



基于注意力机制与多尺度特征融合的人脸表情识别

潘海鹏, 郝 慧, 苏 雯

(浙江理工大学机械与自动控制学院, 杭州 310018)

摘 要: 人脸表情识别一直是人机交互等领域颇为关注的研究,然而当前研究大多忽略了多尺度特征的融合及表情中间特征的改善。针对这一问题,本文提出了一种基于注意力机制与多尺度特征融合学习的人脸表情识别方法,该方法由浅层特征提取模块和多尺度特征融合模块构成,能从深到浅提取更多有价值的信息,并有效改善表情中间特征。首先输入表情图像到浅层网络和骨干网络,分别获取浅层特征和深层特征;然后在浅层特征提取模块中加入注意力机制,对浅层特征进行加强或抑制;最后融合浅层特征与深层特征,构造人脸表情的多尺度融合特征,并通过分类器将人脸表情图像分为7种表情。该方法在两个公开数据集 JAFFE 和 KDEF 上的平均识别准确率分别达到了 96.67% 和 89.29%,表现优于现有的深度学习方法 and 传统方法。实验表明该方法能获取更丰富的人脸表情信息,且具有更强的泛化能力。

关键词: 人脸表情识别;注意力机制;多尺度;特征融合;深度学习

中图分类号: TP183

文献标志码: A

文章编号: 1673-3851 (2022) 05-0382-07

Facial expression recognition based on attention mechanism and multi-scale feature fusion

PAN Haipeng, HAO Hui, SU Wen

(Faculty of Mechanical Engineering & Automation, Zhejiang
Sci-Tech University, Hangzhou 310018, China)

Abstract: Facial expression recognition has been the research of great concern in fields such as human-computer interaction. However, the fusion of multi-scale features and the improvement of intermediate features of facial expressions have been ignored in most of the current studies. To address this problem, a facial expression recognition method based on attention mechanism and multi-scale feature fusion learning was proposed in this paper. The method composed of a shallow layer feature extraction module and a multi-scale feature fusion module can extract more valuable information from deep layer to shallow layer and effectively improve the intermediate features of expression. First, the expression images are input into shallow network and backbone network to obtain shallow layer features and deep layer features respectively. Then an attention mechanism is added to the shallow layer feature extraction module to intensify or suppress the shallow layer features. The shallow layer features and deep layer features are fused to create the multi-scale fusion features of facial expressions. Finally, the facial expression images are divided into 7 expressions by classifier. The average recognition accuracy of this method on two public datasets, JAFFE and KDEF, reaches 96.67% and 89.29%, respectively. The performance outperforms

收稿日期:2021-10-17 网络出版日期:2021-12-17

基金项目:浙江理工大学科研启动基金项目(18022225-Y);浙江省自然科学基金项目(LQ20F020001);浙江理工大学基本科研业务费专项项目(2020Q014);国家自然科学基金项目(62006209)

作者简介:潘海鹏(1965—),男,河南濮阳人,教授,主要从事智能检测、智能控制以及图像信息处理方面的研究。

通信作者:苏 雯, E-mail: wensu@zstu.edu.cn

the current deep learning methods and traditional methods. The experimental results demonstrate that the method can bring about more abundant facial expression information with stronger generalization ability.

Key words: facial expression recognition; attention mechanism; multi-scale; feature fusion; deep learning

0 引言

在日常交流中,人脸表情扮演着非常重要的角色,相比于语言在某些方面它可以传达更准确的信息。近年来,人脸表情识别技术受到了越来越多的关注,广泛应用于增强现实^[1]、虚拟现实^[2]和驾驶员情绪检测^[3]等场景。

传统的人脸表情识别方法一般包括人脸检测、特征提取与处理和特征分类等步骤。其中提取特征的方法主要有局部二值模式(LBP)^[4]、Gabor 小波变换^[5]、方向梯度直方图(HOG)^[6]和主动形状模型(ASM)^[7]等。Shi 等^[8]利用二维 Gabor 小波的多方向和多频率集合来变换表情,并结合变换后的 Gabor 系数来提取相位特征。Han 等^[9]提出了一种结合 LBP 和 SLPP 的模型,该模型使用 LBP 获取图像的直方图,使用 SLPP 来减少维度。传统方法虽然效果不错,但需要手工提取特征,且不能随着任务的变化而主动提取相关特征,费时费力。

不同于传统机器学习算法,深度学习能从海量数据中自主学习所需特征,避免了复杂的人工提取特征过程。近年来深度学习迅速发展,卷积神经网络广泛应用于计算机视觉领域。Ding 等^[10]提出了一种两阶段训练网络,即 FaceNet2ExpNet 网络,该网络由人脸网络和表情网络组成,在 4 个人脸表情数据集上取得了高准确率,但缺少学习多分辨率特征的能力。Wei 等^[11]提出了一种多模式策略,该策略融合了 59 个脸部关键点的外观特征,采用卷积神经网络提取高级自学习特征;但该策略根据经验人工提取了一部分特征,效率较低。Chen 等^[12]构造了一种由 3 个子网络组成的集成卷积神经网络,并利用支持向量机进行最终表情预测。Minaee 等^[13]提出了一种注意力卷积网络来识别表情,并用可视化技术寻找人脸图像中对分类结果影响最强的区域。然而,上述两种方法^[12-13]都没有对局部特征和全局特征进行综合考虑。何志超等^[14]将表情图像输入两个相互独立且卷积层深度不同的网络,得到不同分辨率的特征图。张红颖等^[15]建立了结合 LBP 和梯度特征的双通道表情识别网络,但这两个网络都缺少对人脸表情关键区域的关注。申毫

等^[16]提出了一种融合深层特征与经过筛选的浅层特征的轻量型表情识别方法,提高了表情识别准确率,但该方法没有考虑人脸表情的局部区域。Tian 等^[17]提出了一种判别人脸表情识别方法,利用四重平均损失最大化类内相似度及最小化类间相似度,并利用平衡 Softmax 损失消除不同类别间的不平衡,但该方法的损失计算过程较为复杂。虽然上述方法^[10-17]在识别人脸表情的任务中取得了一定效果,但没有综合考虑浅层特征、深层特征和显著性特征。

较深的神经网络能获取人脸表情的形状、颜色等整体特征,但容易忽略一些细节特征。因此,本文提出了一种基于注意力机制与多尺度特征融合的人脸表情识别方法,对人脸图像的表情进行归类。该方法采用多尺度特征融合的计算结构,将浅层特征提取模块的浅层特征与深层特征相融合,以提取更多细节信息,增强网络鲁棒性;另外,本文改进了通道-空间注意力模块,并将其引入到浅层特征提取模块,使模型能有效学习浅层显著性特征;最后采用人脸表情识别的公开数据集 JAFFE^[18]和 KDEF^[19]对本文的方法进行实验,以验证该方法的有效性。

1 本文方法

本文建立了一种基于注意力机制与多尺度特征融合的网络,用来分类 6 种基础表情(生气、厌恶、恐惧、高兴、悲伤、惊讶)^[20]和 1 种中性表情。网络结构示意图如图 1 所示,由浅层特征提取模块和多尺度特征融合模块构成。首先,通过浅层特征提取模块提取人脸表情的浅层特征,并加入通道-空间注意力模块学习浅层显著性特征;然后,利用多尺度特征融合模块构建多尺度融合特征的计算结构,使模型提取到更丰富的表情特征;最后,使用该多尺度融合特征计算表情类别概率,实现人脸表情识别。

1.1 浅层特征提取模块

在人脸表情识别任务中,浅层特征的提取非常重要。它描述了人脸表情的图像纹理等信息,且分辨率较高,对细节信息较为敏感。

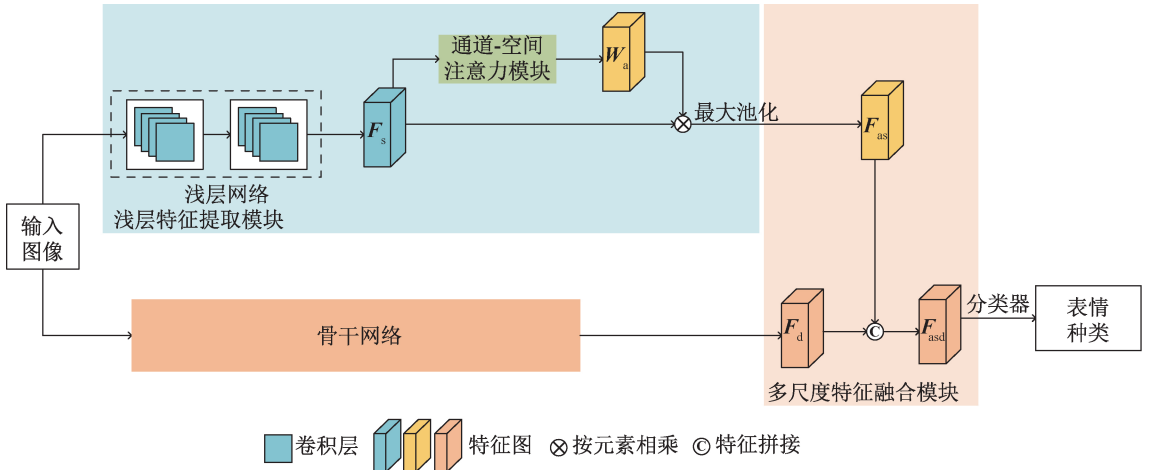


图1 网络结构示意图

本文选择了 Huang 等^[21]提出的 DenseNet121 网络来提取浅层特征。DenseNet 网络是一种具有密集连接的卷积神经网络,具有加强特征传递、缓解梯度消失等优点,有利于提取人脸表情特征。浅层网络由 DenseNet121 网络的 Block1 和 Block2 组成。首先输入图像 $I \in \mathbf{R}^{C \times H \times W}$ 到浅层网络 S ,得到浅层特征 $F_s \in \mathbf{R}^{C_1 \times H_1 \times W_1}$,其中: C 、 H 和 W 分别为输入图像的通道数、高度和宽度, C_1 、 H_1 和 W_1 分别为浅层特征的通道数、高度和宽度。然后经过通道-空间注意力模块得到特征权重图 $W_a \in \mathbf{R}^{C_1 \times H_1 \times W_1}$,并将其与浅层特征 F_s 按元素相乘。该操作对浅层特征 F_s 中的元素赋予不同的权重,加强或抑制其中的元素,以此学习人脸表情的显著性特征。最后经过最大池化层保留更多的纹理特征,得到浅层注意力特征 F_{as} ,其大小为 $C_1 \times (H_1/2) \times (W_1/2)$ 。上述计算过程可用式(1)–(2)表示:

$$F_s = S_{\theta_s}(I) \tag{1}$$

$$F_{as} = \text{Maxpool}(W_a \otimes F_s) \tag{2}$$

其中: θ_s 为浅层特征提取模块参数; $\text{Maxpool}()$ 表示最大池化; \otimes 表示按元素相乘; $F_{as} \in \mathbf{R}^{C_1 \times (H_1/2) \times (W_1/2)}$ 为浅层注意力特征。

1.2 通道-空间注意力模块

浅层特征虽然能得到表情的细节信息,但对表情变化突出的特征区域给予的关注不够,因此本文在提取到的浅层特征上加入通道-空间注意力模块,给其分配不同的权重。本文选择了瓶颈注意力模块(Bottleneck attention module)。BAM 是 Park 等^[22]于 2018 年提出的注意力模块,该模块针对通道间关系及空间关系挖掘图像特征。

注意力模块结构示意图如图 2 所示,其中:通道

间关系的计算,能够对每个通道的全局信息进行提取,如通道分支 M_c ;利用空间关系的计算,强化不同空间位置的显著性特征,即空间分支 M_s ,从而有效利用上下文信息。通道分支 M_c 包括 1 个全局平均池化层和 2 个全连接层,空间分支 M_s 包括 2 个 1×1 卷积层和 2 个 3×3 空洞卷积层。输入浅层特征 $F_s \in \mathbf{R}^{C_1 \times H_1 \times W_1}$ 分别到通道分支 M_c 和空间分支 M_s ,并将输出的两者特征尺寸变为和浅层特征 F_s 相同,然后按元素相加。原始的 BAM 模块在两特征相加后,经过 Sigmoid 函数,得到特征权重图。然而,在表情识别问题中,表情的变化会使面部图像中的五官位置发生轻微改变,因此适当增加感受野,能够更好地获取上下文信息。本文对模型进行改进,两特征融合之后经过 3×3 的卷积层,之后紧跟 Batch Normalization 层和 ReLU 激活函数,再经过 Sigmoid 函数生成尺寸与浅层特征 F_s 尺寸相同的特征权重图 $W_a \in \mathbf{R}^{C_1 \times H_1 \times W_1}$ 。上述计算过程可用式(3)表示:

$$W_a = \text{Sigmoid}(C_{3 \times 3}(M_c(F_s) + M_s(F_s))) \tag{3}$$

其中: Sigmoid 表示 Sigmoid 函数; $C_{3 \times 3}$ 表示卷积核为 3×3 的卷积操作; $W_a \in \mathbf{R}^{C_1 \times H_1 \times W_1}$ 为浅层特征 $F_s \in \mathbf{R}^{C_1 \times H_1 \times W_1}$ 经过注意力模块得到的特征权重图,其中元素 k_{mm}^i 对应浅层特征 F_s 的第 i 通道中第 m 行第 n 列像素所占的权重, $i = 1, 2, \dots, C_1, m = 1, 2, \dots, H_1, n = 1, 2, \dots, W_1$ 。

1.3 多尺度特征融合模块

浅层网络的感受野较小,虽然有较强的几何细节信息表征能力,但是欠缺对表情图像抽象性及整体性的概括。因此本文将其与深层特征相融合,以此学习更全面有效的特征表示,更加完整地提取人脸表情信息。

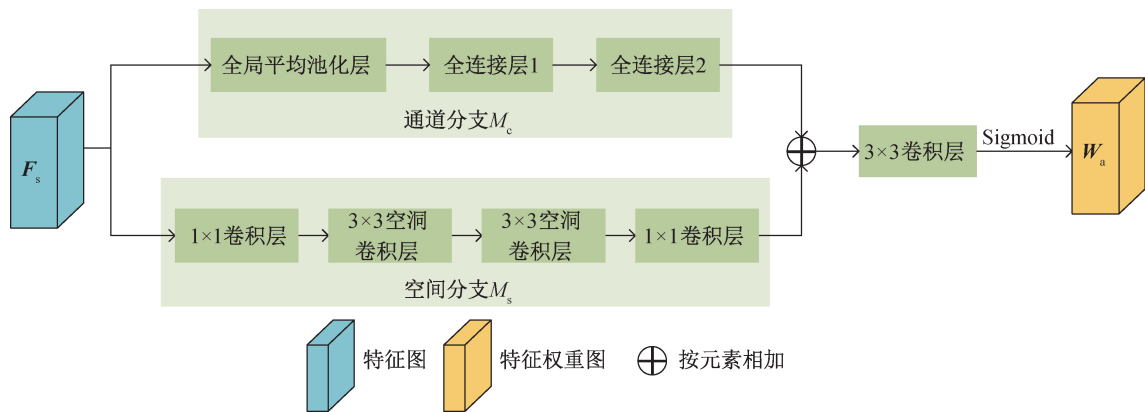


图 2 注意力模块结构示意图

输入图像 $I \in \mathbf{R}^{C \times H \times W}$ 到去除分类层的预训练 DenseNet121 网络 D ，得到深层特征 $F_d \in \mathbf{R}^{C_2 \times H_2 \times W_2}$ ，其中 C_2 、 H_2 和 W_2 分别为深层特征的通道数、高度和宽度。采用双线性插值方法下采样浅层注意力特征图 F_{as} ，使其高度和宽度与深层特征 F_d 相同。然后将两特征进行拼接，得到多尺度融合特征 $F_{asd} \in \mathbf{R}^{(C_1+C_2) \times H_2 \times W_2}$ ，如图 1 中多尺度特征融合模块所示。最后经过分类器进行分类。分类器由全连接层、Batch Normalization 层、激活函数和 Softmax 层组成。上述计算过程可用式 (4) — (5) 表示：

$$F_d = D_{\theta_D}(I) \quad (4)$$

$$F_{asd} = \text{Concat}(F_{as}, F_d) \quad (5)$$

其中： θ_D 为去除分类层的预训练 DenseNet121 网络的参数；Concat 为拼接操作。

1.4 训练策略与损失函数

将真实值与预测值之间的损失记为分类损失 L ，采用交叉熵损失函数来计算， L 可用式 (6) 计算：

$$L = - \sum_{i=1}^k z_i \log(y_i) \quad (6)$$

其中： k 为表情种类数量； z_i 为指示变量，如果该类别和输入图像所属类别相同则取 1，否则为 0； y_i 表示输入图像属于该类别的预测值。分类损失 L 同

时反向传播至两个模块（骨干网络的参数被冻结），不断更新网络参数，直至网络性能达到最优。

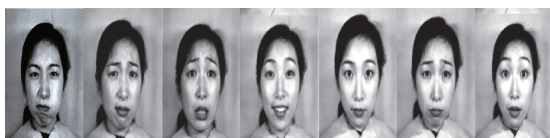
2 实验与分析

2.1 数据集

为了验证提出方法的有效性，本文在公开的人脸表情数据集 JAFFE 和 KDEF 上进行了实验。

a) JAFFE 数据集。该数据集包括 10 位日本女性的 7 类表情图像，一共有 213 张，均为像素为 256×256 的 8 位灰度图，部分样本图像如图 3(a) 所示，从左到右表情分别为生气、厌恶、恐惧、高兴、中性、悲伤和惊讶。实验采用三折交叉验证，即将数据集分为三份，轮流将其中 143 张图像作为训练数据，70 张图像作为测试数据，且测试数据不重复。

b) KDEF 数据集。该数据集包括 70 位业余演员的 7 类表情图像，一共有 4900 张，均为像素为 562×762 的彩色图，每种表情有 5 种角度，分别为 $\pm 90^\circ$ 、 $\pm 45^\circ$ 以及正面角度，部分样本图像（正面角度）如图 3(b) 所示，表情顺序与 JAFFE 数据集样本图像相同。实验选用正面角度图像，一共 980 张。实验采用五折交叉验证，即将数据集分为五份，轮流将其中 784 张图像作为训练数据，196 张图像作为测试数据，且测试数据不重复。



(a) JAFFE 数据集



(b) KDEF 数据集

图 3 JAFFE 和 KDEF 数据集部分样本

2.2 实验过程

本文实验基于深度学习框架 Pytorch 实现，采用 GTX1650 进行 GPU 加速，CPU 为 i5-9300HF@2.40 GHz。在训练过程中，将输入图像的大小调整

为 224×224 ，并进行标准化，其参数为 $[0.485, 0.456, 0.406]$ ， $[0.229, 0.224, 0.225]$ 。网络的 Batchsize 设为 8。对于 KDEF 数据集，在对输入图像调整大小之前，将其中心裁剪为大小 450×450 的

图像。网络学习率为 0.001, 优化器算法采用 Adam, 权重设置为 5×10^{-4} 。本文使用 Kaiming 方法初始化分类器的全连接层和 Batch Normalization 层, 防止激活输出爆炸或消失。

2.3 消融实验

2.3.1 深层特征提取实验

实验选用预训练的 DenseNet121 网络作为主干, 首先输入表情图像到该网络, 得到 7×7 的深层特征 F_d , 其通道数为 1024。然后经过分类层, 将其先降为 256, 再由 256 降到 7, 以此进行分类; 并在最后一层全连接层之前加上 Dropout 层, 随机使 50% 的神经元失活, 防止网络过拟合。深层特征提取实验即本文的 Baseline 实验, 在 JAFFE 数据集和 KDEF 数据集上的平均准确率分别为 92.38% 和 85.20%。

2.3.2 多尺度特征融合实验

该实验提取了人脸表情图像的多尺度特征。首先输入人脸表情图像到浅层网络得到尺寸为 28×28 的浅层特征 F_s , 通道数为 512。采用双线性插值方法下采样浅层特征 F_s , 使其高度和宽度与 2.3.1 实验的深层特征 F_d 相同。然后将上述两特征拼接, 得到尺寸为 7×7 、通道数为 1536 的多尺度融合特征 F_{sd} 。最后经过分类器进行表情分类。实验准确率如表 1 所示, 在 JAFFE 数据集和 KDEF 数据集上的平均准确率分别为 94.76% 和 87.96%, 相比 Baseline 实验分别高了 2.38% 和 2.76%。实验结果表明加入图像的浅层特征是有有效的, 浅层特征的分辨率高, 更容易捕捉到细节信息。

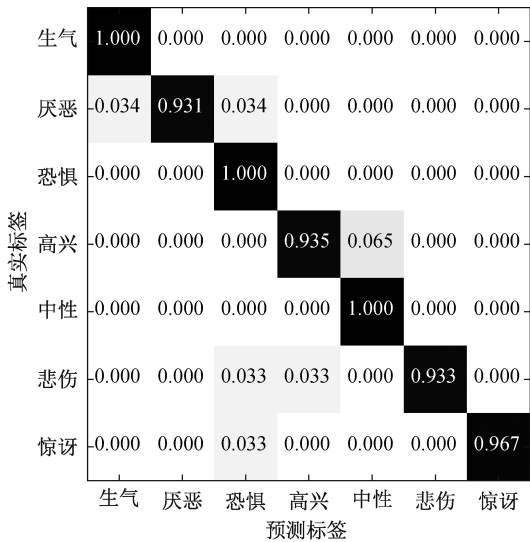
2.3.3 加入注意力模块的多尺度特征融合实验

在 2.3.2 实验的基础上加入注意力机制, 即在

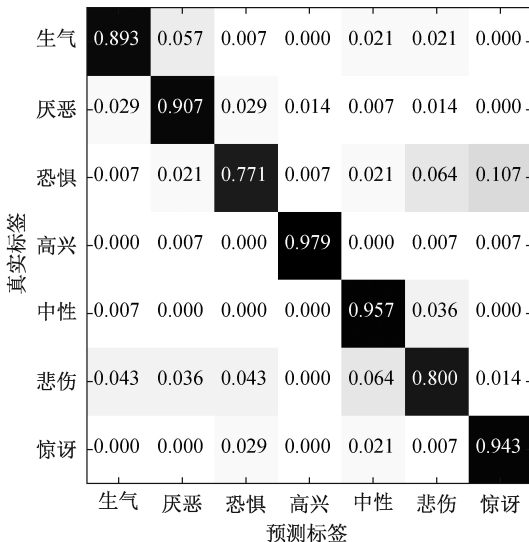
浅层特征提取模块中加入通道-空间注意力模块。首先将浅层特征 F_s 通过通道-空间注意力模块生成特征权重图 W_a , 特征权重图 W_a 的尺寸为 28×28 , 通道数为 512。然后与浅层特征 F_s 按元素相乘, 并经过最大池化层得到尺寸为 14×14 的浅层注意力特征 F_{as} 。采用双线性插值方法下采样 F_{as} , 使其尺寸变为 7×7 。 F_{as} 与 F_d 拼接, 得到尺寸为 7×7 、通道数为 1536 的多尺度融合特征 F_{asd} 。最后经过分类器进行分类, 实验准确率如表 1 所示。当通道-空间注意力模块选用原始的 BAM 模块时, 相比于 2.3.2 实验, 在两个数据集上的准确率都只提升了不到 0.5%。当选用本文改进的 BAM 模块时, 在 JAFFE 数据集和 KDEF 数据集上的平均准确率分别为 96.67% 和 89.29%, 相比 2.3.2 实验分别高了 1.91% 和 1.33%。上述两个数据集的混淆矩阵如图 4 所示。在 KDEF 数据集中, 恐惧和悲伤的识别率较低, 主要原因是有些人的恐惧表情和惊讶表情非常相似, 都是睁大眼睛及张大嘴巴; 有些人的悲伤表情和中性及生气表情比较相似, 都为紧闭双唇或紧皱眉头。这些因素导致网络不能很好地将其区别开来。实验结果表明: 改进的注意力模块可以有效地改善表情浅层特征, 提高浅层特征提取模块的特征提取能力。

表 1 两个数据集的消融实验准确率 %

数据集	方法			
	Baseline	多尺度特征	多尺度特征+ BAM	多尺度特征+ 改进的 BAM
JAFFE	92.38	94.76	95.24	96.67
KDEF	85.20	87.96	88.27	89.29



(a) JAFFE数据集



(b) KDEF数据集

图 4 加入注意力模块的多尺度特征融合实验混淆矩阵

2.4 对比实验与分析

本文的实验结果与现有的几种实验结果进行比较,结果见表 2 和表 3。

表 2 JAFFE 数据集的对比实验准确率	
论文方法	准确率/%
ICLR+AFL ^[23]	78.64
Multi-modal ^[11]	91.80
SFHSA ^[24]	92.21
Deep-Emotion ^[13]	92.80
本文方法	96.67

表 3 KDEF 数据集的对比实验准确率	
论文方法	准确率/%
K-PCANet ^[25]	80.20
Deep-Emotion ^[13]	86.33
Segment+2D Gabor ^[26]	86.84
AlexNet+LDA ^[27]	88.40
本文方法	89.29

在 JAFFE 数据集上,将本文实验结果与文献[23]、[11]、[24]、[13]进行比较。Sun 等^[23]利用先验知识提取和选择活跃特征,并将 DPL 模型应用于特征字典。本文融合了浅层特征与深层特征,增强了网络的泛化能力,准确率比其高了 18.03%。Wei 等^[11]融合了人脸关键点的低级经验特征和卷积神经网络的高级自学习特征。Saha 等^[24]提出了监督滤波器和谐搜索算法(SFHSA),有效去除了冗余特征。与上述两种方法相比,本文加入了注意力模块,使模型更加关注对表情识别影响较大的区域。且 Saha 等^[24]的方法需要手动提取大量特征,费时费力。本文的准确率比上述两种方案分别高了 4.87%和 4.46%。Minaee 等^[13]提出了基于注意力机制的 Deep-Emotion 模型,由特征提取模型和空间变换模型组成。本文提出的多尺度特征融合结构,使网络同时学习细节信息与抽象信息,考虑更加全面。本文的准确率比 Minaee 等^[13]的高了 3.87%。

在 KDEF 数据集上,将本文实验结果与文献[25]、[13]、[26]、[27]进行比较。实验中使用的数据集与以下对比论文的数据集相同,均为 980 张正面角度图像。Sun 等^[25]在利用 PCANet 和 LDANet 提取特征的基础上,加入核空间以此得到特征间的非线性相似性。本文引入了注意力机制,识别率比其高了 9.09%。Minaee 等^[13]为了更关注人脸表情的关键区域,提出了注意力卷积神经网络。Islam 等^[26]手动分割表情区域,并用多频率和多方向的 2D Gabor 滤波器提取其特征。相比以上两种算法,本文的多尺度特征融合机制考虑更加全面,比

其准确率分别高了 2.96%和 2.45%。Fei 等^[27]将预处理的人脸表情图像输入到预训练的 AlexNet 网络,并提取 FC6 层的特征,最后采用 LDA 分类。但该方法没有考虑人脸表情的显著性特征,本文的准确率高於它 0.89%。

3 结 论

本文提出了一种基于注意力机制与多尺度特征融合的人脸表情识别方法。首先输入图像到浅层特征提取模块和骨干网络;然后将得到的浅层注意力特征和深层特征进行融合;最后经过分类器进行分类。浅层特征提取模块中的通道-空间注意力模块提高了模型的特征提取能力。本文提出的方法在 JAFFE 和 KDEF 数据集上进行了实验,识别准确率分别为 96.67%和 89.29%;消融实验及对比实验证明了人脸表情的多尺度融合特征和注意力机制在此项任务中的重要性。

参考文献:

[1] Hynes E, Flynn R, Lee B, et al. An evaluation of lower facial micro expressions as an implicit QoE Metric for an augmented reality procedure assistance application[C]// 31st Irish Signals and Systems Conference. Letterkenny, Ireland; IEEE, 2020:9180173.

[2] Mavridou I, McGhee J T, Hamed M, et al. FACETEQ interface demo for emotion expression in VR[C] // Proceedings of 2017 IEEE Virtual Reality. Los Angeles, CA, USA; IEEE, 2017: 441-442.

[3] Verma B, Choudhary A. A framework for driver emotion recognition using deep learning and Grassmann manifolds [C] // 21st International Conference on Intelligent Transportation Systems. Maui, HI, USA; IEEE, 2018: 1421-1426.

[4] Li X, Hu W M, Zhang Z F, et al. Heat kernel based local binary pattern for face representation[J]. IEEE Signal Processing Letters, 2010, 17(3): 308-311.

[5] Jones J P, Palmer L A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex[J]. Journal of Neurophysiology, 1987, 58(6): 1233-1258.

[6] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C] // Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA; IEEE, 2005: 886-893.

[7] Cootes T F, Taylor C J, Cooper D H, et al. Active

- shape models-their training and application [J]. *Computer Vision and Image Understanding*, 1995, 61(1): 38-59.
- [8] Shi D C, Cai F, Du G Y. Facial expression recognition based on Gabor wavelet phase features[C]//Proceedings of 2013 International Conference on Image and Graphics. Qingdao, China: IEEE, 2013: 520-523.
- [9] Han D, Ming Y. Facial expression recognition with LBP and SLPP combined method[C]//12th International Conference on Signal Processing. Hangzhou, China: IEEE, 2014: 1418-1422.
- [10] Ding H, Zhou S K, Chellappa R. FaceNet2ExpNet: Regularizing a deep face recognition net for expression recognition[C]//Proceedings of 2017 IEEE International Conference on Automatic Face and Gesture Recognition. Washington, USA: IEEE, 2017: 118-126.
- [11] Wei W, Jia Q, Feng Y, et al. Multi-modal facial expression feature based on deep-neural networks[J]. *Journal on Multimodal User Interfaces*, 2020, 14(1): 17-23.
- [12] Chen J, Chu L L, Zhou Y. Facial expression recognition based on the ensemble learning of CNNs [C]//10th IEEE International Conference on Signal Processing, Communications and Computing. Univ Macau, Macau: IEEE, 2020:9259543.
- [13] Minaee S, Minaei M, Abdolrashidi A. Deep-emotion: Facial expression recognition using attentional convolutional network[J]. *Sensors*, 2021, 21(9): 3046.
- [14] 何志超, 赵龙章, 陈闯. 用于人脸表情识别的多分辨率特征融合卷积神经网络[J]. *激光与光电子学进展*, 2018, 55(7): 370-375.
- [15] 张红颖, 王汇三. 结合局部二值模式和梯度特征的双通道表情识别[J]. *激光与光电子学进展*, 2020, 57(14): 62-68.
- [16] 申毫, 孟庆浩, 刘胤伯. 基于轻量卷积网络多层特征融合的人脸表情识别[J]. *激光与光电子学进展*, 2021, 58(6): 148-155.
- [17] Tian Y M, Li M Q, Wang D. DFER-Net: Recognizing facial expression in the wild[C]//2021 IEEE International Conference on Image Processing, Anchorage. AK, USA: IEEE, 2021: 2334-2338.
- [18] Lyons M J, Budynek J, Akamatsu S. Automatic classification of single facial images[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 21(12): 1357-1362.
- [19] Lundqvist D, Flykt A, Öhman A. The Karolinska directed emotional faces (KDEF)[DS/CD]. Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998.
- [20] Ekman P, Friesen W V. Constants across cultures in the face and emotion[J]. *Journal of Personality and Social Psychology*, 1971, 17(2): 124-129.
- [21] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 2261-2269.
- [22] Park J, Woo S, Lee J Y, et al. BAM: Bottleneck attention module[EB/OL]. (2018-07-17) [2020-07-10]. <https://arxiv.org/abs/1807.06514>.
- [23] Sun Z, Chiong R, Hu Z P. Self-adaptive feature learning based on a priori knowledge for facial expression recognition[J]. *Knowledge-Based Systems*, 2020, 204: 106124.
- [24] Saha S, Ghosh M, Ghosh S, et al. Feature selection for facial emotion recognition using cosine similarity-based harmony search algorithm[J]. *Applied Sciences*, 2020, 10(8): 2816.
- [25] Sun Z, Hu Z P, Chiong R, et al. Combining the kernel collaboration representation and deep subspace learning for facial expression recognition[J]. *Journal of Circuits Systems and Computers*, 2018, 27(8): 1850121.
- [26] Islam B, Mahmud F, Hossain A. Facial region segmentation based emotion recognition using extreme learning machine[C]//Proceedings of 2018 International Conference on Advancement in Electrical and Electronic Engineering. Gazipur, Bangladesh: IEEE, 2018: 1-4.
- [27] Fei Z X, Yang E F, Li D D U, et al. Deep convolution network based emotion analysis towards mental health care[J]. *Neurocomputing*, 2020, 388: 212-227.

(责任编辑:康 锋)