

基于扩展局部二值模式的多尺度人脸表情识别方法

胡黄水¹, 戚星烁¹, 王出航², 王玲¹

(1. 长春工业大学 计算机科学与工程学院, 长春 130012;

2. 长春师范大学 计算机科学与技术学院, 长春 130032)

摘要: 针对人脸表情识别在复杂环境下姿态和光照鲁棒性差的问题, 提出一种融合扩展局部二值模式和多尺度网络结构的人脸表情识别方法. 该方法通过扩展传统局部二值模式的感受野并增强像素间的空间联系, 减少光照对人脸表情识别的噪声干扰; 通过将特征图在通道维度均匀分为若干子集并利用不同数量相同卷积块的方式提取特征图的多尺度特征, 有效处理人脸姿态变化. 在数据集 Fer2013 和 RAF-DB 上的实验结果表明, 该方法可有效提高人脸表情识别的准确率和鲁棒性, 为复杂环境下的人脸表情识别提供了有效解决方案.

关键词: 人脸表情识别; 局部二值模式; 多尺度网络; 卷积神经网络

中图分类号: TP183 **文献标志码:** A **文章编号:** 1671-5489(2025)05-1427-10

Multi-scale Facial Expression Recognition Method Based on Extended Local Binary Pattern

HU Huangshui¹, QI Xingshuo¹, WANG Chuhang², WANG Ling¹

(1. College of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China;

2. College of Computer Science and Technology, Changchun Normal University, Changchun 130032, China)

Abstract: Aiming at the problem of poor pose and lighting robustness in complex environments, we proposed a facial expression recognition method that integrated an extended local binary pattern and multi-scale network structure. The method expanded the receptive field of the traditional local binary pattern and enhanced spatial correlations between pixels, reducing noise interference of lighting on facial expression recognition. By evenly dividing the feature maps into several subsets in the channel dimension, and multi-scale features of the feature map were extracted by using different numbers of identical convolutional blocks, it effectively processed facial pose variations. Experimental results on the Fer2013 and RAF-DB datasets show that the proposed method can effectively improve the accuracy and robustness of facial expression recognition, offering an effective solution for facial expression recognition in complex environments.

Keywords: facial expression recognition; local binary pattern; multi-scale network; convolutional neural network

收稿日期: 2024-02-07.

第一作者简介: 胡黄水(1974—), 男, 汉族, 博士, 教授, 从事人工智能和无线传感器网络的研究, E-mail: huhs08@163.com. 通信

作者简介: 王出航(1976—), 女, 汉族, 硕士, 副教授, 从事无线传感器网络和图像处理的研究, E-mail: wangchuhang@ccsfu.edu.cn.

基金项目: 吉林省科技厅产业关键核心技术攻关项目(批准号: 20210201051GX).

人脸表情识别技术作为计算机视觉领域的重要研究内容,近年来取得了显著进展.随着智能系统和人机交互的不断深化,对计算机准确理解人类情感和心理状态的需求日益增长.面部表情是人类交流中一种重要的、自然的情感传达形式,比语言和行为更能体现情感信息^[1].因此,面部表情的捕获、识别和分析目前已广泛应用于人机交互、安全保护、消费支付、医疗诊断等领域.

人脸表情识别主要分为基于传统机器学习方法和基于深度学习方法两类.传统机器学习如 Gabor 小波变换^[2]、局部二值模式(local binary patterns, LBP)模型^[3]、直方图梯度^[4]等具有速度快、易于实现、适合小样本数据等特点,但由于其依赖人为主观选择特征,其方法的泛化性不佳. Krizhevsky 等^[5]在 ImageNet 大赛上提出了 AlexNet,将人工卷积神经网络应用于表情识别领域.之后在 AlexNet 中大量使用卷积池化,并使用多个小卷积核替代大卷积核的方式,提出了 VGG 网络^[6]. VGG 网络在 AlexNet 的基础上,增加了网络深度,并采用 ReLU 函数^[7]作为激活函数,可提取更深层次的图像特征.为克服加深网络深度导致识别准确率下降的问题,新的 ResNet 网络^[8]被提出,其通过将相邻卷积层连接的方式保留了浅层网络提取的特征信息.文献[9]提出了 Inception V1 结构,通过不同尺寸的卷积核提取多尺度特征,有助于网络获得更丰富的特征信息.文献[10]提出了 Res2Net 网络结构,通过将特征图划分为若干子集,并通过不同数量的卷积块,实现了特征信息的多尺度提取,有助于网络识别图片中的小目标特征,提高了网络对目标的识别成功率. Res2Net 存在特征提取不充分、准确率不高等问题,因此本文提出一种融合扩展局部二值模型和多尺度网络的人脸表情识别方法 E-LBP-MSNet(extended local binary pattern and multi-scale networks),通过扩展 LBP 扩大感受野并增加对光照的鲁棒性,结合多尺度网络提取面部表情图片多尺度特征信息降低网络对姿态的敏感性.

1 扩展局部二值模式

1.1 局部二值模式

局部二值模式是一种用于描述图片纹理特征的二值化方法.该方法利用 3×3 的蒙板对图像每个像素进行处理,通过图像灰度化,并比较每个周围像素灰度值与当前像素灰度值的大小,将大于等于当前像素灰度值的像素置 1,小于当前像素灰度值的像素置 0,其表达式为

$$f(x_i) = \begin{cases} 1, & x_i \geq x_c, \\ 0, & x_i < x_c, \end{cases} \quad (1)$$

其中 x_i 表示周围像素灰度值, x_c 表示中心像素灰度值.对周围 8 个像素从左上角开始按顺时针排序可得一个 8 位的二进制数,将此二进制数转化为十进制数作为中心像素的灰度值,如图 1 所示.经处理后的中心像素 LBP 值记为

$$\text{LBP}(x) = \sum_{i=0}^7 f(x_i - x_c) \cdot 2^i. \quad (2)$$

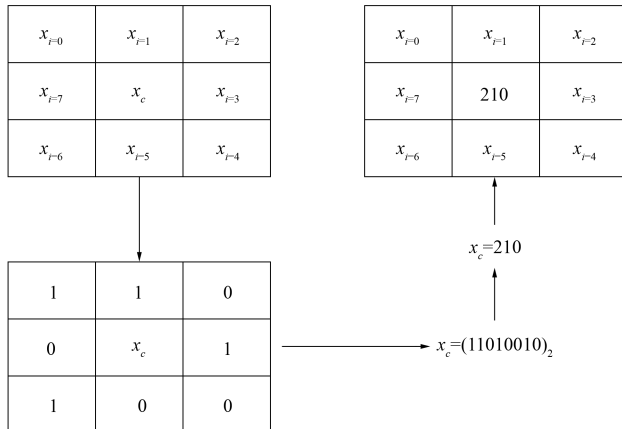


图 1 LBP 算法示意图

Fig. 1 Schematic diagram of LBP algorithm

1.2 扩展局部二值模式

由于像素灰度值的范围为 $[0,255]$, 如果生成的二进制序列长度超过 8 位, 即大于 255, 则会出现数据溢出现象. 所以在 LBP 算法中仅能包括中心像素周围 8 个像素, 即蒙板大小只能是 3×3 . 为扩展 LBP 蒙板的感受野, 本文提出扩展 LBP (extended local binary patterns, E-LBP), 将感受野从 3×3 扩展为 5×5 , 如图 2 所示. 对于 5×5 的蒙板, 其中中心像素周围的 24 个像素按图 2 方式被分为 8 组, 每组像素的灰度值取平均后再与中心像素的灰度值进行比较, 将大于等于中心像素灰度值的像素组置 1, 小于当前中心灰度值的像素组置 0, 并按顺时针排序, 由此得到 8 位二进制数, 将此二进制数转化为十进制数作为中心像素的灰度值.

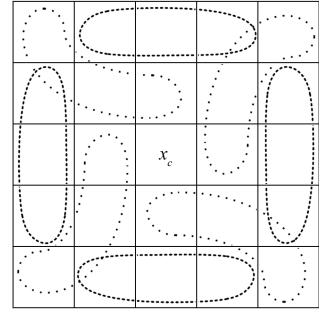


图 2 扩展 LBP 算法示意图

扩展 LBP 在增加感受野的同时考虑了周围像素之间的空间联系, 提取到的局部纹理特征更丰富, 能提供更多的上下文信息, 其表达式为

$$E-LBP(x) = \sum_{i=0}^7 f\left(\frac{1}{3}(x_{i-1} + x_i + x_{i+1}) - x_c\right) \cdot 2^i, \tag{3}$$

其中 x_c 表示中心像素, x_{i-1}, x_i, x_{i+1} 表示同组内 3 个像素, i 表示按顺时针排序第 i 组像素.

2 多尺度网络

传统卷积神经网络一般采用同一卷积核串联的方式提取图片的深层语义特征. 这种方式可自适应提取图片的全局特征, 但由于同一卷积核无法有效提取不同尺度的特征信息, 通常需要人为调整卷积核大小以适应不同尺度的目标检测、图像分类任务. 因此, 本文在 Res2Net 的基础上提出多尺度网络 (multi-scale network, MSNet), 以将特征图分为若干子集并通过不同数目的卷积块提取多尺度特征的方式提高网络的泛化能力和鲁棒性.

2.1 卷积块的感受野

在卷积神经网络中, 感受野是指网络中每一层输出特征图上像素点在输入张量上映射的范围^[11], 如图 3 所示.

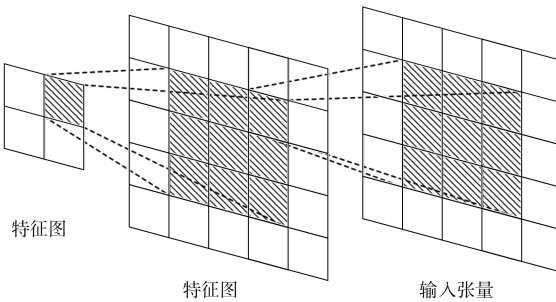


图 3 感受野示意图

Fig. 3 Schematic diagram of receptive field

感受野大小衡量了卷积块对输入张量信息的接受范围和对局部或全局特征的感知能力. 更大感受野可提取输入张量中更广泛的信息, 增强对输入张量的上下文理解. 小感受野主要关注输入张量的局部区域, 在提取细节和小尺度特征上更具优势, 可降低网络的敏感性. 感受野的计算公式为

$$RF_k = RF_{k-1} + \left[(RF_{k-1} - 1) \times \prod_{i=1}^{k-1} s_i \right], \tag{4}$$

其中 RF_k 为第 k 个卷积块的感受野, RF_{k-1} 为第 $(k-1)$ 个卷积块的感受野, s_i 为第 i 个卷积块滑动的步长. 由式(4)可知, 在步长 $s_i = 1$ 的情况下, 2 个 3×3 卷积块串联的感受野和 1 个 5×5 卷积块的

感受野相同,即越多相同步长、相同卷积核大小的卷积块串联,其在输入张量上的感受野越大.

2.2 多尺度网络结构

在步长和卷积核大小相同的情况下,输入张量通过串联卷积块数量的不同,其感受野不同,即提取特征的尺度不同.因此,本文在 Res2Net 的基础上提出了多尺度网络 MSNet. MSNet 网络结构如图 4 所示.输入特征图 X 经过 1×1 卷积操作将通道维度转换为可均匀划分为 n 组的通道数,每组记作 X_i . X_i 通过一个通道注意力模块^[12],分配每个通道特征权值以抑制无关通道信息,通道注意力模块记为 $G(\cdot)$.之后 $G(X_i)$ 经过 3×3 卷积块提取特征并通过残差连接^[8]融合到 X_{i+1} 中.因此最后一个分支包含了所有的 X_i ,且每个 X_i 分别通过不同数目的卷积块,即每个 X_i 具有不同的感受野.

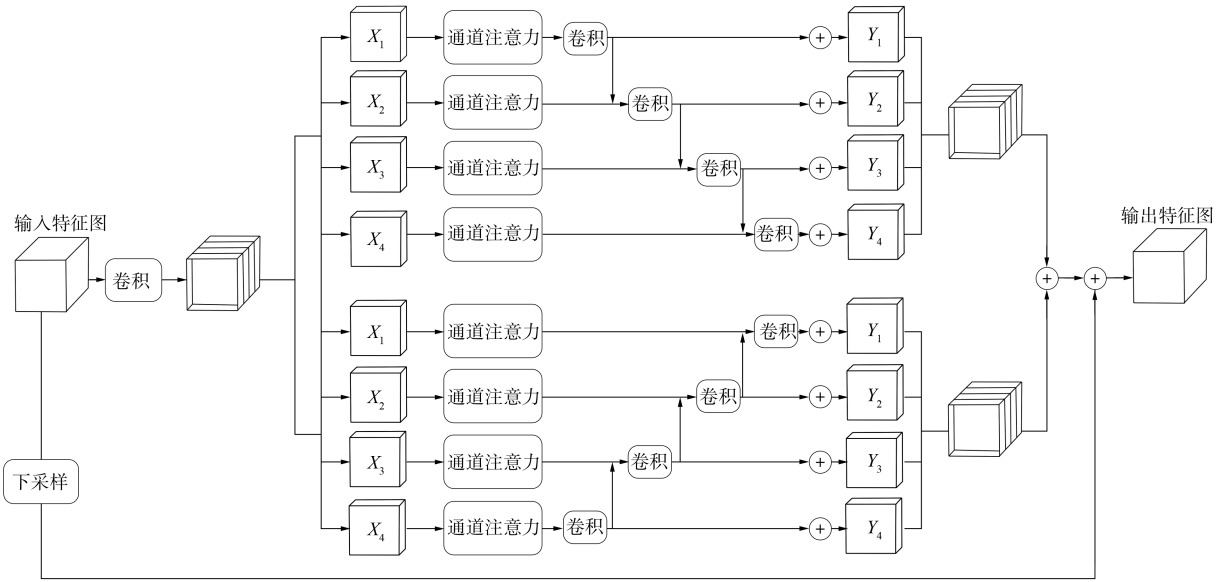


图 4 MSNet 网络结构

Fig. 4 Structure of MSNet network

由于第一组特征图 X_1 只通过了一个卷积块,而最后一组特征图 X_n 通过了 n 个卷积块,因此每组特征图 X_i 对多尺度特征的提取存在不均匀的情况.为提高网络的泛化能力,本文引入对称结构^[13],保证每个 X_i 都可得到相对充分的多尺度特征提取,每个分支的输出为

$$Y_i = \sum_{i=1}^n f^i(G_i(X_i)), \tag{5}$$

其中 i 表示第 i 分支, X_i, Y_i 分别表示第 i 分支的输入和输出, $f^i(\cdot)$ 表示 X_i 通过 i 个 3×3 卷积块, $G(\cdot)$ 表示第 i 分支的通道注意力模块.每个分支的输出 Y_i 在通道维度上依次连接,由对称结构得到的两个输出特征图在相加后取平均,最后通过残差结构将原始输入和网络输出融合得到网络整体输出.网络整体输出为

$$Y = \frac{(Y_1 \oplus \dots \oplus Y_n) + (Y_1 \oplus \dots \oplus Y_n)}{2} + X, \tag{6}$$

其中 \oplus 表示通道维度的连接操作.通过将输入特征图 X 沿通道维度划分为子集,并为每个子集按通道分配不同的权值,增强了网络对特征的提取能力.这种深度语义特征与浅层几何特征的融合提高了网络的泛化能力和鲁棒性.

2.3 融合扩展 LBP 的多尺度网络

由于在卷积神经网络中经常使用分辨率较大的输入图片,通常人为调整面部表情图片的分辨率以保证提取足够充分的特征,因此网络使用经典卷积神经网络 ResNet18 的第 1~6 个卷积块作为特征预提取模块.利用特征预提取模块对张量的空间尺度进行降维,避免使用多个 MSNet 对张量降维导致过拟合,网络结构如图 5 所示.

网络第一个分支先使用原始面部表情作为输入,再使用 4 个 MSNet 串联对预提取模块得到的

特征图进行多尺度特征提取, 最后通过全局平均池化将第一分支提取的全局特征转换为一维全局特征向量用于特征融合.

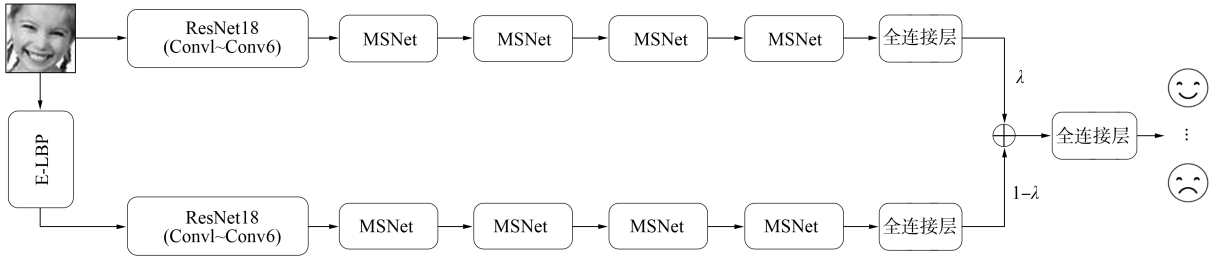


图 5 融合扩展 LBP 多尺度网络结构

Fig. 5 Structure of fusion extended LBP multi-scale network

网络第二个分支使用 E-LBP 处理后的图像作为输入, 先将面部表情图片灰度化处理, 再使用 E-LBP 对灰度图片计算每个像素的 E-LBP 值作为面部表情图片的局部纹理特征输入到网络中. 通过 E-LBP 处理得到的局部纹理特征利用 MSNet 网络提取局部多尺度特征. 将局部多尺度特征通过全局平均池化转换为一维局部特征向量与第一分支的一维全局特征向量通过决策级融合^[14]后经过 FC 模块进行分类. 第一分支全局损失函数和第二分支局部损失函数都使用交叉熵损失函数^[15]衡量预测值与真实值之间的误差, 并采用超参数平衡全局损失和局部损失, 最终的损失函数为

$$L(x, y) = \lambda L_{local}(x, y) + (1 - \lambda) L_{global}(x, y), \tag{7}$$

其中 L_{local} 表示在第二分支使用交叉熵损失函数, L_{global} 表示在第一分支使用交叉熵损失函数, λ 表示可调整的超参数.

3 实验结果与分析

3.1 数据集

本文使用数据集 Fer2013 和 RAF-DB 评估网络性能. 数据集 Fer2013 通过在线平台收集, 由 35 886 张面部图像, 每张图像的分辨率大小为 48×48 的灰度图像组成. 其中, 训练图像 28 708 张, 公开测试图像 3 589 张, 私人测试图像 3 589 张, 以 CSV 文件格式保存. 在面部图像中显示了 7 类基本情绪: 愤怒、厌恶、恐惧、快乐、中性、悲伤和惊讶.

数据集 RAF-DB 包括在真实环境中收集的大量多样化的面部图像, 其中包含 7 类基本情绪和 12 类复合情绪. 本文使用数据集 RAF-DB 中 7 种基本情绪数据集进行实验, 其中 12 271 张图像作为训练数据, 3 068 张图像作为测试数据. 数据集 Fer2013 和 RAF-DB 相对其他表情数据集较大, 图片更接近日常生活环境, 常用于表情识别研究领域. 图 6 为数据集 Fer2013 的表情示例. 图 7 为数据集 RAF-DB 的表情示例.



图 6 数据集 Fer2013 样本示例

Fig. 6 Sample examples of Fer2013 dataset

3.2 实验环境和参数设置

在硬件方面, 实验所用服务器 CPU 为 Intel(R) Core(TM) i9-10920X CPU @ 3.50 GHz, 内存为 64.0 GB, 显卡为 NVIDIA GeForce RTX 3090, 显存为 24 GB. 在软件方面, 实验所用系统为 Windows10 操作系统, 在 Pycharm Community Edition 平台使用 Pytorch1.9 和 Python3.7 作为编程环境.



图 7 数据集 RAF-DB 样本示例

Fig. 7 Sample examples of RAF-DB dataset

在数据预处理阶段, 先将面部表情图片的大小标准化为 224×224 并统一转换为灰度图格式. 将调整后的图片进行图像增强操作, 如随机水平翻转、随机旋转和随机裁剪等, 以扩展数据集, 避免过拟合. 在训练阶段, 设置 batch_size 大小为 64, 总 epoch 数为 100, 每个 epoch 的训练时间约为 180 s. 将 ResNet18 作为网络基准, 并使用 SGD 优化器更新模型的权值和偏差参数. 初始学习率设为 0.001, 使用等间隔步长下降策略, 将下降步长设为 20, Gamma 设为 0.3 调整学习率, 即每 20 个周期将学习率降低至当前值的 30%. 为保证实验的准确性, 所有实验都在相同实验条件下进行.

3.3 实验结果分析

3.3.1 扩展 LBP 实验结果分析

为验证 E-LBP 的有效性, 本文将 E-LBP 算法与传统 LBP 算法的实验结果进行对比, 结果如图 8 所示. 由图 8 可见, E-LBP 算法处理后的表情图片相较于传统 LBP 算法处理后的表情图片其噪点更少, 且眼睛和嘴部等表情特征显著区域的纹理更清晰, 能提供更有效的特征信息.

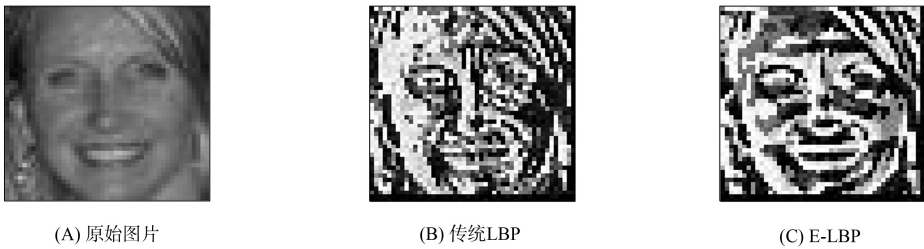


图 8 传统 LBP 和 E-LBP 实验结果对比

Fig. 8 Comparison of experimental results of conventional LBP and E-LBP

3.3.2 融合 E-LBP 的多尺度网络对比实验结果

下面通过对比融合 E-LBP 的多尺度网络和其他网络在数据集 Fer2013 和 RAF-DB 上的识别准确率验证本文构建网络的优越性, 实验结果列于表 1 和表 2.

表 1 不同网络模型在数据集 Fer2013 上的识别准确率

Table 1 Recognition accuracy of different network models on Fer2013 dataset

网络模型	年份	准确率/%	网络模型	年份	准确率/%
VGG16 ^[5]	2015	63.14	RMFANet ^[18]	2024	70.30
ResNet18(基线) ^[7]	2016	63.95	文献 ^[19]	2024	70.57
Fa-Net ^[16]	2019	71.10	本文	2024	73.14
DisEmoNet ^[17]	2021	71.72			

表 2 不同网络模型在数据集 RAF-DB 上的识别准确率

Table 2 Recognition accuracy of different network models on RAF-DB dataset

网络模型	年份	准确率/%	网络模型	年份	准确率/%
VGG16 ^[5]	2015	81.03	A-MobileNet ^[21]	2022	84.49
ResNet18(基线) ^[7]	2016	75.88	CMCNN ^[22]	2022	85.22
gACNN ^[17]	2018	85.07	EmotionNet ^[23]	2024	85.23
MA-Net ^[20]	2021	75.91	本文	2024	85.76

由表 1 可见, 在数据集 Fer2013 上本文 E-LBP-MSNet 模型的表情识别准确率最高, 为 73.14%, 相较于基线的 ResNet18 识别准确率提高了 9.19 个百分点, 相较于 Fa-Net^[16] 和 DisEmoNet^[17] 识别准

准确率分别提高了 2.04, 1.42 个百分点, 与 RMFANet^[18] 相比, 准确率提高了 2.84 个百分点, 与同样是基于 ResNet 进行改进的网络^[19] 相比, 准确率提高了 2.57 个百分点. 由表 2 可见, 在数据集 RAF-DB 上本文 E-LBP-MSNet 模型的表情识别准确率最高, 为 85.76%, 相较于基线的 ResNet18 识别准确率提高了 9.88 个百分点, 相较于 gACNN^[20] 和 MA-Net^[24] 识别准确率分别提高了 0.69, 9.85 个百分点, 相较于 A-MobileNet^[21] 和 CMCNN^[22] 识别准确率分别提高了 1.27, 0.54 个百分点, 与 EmotionNet^[23] 相比, 准确率提高了 0.53 个百分点.

图 9 为 E-LBP-MSNet 模型在数据集 Fer2013 上的混淆矩阵, 其中主对角线上的元素为正确识别的样本, 颜色越深表示该样本识别准确率越高. 由图 9 可见: 本文网络在快乐类别的识别准确率最高, 为 92%; 其次是惊讶、厌恶和中性样本, 其识别准确率分别为 86%, 75% 和 73%; 悲伤和恐惧的样本类别识别准确率最低, 为 58% 和 56%. 此现象与数据集中某类别样本数量和样本特征有关. 在数据集 Fer2013 中快乐表情类别的样本数量最多, 恐惧和悲伤表情类别的样本数量最少. 且快乐表情的特征, 如嘴角、眉毛等更明显, 而悲伤和恐惧的样本特征较相似, 易导致误识别. 由混淆矩阵可知, 有 17% 的恐惧样本被识别为悲伤样本, 有 12% 的悲伤样本被识别为恐惧样本. 图 10 为 E-LBP-MSNet 模型在数据集 RAF-DB 上的混淆矩阵. 由图 10 可见, 本文网络在快乐表情类别的识别准确率最高, 为 94%; 在厌恶和恐惧表情的表情识别成功率最低, 分别为 57% 和 51%.

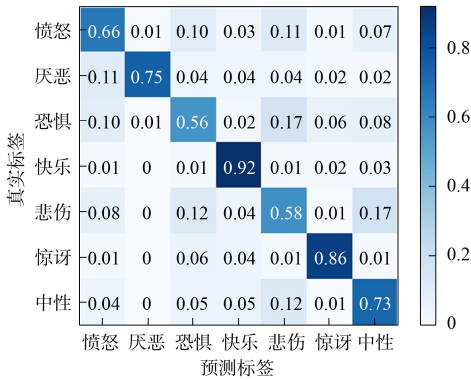


图 9 E-LBP-MSNet 模型在数据集 Fer2013 上的混淆矩阵

Fig. 9 Confusion matrix of E-LBP-MSNet model on Fer2013 dataset

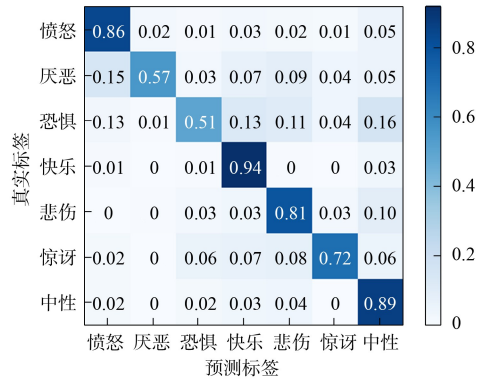


图 10 E-LBP-MSNet 模型在数据集 RAF-DB 上的混淆矩阵

Fig. 10 Confusion matrix of E-LBP-MSNet model on RAF-DB dataset

3.3.3 融合 E-LBP 的多尺度网络消融实验结果

为进一步验证本文 E-LBP-MSNet 网络的有效性, 在数据集 Fer2013 和 RAF-DB 上进行一系列消融实验. 在单分支结构中, 将 ResNet18 作为主干网络, 并使用串联的 MSNet 替换 ResNet18 第 6 个卷积块之后的网络结构, 同时分别将传统 LBP 和 E-LBP 算法处理后的图片作为输入以验证 E-LBP 的有效性, 实验结果列于表 3. 由表 3 可见, 本文提出的多尺度网络可有效提高面部表情识别准确率, 在数据集 Fer2013 和 RAF-DB 上相较于基线 ResNet18 分别提高了 7.30, 7.29 个百分点. E-LBP 处理后的图片作为网络输入相较于传统 LBP 处理后的图片在两个数据集上的表情识别准确率分别提高了 5.30, 4.10 个百分点, 验证了 E-LBP 的有效性.

对于双分支结构, 将所提出的双分支网络的其中一个分支替换为 ResNet18 经典网络以验证在双分支结构下本文 MSNet 网络结构和 E-LBP 算法的有效性. 表 4 列出了双分支结构实验结果. 由表 4 可见: 对于双分支网络采用 MSNet 结合 ResNet18, LBP 和 E-LBP 在数据集 Fer2013 上分别将表情识别准确率提高了 2.60, 5.83, 5.84 个百分点. 在数据集 RAF-DB 上分别将表情识别准确率提高了 6.71, 5.37, 6.28 个百分点; 相较于传统 LBP, ResNet18 和 MSNet 结合 E-LBP 算法处理原始图片在两个数据集上将表情识别准确率分别提高 2.39, 2.40, 2.26, 3.17 个百分点. 实验结果验证了在双分支网络结构中本文提出的 MSNet 网络结构和 E-LBP 算法的有效性.

表 3 单分支结构实验结果

Table 3 Experimental results of single-branching structure

数据集	ResNet18 (基线)	MSNet			准确率/%
		原始图片	LBP	E-LBP	
Fer2013	✓				63.95
		✓			72.25
			✓		67.84
RAF-DB	✓			✓	70.70
		✓			75.88
			✓		82.17
				✓	79.66
				✓	80.89

表 4 双分支结构实验结果

Table 4 Experimental results of double-branching structure

数据集	ResNet18 (基线)	MSNet	局部特征		准确率/%
			LBP	E-LBP	
Fer2013	✓				63.86
	✓	✓			66.46
	✓		✓		64.91
	✓			✓	67.30
RAF-DB		✓	✓		70.74
		✓		✓	73.14
	✓				77.20
	✓	✓			83.91
	✓		✓		77.22
				✓	79.48
			✓	✓	82.59
		✓		85.76	

3.3.4 超参数 λ 的消融实验结果

超参数 λ 作为平衡全局特征和局部特征参数,在调节网络做出整体最优决策中有重要作用.为研究不同 λ 值对网络识别准确率的影响,本文将 λ 的最小值设为 0.1,最大值设为 0.9,取值间隔为 0.1,实验结果如图 11 所示.

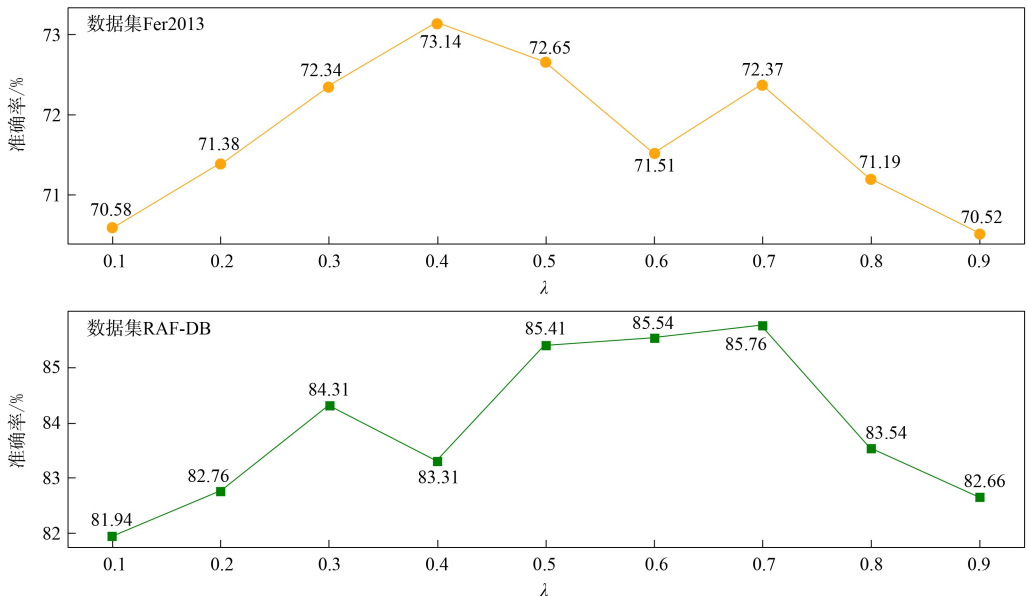


图 11 E-LBP-MSNet 模型在不同 λ 值的识别准确率

Fig. 11 Recognition accuracy of E-LBP-MSNet model at different λ values

由图 11 可见:在数据集 Fer2013 上,当 $\lambda=0.4$ 时,识别准确率最高,为 73.14%,即对于数据集 Fer2013,全局特征的重要性略高于局部特征的重要性;在数据集 RAF-DB 上,当 $\lambda=0.7$ 时,识别准确率最高,为 85.76%,即对于数据集 RAF-DB,全局特征的重要性略低于局部特征的重要性.当 $\lambda=0.1,0.9$ 时,识别准确率最低,即轻全局特征和局部特征都会导致识别准确率下降,与消融实验的结果相符.

综上所述,针对人脸表情识别在复杂环境下姿态和光照鲁棒性差的问题,本文提出了一种扩展 LBP 结合多尺度网络的方法实现人脸表情识别.通过在扩展传统 LBP 的感受野的同时加强相邻像素之间的空间联系,减少传统 LBP 产生的噪声.通过将特征图以通道维度均匀分为若干子集并通过不同数量卷积块的方式提取特征图的多尺度特征以降低网络敏感性.通过对比实验和消融实验,验证了本文提出的 E-LBP 算法和 MSNet 网络结构的优越性和有效性,可有效提高复杂环境下人脸表情识别的准确率.

参 考 文 献

- [1] 蒋斌,钟瑞,张秋闻,等.采用深度学习方法的非正面表情识别综述[J].计算机工程与应用,2021,57(8):48-61. (JIANG B, ZHONG R, ZHANG Q W, et al. Summary of Non-positive Expression Recognition Using Deep Learning Methods [J]. Journal of Computer Engineering & Applications, 2021, 57(8): 48-61.)
- [2] LUAN S Z, CHEN C, ZHANG B C, et al. Gabor Convolutional Networks [J]. IEEE Transactions on Image Processing, 2018, 27(9): 4357-4366.
- [3] OJALA T, PIETIKÄINEN M, HARWOOD D. Performance Evaluation of Texture Measures with Classification Based on Kullback Discrimination of Distributions [C]//Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994). Piscataway, NJ: IEEE, 1994: 582-585.
- [4] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection [C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05). Piscataway, NJ: IEEE, 2005: 886-893.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks [J]. Advances in Neural Information Processing Systems, 2017, 60(6): 84-90.
- [6] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [EB/OL]. (2014-09-04)[2024-01-01]. <https://arxiv.org/abs/1409.1556>.
- [7] GLOROT X, BORDES A, BENGIO Y. Deep Sparse Rectifier Neural Networks [C]//Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. [S.l.]: JMLR, 2011: 315-323.
- [8] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770-778.
- [9] SZEGEDY C, LIU W, JIA Y Q, et al. Going Deeper with Convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 1-9.
- [10] GAO S H, CHEN M M, ZHAO K, et al. Res2net: A New Multi-scale Backbone Architecture [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(2): 652-662.
- [11] 葛延良,孙笑笑,张乔,等.基于循环生成对抗网络的人脸素描合成[J].吉林大学学报(理学版),2022,60(4):897-905. (GE Y L, SUN X X, ZHANG Q, et al. Face Sketch Synthesis Based on Cyclic Generation of Adversarial Network [J]. Journal of Jilin University (Science Edition), 2022, 60(4): 897-905.)
- [12] HU J, LI S, GANG S. Squeeze-and-Excitation Networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 1-13.
- [13] 朱志宇,张冰,刘维亭.一种具有对称结构的神经网络的构造方法研究[J].微电子学与计算机,2005,22(9):173-176. (ZHU Z Y, ZHANG B, LIU W T. A Construction of Neural Network with Symmetric Structure [J]. Microelectronics and Computers, 2005, 22(9): 173-176.)
- [14] PRABHAKAR S, JAIN A K. Decision-Level Fusion in Fingerprint Verification [J]. Pattern Recognition, 2002,

35(4): 861-874.

- [15] 赵晓晴, 李慧盈, 苏安炆, 等. 基于加权损失函数的粘连白细胞分割算法 [J]. 吉林大学学报(理学版), 2021, 59(1): 85-91. (ZHAO X Q, LI H Y, SU A Y, et al. Adiver Leukocyte Segmentation Algorithm Based on Weighted Loss Function [J]. Journal of Jilin University (Science Edition), 2021, 59(1): 85-91.)
- [16] WANG W X, SUN Q, CHEN T, et al. A Fine-Grained Facial Expression Database for End-to-End Multi-pose Facial Expression Recognition [EB/OL]. (2019-07-25)[2023-12-30]. <https://arxiv.org/abs/1907.10838>.
- [17] ZHU Q, GAO L, SONG H, et al. Learning to Disentangle Emotion Factors for Facial Expression Recognition in the Wild [J]. International Journal of Intelligent Systems, 2021, 36(6): 2511-2527.
- [18] 郑浩, 赵光哲. 基于改进卷积的多尺度表情识别 [J]. 计算机工程与应用, 2024, 60(2): 231-243. (ZHENG H, ZHAO G Z. Multi-scale Expression Recognition Based on Improved Convolution [J]. Journal of Computer Engineering & Applications, 2024, 60(2): 231-243.)
- [19] 徐子凡, 程科, 袁雪梅, 等. 基于优化 ResNet 的人脸表情识别 [J]. 计算机与数字工程, 2024, 52(12): 3491-3495. (XU Z F, CHENG K, YUAN X M, et al. Facial Expression Recognition Based on Optimized ResNet [J]. Computer and Digital Engineering, 2024, 52(12): 3491-3495.)
- [20] LI Y, ZENG J B, SHAN S G, et al. Occlusion Aware Facial Expression Recognition Using CNN with Attention Mechanism [J]. IEEE Transactions on Image Processing, 2018, 28(5): 2439-2450.
- [21] NAN Y H, JU J G, HUA Q Y, et al. A-MobileNet: An Approach of Facial Expression Recognition [J]. Alexandria Engineering Journal, 2022, 61(6): 4435-4444.
- [22] YU W M, XU H. Co-attentive Multi-task Convolutional Neural Network for Facial Expression Recognition [J]. Pattern Recognition, 2022, 123: 108401-1-108401-11.
- [23] 罗玉涛, 郭丰瑞. 基于注意力融合特征增强的座舱表情识别模型 [J]. 汽车工程, 2024, 46(9): 1697-1706. (LUO Y T, GUO F R. Cabin Expression Recognition Model Based on Attention Fusion Feature Enhancement [J]. Automotive Engineering, 2024, 46(9): 1697-1706.)
- [24] ZHAO Z Q, LIU Q S, WANG S M. Learning Deep Global Multi-scale and Local Attention Features for Facial Expression Recognition in the Wild [J]. IEEE Transactions on Image Processing, 2021, 30: 6544-6556.

(责任编辑: 韩 啸)