用余弦退火学习率衰减策略, 学习率衰减周期设置为 5, 训练总轮次设置为 50。使用 Adam 优化器优化网络模型, 权重衰减值设置为 0.0001, 并利用交叉熵损失函数评估模型, 实验参数设置如表 1 所示。

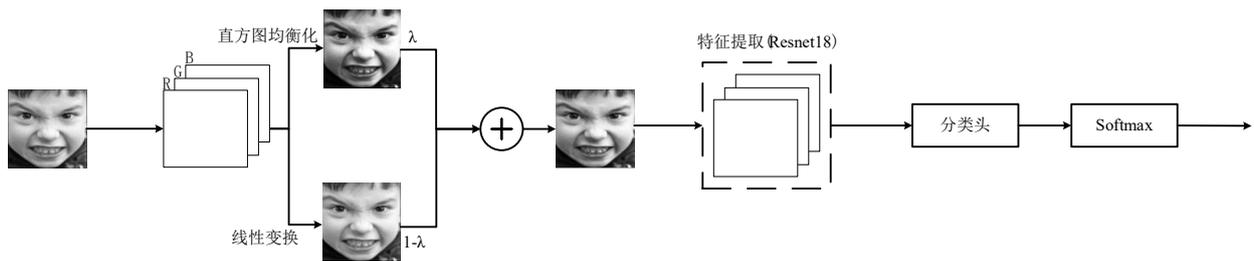


图 1 模型总体架构及模块

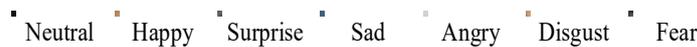


图 2 RAF-DB 数据集 7 种表情示例图像

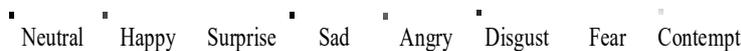


图 3 FERPlus 数据集 8 种表情示例图像

表 1 实验参数

参数类型	参数值
批量大小	16
图像大小	224×224
初始学习率	0.001
学习率衰减周期	5
训练总轮次	50
权重衰减	0.0001

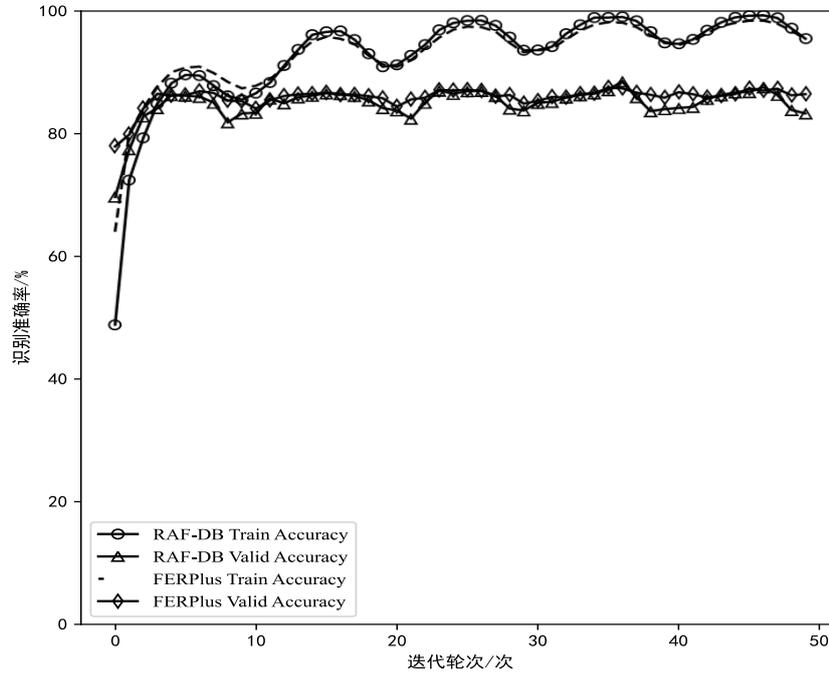


图 4 训练过程中各数据集的识别准确率曲线

表 2 RAF-DB 数据集上的消融实验

编号	ILL	ResNet18*	RAF-DB
a	×	×	61.02
b	×	√	86.15
c	√	×	61.67
h	√	√	87.03

表 3 RAF-DB 数据集上不同 θ 的对比实验

编号	θ	RAF-DB
a	0.3	86.57
b	0.4	86.96
c	0.5	87.03
d	0.6	86.99
e	0.7	86.41

3.3 实验结果可视化

本文的 ILLR 模型在 RAF-DB 和 FERPlus 面部表情数据集上得到的识别准确率曲线如图 4 所示。本文提出的模型在两个公共面部表情数据集上分别取得了 87.03%和 87.46%的识别准确率, 其中 X 轴表示训练迭代轮次, Y 轴表示表情识别准确率。

3.4 消融实验

为验证 ILLR 模型中各模块的有效性, 设置了四组在 RAF-DB 数据集上的不同实验并进行识别性

能对比, 实验配置与结果如表 2 所示。其中, 最优的识别结果以粗体标记, 编号(a)为基线方法(未预训练的 ResNet18), ResNet18*表示经过预训练的网络。

通过(a, b)对比可得, ResNet18 网络经过预训练后, 表情特征提取能力得到大幅度提升。由(a, c), (b, h)对比可知, 添加 ILL 模块时, 识别性能分别提升了 0.65%和 0.88%, 该结果表明 ILL 模块能够有效地中和对比度, 最小化光照对面部图像的影响, 从而提高模型的识别准确率。

该实验在 RAD-DB 数据集上设置了五组不同的 θ 进行识别精度对比, 变量设置与实验结果如表 3 所示, 其中最优的识别结果以粗体标记。由实验结果可知, 识别性能随着 θ 的增大而得到提升, 且当 $\theta = 0.5$ 时达到最优值。对照(c,d,e)三组实验可得, 当 θ 取值过高时对识别性能会有负面影响, 造成该现象可能的原因是 θ 取值过高使得直方图均衡化的比重上升, 从而过度增强图像的局部对比度, 导致部分细节特征难以提取。

4 结论

为解决真实的交互环境下面部图像受光照因素影响而导致准确率降低的问题, 本文提出了结合光照归一化和 ResNet18 的面部表情识别模型, 对光照适中图采用 RestNet18 网络以获得更多的细节面部特征。实验结果表明, 该模型在两个公开的面部表情数据集上均取得了较好的识别准确率。后续, 我们将研究如何把本文所提出的模型进一步应用于情感交互机器人中。

参考文献

- [1] Charles D. The Expression of the Emotions in Man and Animals [M]. New York City: Oxford University Press, 2002.
- [2] Shan Li, Weihong Deng. Deep facial expression recognition: A survey[J/OL]. IEEE Trans on Affective Computing, 2020, 2020(2): Article ID 2981446. (2020-03-17)[2022-01-18]. <http://doi.org/10.1109/taffc.2020.2981446>
- [3] 国家统计局, 国务院第七次全国人口普查领导小组办公室. 第七次全国人口普查公报~(1)(第五号) [N]. 中国信息报, 2021-05-12(002).
- [4] 李勇, 林小竹, 蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别 [J]. 自动化学报, 2018, 44(01): 176-182. (Li Yong, Lin Xiaozu, Jiang Mengying. Facial Expression Recognition with Cross-connect LeNet-5 Network [J]. Acta Automatica Sinica, 2018, 44(01): 176-182)
- [5] SHAO, Jie; QIAN, Yongsheng. Three convolutional neural network models for facial expression recognition in the wild. Neurocomputing, 2019, 355: 82-92.
- [6] Yong Li, Jiabei Zeng, Shiguang Shan, et al. Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism [J]. IEEE Transactions on Image Processing, 2018, 28(5): 2439-2450.
- [7] Kai Wang, Xiaojiang Peng, Jianfei Yang, et al. Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition [J]. IEEE Transactions on Image Processing, 2020(29): 4057-4069.
- [8] Ioannis A K, George T, Georgios E, et al. 3D-2D Face Recognition with Pose and Illumination Normalization [J]. Computer Vision and Image Understanding, 2017(154): 137-151.
- [9] Wuming Zhang, Xi Zhao, Jean-Marie M, et al. Improving Shadow Suppression for Illumination Robust Face Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(3): 611-624.
- [10] Chieh-Ming K, Shang-Hong L, Michel S. A Compact Deep Learning Model for Robust Facial Expression Recognition [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 2018: 2121-2129.
- [11] Haiping Wu, Bin Xiao, Noel C, et al. CvT: Introducing Convolutions to Vision Transformers [C]// Proc of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021: 22-31.
- [12] Shan Li, Weihong Deng, JunPing Du. Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2852-2861.
- [13] Emad B, Cha Zhang, Cristian C F, et al. Training deep networks for facial expression recognition with crowd-sourced label distribution [C]// Proc of the 18th ACM International Conference on Multimodal Interaction. 2016:279-283.
- [14] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks [J]. IEEE signal processing letters, 2016, 23(10): 1499-1503.

收稿日期: 2022 年 8 月 19 日

出刊日期: 2022 年 9 月 7 日

引用本文: 马铭骏, 崔倩芳, 李晓, 基于光照归一化和 ResNet18 的面部表情识别方法[J]. 国际计算机科学进展, 2022, 2(2): 21-25.
DOI: 10.12208/j. aics.20220017

检索信息: RCCSE 权威核心学术期刊数据库、中国知网 (CNKI Scholar)、万方数据 (WANFANG DATA)、Google Scholar 等数据库收录期刊

版权声明: ©2022 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。 <http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS