

文章编号: 2095-2163(2023)05-0009-09

中图分类号: TP391.41

文献标志码: A

改进的残差网络和混合注意力的人脸表情识别

李弟文, 潘伟

(西华师范大学 计算机学院, 四川 南充 637009)

摘要: 针对传统残差网络对人脸表情特征提取存在的泛化能力差、识别准确度低等问题, 本文实现了基于混合注意力机制和残差网络相结合的方法, 从而能够更准确快速的实现人脸表情进行识别。提出使用 CABasicBlock 代替 ResNet34 中的 basic block 作为骨干网络, 并加入空间、通道注意力机制来提取人脸的全局面部特征, 以及与表情重要关联的局部面部特征, 使得网络收敛更快。在网络的残差模块后加入批量归一化和 PReLU 激活函数, 以及在全连接层前加入了 Dropout、全局平均池化, 能有效预防训练过程发生过拟合, 提升其泛化能力; 引入 Focal Loss 函数平衡训练数据样本均衡性的问题, 使用余弦退火策略对训练过程中的学习率进行动态衰减, 从而减少训练时间。通过实验结果数据表明, 在 FER-2013 测试数据集上的准确率为 73.7%, 说明此模型拥有更好的效果。最后, 基于以上方法再结合 OpenCV 和 PyQt5, 构建了一个实时表情识别的可视化检测结果输出。

关键词: 深度残差网络; CABasicBlock; 表情识别; 混合注意力机制; 余弦退火策略

Face expression recognition based on improved residual networks and hybrid attention

LI Diwen, PAN Wei

(School of Computer Science, China West Normal University, Nanchong Sichuan 637009, China)

[Abstract] To address the problems of poor generalization ability and low recognition accuracy of traditional residual networks for face expression feature extraction, this paper implements a method based on a combination of hybrid attention mechanism and residual networks, which can achieve more accurate and fast recognition of face expressions. The CABasicBlock is proposed to replace the basic block in ResNet34 as the backbone network, and the spatial and channel attention mechanisms are added to extract the global facial features of faces and the local facial features associated with expressions to make the network converge faster. The addition of batch normalisation and PReLU activation functions after the residual module, as well as Dropout and global average pooling before the fully connected layer, can effectively prevent overfitting in the training process and improve its generalisation capability. This reduces the training time. The experimental results show that this model has better results, with an accuracy of 73.7% on the FER-2013 dataset respectively. Finally, a visualisation of the detection results based on the above method combined with OpenCV and PyQt5 was constructed for real-time expression recognition.

[Key words] deep residual networks; CABasicBlock; expression recognition; hybrid attention mechanism; cosine annealing strategy

0 引言

通常人类的情感表达主要通过其面部表情来侧面反映, 人类通过说话层面传递的情感信息只占 7% 左右, 而面部表情在情感表达过程中传递了大约 55% 以上的情感信息^[1]。总体而言, 随着计算机技术的进一步发展, 人脸表情的自动识别也成为了计算机视觉方面的重要研究对象, 其主要为人与机器交流互动、人类情感预测、人脸 3D 表情驱动等众多

应用领域提供有效科学的辅助, 在人工智能领域起到基础铺垫的作用。人脸表情识别的过程可以分为 3 个阶段^[2]: 第一阶段为人脸检测, 需要检测视频画面或静态图像中的人脸部分; 第二阶段是对检测的人脸使用特征网络提取人脸表情重要特征, 这些特征是分类网络能够更好区分不同表情的重要依据; 第三阶段是将提取到的特征输入到分类网络中进行输出, 这里输出的就是表情类别矩阵, 包括每类表情的种类及其置信度。

作者简介: 李弟文(1997-), 男, 硕士研究生, 主要研究方向: 深度学习、图像处理; 潘伟(1976-), 男, 博士, 副教授, 主要研究方向: 深度学习、机器学习。

通讯作者: 潘伟 Email: panwei@cwnu.edu.cn

收稿日期: 2022-06-09

为进一步提高面部表情识别的准确率,本文通过在深度残差网络中嵌入空间和通道注意力机制,对重要面部表情特征进行加强,并使用 CABasicBlock 替换原有残差网络中 Block 结构,使网络自动学习每个特征通道的重要性,提升对重要通道特征的关注度,对模型训练中的损失函数和学习率衰减策略进行相应调整,加速模型收敛,并对训练数据进行增强和标签平滑处理来提升模型泛化能力。该方法在 Fer2013 人脸表情数据集上得到了有效验证。

1 相关理论研究

早期的人脸表情识别方法需要人为制定提取规则,然后按照制定的标准来提取特征(如: LBP^[3]、HOG^[4]等),从而进行相应的计算得到表情特征矩阵输出,最后使用支持向量机/KNN 算法/贝叶斯分类算法等方法来对输出的特征矩阵进行分类。此外,有的方法在特征提取完毕后,为了避免特征维数过高,使用 PCA 主成分分析法对其降维,减小算法的时间复杂度。如:成亚丽等人^[5]针对局部外观特征提取造成噪声污染像素等问题,提出了一种改进型的 LDP 编码方案。利用 Sobel 算子代替 Kirsch 掩模来提取图像中的梯度信息,从而避免噪声像素点的影响,并将局部外观特征与全局几何特征相结合后使用 PCA 降维,提升表情识别率并避免敏感噪声影响。

随着深度学习的不断发展,基于卷积神经网络的 AlexNet 在 2012 年的 ImageNet 大赛上一举夺魁,其识别精度高于第二名 10 多个百分点,这让深度学习重新大放异彩,让许多人开始了对其研究。在人脸表情识别方面,基于深度学习的代表算法-卷积神经网络,通过构建多个卷积层来提取人脸表情更深层次的特征,从而能够很好的规避人为制定的单一提取规则,优势明显并且效率高,逐渐成为目前的主流方法。邓楚婕^[6]使用卷积神经网络结合注意力机制,设计了基于金字塔结构的 CNN 来提高准确率。戴蓉^[7]等人提出了一种生成对抗网络的跨域面部表情识别。其中包括特征嵌入、对抗性学习和分类模块,利用联合学习来降低原始数据和目标数据之间的分布差距,并利用无监督生成对抗网络进行优化,根据域自适应方法给出表情分类。陈昌川^[8]等人针对当前表情识别方法通常依赖面部全局特征而忽略局部特征提取等问题,提出一种基于局部表征的表情识别 EAU-CNN 算法。该算法的网络采用 8 个并行的特征提取分支,每个分支支配不

同维全连接层;分支的输出按照注意力自适应地连接,以突出不同局部候选区域的重要程度,最后经 Softmax 函数来进行表情的七分类。He^[9]等人提出的深度残差网络 ResNet,解决了深层次网络训练的性能退化问题,为大模型训练带来了可能。此后,越来越多研究者开始结合残差网络来识别人脸表情。如:王军等人^[10]针对人脸面部遮挡特征难以提取的问题,提取了 VGG16 的主干网络和 ResNet 残差网络,来构建新的双通道遮挡感知神经网络,使得网络能有效综合评判表情分类的精确度。宋玉琴等人^[11]在文献中针对神经网络提取面部表情特征在反向传播训练阶段存在梯度爆炸或消失等问题,提出了一种把残差网络和注意力机制相结合的多尺度深度可分离表情识别网络,通过多层多尺度深度可分离残差单元的叠加进行不同尺度的表情特征提取,并使用 CBAM 注意力机制进行表情特征的筛选,提升有效表情特征权重的表达。陈加敏等人^[12]在文献中针对不同的表情之间具有细微的类间差异信息,由此在残差网络的基础上提出一种注意力金字塔卷积残差网络模型(APRNET50),该模型融合金字塔卷积模块、通道注意力和空间注意力,提供端到端的训练。高涛等人^[13]针对 ResNet 网络残差模块前的卷积层进行改进,提出使用并行的小卷积代替原来的卷积,最后将残差模块的输出送入到全连接层进行特征的融合,使其可以提取到更深层次且不同尺度的图像特征,最后通过 Softmax 分类器进行分类。

2 残差网络与注意力机制

2.1 残差网络

深度残差网络与传统卷积神经网络相比,是在网络中引入残差模块,有效地缓解了网络模型训练时反向传播的梯度消失问题,进而解决了深层网络难以训练和性能退化的问题。残差学习单元结构如图 1 所示。

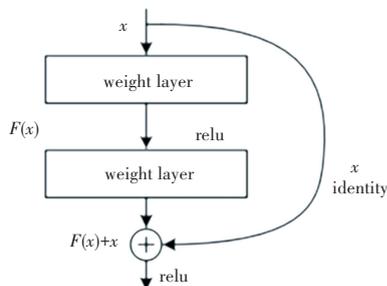


图 1 残差学习单元

Fig. 1 Residual learning units

图中, x 表示输入, $F(x)$ 表示残差映射。当残差 $F(x) = 0$ 时,残差学习单元的作用就是输入为输出的恒等映射。其输出定义为

$$H(x) = F(x) + x \quad (1)$$

2.2 注意力机制

注意力机制是模拟人类大脑观察外界的一种思维方式,其能够较好地选择网络中关键特征信息,聚焦于对当前任务更为关键的特征,并对低关联特征进行过滤或减小权重,提高任务处理的效率和准确性,可以解决网络参数冗余带来的计算效率下降问题。近年来,结合注意力机制的网络模型在 NLP、ASR 及 Computer Vision 等不同的领域中都表现出优越的性能,其有效性得到了广泛的验证。一般来说,注意力机制主要分为空间注意力机制和通道注意力机制两种。

通道注意力机制是根据不同通道的特征赋予其不同的权重,使得经过通道的重要特征信息能够被放大,从而被网络所学习,其结构如图 2 所示。输入一个通道数为 C 、高度 H 、宽度 W 的浅层特征图 $M \in R^{C \times H \times W}$, 通过使用平均池化和最大池化的方法减小特征维度,分别得到输出的 M_{avg} 和 M_{max} 。然后依次

输入到共享网络 MLP 中,MLP 的结构由 1×1 卷积、ReLU 激活函数、 1×1 卷积组成。随后对 MLP 输出的这两个特征向量按元素相加求和,得到 $C \times 1 \times 1$ 大小的特征图,最后经过 Sigmoid 函数激活得到基于通道注意力的新特征权重图 $W_c \in R^{c \times 1 \times 1}$ 。具体计算方式为

$$W_c(F) = \sigma(\text{MLP}(\text{AvgPool}(M)) + \text{MLP}(\text{MaxPool}(M))) = \sigma(W_1(W_0(M_{avg})) + W_1(W_0(M_{max}))) \quad (2)$$

其中, σ 表示 Sigmoid 激活函数, W_0 和 W_1 表示 MLP 网络的权重。

空间注意力机制在图像空间上分配权重,利用特征间的关系构建空间注意力机制,来寻找对于表情较有区分力的局部位置特征信息,其结构如图 3 所示。输入一个特征图 $M \in R^{C \times H \times W}$, 使用平均池化和最大池化分别输出 $M_{avg}^s \in R^{1 \times H \times W}$ 和 $M_{max}^s \in R^{1 \times H \times W}$ 特征图,将两个特征图使用 Concat 融合拼接后得到新的特征图,再将其输入到卷积核为 7×7 的卷积层,经 Sigmoid 函数激活得到空间特征权重图。 W_s 的计算公式为

$$W_s(F) = \sigma(\text{Conv}_{7 \times 7}([\text{AvgPool}(M); \text{MaxPool}(M)])) = \sigma(\text{Conv}_{7 \times 7}(M_{avg}^s; M_{max}^s)) \quad (3)$$

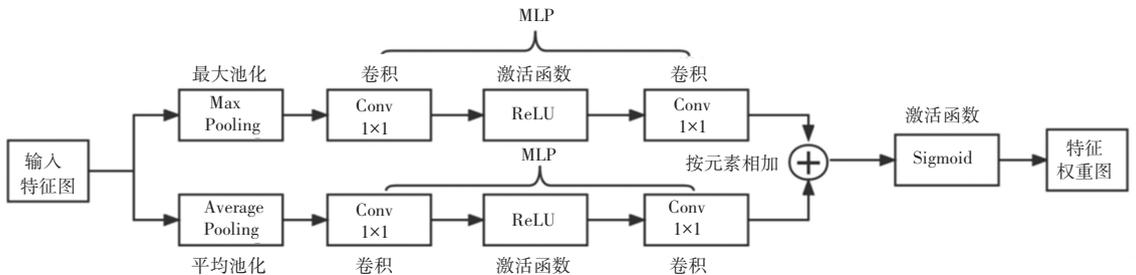


图 2 通道注意力模块

Fig. 2 Channel attention module

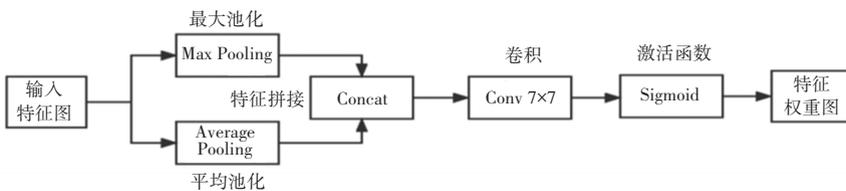


图 3 空间注意力模块

Fig. 3 Spatial attention module

3 基于深度残差网络和混合注意力的表情识别

本文使用深度残差网络作为人脸表情识别的骨干网络,该网络相比普通网络多了由残差学习单元组成的残差模块,解决了网络层数堆叠过深导致分类性能和准确率不能提高的问题。该模块的引入有效地缓解模型训练时反向传播梯度消失问题,从而解决了

对深层网络训练的退化问题。使用的骨干网络模型是以 ResNet34 网络作为基础模型,残差网络由若干个 Block 残差模块组成,其拥有跨层直连的 shortcut,而残差模块又通过不同数量的残差学习单元组成。ResNet34 经过了 4 个 Block,也叫 4 个 layer 层,分别对应有 3、4、6、3 个 BasicBlock 的残差模块。输入图片经过第一个卷积层和最大池化,进入 Block 块结构,如图

4所示。每一个 Block 包含两种不同的 BasicBlock: Conv Block(卷积块)和 Identity Block(残差块)。

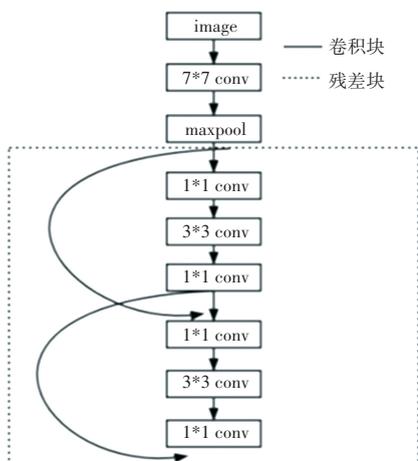


图4 ResNet-34网络的第一个Block结构

Fig. 4 The first Block structure of the ResNet-34 network

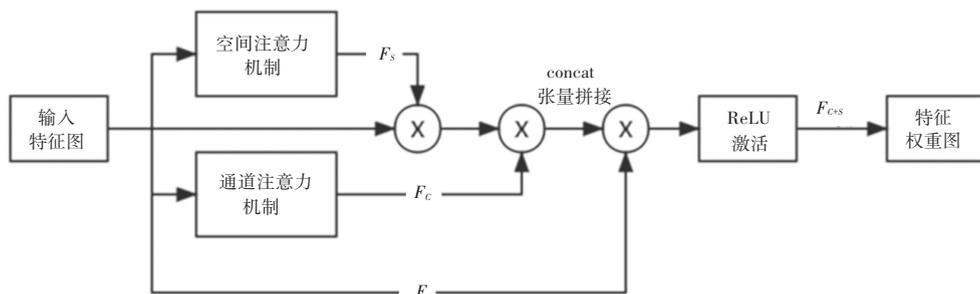


图5 混合注意力模块

Fig. 5 Mixed attention module

3.2 Focal Loss 损失函数

由于在 FER2013 训练集数据中,数量最少的厌恶表情只有 436 张,而数量最多的开心表情却有 7 215 张,厌恶与开心两种表情分别占训练总样本的占比差(23.6%)超过了 20%。为了避免样本不平衡带来的损失震荡问题,在训练过程中使用 Focal Loss 函数来计算预测值和标签值的损失。该损失函数是基于多分类交叉熵损失函数的改进版本,并自带 Softmax 操作,其作用在于通过赋予易分类样本较小的权重,赋予难分类样本较大的权重,使模型在训练时更专注于难分类的样本,在一定程度上解决了样本分布不平衡的问题。在表情识别多分类任务下,通过 α 可以抑制正负样本的数量失衡,通过 γ 可以控制简单样本、难区分样本的数量失衡, Focal Loss 损失函数计算公式为^[14]

$$FL(p, y) = - \sum_n \sum_i [\alpha (1 - p_i^n)^\gamma y_i^n \ln p_i^n + (1 - \alpha) (p_i^n)^\gamma (1 - y_i^n) \ln(1 - p_i^n)] \quad (4)$$

其中, $\alpha \in [0, 1]$ 为权重因子; γ 为一个调制因

3.1 混合注意力模块

为了使表情局部特征区域更加突出,使用注意力机制让网络学会关注重点信息,给予更大权重得到更细化的信息。本文在提取到的浅层特征上加入通道和空间注意力的混合模块给其分配不同的权重。

将通道注意力与空间注意力机制进行混合,构成新的权重分配模块,实现通道和空间上特征相互融合后的权重特征图。如图 5 所示, F 作为输入特征,经过通道注意力机制加权和空间注意力机制加权后,分别得到 F_c 和 F_s , 再将 F_c 与 F_s 进行特征拼接融合,融合后的特征与原输入特征图相乘,最后通过 ReLU 激活获得新特征权重图 F_{C+S} 。

子参数,范围在 $[0, 5]$; p_i^n 表示第 n 个样本为类别 i 的概率。

3.3 基于混合注意力残差网络的表情特征提取

本文在 ResNet34 的基础上对 basic block 进行了改进。融合 CoordAttention 坐标注意力机制,提出一个新的 CABasicBlock,其通过嵌入位置信息到通道注意力模块中,从而使网络获取更大的感受野并减少计算开销。为避免 Conv2d 和全局池化导致位置信息丢失,坐标注意力机制使用两个并行通道,从而高效地将特征编码整合空间坐标信息到生成的特征图中。如图 6 所示,该模块是通过左边的一个残差支路和右边的一个 CoordAttention 结构组成,沿着水平和垂直两个空间方向进行特征聚合,分别对应下方的 X AvgPool 和 Y AvgPool,分别返回一对方向感知注意力图;在空间维度上通过 Concat 和 Conv2d 来压缩通道,后面接批量归一化 BN 层和 Non-linear 层来编码垂直方向和水平方向的空间信息,再使用分割操作后,各自通过 1×1 的 Conv2d 卷积得到与输

入特征图一样的通道数,然后使用 Sigmoid 激活函数输出,与残差之路一起合并归一化加权,这样可以充分捕获到人脸中的表情位置信息。

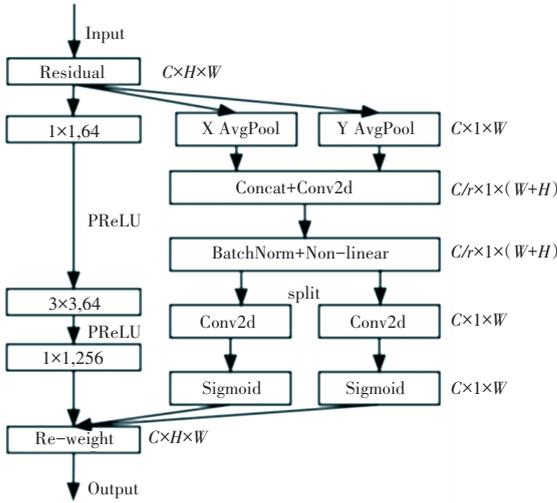


图 6 CABasicBlock 结构

Fig. 6 CABasicBlock structure

基于 ResNet34 提出的混合注意力残差网络,主要由空间注意力模块、通道注意力模块、残差模块、

3x3卷积层、最大池化的组合层、自适应平均池化,以及 Dropout 层、全连接层、Softmax 层构成,网络具体结构如图 7 所示。为减小网络的计算量,需要将输入图片进行下采样后再输入进网络层,使图像的分辨率从 48 * 48 下采样至 24 * 24。

首先使用一个 3x3 卷积提取人脸图像特征信息,然后通过通道注意力和空间注意力模块进行表情特征的提取融合获得输出的混合权重特征图;经过最大池化下采样处理后去除冗余信息,把混合后的特征送入到模型的 4 个 Layer 层;每个 Layer 由两个 CABasicBlock 组成,然后经过自适应平均池化层加强特征图与类别的一致性,对空间特征信息进行求和实现降维。在接入全连接层前,尽可能在保证主要分类特征信息基础上减少网络参数,并使用 Dropout 使网络训练过程中随机丢弃神经元,降低特征信息之间太过密切的交互作用,避免过拟合。最后,送入全连接层输出 7x1 的特征向量,分别代表 7 种表情分类,再使用 Softmax 输出概率最高的表情类别标签。

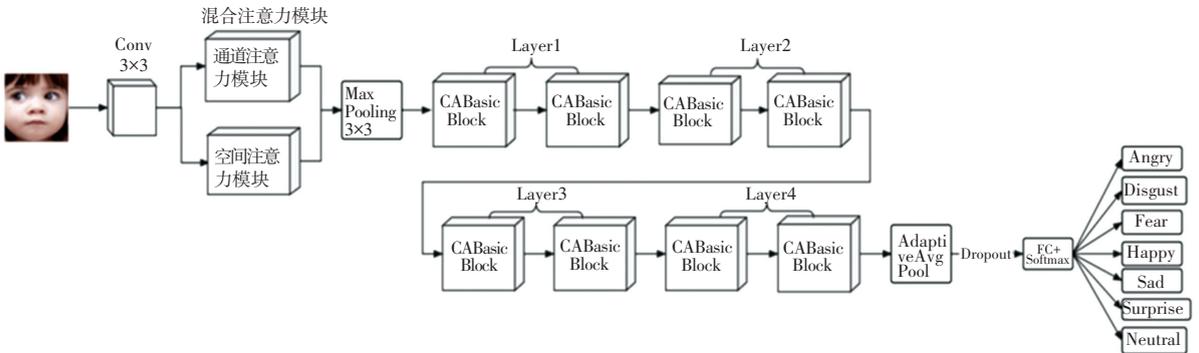


图 7 人脸表情识别网络模型示意图

Fig. 7 Schematic diagram of the face expression recognition network model

4 实验结果与分析

实验使用 python3.7 语言进行编程实现,在 Pycharm 中使用 Pytorch 深度学习框架搭建神经网络模型。实验硬件环境:CPU 为 Intel(R) Xeon(R) CPU E5-2670 v3,GPU 显卡为 NVIDIA GeForce RTX 3060,内存为 16 GB。在 Fer2013 数据集实验中的迭代轮数(epoch)均设为 300 轮,学习率初始为 0.001,在迭代 50 轮后学习率以 0.9 倍衰减,批次大小设为 128,使用 Adam 优化器,权重衰减系数为 5E-4。

4.1 数据集

为了对本文方法进行验证,本次实验采用 Fer2013^[15] 人脸表情数据集进行训练和测试。Fer2013 数据集是由大小为 48x48 像素的灰度图片和情感标签组成,情感标签包含 7 类基本表情,由 35 886 张采集于不同复杂环境因素下的人脸表情图像组成。其中,训练数据集 28 709 张图片,验证数据集图片 3 589 张,测试集 3 589 张。如图 8 所示,情感标签分别对应于数字标签 0~6。其中,0 代表生气,1 代表厌恶,2 代表恐惧,3 代表开心,4 代表伤心,5 代表惊讶,6 代表中性。



图8 Fer2013人脸表情数据集中对应七种表情图

Fig. 8 Fer2013 face expression dataset corresponds to seven expression maps

每种表情具体数量的柱状图如图9所示,横轴为每种表情的类别,纵轴为其对应的数量。

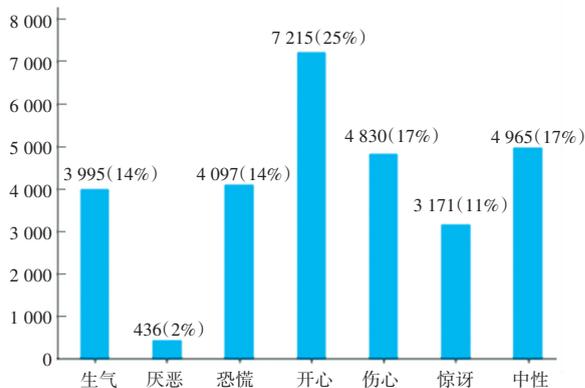


图9 Fer2013表情数据图片数量分布情况图

Fig. 9 Fer2013 expression data image count distribution chart

4.2 数据增强预处理

由于原始 Fer2013 数据集存在数据样本分布不均匀问题,在训练表情识别网络之前,为了防止网络过快地过拟合^[16],需要进行训练数据增强。数据增强后的图像信息的细节特征也能被深度卷积神经网络所提取,并学习到更深层次的高维有效特征空间,提高了模型分类的准确率。本文实验对训练图像作翻转、镜像、Gamma 变换、随机擦除、噪声、随机旋转角度、随机裁剪等图像增强处理方法,其中主要的增强方法参数设置见表1,增强后的数据集各表情类别数量分布如图10所示。

表1 四种数据增强方法的参数设置

Tab. 1 Parameter settings for the four data enhancement methods

数据增强方法	参数设置
随机旋转	30°~90°
高斯噪声	均值为0,方差为0.001
随机裁剪	原面积50%~100%
Gamma 变换	γ为2

经过试验发现,当只使用基本的 ResNet34 网络且不使用数据增强策略进行分类时,在验证集上的准确率不到60%。但在增加随机裁剪等数据增强策略进行基础网络训练时,最终模型的准确率有

8%的提升,并且配合使用这6种数据增强策略,在未改进的 ResNet 基本分类网络上最终能达到68%左右的验证集准确率。

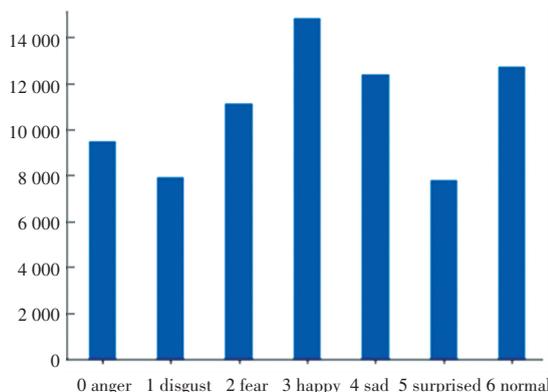


图10 数据增强后的人脸表情数据图片数量分布情况图

Fig. 10 Distribution of the number of images of face expression data after data enhancement

4.3 评价标准

在计算机视觉领域中的图像分类问题,一般采用的分类评价标准包括 Accuracy(准确率)、Confusion matrix(混淆矩阵)、召回率等指标,用来评价分类方法性能的好坏。针对人脸表情识别而言,由于要计算7种表情每类的识别精度,所以评价标准使用准确率和混淆矩阵来进行对比实验。分类准确率的计算公式为

$$Accuracy = \frac{TP + TN}{P + N} \quad (5)$$

其中,TP代表正例分类正确的数量;TN代表负例分类正确的数量;P为所有正例;N为所有负例。

分类混淆矩阵是将正例分类正确(TP)、正例分类错误(FP)、负例分类正确(TN)和负例分类错误(FN)的4个指标一起呈现在表格中,可以更加清晰的展示分类模型预测各个类别的概率。混淆矩阵的对角线为该类别预测准确率,值越高表明分类正确的效果越好,对角线之外为分类成其它类别的概率。

4.4 模型训练和实验结果分析

模型优化器采用 Adam 算法,相比之前 SGD 随

机梯度^[17]下降算法具有计算效率高、所需内存少等优点。在网络训练过程中,使用动态学习率衰减来使权重更新更加科学合理,使用 Focal Loss 来计算模型预测值和真实标签的损失,从而回传梯度调整模型权重参数,完成反向传播的迭代训练。模型训练时设置训练批次 Epoch 为 300,在全连接层之后使用 Dropout,概率值为 0.3。训练初始网络各层权重参数初始化使用 Kaiming 的权重初始化方法,批量训练的 batch_size 大小为 256,使用批量归一化防止梯度爆炸。通过在训练过程中不断调节各个超参数的值,最后经过约 200 个批次的迭代训练,模型的 Loss 损失值逐渐下降并趋近于平滑曲线,说明这次模型训练取得了收敛。

图 11 展示了训练过程中验证集损失值和准确率随着训练批次 Epoch 增加的变化曲线。图形曲线反映了模型在 Fer2013 数据集的验证集上识别准确率达到 72.86%,并可以看出在网络训练前面 200 个 epoch 内的准确率波动较大,后面逐渐趋向于训练稳定的状态。图 12 为训练集的损失值变化曲线,随着训练批次的增加逐渐趋于平滑曲线。在测试集上分类的混淆矩阵如图 13 所示,纵坐标为真实标签,横坐标为预测标签,总体识别准确率达到 73.7%。其在识别开心表情的准确率为 89%,而在识别恐惧表情的准确率仅为 57%,这是受到测试集数据样本数量不均衡及质量不佳的影响,以及表情本身的相似性,影响了模型的预测。

率),当训练过程中该指标不再增大(或减小),则适当的降低学习率。

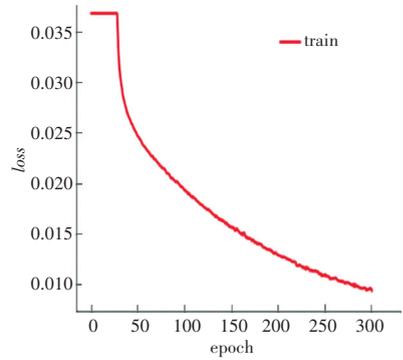


图 12 改进模型训练损失值曲线

Fig. 12 Improved model training loss value curve

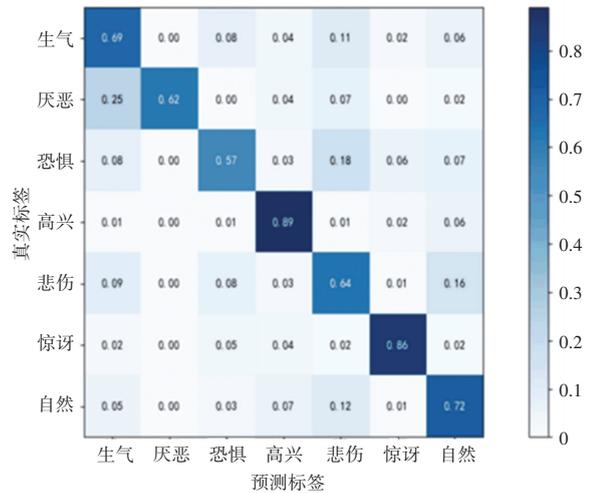


图 13 改进模型在 fer2013 测试集上的分类混淆矩阵

Fig. 13 Classification confusion matrix of the improved model on the fer2013 test set

为了评估改进模型的效果,将本文模型与并行卷积神经网络、ResNet18、ResNet34、VGG19、DenseNet、SENet18 等网络模型在 Fer2013 数据集上进行对比,对比结果见表 2。

表 2 FER-2013 私有测试集的准确率对比

Tab. 2 Accuracy comparison of the FER-2013 private test set

方法	准确率/%
并行卷积神经网络	65.6
ResNet18	66.7
ResNet34	70.61
VGG19	69.47
DenseNet	71.06
SENet18	70.32
本文模型	73.7

由此可见,在网络中加入混合注意力机制和

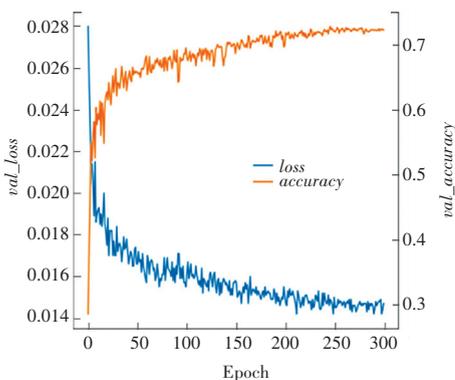


图 11 改进模型在验证集上的损失值和准确率曲线

Fig. 11 Loss value and accuracy curve of the improved model on the validation set

在网络训练时,当发现验证集的损失值在连续 10 轮的训练中,仍然没有下降,则提前终止训练,避免网络过拟合。选择使用余弦退火学习率衰减策略 (CosineAnnealingLR),该策略能让学习率随训练 epoch 的变化图像类似于余弦函数图像,在使用时可以提前指定某一个性能评价指标(如验证集准确

CABasicBlock 模块,可以使网络更加关注有助判别的显性表情特征,提高了表情识别的准确率,验证了本文设计模型在人脸表情识别上的有效性。

4.5 人脸表情实时识别

为了验证本文算法在真实场景中的有效性,通过使用 OpenCv 来调取本地计算机的摄像头。对摄像头捕捉的画面进行人脸检测获得脸框坐标,并依据坐标对其画面进行裁剪获得人脸图像,然后对图像进行灰度化操作,使视频帧图像由三通道的彩色图像转化为灰度图像,便于后续送入训练好的模型网络中进行预测输出,实时获取表情识别结果。使用 Python 当中的 PyQt5^[18] 模块,搭建了基于混合注意力和残差网络的高效人脸表情识别系统,该系统通过调用训练好的模型对用户输入的图片或者视频

进行表情识别。人脸表情识别系统界面整体有打开摄像头、选择人脸图片、拍照识别、关闭摄像头等 4 个按钮,分别对应 4 种不同的功能。打开摄像头其调用的是连接电脑的摄像头或者笔记本自带的摄像头,以实现人脸表情实时拍摄识别;选择人脸图片功能就是上传本地电脑上的图片进行检测,从而得出图片中人脸的表情;拍照识别是打开摄像头用户点击拍照,把此时的图像作为输入进行表情识别分类。

表情识别效果如图 14、图 15 所示。其中,绿色框标为人脸区域,其上边的英文单词就是模型预测的表情结果类别标签;由 SoftMax 公式分别计算出各个表情类别的概率^[19],并返回概率最高的分类,将其作为输出预测标签值放到界面上进行显示。

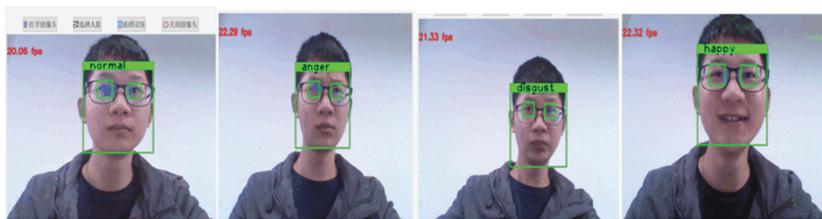


图 14 正常、生气、厌恶、开心表情示例

Fig. 14 Examples of normal, angry, disgusted and happy expressions

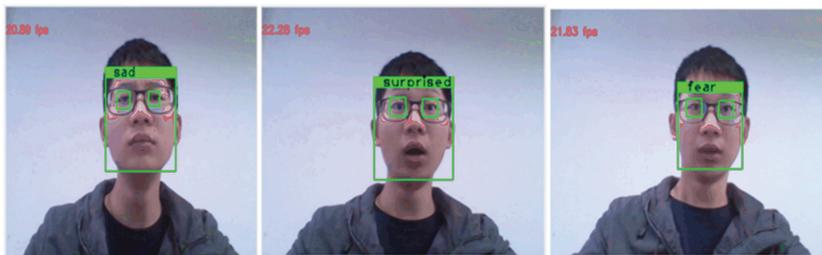


图 15 伤心、惊讶、害怕表情示例

Fig. 15 Examples of sad, surprised and scared expressions

5 结束语

本文提出了一种优化残差网络和混合注意力机制的人脸表情识别模型,并通过使用该网络模型搭建了实时人脸表情识别系统。从以上对比实验可以看出,本文所提方法在 Fer2013 数据集上的识别准确率有所提高。由此可见,本文所提方法在人脸表情重点区域的特征提取有一定的提升效果,使得网络的分类性能得到了改善。

参考文献

[1] 李旻择,李小霞,王学渊,等. 基于多尺度核特征卷积神经网络的实时人脸表情识别[J]. 计算机应用, 2019, 39(9): 2568-2574.

[2] 魏为民,孟繁星,才智,等. 人脸表情识别综述[J]. 上海电力大学学报, 2021, 37(6): 597-602.

[3] Dong Feiyan. Application of distributed training of convolutional neural network in facial expression recognition [J]. Software, 2020, 3(1): 160-164.

[4] 王信,汪友生. 基于深度学习与传统机器学习的人脸表情识别综述[J]. 应用科技, 2018, 45(1): 65-72.

[5] 成亚丽,秦飞龙,李政文. 改进 LDP 结合几何特征融合的人脸表情识别[J]. 计算机工程与设计, 2021, 42(9): 2577-2584.

[6] 邓楚婕. 基于卷积神经网络的人脸表情识别方法研究[D]. 广州: 华南理工大学, 2020.

[7] 戴蓉. 一种基于条件生成对抗网络的面部表情识别技术[J]. 计算机应用与软件, 2020, 37(8): 166-170, 232.

[8] 陈昌川,王海宁,黄炼,等. 一种基于局部表征的面部表情识别算法[J/OL]. 西安电子科技大学学报, 1-10[2021-09-26].

[9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

- [10]王军,赵凯,程勇. 基于遮挡感知卷积神经网络的面部表情识别方法[J/OL]. 计算机工程;1-11[2021-08-16]. <https://doi.org/10.19678/j.issn.1000-3428.0059166>.
- [11]宋玉琴,高师杰,曾贺东,等. 嵌入注意力机制的多尺度深度可分离表情识别[J/OL]. 北京航空航天大学学报;1-10[2021-09-23]. <https://doi.org/10.13700/j.bh.1001-5965.2021.0114>.
- [12]陈加敏,徐杨. 注意力金字塔卷积残差网络的表情识别[J/OL]. 计算机工程与应用;1-11[2021-09-28]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20210702.1749.004.html>.
- [13]高涛,邵倩,张亚南,等. 基于深度残差网络的人脸表情识别研究[J]. 电子设计工程,2020,28(23):101-104.
- [14]LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] //Proceedings of the IEEE international conference on computer vision. 2017; 2980-2988.
- [15]FER2013 人脸表情数据库[EB/OL].(2019-04-09)[2020-03-17]. <http://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-challenge/data>.
- [16]亢洁,李佳伟,杨思力. 基于域适应卷积神经网络的人脸表情识别[J]. 计算机工程,2019,45(12):201-206.
- [17]陈俊名. 基于深度卷积神经网络的人脸表情识别研究[D]. 重庆:重庆邮电大学,2020.
- [18]何永强,秦勤,王俊鹏. 改进深度学习块卷积神经网络的人脸表情识别[J]. 计算机工程与设计,2019,40(3):850-855.
- [19]卢官明,何嘉利,闫静杰,等. 一种用于人脸表情识别的卷积神经网络[J]. 南京邮电大学学报(自然科学版),2016,36(1):16-22.

(上接第8页)

- [37]ZHANG Fuguo, QI Shumei, LIU Qihua, et al. Alleviating the data sparsity problem of recommender systems by clustering nodes in bipartite networks [J]. Expert Systems with Applications, 2020, 149: 113346.
- [38]HU G, ZHANG Y, YANG Q. Conet: Collaborative cross networks for cross-domain recommendation[C]//Proceedings of the 27th ACM international conference on information and knowledge management. 2018: 667-676.
- [39]ZHU Y, LIN J, HE S, et al. Addressing the item cold-start problem by attribute-driven active learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 32(4): 631-644.
- [40]LU Y, FANG Y, SHI C. Meta-learning on heterogeneous information networks for cold-start recommendation [C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020: 1563-1573.
- [41]FENG Junmei, XIA Zhaoqiang, FENG Xiao, et al. RBPR: A hybrid model for the new user cold start problem in recommender systems[J]. Knowledge-Based Systems, 2021, 214: 106732.
- [42]RIYABI M, SOHRABI M K. Providing effective recommendations in discussion groups using a new hybrid recommender system based on implicit ratings and semantic similarity [J]. Electronic Commerce Research and Applications, 2020, 40: 100938.
- [43]ABBASI-MOUD Z, HOSSEINABADI S, KELARESTAGHI M, et al. CAFOB: Context-aware fuzzy-ontology-based tourism recommendation system[J]. Expert Systems with Applications, 2022, 199: 116877.
- [44]ZHANG Q, ZHANG G, LU J, et al. A framework of hybrid recommender system for personalized clinical prescription [C]//2015 10th international conference on intelligent systems and knowledge engineering (ISKE). IEEE, 2015: 189-195.
- [45]HANSEN C, HANSEN C, MAYSTRE L, et al. Contextual and sequential user embeddings for large-scale music recommendation [C]//Proceedings of the 14th ACM Conference on Recommender Systems. 2020: 53-62.
- [46]JIN B, CHENG K, ZHANG L, et al. Partial relationship aware influence diffusion via a multi-channel encoding scheme for social recommendation[C]//Proceedings of the 29th ACM international conference on information & knowledge management. 2020: 585-594.
- [47]LIU D, LIAN J, WANG S, et al. KRED: Knowledge-aware document representation for news recommendations [C]//Proceedings of the 14th ACM Conference on Recommender Systems. 2020: 200-209.
- [48]HU Y, XIONG F, LU D, et al. Movie collaborative filtering with multiplex implicit feedbacks [J]. Neurocomputing, 2020, 398: 485-494.
- [49]DARBAN Z Z, VALIPOUR M H. GHRS: Graph-based hybrid recommendation system with application to movie recommendation [J]. Expert Systems with Applications, 2022, 200: 116850.
- [50]SYMEONIDIS P, JANES A, CHALTSEV D, et al. Recommending the video to watch next: An offline and online evaluation at you.tv. de [C]//Proceedings of the 14th ACM Conference on Recommender Systems. 2020: 299-308.
- [51]CHEN L, WU Z, CAO J, et al. Travel recommendation via fusing multi-auxiliary information into matrix factorization [J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2020, 11(2): 1-24.
- [52]CHEN Lei, CAO Jie, ZHU Guixiang, et al. A multi-task learning approach for improving travel recommendation with keywords generation [J]. Knowledge-Based Systems, 2021, 233: 107521.