

引用格式:史浩,邢瑜航,陈炼. 基于多尺度融合注意力机制的人脸表情识别研究[J]. 微电子学与计算机, 2022, 39(3): 34-40.
[SHI H, XING Y H, CHEN L. Facial expression recognition based on multi-scale feature fusion and attention mechanism[J].
Microelectronics & Computer, 2022, 39(3): 34-40.]DOI:10. 19304/J. ISSN1000-7180. 2021. 0799

基于多尺度融合注意力机制的人脸表情识别研究

史 浩,邢瑜航,陈 炼
(南昌大学 信息工程学院,江西 南昌 330031)

摘 要: 针对传统卷积神经网络在表情特征提取阶段容易丢失大量有用信息,无法提取到高判别性表情特征,从而导致表情识别率低的问题,提出一种基于多尺度特征融合注意力机制的人脸表情识别方法. 首先,采用 VGGNet16 来提取卷积特征. 为了避免表情特征信息的丢失,将网络中不同层次卷积层的输出特征图进行多尺度特征融合,引入上下文信息的同时提取更加丰富的表情特征信息;为了能够着重关注关键表情特征,在网络中引入了注意力机制. 该机制利用分组卷积操作对通道注意力模块进行改进,学习不同通道的权重信息,获取注意力特征图,增强特征的表达能力,抑制冗余信息的影响. 为了进一步提高提取到表情特征的可判别性,引入孤岛损失函数,并与 Softmax 分类损失函数联合使用构成新的损失函数. 最后,由于对全连接层进行了删减. 为防止网络出现过拟合问题,在卷积层引入了 DropBlock 策略. 实验结果表明,该模型在 Fer2013 和 CK+数据集上分别取得了 73. 32% 和 97. 40% 的平均准确率.

关键词: 计算机视觉;深度学习;人脸表情识别;特征提取;多尺度特征融合;注意力机制
中图分类号: TP391. 41 **文献标识码:** A **文章编号:** 1000-7180(2022)03-0034-07

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

SHI Hao, XING Yuhang, CHEN Lian

(School of Information Engineering, Nanchang University, Nanchang 330031, China)

Abstract: Aiming at the problem that traditional convolutional neural networks easily lose a lot of useful information in the expression feature extraction stage, and cannot extract high discriminative expression features, which leads to the problem of low expression recognition rate, a facial expression recognition method based on multi-scale feature fusion and attention mechanism is proposed. First, use VGGNet16 to extract convolutional features, multi-scale feature fusion is performed on the output feature maps of different layers of convolutional layers in the network, and context information is introduced while extracting richer expression feature information. In order to focus on key expression features, an attention mechanism is introduced in the network, the channel attention module is improved by grouping convolution operations, learning the weight information of different channels, obtaining attention feature maps, enhancing the expression ability of features. In order to further improve the discriminability of the extracted expression features, an island loss function is introduced, and combined with the Softmax classification loss function to form a new loss function. Finally, due to the deletion of the fully connected layer, the DropBlock strategy is introduced in the convolution layer to prevent the network from over fitting. The experimental results show that the model has achieved average accuracy rates of 73. 32% and 97. 40% on the Fer 2013 and CK+ datasets.

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

1 引言

人脸表情能通过非言语形式直观的表达人类的真实情感和意图,是传递情感信息、协调双方关系的重要方式.研究表明,在人们日常交流过程中,通过表情传递的信息量占比高达 55%.由此可知,人脸表情识别(Facial Expression Recognition, FER)是一个极具研究价值和意义的重要课题.

早期的人脸表情识别算法主要基于人为设计的特征提取方法,如 LBP^[1]、HOG^[2]、SIFT^[3]等,这些算法都受到研究人员先验知识的限制,泛化性不足,无法满足现实的应用需求.随着深度学习技术的快速发展,各种利用深度神经网络模型进行人脸表情识别的研究都取得了进一步的发展.其中卷积神经网络(CNN)凭借其强大的特征提取能力,被广泛应用于图像识别与分类领域.为了提高卷积神经网络对表情识别的准确率,研究人员对网络模型进行了多个方面的探索:

(1)在网络深度方面,通过不断增加网络层数,提高网络对图像的特征提取能力. VGGNet^[4]通过叠加多个相同卷积模块搭建深层的网络模型,使网络性能得到提升.何凯明等人^[5]提出 ResNet 网络,通过引入残差单元使网络在性能得到提升的同时解决了梯度消失的问题.但是网络深度的不断增加也带来了训练难度和能耗极具增加的问题.

(2)在网络宽度方面,GoogleNet 提出了 Inception 结构^[6],在卷积阶段使用了多个不同尺度的卷积核,通过整合不同尺度卷积核提取的图像特征信息,获取更具判别性的特征. Zagoruyko^[7]等提出一种基于 ResNet 的增加网络宽度的模型,得出经过加宽的 28 层 ResNet 网络具有比 1001 层 ResNet 网络更高的性能.但是在一定程度上使网络中的参数大幅增加,提升了训练难度,模型的泛化性能也大幅降低.

综上所述,研究人员将注意力机制引入卷积神经网络,通过学习自适应的为特征图上不同区域赋予不同的权重系数,使网络在较少能耗的条件下获取表现力更强的特征,提高特征提取的效率与准确率. Hu 等人^[8]提出了 SENet 网络,获取特征的通道依赖关系,为卷积神经网络带来显著的性能提升, Woo 等人^[9]提出 CBAM 网络,同时在空间与通道

两个维度进行特征的注意力操作,得到良好的识别效果.

本文设计了基于多尺度特征融合注意力机制的网络模型(Multi-scale ChannelAttention Net, MSCANet).使用改进的 VGGNet 作为主干网络提取人脸表情特征,同时采用多尺度特征融合策略融合不同层次的特征图,以提高低层次特征的利用率.为了防止出现过拟合,在网络中引入批处理归一化、DropBlock^[10]技术.本文在 Fer2013 及 CK+人脸表情数据集上进行了多组对比实验,实验结果表明,改进后的网络能够有效提高表情识别准确率,实现更好地分类效果.

2 多尺度特征融合注意力机制模型

2.1 改进的通道注意力模块

在卷积神经网络中,不同特征通道具有不同的响应,若将各通道赋予相同权重,则没有充分考虑不同通道对特征提取的重要程度.为了充分利用各个特征通道,本文设计了一个基于分组卷积^[11]的通道注意力模块(Group Convolution Channel Attention, GCA),模块结构如图 1 所示.

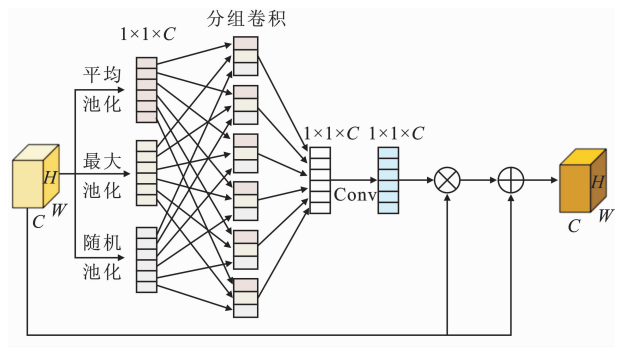


图 1 通道注意力模块

Fig. 1 Group Convolution Channel Attention

该模块在 SENet 仅使用最大池化的基础上添加了平均池化和随机池化,通过联合使用三种不同池化方法更加全面的获取不同通道的全局特征.经过三种池化处理后,特征图转化为三个通道描述符,且与输入特征图维数保持一致.

假设输入特征图 $F \in R^{H \times W \times C}$,经过上述步骤压缩成三个维度为 $1 \times 1 \times C$ 的通道描述符; F_{avg}^c 、 F_{max}^c 和 F_{sto}^c .为了进一步学习各通道间的相关性,引入分

组卷积操作,首先,将这三个通道描述符按照通道分组,相同通道的全局信息拼接在一起组成新的特征向量,每个新特征向量均包含 3 种全局信息,然后,分别使用卷积核大小 1×1 的卷积层对特征向量进行卷积,从而将 3 种全局信息自适应融合在一起,最终得到一个维度为 $1 \times 1 \times C$ 的通道描述符. 之后将其送入两个卷积核大小为 1×1 的卷积层进行学习,前一个卷积层的通道数为 $C/16$,后一个卷积层的通道数为 C ,学习不同通道的权重系数,计算公式如式(1);

$$P(F) = \sigma(W_2(\delta(W_1(\delta(G(F_{\text{avg}}; F_{\text{max}}; F_{\text{sto}})))))) \quad (1)$$

式中, σ 和 δ 分别表示 Sigmoid 函数和 Relu 函数, G 表示分组卷积, W_1 和 W_2 分别表示两个卷积层的卷中参数.

由于 Sigmoid 激活函数使每个元素值都被限定在 $[0, 1]$ 区间内,若直接与输入特征图进行点乘,会使得特征图输出响应变弱. 为此,在本实验中,通过点乘运算将注意力权重系数加权到输入特征图上,从而使有效特征得到增强,冗余特征得到抑制. 为了防止输出响应减弱,将输入特征图与加权后注意力特征图相加,增强模型的稳定性. 最终通道注意力模块输出:

$$F' = M_c(F) \otimes F + F$$

式中, F' 为经过注意力模块输出后的特征图, $M_c(F)$ 为注意力权重系数.

2.2 DropBlock

为防止网络出现过拟合,往往会在网络全连接层添加 Dropout,抑制部分神经元激活,阻止信息在网络中流动. 但 Dropout 无法实现对卷积层的正则化,为此本文引入了结构化丢弃策略——DropBlock,通过随机丢弃特征图中一部分相邻的块状区域,迫使网络利用人脸其它部位语义信息进行决策分类.

DropBlock 使用伯努利分布生成特征图二进制掩码矩阵,仅得到 0,1 两值,其中元素值为 0 的点作为基点,并将其周围方形区域也设置为 0,然后将所得矩阵与特征图进行哈达玛积,得到输出特征图,最后进行归一化操作确保 DropBlock 前后数据分布相同. DropBlock 主要有两个参数: block_size 和 γ ,前者表示随机丢弃方块的大小,后者表示伯努利函数概率值,即丢弃基点的概率,具体计算公式如式(2)

所示:

$$\gamma = \frac{1 - \text{keep_prob}}{\text{block_size}^2} \frac{\text{feat_size}^2}{(\text{feat_size} - \text{block_size} + 1)^2} \quad (2)$$

式中, keep_prob 为神经元保留概率,有效种子采样区域大小为 $(\text{feat_size} - \text{block_size} + 1)^2$,其中 feat_size 表示输入特征图大小. 由于在 DropBlock 中会存在一定的重叠,因此 γ 只是一个近似值. 实验过程中使用线性下降的 keep_prob 值可以获得更好的训练效果,即从初始值 1 下降到指定目标值.

DropBlock 掩码采样如图 2. 图 2 左方框区域表示有效的采样区域,灰色 $\times$ 表示所选择的基点,右框黑色 $\times$ 填充区域为要丢弃的特征块,这部分区域元素值都将重置为 0.

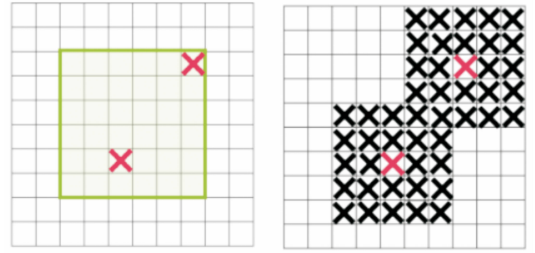


图 2 DropBlock 掩码采样

Fig. 2 Mask sampling in DropBlock

2.3 联合损失函数

Softmax 损失函数通过将神经网络计算后的多个神经元输出,映射到 $(0, 1)$ 区间,给出各种分类的概率情况,然后比较概率大小实现多分类. Softmax 对类间间距优化效果明显,但对于人脸表情同一类表情不同样本之间存在较大的差异的情况容易产生错误判断. 针对此问题,本文引入了 Island loss^[12],通过两者的联合使用,实现增大类间距离同时减小类内距离.

Island loss 是在中心损失的基础上改进而来,首先对每个聚类中心求余弦距离,然后对其 +1,使范围变为 $(0, 2)$,使得不同类别之间的间距变大. 聚类中心余弦距离为:

$$L_{\text{ILCOS}} = \sum_{c_j \in N} \sum_{\substack{c_k \in N \\ c_j \neq c_k}} \left(\frac{c_k \cdot c_j}{\|c_k\|_2 \|c_j\|_2} + 1 \right) \quad (3)$$

式中, N 为样本分类标签的集合, c_k 和 c_j 表示第 k 类和第 j 类表情的聚类中心, $\|c_k\|_2$ 和 $\|c_j\|_2$ 表示聚类中心到坐标原点的欧氏距离. Island loss 的计算公式为:

$$L_{\text{IL}} = L_C + \lambda_1 L_{\text{ILCOS}}$$

$$= \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 + \lambda_1 \sum_{c_j \in N} \sum_{c_k \in N} \left(\frac{c_k \cdot c_j}{\|c_k\|_2 \|c_j\|_2} + 1 \right) \quad (4)$$

式中, L_C 为中心损失函数, 用于优化类内间距, λ_1 为超参数, 表示占 Island loss 的权重。

本文通过对 Softmax loss 和 Island loss 联合使用, 来优化网络的训练, 使同类特征相互靠近, 不同类特征间距增大, 以实现更好地分类效果。调整后的损失函数计算公式如(5)所示:

$$L = L_s + \lambda L_{\text{Island}} \quad (5)$$

式中, λ 表示 Island loss 占联合损失函数的权重。参考 island loss 参数取值将 λ 设置为 0.01, 将 λ_1 设置为 10。

2.4 基于多尺度融合注意力机制网络模型

深度神经网络不同层次提取到的特征信息会存在较大差异, 浅层网络提取到的特征包含大量图像内容的细节信息, 深层网络则包含大量高层次的语义特征。随着网络不断加深, 每一层都会丢失部分信息, 当到达最后一层时往往只包含少量信息, 为解决这一问题本文提出了多尺度特征图融合的策略, 将不同卷积层产生的特征图进行融合, 引入上下文信息, 使得网络可以保留更多的特征信息。具体网络结构如图3所示。

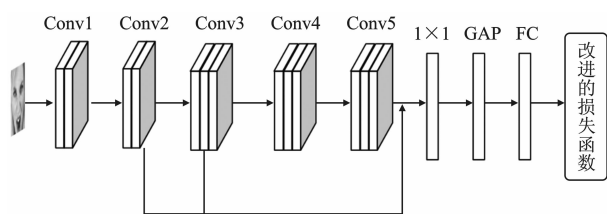


图3 多尺度特征图融合网络结构

Fig. 3 The structure of multi-scale feature map fusion

为实现不同尺度的人脸图像输入网络之后将卷积模块 conv2 和 conv3 在进入 maxpooling 前产生的特征图作为支路特征提取出来, 与 conv5 的特征图进行融合和尺度缩放, 然后通过 1×1 的卷积层进行通道降维, 使融合后结果与原网络相匹配。最后使用全局平均池化层 (GAP) 代替网络的前两个全连接层并将节点数设置为 7, 在大幅降低参数量同时实现与输出类别的直接关联。

同时, 为了更好地提取表现力更强的表情特征, 本文在网络中引入了注意力机制。在骨干网络的第二个卷积模块 conv_2 之后所有卷积层后添加 GCA 模块, 以实现输入特征图中有用信息的增强和对

冗余信息的抑制。注意力网络结构如图4所示。



图4 注意力网络结构

Fig. 4 The structure of Attention network

3 实验

3.1 人脸表情数据集与预处理

Fer2013 数据集由 35 887 张带表情标签的人脸图片组成, 其中训练集有 28 709 张, 公共测试集和私有测试集均有 3 589 张, 共包含愤怒、厌恶、恐惧、高兴、悲伤、惊讶、中立七种表情。

CK+ 表情数据集包含 123 人共 593 个由自然到高峰的表情序列, 其中 327 个表情序列添加了表情标签。其中包含 8 种基本的表情: 愤怒、蔑视、高兴、悲伤、惊讶、厌恶、恐惧、中立。为了与 Fer2013 数据集兼容, 本实验中去掉样本量较少的蔑视表情。

本文所用数据集的人脸表情图像并非仅包含人脸部分, 还包括与实验无关的其他身体部分及背景信息。为避免其他无关信息的影响, 本文使用 MTCNN 网络模型进行人脸检测和裁剪, 得到几乎不含背景信息的人脸图像, 最后将所得人脸图像通过双线性内插值算法进行尺度归一化, 输出保存为 224×224 的人脸图像。

为增强模型抗干扰稳定性, 对所使用的样本进行数据增强, 扩充样本数量, 对每张图像都使用翻转与旋转操作, 旋转角度为 $\pm 10^\circ$, 每隔 5° 旋转一次, 最终使样本数量扩大为原来 15 倍。数据增强效果如图5所示。

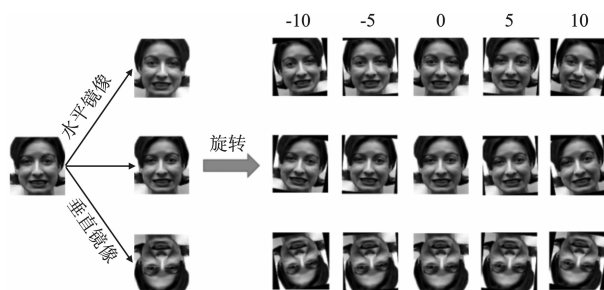


图5 数据增强

Fig. 3 Data Augmentation

3.2 实验环境

实验基于 Ubuntu16.04 操作系统, 使用 Py-

torch 深度学习框架进行实验,Python 版本为 3. 6. 5. 实验所用显卡为 NVIDIA GTX-1080 8G,内存为 16 G.

本实验采用随机梯度下降算法对模型参数进行更新,批尺度设为 32,动量设为 0. 9,初始学习率设置为 0. 01,采用学习率衰减策略逐步降低学习率.

3. 3 实验结果及分析

(1)不同 DropBlock 参数对模型的提升比较

在 DropBlock 中, γ 一般设置为 0. 9,而 block_size 大小则需根据具体网络进行调整,一般取 1、3、5、7、9 等,当取 1 时,DropBlock 效果与 Dropout 等价. 在 CK+ 及 Fer2013 数据集进行其他取值实验,探究不同丢弃块尺寸对准确率的影响,实验结果如表 1 所示.

表 1 不同丢弃块表情识别率对比

Tab. 1 Comparison of facial expression recognition rates of different sizes

数据集	block_size 尺寸				
	1	3	5	7	9
Fer2013	72. 19	72. 52	72. 90	73. 11	72. 64
CK+	94. 92	95. 35	95. 75	96. 27	95. 53

从表 1 可看出,表情识别率随丢弃块增大,先上升后下降,当丢弃块大小取值为 7 时,表情识别率最高. 当取值过小时,无法阻止信息传递,当取值过大时,则影响特征提取. 实验在 CK+ 数据集上取得了更大的提升,主要是由于 CK+ 数据集样本量较小,加入防止过拟合策略后提升了网络的泛化性能. 故本文选取 block_size=7 时来进行后续对比实验.

(2)不同改进策略对模型性能的影响

为了进一步探究本文所提方法的效果,本文进行了多组消融对比实验,以此来检验所提出不同改进策略的有效性与可靠性,在 Fer2013 和 CK+ 数据集上进行实验,多组不同实验设置相同的实验参数,实验结果如表 2 所示,其中×代表未采用,√代表采用.

表 2 不同改进策略对比实验结果

Tab. 1 Comparison of experimental results with different improvement strategies

编号	多尺度特征融合	Drop Block	联合损失函数	注意力模块	Fer 2013	CK+
1	×	×	×	×	70. 65	93. 28
2	×	√	√	×	72. 10	94. 35
3	√	×	√	×	72. 19	94. 92
4	√	√	×	×	72. 47	95. 82
5	√	√	√	×	73. 11	96. 27
6	√	√	√	√	73. 32	97. 40

从表 2 可以看出,与原始 VGG16 模型相比,本文所采用不同的改进策略对于模型识别率都有一定程度的提升,其中,添加改进的通道注意力模块 GCA 后,模型识别率提升较为明显,最终将多种策略共同应用后,在 Fer2013 和 CK+ 数据集上分别取得了 73. 32%与 97. 40%的准确率.

为了进一步分析本文所提模型对各类表情的识别准确率,本文分别绘制了在 Fer2013 和 CK+ 数据集上识别结果的混淆矩阵,混淆矩阵如图 6、图 7 所示.

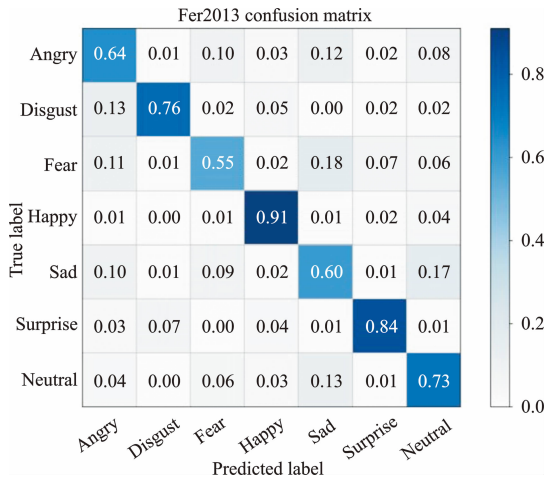


图 6 Fer2013 数据集混淆矩阵

Fig. 6 Fer2013 confusion matrix

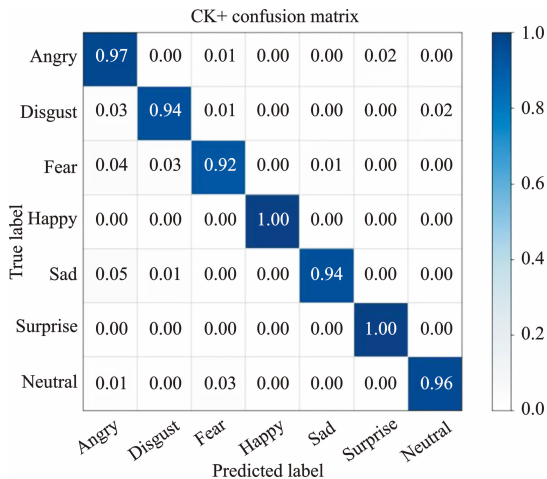


图 7 CK+ 数据集混淆矩阵

Fig. 7 CK+ confusion matrix

在 Fer2013 数据集中,本文模型对于高兴和惊讶这两类表情具有较高的识别准确率,而对于恐惧和悲伤这两类表情则无法准确的进行识别,识别准确率仅为 0. 55 和 0. 60. 这是因为 Fer2013 数据集

上各类样本分布极不平衡,训练集中悲伤表情样本量仅有约五百张,而高兴表情样本量则高达七千以上.且对于悲伤、厌恶这两类表情,都具有相似的嘴部区域或眼部区域的变化,表情特征辨识度不够高,从而容易出现识别错误的情况.

在 CK+数据集上,对于各类表情的识别准确率与上表中 Fer2013 的准确率相比都有了较大幅度的提升,且各类表情的识别准确率差距不大,这是因为该数据集取自可控的实验室条件下,人脸图像清晰,无遮挡和环境因素干扰.恐惧、厌恶和悲伤表情同样由于具有一定程度的相似性,增加了提取识别不同表情的那部分特征的难度.

(3)本文模型与其他经典模型的对比

表 3 为不同算法在 Fer2013 与 CK+数据集上的准确率对比结果.实验结果表明,本文算法在 Fer2013 数据集上的识别率相比 Kaggle 人脸表情识别大赛冠军 Tang 的识别率高出 2%.Lv 结合 FsNet 与 TcNet 对人脸图像进行分割,提取局部特征进行分类,取得了 71.37%的准确率.Xin 提出使用特征图聚类的方法对特征提取去冗余,也取得了较高的准确率,但不同类表情具有较高相似性影响识别结果的问题依旧无法解决.而在 CK+数据集上,Feng 设计了小尺度核卷积模块,在降低了网络参数的同时获取了 94.65%的准确率.Wu 使用基于面部特征点的几何特征向量进行表情识别,取得了 95.16%的准确率.Meng 为了缓解个人属性带来的影响,提出了对身份敏感的 IACNN 模型,在多个数据集上获得较好效果,但与本文所提算法相比仍有一定差距.本文算法与近来提出的多种算法相比都取得了更好地表现.

表 3 不同方法的准确率对比

Tab. 1Comparison of accuracy of different methods

数据集	方法	准确率/%
Fer2013	Tang ^[13]	71.20
	Lv ^[14]	71.37
	Liu ^[15]	72.10
	Xin ^[16]	71.67
	本文方法	73.32
	Feng ^[17]	94.65
	Meng ^[18]	95.37
CK+	Wu ^[19]	95.16
	Liu ^[15]	95.66
	本文方法	97.40

4 结束语

本文针对传统人脸表情识别算法在特征提取阶段随网络深度增加容易忽略一些重要特征的问题,提出了一种多尺度特征融合注意力机制的人脸表情识别算法.通过改进的 VGGNet 提取人脸图像特征,并采用多尺度特征图融合的策略,引入上下文信息,从而提高识别准确率.此外,文本还引入了注意力机制,基于分组卷积对通道注意力模块进行改进,提取表现力更强的表情特征.实验结果表明,本文模型对人脸表情识别具有较高准确率.在后续研究中,将进一步优化网络结构,提高检测效率,同时使用更接近自然环境的数据集进行研究,提高在实际应用中的价值.

参考文献:

[1] DING Y Y, ZHAO Q, LI B Q, et al. Facial expression recognition from image sequence based on LBP and Taylor expansion[J]. IEEE Access, 2017(5): 19409-19419. DOI: 10.1109/ACCESS.2017.2737821.

[2] WANG C L, LI Z R, DEY N J, et al. Histogram of oriented gradient based plantar pressure image feature extraction and classification employing fuzzy support vector machine[J]. Journal of Medical Imaging and Health Informatics, 2018, 8(4): 842-854. DOI: 10.1166/jmhi.2018.2310.

[3] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110. DOI: 10.1023/B:VISI.0000029664.99615.94.

[4] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//3rd International Conference on Learning Representations. San Diego:ICLR, 2015.

[5] 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. Las Vegas: IEEE, 2016: 770-778. DOI: 10.1109/CVPR.2016.90.

[6] 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. Boston: IEEE, 2015: 1-9. DOI: 10.1109/CVPR.2015.7298594.

[7] ZAGORUYKO S, KOMODAKIS N. Wide residual networks[C]//Proceedings of the British Machine Vi-

- sion Conference (BMVC). York, UK: BMVA Press, 2016. DOI:10.5244/C.30.87.
- [8] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132-7141. DOI: 10.1109/CVPR.2018.00745.
- [9] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the 15th European Conference ON Computer Vision (EC-CV). Munich, Germany: Springer, 2018: 3-19. DOI: 10.1007/978-3-030-01234-2_1.
- [10] GHIASI G, LIN T Y, LE Q V. DropBlock: a regularization method for convolutional networks[C]//Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018. Montréal, Canada: NeurIPS, 2018: 10750-10760.
- [11] ZHANG T, QI G J, XIAO B, et al. Interleaved group convolutions for deep neural networks[J]. arXivpreprint arXiv:1707.02725, 2017.
- [12] CAI J, MENG Z B, KHAN A S, et al. Island loss for learning discriminative features in facial expression recognition[C]//Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition. Xi'an, China: IEEE, 2018: 302-309. DOI: 10.1109/FG.2018.00051.
- [13] TANG Y C. Deep learning using linear support vector machines[J]. arXivpreprint arXiv:1306.0239, 2013.
- [14] 吕海, 童倩倩, 袁志勇. 基于人脸分割的复杂环境下表情识别实时框架[J]. 计算机工程与应用, 2020, 56(12): 134-140. DOI: 10.3778/j.issn.1002-8331.1903-0416.
- LV H, TONG Q Q, YUAN Z Y. Realtime architecture for facial expression recognition in complex scenes based on face region segmentation[J]. Computer Engineering and Applications, 2020, 56(12): 134-140. DOI: 10.3778/j.issn.1002-8331.1903-0416.
- [15] LIU X Q, ZHOU F Y. Improved curriculum learning using SSM for facial expression recognition[J]. The Visual Computer, 2020, 36(8): 1635-1649. DOI: 10.1007/s00371-019-01759-7.
- [16] 刘全明, 辛阳阳. 基于卷积神经网络特征图聚类的人脸表情识别[J]. 计算技术与自动化, 2020, 39(1): 106-111. DOI: 10.16339/j.cnki.jsjsyzdh.202001022.
- LIU Q M, XIN Y Y. Facial expression recognition based on convolutional neural network feature map clustering[J]. Computing Technology and Automation, 2020, 39(1): 106-111. DOI: 10.16339/j.cnki.jsjsyzdh.202001022.
- [17] 冯杨, 刘蓉, 鲁甜. 基于小尺度核卷积的人脸表情识别[J]. 计算机工程, 2021, 47(4): 262-267. DOI: 10.19678/j.issn.1000-3428.0056775.
- FENG Y, LIU R, LU T. Facial expression recognition based on small-scale kernel convolution[J]. Computer Engineering, 2021, 47(4): 262-267. DOI: 10.19678/j.issn.1000-3428.0056775.
- [18] MENG Z B, LIU P, CAI J, et al. Identity-aware convolutional neural network for facial expression recognition[C]//2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). Washington: IEEE, 2017: 558-565. DOI: 10.1109/FG.2017.140.
- [19] 吴珂, 周梦莹, 李高阳, 等. 基于角度几何特征的人脸表情识别[J]. 计算机应用与软件, 2020, 37(7): 120-124. DOI: 10.3969/j.issn.1000-386x.2020.07.019.
- WU K, ZHOU M Y, LI G Y, et al. Facial expression recognition based on geometrical features of angles[J]. Computer Applications and Software, 2020, 37(7): 120-124. DOI: 10.3969/j.issn.1000-386x.2020.07.019.

作者简介:

史浩男, (1994-), 硕士研究生. 研究方向为图像处理、深度学习.

邢瑜航 女, (1995-), 硕士研究生. 研究方向为深度学习、数据挖掘.

陈炼(通讯作者) 男, (1963-), 硕士研究生, 教授. 研究方向为数据挖掘、细粒度. E-mail: nc_chenlian@163.com