闫婷. 嵌入注意力机制的多尺度 Mini-Xception 人脸表情识别系统[J]. 智能计算机与应用,2025,15(6):67-72. DOI:10.20169/j.issn.2095-2163.240222

嵌入注意力机制的多尺度 Mini-Xception 人脸表情识别系统

闫 婷

(皖西学院 信息化建设与管理中心,安徽 六安 237000)

摘 要:卷积神经网络在人脸表情识别领域应用广泛。本文针对基于卷积神经网络的人脸表情识别算法在满足高识别率的同时实现模型的轻量化的问题,提出一种嵌入注意力机制的多尺度 Mini-Xception 人脸表情识别网络。本文引用 Mini-Xception 卷积神经网络作为主干网络,删除了传统模型的全连接层并使用深度可分离卷积替代,减少了模型参数。通过多尺度混合深度卷积的叠加提升卷积感受野,添加 CBAM 注意力机制提高对关注目标的检测效果,实现在不增加网络参数的同时提升网络性能。实验结果表明,本文方法在公开人脸表情数据集 FER-2013 上识别率达到 72.77%,同目前先进的表情识别网络对比本文所提方法有较高的准确率。此外,利用本文提出的模型对视频及图片中的人脸进行面部表情识别测试,识别结果表明本文所提网络模型具有良好的识别效果,可以更好应用于日常表情识别场景。

关键词:人脸表情识别;深度可分离卷积;多尺度卷积核;注意力机制

中图分类号:TP391.4 **文献标志码**:A **文章编号**:2095-2163(2025)06-0067-06

Multiscale Mini-Xception facial expression recognition integrating attention mechanism

YAN Ting

(Information Construction and Management Center, West Anhui University, Lu'an 237000, Anhui, China)

Abstract: Convolutional neural networks are widely used in the field of facial expression recognition. In order to solve the problem that the face expression recognition algorithm based on convolutional neural network can achieve the lightweight of the model while satisfying the high recognition rate, this paper proposes a multi-scale Mini-Xception face expression recognition network embedded with attention mechanism. In this paper, Mini-Xception convolutional neural network is used as the backbone network, which removes the full connection layer of the traditional model and uses depth-separable convolution instead to reduce the model parameters. The superposition of multi-scale mixed depth convolution is used to improve the convolutional receptive field, and the CBAM attention mechanism is added to improve the detection effect of the target of concern, so as to improve the network performance without increasing the network parameters. Experimental results show that the recognition rate of the proposed method on the open facial expression data set FER-2013 reaches 72.77%, which has a higher accuracy compared with the current advanced facial expression recognition networks. In addition, the model proposed in this paper is used to test facial expression recognition of faces in videos and pictures, and the recognition results show that the network model proposed in this paper has good recognition effect, and can be better applied to daily expression recognition scenes.

Key words: facial expression recognition; depth-separable convolution; multiscale convolution kernel; attention mechanism

0 引 言

随着人工智能的快速发展,人脸表情识别(Facial Expression Recognition,FER)技术在安全驾驶、医疗辅助、教育和娱乐等多个领域有着广泛应用。表情识别是在给定图像或视频序列中检测到人脸的位置后,采用有效算法构建合适模型来提取表情特征,最后利用提取的特征来识别相应的情感状态[1]。人脸表情主要被定义为以下 7 类:生气(Angry)、厌恶(Disgust)、恐惧(Fear)、高兴(Happiness)、悲伤(Sadness)、惊讶

基金项目:皖西学院校级自然重点项目(WXZR202220);安徽大学信息材料与智能感知安徽省实验室开放基金(IMIS202010);2022 年安徽省高校杰出青年科研项目(2022AH020091)。

作者简介:闫 婷(1990—),女,助理工程师,主要研究方向:表情识别。Email:48000001@wxc.edu.cn。

收稿日期:2024-02-22

(Surprise)和正常(Neutral)。

传统表情识别过程为图像处理和预处理、特征提取和分类识别。Gabor 小波变换、主成分分析(PCA)和局部二值模式(LBP)等是早期经典的特征提取算法。传统的表情识别方法对数据和设备的依赖较小,可以取得较高的识别率。但传统方法应用于现实场景中存在很多弊端,主要表现为传统算法很难融合到一个端到端(End-to-End)的模型中,同时设计的算法、环境干扰(如光照、复杂背景、角度等)和年龄性别等特征因素对识别效果均有很大影响,导致最终识别效果不佳。

随着人工智能的快速发展,卷积神经网络(CNN)在计算机视觉研究中得到广泛应用,并在图像处理任务中取得显著成效。基于深度学习的方法可以实现直接从输入数据到分类结果的"端到端"学习,受环境影响较小,在表情识别上具有明显优势。标准的卷积神经网络由输入层、卷积层、池化层、全连接层组合而成,并通过反向传播确定参数的方式来揭示海量数据之间的关系,该网络的优点在于可以直接输入图像,自行抽取图像特征,是一个可以实现端到端(End-to-End)的模型,在实际应用上具有良好的鲁棒性和运算效率。经典的卷积神经网络模型有 VGGNet[2]、GoogleNet[3]和 ResNet[4]等,越来越多的研究者运用这些优秀的模型完成图像分类、图像分割、图像检测等任务,人机交互领域得到高速发展。李勇等学者[5]提出一种改进的 LeNet-5 卷积神经网络,通过跨连接的方式将低层次特征和高层次特征连接组成分类器来进行面部表情识别。马中启等学者[6]设计一种基于多特征融合密集残差卷积神经网络,该方法提升了分类准确率,同时具有较强的鲁棒性。Zhou 等学者[7]提出多尺度卷积神经网络,利用多个尺度的图像输入到神经网络中进行学习得到优化后的参数,实验结果表明该方法比单一尺度卷积神经网络准确率更高。张波等学者[8]提出一种改进的卷积神经网络,通过添加可分离卷积网络和输出位置加入通道注意力算法,参数量和运算成本明显降低,同时输出通道的权值可以按不同特征的贡献大小进行调整,该方法在保证分类准确率的同时实现更轻的网络结构。张鹏等学者[9]提出一种将多尺度特征和通道注意力机制相结合的算法,基于 Inception 网络结构同时引入空洞卷积来提取不同尺度上人脸表情特征信息,再引入通道注意力机制来增强网络提取关键目标的能力,最终识别准确率优于许多经典算法。

由于卷积神经网络在局部特征提取上有明显优势,层次具有多样性,因此对人脸表情的识别与分类具有较好的效果[10]。各项研究结果证明,采用卷积神经网络的表情识别算法识别率较高同时受环境影响较小,较传统模型能更好地应用于社会生活多个领域。但是卷积神经网络模型仍存在一些问题,如深层卷积网络计算机量大、参数多,在移动端和嵌入式设备等应用场景中推广困难。

针对上述问题,本文提出一种嵌入注意力机制的多尺度 Mini-Xception 人脸表情识别网络,共进行了 3 个方面的改进:

(1)本文引用 Mini-Xception 卷积神经网络作为主干网络,删除了传统模型的全连接层并使用深度可分离卷积替代,减少了模型参数。

(2)通过多尺度卷积核的叠加提升卷积感受野。

(3)同时在网络架构中添加卷积块注意力模型 CBAM,实现在不增加网络参数的同时提升网络性能[11]。

最后,将输入图像经过图像归一化、数据增强等预处理操作,输入到改进后的网络模型,经过多次实验证明,改进后网络模型可以取得较高的识别率,同时更好地应用于日常的表情识别应用场景。

1 相关算法

1.1 Mini-Xception 网络

Arriaga 等学者[12]于 2017 年提出 Mini-Xception 卷积神经网络,该模型结合深度可分离卷积和残差模块组成 4 个残差深度可分离卷积块,去除最后全连接层,模型参数量进一步减少同时训练时间也显著降低,模型架构的参数约 60 000 个,与标准卷积相比减少了约 80 倍。在 FER-2013 数据集上进行情感分类测试获得了 66%的准确率,并且最终的架构可以存储在一个 855 kb 的文件中,使得在移动端和嵌入式设备等应用场景中推广成为可能。因此本文利用 Mini-Xception 卷积神经网络进行人脸表情识别。

Mini-Xception 卷积神经网络架构如图 1 所示。每个深度可分离卷积之后对卷积层进行批量归一化操作、并采用 ReLU 激活函数,随后用全局平均池化层进行下采样,最后利用 Softmax 激活函数进行分类。

1.2 CBAM 注意力机制

为了提升 Mini-Xception 网络获取关键目标的能力,本文在深度可分离卷积模块后添加 CBAM 注意力机制模块,提升网络提取关键特征提取能力。注意力机制是根据训练得到不同特征的贡献大小调

整通道权重，然后将训练的权重分布作用在原图像特征上，增强图像中有用特征权重并减少非显著性特征的权重。

CBAM 注意力机制包括 2 个独立的子模块，分别是通道注意力机制和空间注意力机制，由两者组合连接而成[13]。添加 CBAM 的过程共 2 步，具体步骤如下。

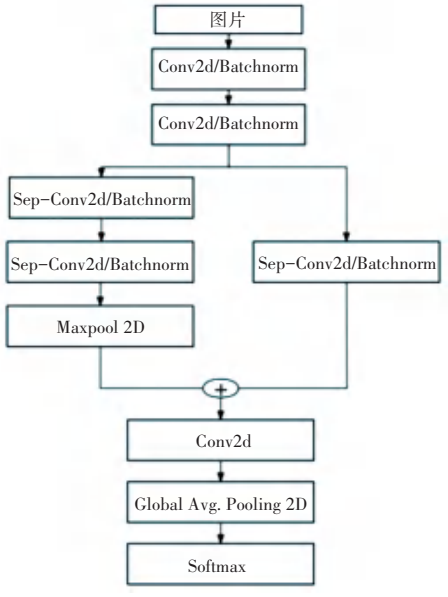


图 1　Mini-Xception 网络模型

Fig. 1　Mini-Xception model structure

步骤 1　计算通道注意力特征图。将输入的特征图 $F(F \in R^{C\times H\times W})$ 分别进行全局最大池化(Global Max Pooling)和全局平均池化(Global Average Pooling)操作，得到 2 个 $1\times1\times C$ 的特征图，然后将得到的特征图分别送入多层感知机(MLP)中。将 MLP 输出的特征进行卷积、逐元素的加和操作，通过 Sigmoid 激活操作，继而生成最终的通道注意力特征图，即 $M_C(F)$。其计算公式为：

$$M_C(F)=\sigma(MLP(AvgPool(F))+MLP(MaxPool(F))) \tag{1}$$

其中，σ 表示 Sigmoid 函数。

最后与输入特征图 F 逐元素相乘，得到融合通道注意力后的特征图。研究用到的公式如下：

$$F'=M_C(F)\otimes F \tag{2}$$

步骤 2　计算空间注意力特征图。将步骤 1 得到的特征图 $M_C(F)'$ 作为本模块的输入特征图。首先分别进行基于通道的平均池化和最大池化操作，得到 2 个 $H\times W\times1$ 的特征图，然后将得到的特征图基于通道进行拼接(Concat)。最后经过一个 7×7 卷积操作，降维为 1 个通道，即 $H\times W\times1$，再经过 Sgmoid 函数得到空间注意力特征图，推得的公式为：

$$M_S(F)=\sigma(f^{7\times7}([AvgPool(F);MaxPool(F)])) \tag{3}$$

最后与该步骤的输入特征做乘法，得到最终生成的特征。计算公式如下：

$$F''=M_S(F')'\otimes F' \tag{4}$$

1.3　多尺度深度可分离卷积

原深度可分离卷积由深度卷积(DepthWise Convolution)和逐点卷积(PointWise Convolution)两部分组成，深度卷积网络架构如图 2 所示。本文提出多尺度深度可分离卷积，在深度可分离卷积中融合不同尺度深度卷积，融合不同尺度的多样性特征扩大卷积感受野，进而提高网络特征学习能力和网络的识别率。

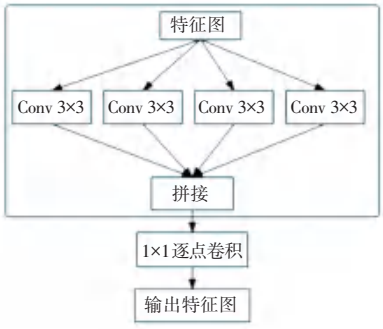


图 2　深度可分离卷积结构

Fig. 2　Depth-separable convolution structure

本文中将特征图共分为 3 组，卷积核尺寸选取{3×3，5×5，7×7}，然后将每组输出进行拼接。多尺度深度可分离卷积结构见图 3，本文利用 2 层 3×3 卷积替代 5×5 卷积核，利用 3 层 3×3 卷积替代 7×7 卷积核，降低计算量和参数量。

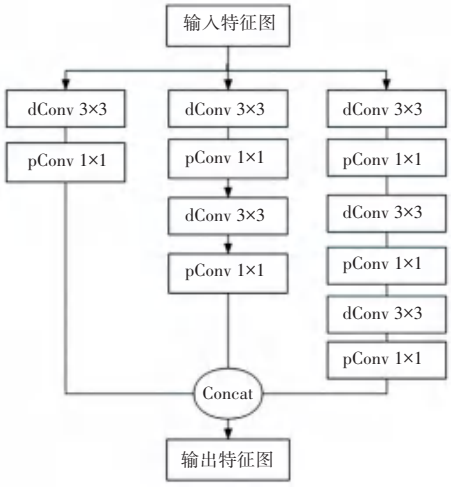


图 3　多尺度深度可分离卷积单元结构

Fig. 3　Multiscale depth-separable convolution structure

为了对比不同尺寸卷积核对表情识别分类准确率的影响,选取不同尺寸组合的网络结构在 FER-2013 数据集[14]上进行对比实验,实验结果见表 1。通过表 1 实验结果分析可知,多尺度卷积核选择{3×3,5×5,7×7}组合,准确率较基础网络提高了 0.98%。

表 1 不同尺寸卷积组合实验结果对比

Table 1 Comparison of experimental results of convolution with different size

卷积核组合	识别率/%
3×3	71.10
3×3,5×5	72.12
1×1,3×3,5×5	72.02
3×3,5×5,7×7	72.08

2 嵌入注意力机制的多尺度 Mini-Xception 人脸表情识别

本文在 Mini-Xception 网络上进行设计和改进,首先采用多尺度深度可分离卷积来提取多样性特征,扩大卷积感受野,卷积后分别加入 CBAM 注意力模块,CBAM 注意力机制结合通道和空间注意力机制模块,通过学习重要的特征、抑制无效的特征来有效地提高网络模型的识别率和鲁棒性。网络模型结构如图 4 所示。从图 4 可以看出,本文网络结构由 4 个残差多尺度深度可分离卷积模块和 4 个 CBAM 模块组成,每一次卷积操作后都经过批量归一化操作和激活函数 ReLU,最后经过全连接层后,利用 Softmax 激活函数进行分类。

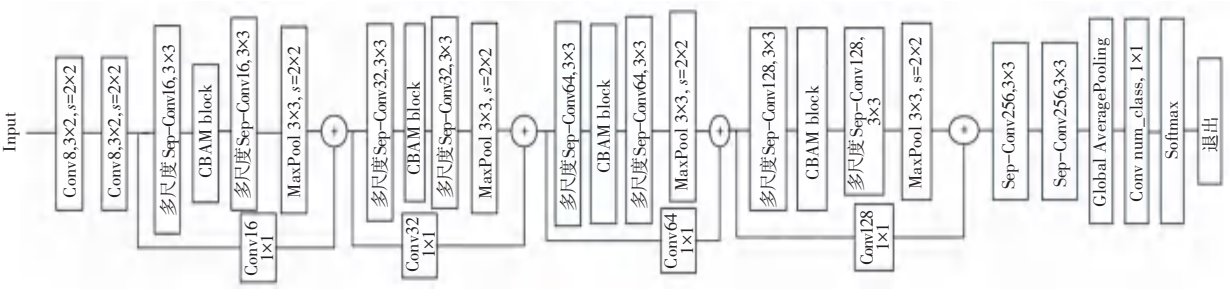


图 4 网络模型结构

Fig. 4 Network model structure

3 实验与分析

本文的研究是基于 Keras 深度学习框架来搭建网络模型,操作系统为 Windows 10 专业版,运行内存为 16 G。

3.1 数据集介绍

本文使用 Fer2013 公开人脸表情数据集作为实验数据集,数据集共包含 35 886 张人脸照片,分为训练集 28 709 张、验证集和测试集各 3 589 张,图像均为 48×48 像素。该数据集图像均来自网上获取,图片样本较丰富,包括不同年龄、表情、场景及性别样本,满足自然环境下的表情识别分类,人脸表情识别具有一定的难度。

3.2 数据预处理及实验设置

3.2.1 人脸检测

人脸检测是判断图片中是否存在人脸的操作,是表情识别系统中的首要步骤,本文采用 Haar 级联分类器进行人脸检测识别,分类采用主流 AdaBoost 算法。主要步骤如下:

(1)通过图片上滑动来提取检测区域内图片 Haar 特征值。

(2)采用 AdaBoost 算法作为分类算法,通过串联的方式组成强分类器,判断检测区域是否存在人脸。

(3)如果检测区域存在人脸则返回该区域坐标,不存在则重复步骤(1)。

(4)待图片全部扫描完毕,结束检测。

3.2.2 数据预处理和数据增强

由于原始图像存在角度偏差、光照不均匀等问题,本文图像数据输入到构建完整的神经网络模型前,需对其进行预处理操作。图像数据输入到网络模型前,利用图像增强技术对训练图像进行随机裁剪、旋转等增广操作,扩充数据集进而增强模型的表达能力。数据增强操作的参数设置见表 2,训练参数设置见表 3。

表 2 数据增强参数设置

Table 2 Data enhancement parameter settings

增广类型	图片随机转动	水平偏移	竖直偏移	水平翻转	缩放
参数	[-10, 10]	0.1	0.1	是	0.1

表 3　训练参数设置

Table 3　Training parameter settings

参数	取值
Batch Size	32
Epoch	100
Patience	10
Num Classes	7

3.3　消融实验

为验证本文所提出的改进模型的有效性，本文在 Fer2013 数据集上进行消融实验，消融实验结果见表 4。表 4 中，Mini-Xception 为没有进行任何改进的基础网络，Mini-Xception+CBAM 表示在 Mini-Xception 网络上添加 CBAM 注意力模块，多尺度卷积核为原始网络中采用多尺度深度可分离卷积核，最后为本文提出的改进网络模型。

表 4　消融实验

Table 4　Ablation experiment

方法	识别率/%
M-Xception	71.14
M-Xception+CBAM	72.04
多尺度卷积核	72.08
Mini-Xception+CBAM+多尺度卷积核	72.77

由表 4 可以看出，添加注意力模型和多尺度深度可分离卷积相比 Mini-Xception 网络的表情识别准确率提高了 1.63%。

3.4　其他算法对比

为验证本文所提出改进网络模型的有效性，本文在 Fer2013 数据集上同目前先进的表情识别网络进行实验对比，以此验证本文方法的有效性。表 5 是不同方法在 Fer2013 数据集上的识别率对比结果，由实验结果对比可以得出本文所提出的改进网络模型有较高的准确率。

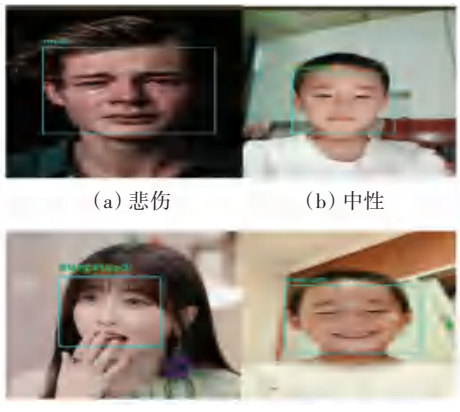
表 5　不同算法在 FER-2013 数据集上的识别率

Table 5　The recognition rate of different methods on FER-2013 dataset

方法	准确率/%
张俞晴等学者[15]	68.10
Khemakhem 等学者[16]	70.59
Liu 等学者[17]	72.11
Zhou 等学者[18]	70.91
严春满等学者[19]	68.90
Chen 等学者[20]	72.36
本文方法	72.77

4　现实应用能力测试

为验证本文提出的人脸表情识别网络模型在现实场景中的应用推广能力，本文设计一个人脸表情识别系统用于人脸表情检测。本文选择在 FER2013 数据集上训练的识别率最高的网络模型，利用该模型对视频及图片中的人脸进行面部表情识别，结果如图 5 所示。从图 5 的识别结果可以看出，本文所提网络模型具有良好的识别效果，可以更好应用于日常表情识别场景。



(a) 悲伤　(b) 中性　(c) 惊讶　(d) 开心

图 5　人脸表情识别图

Fig. 5　Facial expression recognition images

5　结束语

本文针对基于卷积神经网络的人脸表情识别算法在提高识别率的同时兼顾轻量化的问题，提出一种融合通道注意力多尺度卷积核的实时人脸表情识别网络。首先，采用 Mini-Xception 卷积神经网络，删除了传统模型的全连接层并使用深度可分离卷积替代，减少了模型参数，利用多尺度混合深度卷积的叠加提升卷积感受野，通过添加 CBAM 注意力机制提高对关注目标的检测效果，最后将输入图像经过图像归一化、数据增强等预处理操作，输入到改进后的网络模型进行分类。在公开人脸表情数据集 FER-2013 上经过多次实验证明，本文改进后网络模型可以取得较高的识别率。此外搭建系统模型进行应用场景测试，结果验证本文所提模型具有良好的识别效果。

参考文献

[1] 何超，侯明. 基于改进卷积神经网络的人脸表情识别方法[J]. 信息技术，2022(5)：107-111.

[2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint

arXiv,1409.1556, 2015.
[3] SZEGEDY C , LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[J]. arXiv preprint arXiv,1409.4842, 2014.
[4] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, 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
[5] 李勇,林小竹,蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别[J]. 自动化学报,2018,44(1):176-182.
[6] 马中启,朱好生,杨海仕,等. 基于多特征融合密集残差 CNN 的人脸表情识别[J]. 计算机应用与软件,2019,36(7):197-201.
[7] ZHOU Shuai, LIANG Yanyan, WAN Jun, et al. Facial expression recognition based on multi-scale CNNs [C]// Proceedings of the 11th Chinese Conference on Biometrics, (CCBR2016). Cham:Springer,2016:503-510.
[8] 张波,兰艳亭,李大威,等. 基于卷积网络通道注意力的人脸表情识别[J]. 无线电工程,2022,52(1):148-153.
[9] 张鹏,孔韦韦,滕金保. 基于多尺度特征注意力机制的人脸表情识别[J]. 计算机工程与应用,2022,58(1):182-189.
[10] 董翠,罗晓曙,蒙志明,等. 一种基于改进 VGG 网络的表情识别算法[J]. 现代电子技术,2022,45(10):63-68.
[11] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]// Proceedings of European Conference on Computer Vision. Cham:Springer, 2018: 3-19.
[12] ARRIAGA O, VALDENEGRO-TORO M, PLÖGER P . Real-time convolutional neural networks for emotion and gender classification[J]. arXiv preprint arXiv,1710.07557,2017.
[13] 谢银成,黎曦,李天,等. 基于改进 ResNet 和损失函数的表情识别[J]. 自动化与仪表,2022,37(4):64-69.
[14] DHALL A, GOECKE R, LUCEY S, et al. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark [C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway, NJ: IEEE,2011: 2106-2112.
[15] 张俞晴,何宁,魏润辰. 基于卷积神经网络融合 SIFT 特征的人脸表情识别[J]. 计算机应用与软件,2019,36(11):161-167.
[16] KHEMAKHEM F, LTIFI H. Facial expression recognition using convolution neural network enhancing with pre-processing stages [C]//Proceedings of 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA). Piscataway,NJ:IEEE, 2019:1-7.
[17] LIU Xiaoqian, ZHOU Fengyu. Improved curriculum learning using SSM for facial expression recognition [J]. The Visual Computer, 2020,36:1635-1649.
[18] ZHOU Jiancan, JIA Xi, SHEN Linlin, et al. Improved softmax loss for deep learning based face and expression recognition[J]. Cognitive Computation and Systems, 2019, 1(4): 97-102.
[19] 严春满,张翔,王青朋. 基于改进 MobileNetV2 的人脸表情识别[J]. 计算机工程与科学,2023,45(6):1071-1078.
[20] CHEN Yizhen, HU Haifeng. Facial expression recognition by inter-class relational learning[J]. IEEE Access, 2019,7: 94106-94117.