基于时空注意力的多视角人脸表情识别算法

杜睿山，王紫珊

（东北石油大学 计算机与信息技术学院，黑龙江 大庆 163318）

摘 要：首先，利用肤色分割技术定位学生图像中的脸部区域，并将定位的脸部区域输入到时空注意力模块中，以获得脸部多视角的关键信息。其次，通过带权重衰减的自适应梯度下降算法对卷积神经网络中的参数展开优化，并将脸部关键信息输入到优化后的网络中，以确定学生脸部表情类型，完成多视角人脸表情识别。实验结果表明，应用本文算法可以精准地提取人脸的关键信息，且表情识别准确率为100%，即本文算法可以有效识别人脸，并提高人脸表情识别精度。

关键词：时空注意力；人脸表情识别；肤色分割；人脸定位；卷积神经网络

中图分类号：TP391.41 **文献标志码**：A **文章编号**：1671-5497(2025)06-2097-06

DOI：10.13229/j.cnki.jdxbgxb.20240582

Multi perspective facial expression recognition algorithm based on spatiotemporal attention

DU Rui-shan，WANG Zi-shan

（*School of Computer and Information Technology*，*Northeast Petroleum University*，*Daqing* 163318，*China*）

Abstract：Firstly，skin color segmentation technology was used to locate facial regions in student images，and the located facial regions were input into the spatiotemporal attention module to obtain key information from multiple perspectives of the face. Secondly，the parameters in the convolutional neural network were optimized using an adaptive gradient descent algorithm with weighted decay，and key facial information was input into the optimized network to determine the types of facial expressions of students and complete multi view facial expression recognition. The experimental results show that the proposed algorithm can accurately extract key information of the face，and the accuracy of facial expression recognition is 100%. Therefore，the proposed algorithm can effectively recognize faces and improve the accuracy of facial expression recognition.

Key words：spatiotemporal attention；facial expression recognition；skin color segmentation；facial localization；convolutional neural networks

收稿日期：2024-05-24.

基金项目：黑龙江省教育科学规划重点项目(GJB1320039)；国家重点研发计划项目(2022YFE0206800).

作者简介：杜睿山(1977-)，男，副教授，博士. 研究方向：知识图谱，机器学习. E-mail：ruishan_du@163.com

0 引 言

人脸表情识别技术是指通过分析和解读人脸图像中的表情信息,来理解和推断人的情感状态。该技术已广泛应用于人机交互、智能监控、医疗诊断和远程教育等多个领域[1]。然而,由于人脸表情的变化复杂多样,并且表情识别经常受到光照、遮挡以及姿态变化等环境因素的干扰,导致人脸图像的清晰度较低,从而影响了表情识别的精度[2],为此,亟需对人脸表情识别技术展开研究。

文献[3]首先通过人脸对齐网络定位人脸表情区域;然后,基于欧拉视频放大技术获取面部动作变化,并对光流信息展开提取以得到视频序列特征;最后,利用双分支分类网络实现表情识别。但是,该算法中人脸对齐网络的泛化能力有限,增大了人脸定位的误差。文献[4]首先构建包含不同年龄段人脸表情的数据库,并根据该数据库获取人脸、眼睛和嘴巴区域,其次对这些区域展开特征提取,最后根据特征对表情识别的影响计算各特征权值,通过对特征展开加权融合完成表情识别。但是该算法构建数据库的计算成本较大,影响了人脸定位的效率。文献[5]对K5_Light_ShuffleNet网络展开分析和剪裁以优化网络性能,将轻量化通道空间关键权重推断模块加入网络中,利用该网络获取人脸特征,最后采取标签平滑学习算法实现人脸识别。但是,该算法中K5_Light_ShuffleNet网络受环境因素干扰较大,导致特征提取的性能较低。

由于人脸是3D生物体征,而大多数人脸识别系统基于2D图像进行识别,当人脸角度与事先采集的数据库中的人脸角度相差较大时,识别效果可能会受到影响。因此,本文提出基于时空注意力的多视角人脸表情识别算法,通过考虑多视角,利用时空注意力模型获取人脸在不同角度下的特征信息,从而提高识别的准确度。

1 人脸区域定位

由于人脸是一个3D对象,当人脸的角度相对于摄像头发生变化时,人脸在图像中的形状、大小和纹理等特征都会发生变化。特别是当人脸偏转角度较大时,表情特征不完整,会增加识别的难度。由于肤色特征在不同视角下变化较小,且人脸肤色和图像中的其他区域颜色相差较大[6],通过肤色分割可以在预处理阶段快速定位人脸区域,降低后续特征提取和识别的计算复杂度。因此,通过肤色区域的分割[7]来实现人脸区域定位。

YcbCr颜色空间是一种亮度和色度分离的空间,该空间能够限制学生脸部肤色的分布区域,提高脸部肤色的聚类特性。因此,将采集到的课堂学生图像从RGB空间转换到YcbCr空间,并在YcbCr空间完成脸部定位。

将课堂学生图像从RGB空间转换到YcbCr空间[8]的过程为:

$$\begin{bmatrix} Y \\ Cb \\ Cr \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} R \\ G \\ B \\ 1 \end{bmatrix} \tag{1}$$

式中:Y为亮度;Cb和Cr分别为蓝色和红色色度分量;λ为像素范围;R、G、B分别为红色、绿色和蓝色分量。

在实际应用中很难保证所有视角下的光照条件和遮挡情况都相同。肤色高斯分布模型在一定程度上可以克服光照和遮挡对识别结果的影响,所以利用肤色高斯分布模型[9]对YcbCr空间中的课堂学生图像展开转换。学生脸部肤色的高斯分布模型$W(q, V)$为:

$$\begin{cases} q = R\{c\} \\ V = R\{(c-q)(c-q)Y\} \end{cases} \tag{2}$$

式中:$c=(Cr, Cb)^Y$为YcbCr空间中学生图像的任意像素;q和V分别为YcbCr空间中学生图像像素的平均值和协方差。

图像中像素点c属于学生脸部皮肤区域的概率为$A(Cb, Cr)$,该概率值反映了灰度图像中像素点c和学生脸部肤色的相似程度,其表达式为:

$$A(Cb, Cr) = r^{\left[-0.5(c-q)^Y V^{-1}(c-q)\right]} \tag{3}$$

式中:r为固定值。

设置概率阈值α,将$A(Cb, Cr)\geqslant\alpha$的像素点判定为学生脸部肤色像素点,$A(Cb, Cr)<\alpha$的像素点判定为非脸部肤色像素点,以此获得学生脸部肤色和非脸部肤色分离后的二值图像。

在不同视角下,人脸的角度和位置可能有所不同,膨胀处理能够增强人脸区域的连通性,使其更易于被识别和定位。同时,人脸的某些细节特征可能会变得模糊或不可见,而腐蚀处理能够减少这些不必要的细节信息,突出人脸的主要特征。因此,对$A(Cb, Cr)\geqslant\alpha$时的二值图像展开形态学[10]的膨胀与腐蚀处理,得到学生脸部区域面积

S，由此完成人脸区域定位。人脸区域面积可以表示为：

$$S=A(Cb,Cr)(Z\times E)N[c,u] \tag{4}$$

式中：Z为人脸区域的长度；E为人脸区域的宽度；$N[c,u]$为像素点坐标(c,u)处的像素值。

2　脸部关键点信息提取

人脸表情识别通常依赖于脸部关键点（如眼睛、鼻子、嘴巴等）的准确位置。然而，由于人脸是三维的，在多视角环境下，人脸会因为角度、距离等因素而发生形变，面部特征的可见性和形状也会发生变化，影响关键点定位的准确性，导致表情识别算法的性能下降。时空注意力机制能够综合考虑图像的空间和时间信息[11]，通过自动学习并关注不同视角下的关键特征[12]，从而提高对视角变化和表情复杂性的鲁棒性，使得即使在复杂的光照条件下，也能稳定地提取出脸部关键点。

2.1　空间注意力信息

设学生脸部区域S的维度为$M\times U\times Q$，其中M、U、Q分别为通道数量、课堂学生图像序列的帧数以及脸部图像的数量，$S^d_{i=1,2,\cdots,Q}$为定位到的学生脸部图像，其中d为脸部图像的维数，$S^d=\left[S^d_1,S^d_2,\cdots,S^d_Q\right]\in R^{M\times U\times Q}$，将$S^d_{i=1,2,\cdots,Q}$输入到空间注意力模块中。设$g$为空间注意力中的高斯函数，将$S^d_{i=1,2,\cdots,Q}$嵌入到$g$中，得到第$i$张和第$j$张学生脸部图像之间的相关性$o^d_i$为：

$$o^d_i=\frac{g(S^d_{i=1,2,\cdots,Q})}{M(S^d_i,S^d_j)} \tag{5}$$

式中：$M(\cdot)$为归一化因子；S^d_i与S^d_j分别为第i张与第j张学生脸部图像。

由此得到的学生脸部图像的空间注意力信息p^d_i为：

$$p^d_i=E^d_p o^d_i \tag{6}$$

2.2　时间注意力信息

设$S^y_{i=1,2,\cdots,U}$为连续的学生脸部图像序列，且$S^y=\left[S^y_1,S^y_2,\cdots,S^y_U\right]\in R^{M\times Q\times U}$，其中$y$为时间步长，将$S^y_{i=1,2,\cdots,U}$输入到时间注意力模块中，对其展开和空间注意力模块同样的操作，得到学生脸部图像的时间注意力信息$p^y_i\in R^{M\times U\times Q}$。

将空间注意力模块和时间注意力模块中的学生脸部关键点信息p^d_i和p^y_i进行融合，得到学生脸部的关键点信息p_i为：

$$p_i=p^d_i+p^y_i+S_i \tag{7}$$

由此可知，时空注意力可以根据多视角人脸图像之间的相关性动态地控制对于不同时间点的关注度，从而提取出多视角人脸图像中的重要信息，完成脸部关键点信息提取。

3　多视角人脸表情识别

脸部关键点信息提取后，由于人脸表情的变化非常复杂，不同人表达相同表情的方式可能存在差异，而同一个人在不同时间、不同情境下表达相同表情的方式也可能不同，增加了表情识别的难度。卷积神经网络[13]可以通过学习不同视角下的图像数据，自动适应不同视角下的特征差异。在训练过程中，卷积神经网络会学习到如何从不同角度的图像中提取出有用的特征，并将其与对应的表情类型进行关联。

已知学生脸部的关键点信息是p_i，对应的表情类型为a_i，利用p_i和a_i建立卷积神经网络的训练集$\boldsymbol{D}=\left\{(p_1,a_1),(p_2,a_2),\cdots,(p_m,a_m)\right\}$，其中$m$为学生脸部关键点信息的数量，则卷积神经网络的决策函数$k(p)$为：

$$k(p)=l\sum_{m=1}^{n}\vartheta(p_m,a_m) \tag{8}$$

式中：$k(p)$为学生脸部特征；p属于各种表情类型的概率分布；l为表情阈值；ϑ为模型参数。

为了提高上述卷积神经网络的泛化能力，降低网络的训练成本，通过带权重衰减的自适应梯度下降算法[14]对该网络中的参数ϑ展开训练。

假设x_b为m个学生脸部关键点信息的平均梯度，其表达式为：

$$x_b=\frac{X\left(\xi\cdot\vartheta_{b-1}\right)}{m} \tag{9}$$

式中：b为迭代次数；X为损失函数；ξ为权重衰减系数，且$\xi=\xi_{\text{norm}}\sqrt{e/TI}$，其中$\xi_{\text{norm}}$为标准权重衰减系数，$e$为批量大小，$T$为每个迭代周期的学生脸部关键点信息的数量，$I$为迭代总次数。

人脸是一个柔性体，不同脸型、不同性别和不同年龄的人的表情很难用一个精确的模型来表征。通过计算x_b的一阶矩和二阶矩可以分析这些关键点的位置、方向和形状变化，从而推断出学生的表情和头部姿态，进而提取出每个学生的个性化特征。x_b的一阶矩w_b与二阶矩f_b为：

$$\begin{cases} w_b = \chi_1 \cdot w_{b-1} + (1-\chi_1) \cdot x_b \\ f_b = \chi_2 \cdot f_{b-1} + (1-\chi_2) \cdot x_b^2 \end{cases} \tag{10}$$

式中:χ_1与χ_2分别为对应的指数衰减速率。

基于上述一阶矩和二阶矩,获得卷积神经网络参数ϑ的优化量$\Delta\vartheta_b$为:

$$\Delta\vartheta_b = -\iota \frac{\tilde{w}_b}{\sqrt{\tilde{f}_b} + \phi} \tag{11}$$

式中:ι为学习率[15];$\tilde{w}_b$与$\tilde{f}_b$分别为一阶矩与二阶矩的偏差补偿量,且$\tilde{w}_b = w_b/(1-\chi_1^b)$,$\tilde{f}_b = f_b/(1-\chi_1^b)$。

根据优化量$\Delta\vartheta_b$对卷积神经网络的参数ϑ展开更新,将更新后的参数$\vartheta' = \vartheta + \Delta\vartheta_b$代入决策函数$k(p)$的表达式中,则$k'(p)$为:

$$k'(p) = l \sum_{m=1}^{n} \vartheta'(p_m, a_m) \tag{12}$$

将学生脸部关键点信息p_i代入式(12)中,得到对应的表情类型概率分布$k(p_i)$。当$k(p_i) > l$时,学生面部表情为兴奋;当$k(p_i) = l$时,学生面部表情为专注;当$k(p_i) < l$时,学生面部表情为疲劳。由此,完成多视角人脸表情识别。

4 实验分析

4.1 实验设置

为了验证基于时空注意力的多视角人脸表情识别算法的整体有效性,以某学校阶梯教室中采集到的课堂学生图像作为实验对象,具体如图1所示。

图1 课堂学生图像

Fig. 1 Classroom student images

图1中学生脸部图像序列共120帧;灰度图像大小为48×48;学生多角度表情图像共40张。实验过程中相关参数的设置如表1所示。

在上述设置的基础上,使用基于时空注意力的多视角人脸表情识别算法对学生脸部展开精准定位,结果如图2所示。

表1 实验参数设置

Table 1 Experimental parameter settings

参数名称	参数值
卷积核大小	3×3
卷积层数	2
池化层核大小	2×2
学习率	0.001
权重衰减	0.000 1
训练轮次	1 000

图2中方框内为定位到的学生脸部区域。根据图2可知,利用本文算法能够将图像中所有学生的脸部区域精准定位。因此,本文算法可以有效分割背景区域和人脸区域,更利于后续的关键信息提取。

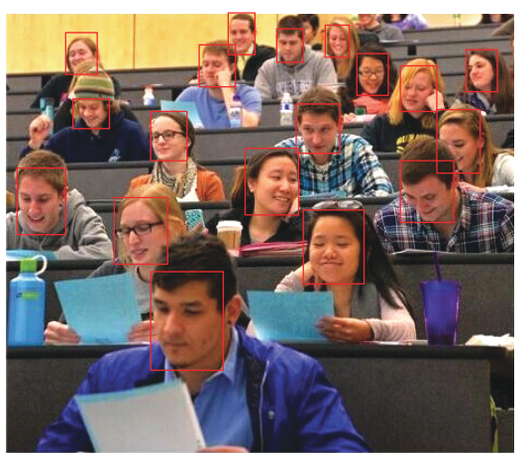

图2 人脸定位

Fig. 2 Facial localization

4.2 结果与分析

4.2.1 关键信息提取

学生脸部关键信息是识别学生脸部表情的重要依据,关键信息提取效果越好,表情识别的精度越高。在图1中选取低头男孩、转头女孩和正视戴眼镜(有遮挡)女孩的图像,利用基于时空注意力的多视角人脸表情识别算法、文献[3]算法和文献[4]算法对其眼部、鼻子和嘴部的关键信息展开提取。提取结果如图3所示。

通过图3能够发现,本文算法提取到的眼部关键信息和实际图像中眼部的关键信息完全吻合,文献[3]算法提取的关键信息中存在冗余信息,这会影响表情识别的精度,而文献[4]算法的提取结果出现部分关键信息未提取的问题。因此,本文算法的关键信息提取效果最好。

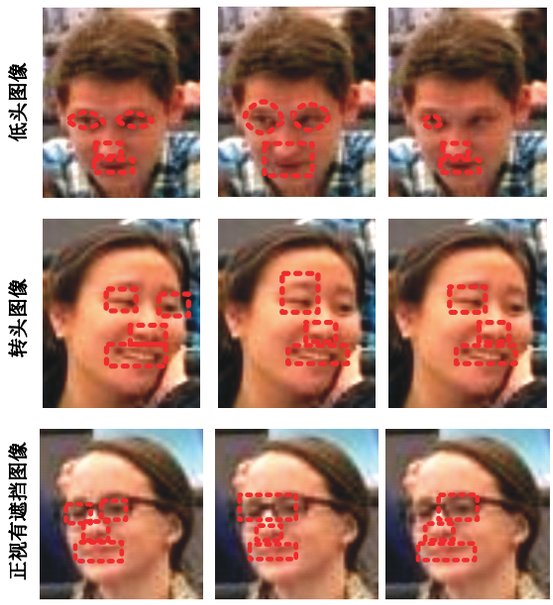

图 3　关键信息提取

Fig. 3　Key information extraction

4.2.2　表情识别

为了更直观地比较本文算法、文献[3]算法和文献[4]算法的学生脸部表情识别准确性，在图 1 中随机选取 10 名学生，利用上述 3 种算法对这 10 名学生的脸部表情展开识别，定义表情类型为兴奋、专注、疲劳 3 种，并将这 3 种类型分别通过 1、2、3 来表示，3 种算法的表情识别结果如表 2 所示。

表 2　表情识别

Table 2　Expression recognition

学生序号	实际类型	本文算法	文献[3]算法	文献[4]算法
1	1	1	2	1
2	3	3	2	3
3	3	3	3	2
4	2	2	2	1
5	2	2	2	2
6	3	3	3	2
7	1	1	1	1
8	1	1	1	1
9	1	1	3	1
10	3	3	3	1

分析表 2 可得，实际情况中，共有 4 名学生表情为兴奋，2 名学生表情为专注，4 名学生表情为疲劳，本文算法识别出的学生表情和实际情况完全一致，识别准确率为 100%。文献[3]算法和文献[4]算法的识别结果分别存在 3 处和 4 处误差，表明本文算法的人脸表情识别准确性更高。

5　结束语

为了提高多视角人脸的表情识别精度，本文提出了一种基于时空注意力的多视角人脸表情识别算法。实验结果表明，该算法能够有效提高人脸定位和关键信息提取的效果，且具有较高的表情识别精度。这表明通过时空注意力机制，算法能够精确地定位人脸并提取出与表情相关的关键信息，从而提高识别精度。

参考文献：

[1] 王军杰，王泉，蒋平，等. 一种孤立中心损失方法及其在人脸表情识别中的应用[J]. 西安交通大学学报，2022，56(4)：119-126.
Wang Jun-jie，Wang Quan，Jiang Ping，et al. An isolated central loss method applied in facial expression recognition[J]. Journal of Xi'an Jiaotong University，2022，56(4)：119-126.

[2] 周丽芳，刘俊林，李伟生，等. 深度二值卷积网络的人脸表情识别方法[J]. 计算机辅助设计与图形学学报，2022，34(3)：425-436.
Zhou Li-fang，Liu Jun-lin，Li Wei-sheng，et al. Facial expression recognition based on deep binary convolutional network[J]. Journal of Computer-Aided Design & Computer Graphics，2022，34(3)：425-436.

[3] 李召峰，朱明. 基于视频放大和双分支网络的微表情识别[J]. 液晶与显示，2022，37(3)：386-394.
Li Zhao-feng，Zhu Ming. Micro-expression recognition based on video magnification and dual-branch network[J]. Chinese Journal of Liquid Crystals and Displays，2022，37(3)：386-394.

[4] 虞苏鑫，贺俊吉. 基于子区域加权的不同年龄段人脸表情识别[J]. 计算机工程与科学，2022，44(8)：1426-1432.
Yu Su-xin，He Jun-ji. Facial expression recognition of different age groups based on face sub-region weighting[J]. Computer Engineering & Science，2022，44(8)：1426-1432.

[5] 唐宏，向俊玲，陈海涛，等. 多区域融合轻量级人脸表情识别网络[J]. 激光与光电子学进展，2023，60(6)：71-79.
Tang Hong，Xiang Jun-ling，Chen Hai-tao，et al. Multi region fusion lightweight facial expression recognition network[J]. Progress in Laser and Optoelectronics，2023，60(6)：71-79.

[6] 黄兴禄，苟小珊，陈希. 基于混合特征与信息熵的人脸微表情识别算法[J]. 计算机仿真，2023，40(6)：

197-201.
Huang Xing-lu, Gou Xiao-shan, Chen Xi. Face micro-expression recognition algorithm based on hybrid features and information entropy[J]. Computer Simulation, 2023, 40(6): 197-201.

[7] 戴嫣然, 戴国庆, 袁玉波. 基于肤色学习的多人脸前景抽取方法[J]. 计算机应用, 2021, 41(6): 1659-1666.
Dai Yan-ran, Dai Guo-qing, Yuan Yu-bo. Multi-face foreground extraction method based on skin color learning[J]. Journal of Computer Applications, 2021, 41(6): 1659-1666.

[8] 王超, 刘文超, 翟海祥, 等. 基于色彩空间和暗原色先验图像融合去雾算法[J]. 电光与控制, 2022, 29(10): 44-50.
Wang Chao, Liu Wen-chao, Zhai Hai-xiang, et al. An image fusion defogging algorithm based on color space and dark primary color priori[J]. Electronics Optics & Control, 2022, 29(10): 44-50.

[9] 朱帅康, 董龙雷, 官威, 等. 基于高斯混合模型的非高斯振动疲劳频域求解方法[J]. 振动与冲击, 2022, 41(16): 93-99.
Zhu Shuai-kang, Dong Long-lei, Guan Wei, et al. A frequency method for fatigue life estimation under non-Gaussian random loading based on a Gaussian mixture model[J]. Journal of Vibration and Shock, 2022, 41(16): 93-99.

[10] 花胜强, 陈意, 郑慧娟, 等. 和声搜索改进的形态学分析在库区漂浮物体量预估中应用的研究[J]. 水力发电, 2022, 48(9): 108-113.
Hua Sheng-qiang, Chen Yi, Zheng Hui-juan, et al. Research on the estimation of floating objects in the reservoir based on harmony search improved morphological analysis[J]. Water Power, 2022, 48(9): 108-113.

[11] 彭向东, 潘从成, 柯泽浚, 等. 基于并行架构和时空注意力机制的心电分类方法[J]. 浙江大学学报: 工学版, 2022, 56(10): 1912-1923.
Peng Xiang-dong, Pan Cong-cheng, Ke Ze-jun, et al. Classification method for electrocardiograph signals based on parallel architecture model and spatiol-temporal attention mechanism[J]. Journal of Zhejiang University (Engineering Science), 2022, 56(10): 1912-1923.

[12] 张云峰, 张超, 吕钊. 基于关键点的残差全连接网络动态手势识别方法[J]. 安徽大学学报: 自然科学版, 2022, 46(2): 30-38.
Zhang Yun-feng, Zhang Chao, Lv Zhao. Research on continuous gesture recognition based on residual fully connected network in vehicle scenes[J]. Journal of Anhui University (Natural Science Edition), 2022, 46(2): 30-38.

[13] 张蕾, 窦宏恩, 王天智, 等. 基于集成时域卷积神经网络模型的水驱油田单井产量预测方法[J]. 石油勘探与开发, 2022, 49(5): 996-1004.
Zhang Lei, Dou Hong-en, Wang Tian-zhi, et al. A production prediction method of single well in water flooding oilfield based on integrated temporal convolutional network model[J]. Petroleum Exploration and Development, 2022, 49(5): 996-1004.

[14] 葛泉波, 张建朝, 杨秦敏, 等. 带有微分项改进的自适应梯度下降优化算法[J]. 控制理论与应用, 2022, 39(4): 623-632.
Ge Quan-bo, Zhang Jian-chao, Yang Qin-min, et al. Adaptive gradient descent optimization algorithm with improved differential term[J]. Control Theory & Applications, 2022, 39(4): 623-632.

[15] 高涛, 杨朝晨, 陈婷, 等. 深度多尺度融合注意力残差人脸表情识别网络[J]. 智能系统学报, 2022, 17(2): 393-401.
Gao Tao, Yang Chao-chen, Chen Ting, et al. Deep multiscale fusion attention residual network for facial expression recognition[J]. Journal of Intelligent Systems, 2022, 17(2): 393-401.