



深度残差网络和 LSTM 结合的图像序列表情识别

马玉环¹, 张瑞军¹, 武 晨¹, 屈军锁²

(1. 西安邮电大学 通信与信息工程学院, 西安 710121;

2. 西安邮电大学 自动化学院, 西安市先进控制与智能处理重点实验室, 西安 710121)

摘 要: 为了改善图像表情和图像序列表情识别效果, 针对传统表情识别特征提取复杂和效果不理想问题, 提出了一种深度残差网络和局部二值模式(local binary patterns, LBP)相结合的特征提取方法, 利用深度残差网络提取数据集的空域特征, 长短期记忆网络(long short-term memory, LSTM)处理时域特征, 实现空域与时域特征的结合。研究了不同层数的残差网络、不同形式的 LBP 算子以及其他网络结构对人脸表情识别的影响, 对比了支持向量机和随机森林实现的序列表情识别算法。在 Cohn-Kanade 数据集和 AFEW6.0 数据集上进行了验证, 实验结果表明, 算法在验证集上的准确率分别为 73.1% 和 58.4%, 相比其他算法有一定程度的提升。

关键词: 人脸表情识别; 深度残差网络; 长短期记忆网络

中图分类号: TP183

文献标志码: A

文章编号: 1673-825X(2020)05-0874-10

Expression recognition of image sequence based on deep residual network and LSTM

MA Yuhuan¹, ZHANG Ruijun¹, WU Chen¹, QU Junsuo²

(1. School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, P. R. China;

2. Xi'an Key Laboratory of Advanced Control and Intelligent Process, School of Automation, Xi'an University of Posts and Telecommunications, Xi'an 710121, P. R. China)

Abstract: In order to improve the effect of image expression and image sequence expression recognition, in view of the complexity and low efficiency of traditional expression recognition feature extraction, a feature extraction method combining deep residual network and local binary patterns (LBP) is proposed. The deep residual network extracts the spatial features, and the long short-term memory (LSTM) network processes the temporal features, which combines spatial and temporal features. The effects of different layers of residual networks, different forms of LBP operators and other network structures on facial expression recognition are studied. The support vector machine and random forest implementation are compared. The algorithm is verified on the Cohn-Kanade data set and the AFEW6.0 data set. The experimental results show that the accuracy of our algorithm on the Cohn-Kanade and AFEW6.0 verification sets are 73.1% and 58.4%, respectively. Compared with other algorithms, this algorithm has a certain degree of improvement.

Keywords: facial expression recognition; deep residual network; long short-term memory network

收稿日期: 2020-07-22 修订日期: 2020-09-18 通讯作者: 屈军锁 qujunsuo@xupt.edu.cn

基金项目: 陕西省重点研发计划国际合作项目(2018KW-026); 陕西省自然科学基金(2019JM-606)

Foundation Items: The Shaanxi Provincial Key R&D Program International Cooperation Project (2018KW-026); The Shaanxi Provincial Natural Science Foundation Project (2019JM-606)

©1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

0 引言

随着计算机技术的飞速发展,不同研究领域借助计算机的高运算能力都发生了翻天覆地的变革,当前都向着智能化的方向发展。其中,实现智能化的一个重要环节就是尽可能地模拟人类的思想活动过程以及洞察人类的情感状态,只有这样才能实现真正的智能化以及更好地实现人机交互。所以为了准确解读人类的情感,首先要完成的便是对人类情感的准确识别。

1971年,美国心理学家 Albert Mehrabian 教授提出在人类日常人际交往中,肢体语言在人际交往中占 55% 的比重^[1],这个数据说明人脸表情在人际交往中起着十分重要的地位。在日常交流中,我们通过面部表情可以表达自己的实际感受,并且推断别人的目的和态度。因此,面部表情识别(facial expression recognition, FER)是动态情感分析中的重要组成部分。面部表情识别方法将给定图像或图像序列中的面部表情分为 6 种基本情绪(愤怒、厌恶、恐惧、快乐、悲伤、惊讶或中性)^[2]。由于面部表情的复杂性和微妙性,准确识别面部表情仍然面临很大的困难。

到目前为止,人脸表情识别领域已经有了众多突出的研究成果,应用在生活中的人脸表情识别技术的身影随处可见,例如,安全驾驶、体感游戏、刑事侦查、远程教学等各个方面^[3]。关于人脸表情识别的早期研究主要集中在从静止帧识别表情^[4],这些方法有效地提取了空间信息,但是无法对形态和上下文因素的可变性进行建模。最近,一些研究尝试基于手工制作的特征或深度学习方法(例如 LBP-TOP^[5](local binary patterns from three orthogonal planes),HOG^[6](histogram of oriented gradient),STM^[7](spatio temporal manifold))从连续帧中捕获面部物理结构的动态变化和 DTAGN^[8](deep temporal appearance-geometry network)。

但是面部表情具有其自身的特征。特别地,面部表情可以被认为是关键部分(例如,眼睛、鼻子和嘴巴)的动态变化,这些关键部分融合在一起形成局部部分和整个面部的变化,而关键的挑战就变成了捕捉面部的这种动态变化。连续帧的物理结构。对于传统方法,很难基于手工描述符提取隐藏在面部图像中的强大时态特征。而且,将面部图像直接输入到深度神经网络中不能有效地利用如上所述的先验知识,这不利于学习表达的进化特性。

本文的主要内容是研究人脸表情的识别问题,提出了一种基于深度残差网络的静态图像人脸表情识别算法,并在静态表情识别的基础上实现了图像序列人脸表情的识别。本文的主要贡献有 3 方面:①研究了输入图像的预处理操作来提高人脸表情的识别率;②提出了一种使用深度残差网络作为骨干网络对静态人脸表情识别的算法;③提出了一种深度残差网络结合长短期记忆(long short-term memory, LSTM)网络实现图像序列人脸表情识别的算法。与其他图片序列表情识别算法进行了对比,实验结果表明,本文算法识别率有一定的提升。图 1 展示了面部表情识别的流程。



图1 面部表情识别流程

Fig.1 Facial expression recognition process

1 图像表情识别

1.1 算法框架

本文提出利用卷积神经网络将特征提取与表情分类等操作通过端对端的学习方式来实现,相比于传统的算法具有显著的优势。本节介绍的面部表情识别系统只在一个分类器上完成 FER 3 个学习阶段,该系统的整体框架如图 2。

1.2 图像预处理

1.2.1 人脸检测

本文使用 RetinaFace 来进行人脸检测。RetinaFace 对目前目标检测方向效果最好的 Faster RCNN 方法进行了改进,它使用了预训练模型,在训练过程中使用了多特征组合和多尺度图像策略来提高人脸检测的准确率^[3]。Retinaface 的网络结构如图 3。

1.2.2 人脸对齐

为了提取更加有效的人脸特征,在检测到人脸后采用基于面部 4 点的人脸对齐法进行对齐,这 4 个点分别是人脸的左眼中心和右眼中心以及左嘴角和右嘴角。

使用 OpenCV 自带的 Dlib 库中的人脸特征点检测算法得到人脸的 68 个特征点,然后使用眼部和嘴部特征点坐标进行人脸对齐,图 4 为人脸特征点检测结果。左眼外角标号为 36,左眼内角标号为 39,右眼外角标号为 45,右眼内角标号为 42,嘴左角标号为 48,嘴右角标号为 54。

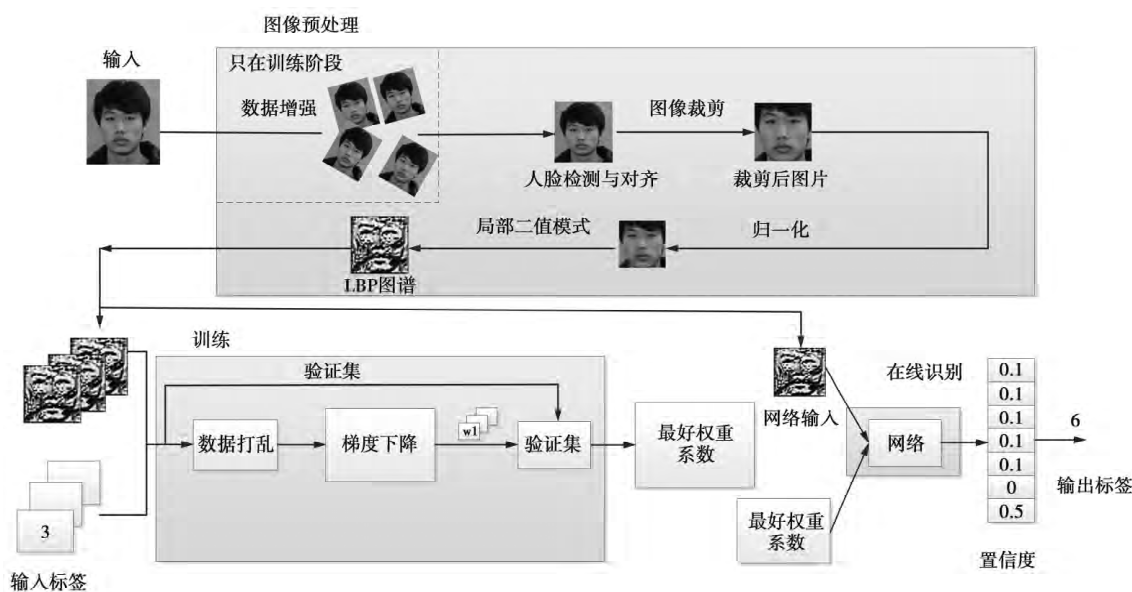


图 2 面部表情识别系统整体框架

Fig.2 Framework of facial expression recognition system

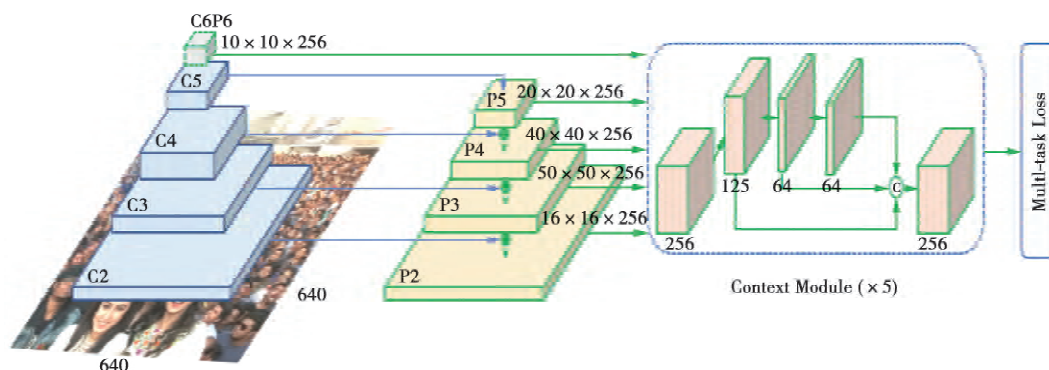


图 3 RetinaFace 网络结构

Fig.3 RetinaFace network structure



图 4 人脸的特征点

Fig.4 Feature points of the face

设左眼外角坐标为 (a_1, b_1) , 左眼内角坐标为 (a_2, b_2) , 右眼外角坐标为 (a_3, b_3) , 右眼内角坐标为 (a_4, b_4) , (x_1, y_1) 与 (x_2, y_2) 分别为左眼中心点与右眼中心点的空间坐标, 根据左眼中心点与右眼中心点的坐标, 脸部的倾斜角 θ 可由表示为

$$\tan \theta = \frac{y_2 - y_1}{x_2 - x_1} \quad (1)$$

为了使人脸图像中的左眼中心右眼中心位于同一条水平线上, 根据计算出的倾斜角 θ 对图像进行旋转, 最后计算出旋转后对应特征点的坐标。

按照上述算法流程对 FER2013 数据集进行人脸对齐, 图 5 为经过上述操作之后的示例。

1.2.3 人脸裁剪

人脸图像中除人脸之外的背景信息对于表达分类过程并不重要。裁剪没有表达特定信息的图像部分。因此, 定义了感兴趣的区域基于眼间距离的比率, 从而能够在没有人类干预的情况下处理不同的人脸图像大小。按图 6 所示进行裁剪, 裁剪系数为 0.2, 因为人脸具有对称性, 所以最终人脸图像只保留宽度的 0.6 倍。



图 5 人脸对齐示例

Fig.5 Face alignment example

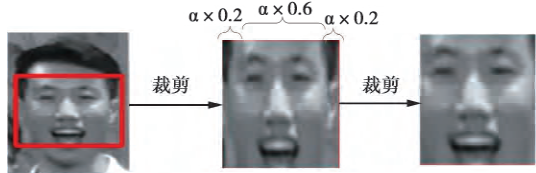


图 6 人脸裁剪示例图

Fig.6 Example image of face cropping

1.2.4 强度归一化

为了减少可变性从而使问题简单化,采用了强度归一化。根据生物激发技术,采用了一种对比均衡的方法^[9]。归一化的操作如下:首先,对图像进行了减法的局部对比度归一化;其次,对图像进行了分割的局部对比度归一化。

强度归一化后每个新的像素值表示为

$$y' = \frac{y - \mu_{nhy}}{\sigma_{nhy}} \quad (2)$$

(2) 式中: y 表示图像原始像素值; μ_{nky} 表示 y 的近邻的高斯加权平均; σ_{nky} 表示 y 的近邻标准差。

1.3 深度残差网络设计

深度残差网络是一种极深卷积神经网络模型,该网络模型可以避免简单堆叠卷积神经网络产生的梯度消失或爆炸以及精度退化问题,性能提升明显。本文将深度残差网络作为人脸表情识别的骨干网络,分别设计了 20、36、52 和 70 层的网络模型来进行实验分析。网络结构的详细信息如表 1,图 7 为 ResNet36 的网络结构。

表 1 网络结构详细信息

Tab.1 Network structure details

层数	输出尺寸/像素	20 层	36 层	52 层	70 层
Conv1	48×48		3×3 64, 步长 1		
Conv2_x	24×24	$\begin{bmatrix} 3 \times 3 & 64 \\ 3 \times 3 & 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 64 \\ 3 \times 3 & 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} \times 3$
Conv3_x	12×12	$\begin{bmatrix} 3 \times 3 & 128 \\ 3 \times 3 & 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 128 \\ 3 \times 3 & 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} \times 4$
Conv4_x	6×6	$\begin{bmatrix} 3 \times 3 & 256 \\ 3 \times 3 & 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 256 \\ 3 \times 3 & 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{bmatrix} \times 12$
Conv5_x	3×3	$\begin{bmatrix} 3 \times 3 & 512 \\ 3 \times 3 & 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 & 512 \\ 3 \times 3 & 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2048 \end{bmatrix} \times 3$
Fc1	1×1		均值池化 4 096-d fc		
Fc2	1×1		均值池化 2 048-d fc		
Fc3	1×1		均值池化 7-d fc, Softmax		

本文中的 CNN 框架包含多个卷积层、残差块和全连接层。每一个卷积层之后加入了 BatchNorm 层,用来增强模型的泛化能力。3 个全连接层是 Fc1、Fc2 和 Fc3 层。其中,在 Fc1 层和 Fc2 层后分

别加入了 Dropout 层,用来防止网络的过拟合;在 Fc3 层之后是 Softmax 回归层,用于得到每一类的置信度,然后根据置信度输出最终分类结果。Softmax 的数学表达式为

该 LSTM 网络结构包括 1 个 LSTM 网络单元,1 个 Dropout 层 2 个全连接层。流程如下:首先,将输入的图像序列提取序列特征后连接一个 LSTM 网络单元;然后,使用平坦操作将 LSTM 网络单元输出的序列特征展开为一维特征,并且在后序操作中添加 Dropout 层来防止过拟合;最后,将全连接层得到的特征向量使用 Softmax 进行分类得到表情类别。

3 实验分析

3.1 ResNet 网络参数设置

在静态图像人脸表情识别算法中使用的是深度残差网络,其网络结构中使用的是 Relu 非线性激活函数^[10]。所以,采用专门为网络中使用 Relu 作为激活函数设计的 MSRA 初始化方法^[11]。模型的优化过程使用随机梯度下降法进行迭代,部分超参数设置如表 2。

表 2 ResNet 网络参数设置
Tab.2 ResNet network parameter settings

参数	参数值
Learning rate	0.001
Epochs	100
Loss function	Cross entropy
Metrice	Accuracy
Batch_size	64

3.2 LSTM 网络参数设置

在训练 LSTM 模型常用的算法是时间反向传播^[11](back-propagation through time,BPTT),本质上仍为 BP 算法。

在对网络进行优化求解的过程,时间反向传播算法的核心是沿着需要优化的参数的负梯度方向不断迭代来寻找最优的解,直到模型收敛或到达最终条件。网络训练过程的参数影响模型优化效果的好坏,本文中 LSTM 网络训练过程的部分超参数具体设置如表 3。

表 3 LSTM 网络参数设置
Tab.3 LSTM network parameter settings

参数	参数值
Learning rate	0.001
Epochs	100
Loss function	Cross entropy
Metrice	Accuracy
Batch_size	16

3.3 图像表情算法比较

3.3.1 不同层数残差网络的影响实验与分析

为了研究不同层数的深度残差网络对人脸表情识别的影响,分别选取了 20,36,52 和 70 层的残差网络进行模型的训练。在实验过程中,使用 FER2013 数据集,图片经预处理后输入到深度神经网络。不同层数的网络训练过程的损失函数如图 10,可以从迭代次数与损失函数图像看出,36 层的深度残差网络模型收敛效果最好。

将训练好的模型在测试集上进行模型验证,图 11 为不同层数残差网络在 FER2013 测试集上的表情分类结果,图 11 中给出了不同层数的网络的每一类准确率以及整体准确率。

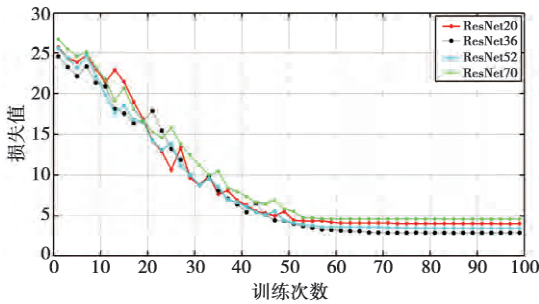


图 10 训练过程损失值变化曲线

Fig.10 Change curve of loss value during training

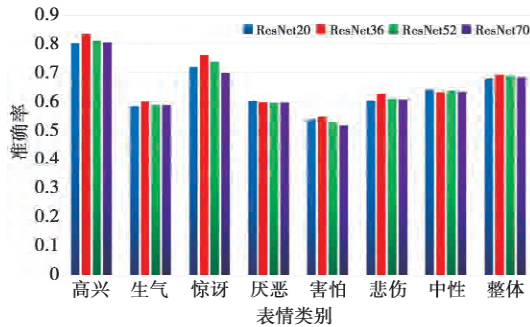


图 11 不同层数残差网络表情分类准确率对比

Fig.11 Comparison of the accuracy of different layers of residual network expression classification

从图 11 可以看出,36 层的残差网络比 20 层的残差网络的准确率高,但随着网络层数的加深,表情识别的准确率出现了下降的退化现象。实验结果表明,36 层的深度残差网络在 FER2013 数据集上具有较高的识别率。

3.3.2 不同 LBP 算子的影响实验与分析

1994 年,Ojala 等首次提出了 LBP 算法,该算法可以表示区域图像的的平坦、边缘、角点等纹理特征。由于 LBP 对光照的变化不敏感、求解过程计算

量小和特征辨别力强等,所以在人脸表情识别的特征提取中经常用 LBP 算法来提取脸部纹理特征。

不同的 LBP 算子具有不同的功能并在特征表示方面有各自的优势,为了选取合适的 LBP 算子,对不同 LBP 算子作为神经网络的输入对实验结果的影响进行分析。发现 36 层的残差网络特征提取能力较强,具有较高的识别率;所以,选取 36 层的残差网络作为骨干网络,分析了不同的 LBP 算子对人脸表情分类结果的影响,其中,圆形 LBP 的 $R=1$, $P=8$,实验仍在 FER2013 数据集上进行。

不同 LBP 算子表情分类准确率对比如图 12,由图 12 的实验结果表明,使用原始 LBP 算法,圆形 LBP 算法和旋转不变 LBP 算法的识别率分别为 69.80%、70.56% 和 72.40%。在其他预处理操作相同的情况下,使用 36 层残差网络作为人脸表情识别的骨干网络,旋转不变 LBP 算子相比较传统的 LBP 算子和圆形 LBP 算子在 FER2013 上具有更高的识别率。

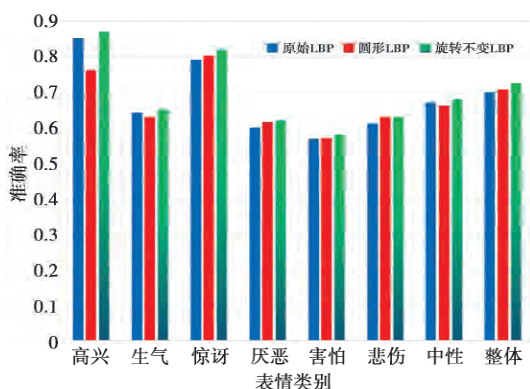


图 12 不同 LBP 算子表情分类准确率对比

Fig.12 Comparison of expression classification accuracy of different LBP operators

3.3.3 不同卷积神经网络的影响实验与分析

为了对比其他算法对识别结果的影响,选取 VGG19 和 ResNet36 进行实验对照,同时加入 Dropout 层分析对实验的影响,分类器的选取使用支持向量机(support vector machine, SVM) 和 Softmax。不同算法在 FER2013 的识别率如表 4,由表 4 可见,使用 ResNet36 作为骨干网络比使用 VGG19 的准确率高;使用 Dropout 策略可以提高模型的准确率,有效地避免过拟合的问题;使用 Softmax 进行分类比使用 SVM 的分类效果好。

3.3.4 FER2013 数据集上的实验与分析

通过上述分析,本文中人脸表情识别系统选取 36 层的残差网络作为骨干网络,选取旋转不变 LBP

算子,使用数据增强将数据集扩大 10 倍,模型加入 Dropout 策略,人脸表情分类使用 Softmax 函数。本算法与其他相关算法的对比如表 5,表 5 显示了本算法对 FER2013 基线的性能,基线包括深度神经网络(deep neural networks, DNN) 和时延神经网络(time-delay neural network, T-DNN) 等,可以发现,本算法静态表情的识别率最高,证明了本算法的准确率高、鲁棒性好,在未来还有一定的提升空间。

表 4 不同算法在 FER2013 的识别率

Tab.4 Recognition rate of different algorithms in FER2013 dataset

模型	准确率/%
VGG19+Softmax	71.20
VGG19+Dropout+Softmax	71.88
VGG19+SVM	70.54
VGG19+Dropout+SVM	71.65
ResNet36+Softmax	72.40
ResNet36+Dropout+Softmax	73.00
ResNet36+SVM	72.12
ResNet36+Dropout+SVM	72.50

表 5 FER2013 数据集上最新方法的识别率

Tab.5 Recognition rate of the latest method in the FER2013

方法	准确率/%
Unsupervised ^[12]	69.26
Maxim ^[13]	68.82
Tang ^[14]	71.16
Mollahosseini ^[15]	66.40
Liu ^[16]	65.03
DNNRL ^[17]	70.60
FC3072 ^[18]	70.58
CPC ^[19]	71.35
本文算法	73.10

3.4 图像序列算法比较

3.4.1 Cohn-Kanade 数据集上的实验与分析

为验证本文算法的有效性,选取 Cohn-Kanade 数据集来进行图像序列人脸表情识别实验,该数据集是国际常用标准表情测试数据库。在 Cohn-Kanade 表情库上选取了 700 幅完整的图像序列对本文算法进行实验,随机选取 70% 作为训练集,其余 30% 作为测试集。

为了可视化分析各类表情的识别情况,图 13 绘制了 Cohn-Kanade 测试集序列表情识别结果的混淆矩阵,可以看出,“愤怒”、“恐惧”和“悲伤”之间容

易出现混淆。

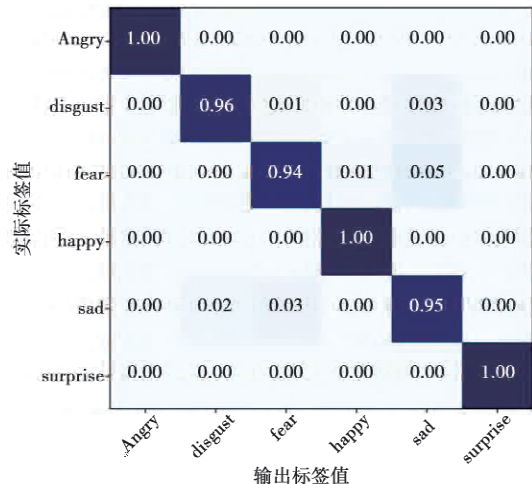


图 13 Cohn-Kanade 数据集的测试集识别结果混淆矩阵
Fig.13 Confusion matrix of the test set recognition result of the Cohn-Kanade dataset

此外,为了对比分析实验结果,将本文提出的算法与其他几种算法进行比较,在分类时序特征还选取了 SVM 和随机森林算法进行人脸表情的识别。在使用 SVM 算法和随机森林算法时,首先将时序特征展开为一维特征,得到 20 480 的特征维度,然后使用主成分分析法将数据降维到 1 000 特征维度,最后将提取到的 1 000 特征维度作为时序特征。对于支持向量机,选取核函数为高斯核;对于随机森林算法,森林中树的个数最大为 100 颗。3 种算法的实验结果对比如图 14。

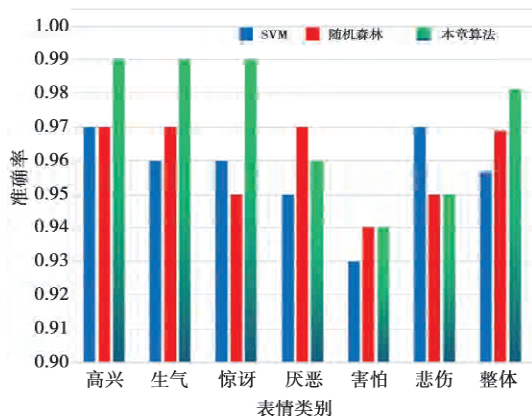


图 14 不同分类算法每种表情的识别准确率对比
Fig.14 Comparison of the recognition accuracy of each expression of different classification algorithms

由图 14 可知,LSTM、SVM 和随机森林的平均准确率分别为 98.12%、95.68% 和 96.88%,LSTM 相对于使用 SVM、随机森林进行序列特征的分类具有更

高的准确率。由此可以说明本算法的有效性,在处理时序问题时使用长短期记忆网络相对于只是简单的使用静态特征具有更好的效果。

3.4.2 AFEW6.0 数据集上的实验与分析

为了验证算法的有效性,在 AFEW6.0 数据集数据集上进行了相关实验。本节的实验结果都是在验证集上进行分析对比。经过模型迭代训练,实验结果表明,本文算法在 AFEW6.0 数据集的验证集上的表情识别准确率为 58.40%。

图 15 为 AFEW6.0 数据集中的验证集识别结果混淆矩阵,从图 15 可以看出,“高兴”、“生气”识别率较高,“厌恶”的识别率最低。其中,“愤怒”、“恐惧”和“悲伤”在识别过程中经常会出现混淆,这与现实中的情形是相吻合,这 3 类都是使我们容易混淆的表情。同样,从混淆矩阵中可以看出,“自然”很容易被归类为“悲伤”。

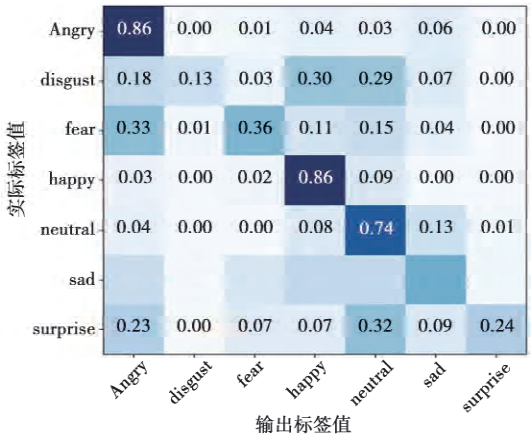


图 15 AFEW6.0 数据集的验证集识别结果混淆矩阵
Fig.15 Confusion matrix of the recognition result of the validation set of the AFEW6.0 dataset

除了本文算法以外,对分类时序特征还选取了 SVM 和随机森林算法来验证不同的分类器对实验结果的影响。经过模型训练后,使用 SVM 算法在 AFEW6.0 验证集上的准确率为 53.53%,使用随机森林算法在 AFEW6.0 验证集上的准确率为 55.36%,图 16 为 3 种算法的实验结果对比图。

为了分析本算法的有效性,选取了在 EmotiW2016 挑战赛中取得高准确率的算法与本算法进行比较,表 6 为不同算法的对比结果。分析可见,本文算法相较于基线准确率提高了 19.59%。回顾当年比赛的排行榜可以发现,本文算法在验证集上相比第 1 名(Fan)高出了 6.44%,相比第 2 名(Yao)也高了 6.44%。

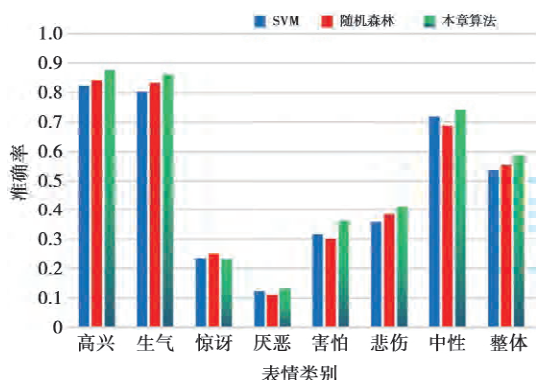


图 16 不同分类算法每种表情的识别准确率对比

Fig.16 Comparison of the recognition accuracy of each expression of different classification algorithms

表 6 AFEW6.0 数据集中验证集的准确率

Tab.6 Accuracy of the validation set in the AFEW6.0 dataset

方法	准确率/%
Baseline	38.81
Fan ^[20]	51.96
Yao ^[21]	51.96
Li M ^[22]	53.27
Bargal ^[23]	59.42
SVM	53.53
随机森林	55.36
本文算法	58.40

该算法与 Bargal 使用的算法进行比较还是处于一定的劣势,究其原因发现其使用了 3 种网络进行融合并且使用了大量的额外数据,而本文算法没有额外扩充数据集并且只是使用了单一的网络,所以其模型的数据维度以及模型的复杂度都不及 Bargal 使用的算法。但是从目前的结果看,本文算法相对其他图像序列表情的识别算法还是有一定的提升,说明将图像序列人脸表情提取时序特征后使用 LSTM 进行分类处理具有一定的准确率优势。

4 结 论

本文研究了使用深度残差网络来进行人脸表情的识别问题,在静态图像人脸表情识别的基础上提出了图像序列人脸表情识别的算法,提出了基于深度残差网络和 LSTM 结合图像序列人脸表情识别算法,分别在 FER2013、Cohn-Kanade 和 AFEW6.0 数据集上进行了算法验证,实验结果表明,在平均识别精度方面,本文算法在处理序列特征上具有较高的准确率。

参考文献:

- [1] MEHRABIAN A, RUSSELL J A. An approach to environmental psychology [M]. Massachusetts: MIT, 1974.
- [2] EKMAN P. Contacts across cultures in the face and emotion [J]. J Pers Soc Psychol, 1971, 17(2): 124-129.
- [3] 王玉琪. 基于深度学习的人脸情绪识别研究 [J]. 电子制作, 2020(14): 36-37, 63.
WANG Y Q. Research on facial emotion recognition based on deep learning [J]. Electronic Production, 2020(14): 36-37, 63.
- [4] PANTIC M, ROTHKRANTZ L J M. Automatic analysis of facial expressions: The state of the art [J]. IEEE Trans Pattern Anal Mach Intell, 2000, 22(12): 1424-1445.
- [5] SHAN C, GONG S, MCOWAN P W. Facial expression recognition based on local binary patterns: A comprehensive study [J]. Image Vis Comput, 2009, 27(6): 803-816.
- [6] KLASER A, MARSZALEK M, SCHMID C. A Spatio-Temporal Descriptor Based on 3D-Gradients [C]//BMVC 2008-19th British Machine Vision Conference. Leeds, United Kingdom [s.n.], 2008: 1-10.
- [7] LIU M, SHAN S, WANG R, et al. Learning expression-lets on spatio-temporal manifold for dynamic facial expression recognition [C]//in Proc. IEEE Conf Comput Vis Pattern Recognit (CVPR). New York: IEEE, 2014: 1749-1756.
- [8] JUNG H, LEE S, YIM J, et al. Joint fine-tuning in deep neural networks for facial expression recognition [C]//in Proc. IEEE Int Conf Comput Vis (ICCV). New York: IEEE, 2015: 2983-2991.
- [9] MASE K, PENTLAND A. Automatic lipreading by optical-flow analysis [J]. Systems and Computers in Japan, 1991, 22(6): 67-76.
- [10] CHANG T, WEN G, HU Y, et al. Facial Expression Recognition Based on Complexity Perception Classification Algorithm [J]. ArXiv Preprint ArXiv, 2018(1803): 00185.
- [11] SHIN M, KIM M, KWON D S. Baseline CNN structure analysis for facial expression recognition [C]//2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). New York: IEEE, 2016: 724-729.
- [12] YANG H, CIFTCI U, YIN L. Facial expression recognition by de-expression residue learning [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 2168-2177.
- [13] 张璐璐. 基于卷积神经网络的人脸面部表情识别方法

- 研究[D].石家庄:河北科技大学,2019.
- ZHANG L L. Research on facial expression recognition method based on convolutional neural network [D]. Shijiazhuang: Hebei University of Science and Technology, 2019.
- [14] GOODFELLOW I J, ERHAN D, CARRIER P L, et al. Challenges in representation learning: A report on three machine learning contests [C]//International Conference on Neural Information Processing. Berlin: Springer, 2013: 117-124.
- [15] MENG YU X, ZHENMIN T, YAZHOU Y, et al. Deep Learning for Person Reidentification Using Support Vector Machines [J]. Advances in Multimedia, 2017(2017): 1-12.
- [16] MOLLAHOSSEINI A, CHAN D, MAHOOR M H. Going deeper in facial expression recognition using deep neural networks [C]//2016 IEEE Winter conference on applications of computer vision (WACV). New York: IEEE, 2016: 1-10.
- [17] LIU K, ZHANG M, PAN Z. Facial expression recognition with CNN ensemble [C]//2016 international conference on cyberworlds (CW). New York: IEEE, 2016: 163-166.
- [18] GUO Y, TAO D, YU J, et al. Deep neural networks with relativity learning for facial expression recognition [C]//2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). New York: IEEE, 2016: 1-6.
- [19] KIM B K, ROH J, DONG S Y, et al. Hierarchical committee of deep convolutional neural networks for robust facial expression recognition [J]. Journal on Multimodal User Interfaces, 2016, 10(2): 173-189.
- [20] FAN Y, LU X, LI D, et al. Video-based emotion recognition using CNN-RNN and C3D hybrid networks [C]//Proceedings of the 18th ACM International Conference on Multimodal Interaction. New York: IEEE, 2016: 445-450.
- [21] YAO A, CAI D, HU P, et al. HoloNet: towards robust emotion recognition in the wild [C]//Proceedings of the 18th ACM International Conference on Multimodal Interaction. New York: IEEE, 2016: 472-478.
- [22] LI M, WANG Z. Dynamic Facial Expression Recognition Based on Trained Convolutional Neural Networks [C]//Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Cham: Springer, 2018: 218-226.
- [23] BARGAL S A, BARSOUM E, FERRER C C, et al. Emotion recognition in the wild from videos using images [C]//Proceedings of the 18th ACM International Conference on Multimodal Interaction. New York: IEEE, 2016: 433-436.

作者简介:



马玉环(1995-)男,陕西延安人,硕士研究生,主要研究方向为物联网技术与应用。
E-mail: 1012424062@qq.com。



张瑞军(1992-)男,陕西西安人,硕士研究生,主要研究方向为物联网技术与应用。
E-mail: 472780551@qq.com。



武晨(1993-)男,陕西西安人,硕士研究生,主要研究方向为物联网技术与应用。
E-mail: 329277047@qq.com。



屈军锁(1968-)男,陕西澄城人,教授,硕士。西安邮电大学自动化学院副院长,陕西省科技厅“双导师制”科技创新创业平台主任,西安市先进控制与智能处理重点实验室主任,中移物联网有限公司-西安邮电大学物联网创新联合实验室主任。1991年毕业于重庆邮电大学,获学士学位。主要研究方向为物联网技术、视频语义理解。
E-mail: qujumsuo@xupt.edu.cn。

(编辑:王敏琦)