吉林大学学报(工学版)
Journal of Jilin University(Engineering and Technology Edition)
ISSN 1671-5497,CN 22-1341/T

《吉林大学学报(工学版)》网络首发论文

题目：基于卷积网络注意力机制的人脸表情识别
作者：郭昕刚，程超，沈紫琪
DOI：10.13229/j.cnki.Jdxbgxb20221345
收稿日期：2022-10-20
网络首发日期：2023-02-15
引用格式：郭昕刚，程超，沈紫琪．基于卷积网络注意力机制的人脸表情识别[J/OL]．吉林大学学报(工学版). https://doi.org/10.13229/j.cnki.Jdxbgxb20221345

网络首发时间：2023-02-15 11:21:12
网络首发地址：https://kns.cnki.net/kcms/detail//22.1341.T.20230214.1705.003.html

基于卷积网络注意力机制的人脸表情识别

郭昕刚，程超，沈紫琪

（长春工业大学 计算机科学与工程学院，长春 130000）

摘　要：针对表情识别时出现参数量大和识别能力弱等问题，提出一种基于卷积网络人脸表情识别方法。引入改进型残差模块，减少参数量的同时增强对表情区域的关注；利用通道-空间注意力机制对网络提取的表情区域实现不同维度和位置上的权重分配，专注于表情关键点中细微差别特征信息；利用细节模块进一步提取深度特征信息。为得到更高准确度，引入联合损失函数增加类外距离，减少类内距离以提高表情识别准确度。本文将此网络运用到数据集 FER2013，CK+中，实验结果表明：本算法平均识别率分别为 63.91%，97.98%，参数量为 11.34M。与 VGG 网络、残差网络等对比，该模型不仅识别率提高还减少冗余参数量。

关键词：面部表情识别；残差模块；通道-空间注意力机制；细化模块

中图分类号：TP391　　**文献标志码：**A

DOI：10. 13229/j. cnki. Jdxbgxb20221345

Face expression recognition based on attention mechanism of convolution network

GUO Xin-gang, CHENG Chao, SHEN Zi-qi

(*School of Computer Science and Engineering, Changchun University of Technology, Changchun 130000, China)*

Abstract: A convolutional network based facial expression recognition method was proposed to solve the problems of large reference number and weak recognition ability in facial expression recognition. The improved residual module is introduced to reduce the parameters and enhance the attention to the expression area; The Channel-space attention mechanism was used to assign the weights of different dimensions and positions to the expression regions extracted from the network, and the subtle feature information of the key points of expression was focused on; The refinement module was used to further extract the depth feature information. In order to obtain higher accuracy, the joint loss function was introduced to increase the out-of-class distance and reduce the in-class distance to improve the accuracy of expression recognition. The experimental results showed that the average recognition rate was 63.91% and 97.98% respectively, and the parameter was 11.34 M. Compared with VGG network and residual network, the model not only improves the recognition rate but also reduces the redundant parameters.

Key words: Facial expression recognition; Residual module; Channel-spatial attention module; Refinement module

0 引言

面部表情是一种非言语性的表达方式，且在大多数场景下比语言更能判断人们内心的真实感受[1]。1971 年，Ekman 等专家系统性的将面部表情分为：生气、害怕、厌恶、开心、悲伤、惊讶六类[2-3]。现代，人脸表情识别技术与多种领域交叉融合，并在谎言检测、医学、智能驾驶等领域都有广泛应用。人在说谎时，会存在一些人眼不易察觉的细微动作，警方可以通过表情识别系统检测人脸区域中的细小变化判断嫌疑人是否说谎；医生利用人脸表情识别系统检测存在心理疾病的患者微表情判断其情绪状况，以保证及时加大或者减轻药物；在车辆上安装面部表情识别系统，当司机出现疲劳驾驶症状时采取相应的措施制止[4-5]。传统人脸表情识别基本是在几何特征的基础上特征提取和识别[6]。目前更多的是使用深度学习网络实现自主学习特征。谢银成等学者[7]在 ResNet 网络中嵌入自注意力机制并在损失函数中加入权重系数，以此针对类别不均衡数据集；何超等[8]在 AlexNet 网络的基础上设计 UCNN 网

收稿日期：2022-10-20.
基金项目：吉林省教育厅基金项目（JKH20210754KJ）；长春市科技局重大专项项目（21GD05）；吉林省科技厅重点攻关项目（20210201113GX）.
作者简介：郭昕刚（1979-），男，副教授，硕士，硕士生导师。研究方向：数字图像处理。E-mail：6889068@qq.com
通信作者：程超（1984-），男，教授，博士生导师。研究方向：人工智能与数据驱动方法、人工智能与数据驱动方法。E-mail：125725673@qq.com

络，利用小卷积核组成卷积组代替大卷积核进行表情判定；崔子越等[9]对 VGG 网络进行改进，并且对损失函数设置概率阈值避免错误样本对模型分类产生影响；亢洁等 [10]在 AlexNet 网络的基础上引入 SE 模块和借助域适应方法进行表情识别；张波等 [11]在卷积神经网络中加入可分离卷积和通道注意力机制从而表情判定；Jiang Dahong 等[12]就针对瓶颈问题上提出 RexNet 网络，从而进行人脸表情识别。

上述文献中所述的网络模型各有特色，但是略有不足，具有模型过大，识别精确度不高等缺点。为解决以上问题，本文提出一种基于卷积网络注意力机制的模型。该方法改进残差模块，减少网络参数量，使模型轻量化；通过从通道和空间两个方向的注意力机制增加关键区域权重；引入细节模块利用不同深度多尺度特征提取和特征拼接等方式将支路网络与主网络融合对人脸关键点做更细致化操作，最后使用联合损失函数[13]增加类外距离，减少类内距离以提高表情识别准确度。

1 网络设计

设计网络架构时，网络层数太多会出现过拟合现象，网络层数太少会对图像特征处理的不充分。本文综合考虑这些因素设计了一种在卷积网络的基础上融合注意力机制和细化模块的模型，其中包括：浅层特征提取层、注意力机制模块、中间特征提取层、末端特征提取分类层。其网络结构如图 1 所示：

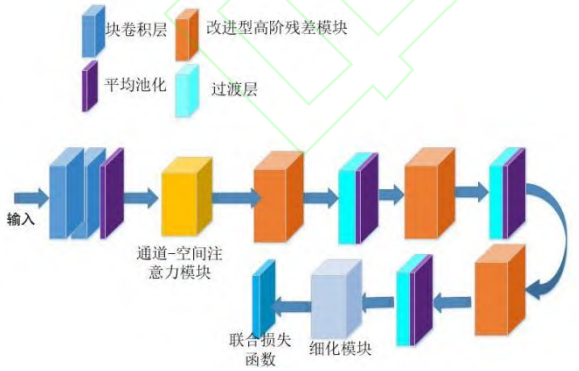

图 1 网络模型结构

Fig .1 Network model structure

本文网络模型的输入图像尺寸是 48×48，其中浅层特征提取层包含了 2 个块卷积层（卷积+归一化+激活），即 2 个卷积、2 个归一化（Batch normalization, BN）、2 个激活函数，再加上 1 个平均池化层；注意力机制模块是将通道和空间分为两个分支并联相加，利用空间和位置这两种不同的方式得到更准确的表情区域关键点；中间特征提取层包含 3 个改进型残差模块、3 个过渡层，连接方式如图 1 所示。中间特征提取层中将部分普通卷积层换成深度可分离卷积，以减少冗余参数量并更好的提取特征信息；末端特征提取分类层包含细化模块和联合损失函数；利用细化模块进一步提取深度的表情区域特征，并加入联合损失函数增加类与类之间的距离，减少类内距离，从而得到更精准的分类。

1.1 浅层特征提取层

浅层特征提取层是最靠近输入图片的网络层，该层提取出的特征丰富性直接影响到中间末尾特征提取层的发展以及分类的准确性。提取层中卷积层的目的是为得到局部区域的特征，卷积核相当于是“特征提取器”，用来得到特定区域的局部信息。例：模型开始的图像特征为$X \in K^{M\times N\times D}$，其中每一个通道 D 的输入特征为$X^d \in K^{M\times N}$ $(1 \le d \le D)$，卷积核$W^{P,1}$，$W^{P,2}$，$W^{P,3}$，…，$W^{P,D}$分别与输入特征X^1，X^2，X^3，…，X^D相卷积，得到特征$Z^{P,d}$，将其相加，若存在偏置 b，则直接在其后加入，得输出特征为Z^P。使用非线性激励得到最后结果Y^P。具体如公式 1,2,3 所示：

$$Z^{P,d} = W^{P,d} \otimes X^d \tag{1}$$

$$Z^P = W^P \otimes X + b^P = \sum_{d=1}^{D} Z^{P,d} + b^P \tag{2}$$

$$Y^P = f\left(Z^P\right) \tag{3}$$

其中f(x)表示非线性激活函数。

提取层中有 2 个卷积层，其输出的通道数都是 64，卷积核大小是 3；为保持输入图片尺寸和输出图片尺寸相同，选择 padding=same 操作；并利用归一化减轻梯度消失问题；采取修正线性单元（Rectified Linear uint，ReLU）引入非线性因素使深度神经网络输出有界，其中 ReLU 函数如公式 4 表示：

$$f(X) = \begin{cases} 0 & X \le 0 \\ X & X > 0 \end{cases} \tag{4}$$

采用平均池化层对特征进行最后处理。其中平均池化层的步长为 2，表示经过平均池化层后，输出图片的尺寸相较于原来输入图片缩小了 1/4，但维度并未发生改变。

1.2 注意力机制模块

传统卷积网络进行人脸表情识别是将整张人

脸输入其中，直接特征提取实现表情状态的预测。这会受到除表情特征区域以外的干扰，得不到较好的表情识别结果[14]。因此，本文以 CBAM[15]的方式设计一种新的注意力机制：通道-空间注意力模块（Channel-spatial attention module, CSAM），利用通道和空间两种注意力机制丰富网络结构并增强关键特征信息。与文献[15]不同的是：通道注意力机制与空间注意力机制是并联相加的，其中通道注意力机制使用全局平均池化层（Global Average Pool，GAP）压缩提取实数列，空间注意力机制在平均池化和最大池化后分别使用一次卷积，保留更多的通道信息后再级联相加。具体流程如图 2 所示。

输入是经过浅层特征提取层后的特征图 F。F 的宽高分别为 W 和 H，通道数为 D。随后将 F 分别送入图 3 所示的通道注意力支路（Channel attention branch）和空间注意力支路（Spatial attention branch）上。

通道注意力支路思路是：将输入特征 $F\in R^{H\times W\times D}$ 送入 GAP 压缩，提取长度为 D 的实数列，聚合每个通道的特征，使生成向量是在全局信息上对通道的软编码，经过两次全连接层（Fully Connected Layer，FC）建立通道间的相关性，具体操作为：利用一个 FC 层将特征维度下降到输入的 1/r 倍（r 为压缩率）,使用 ReLU 激活增加非线性度，更好的拟合通道相关性，再利用一个 FC 层恢复成原来的维度后通过 ReLU 函数增加网络稳定性。其中 FC 层参数为每个特征通道生成权重，通过 Sigmoid 将权重限制在 0~1 之间，最后用乘法将得到的权重系数加权到经过一次 3×3 卷积的输入特征上，增加特征的可辨性。通道注意力支路具体过程 M_c（F）用公式 5 可表示为：

$$M_c(F)=f^{3\times3}(F)\times\sigma\left(\delta\left(FC\left(\delta\left(FC\left(GAP(F)\right)\right)\right)\right)\right) \tag{5}$$

其中：f 表示卷积操作，3×3 表示卷积核大小，δ 表示 ReLU 激活函数，σ 表示的是 Sigmoid 激活函数。

空间注意力支路思路是：将输入特征 $F\in R^{H\times W\times D}$ 沿着两条支路分别做最大池化（Maximum Pool，MP）加卷积和平均池化（Average Pool, AP）加卷积，卷积层的卷积核都为 3，步长为 1，输出通道数为 3，使输出特征图为 H×W×3，保留更多的通道信息，采用通道级联的方式合并形成 H×W×6 的聚合特征图，经过一次 3×3 的卷积，设步长为 1，填充数为 same，减少参数量并保证特征图的大小不变，利用 Sigmoid 函数生成相应的空间注意力图，与经过 3×3 卷积的输入特征图相乘，以便更好的在空间上突出定位目标，获得空间位置关键特征表征图。这里，所有的卷积后面都加入了归一化层(Batch normalization, BN)和 ReLU 激活函数，避免梯度消失，减少正则化的使用。该过程 M_s（F）可以用公式 6 表示为：

$$M_s(F)=F\times\sigma\left(f^{3\times3}\left(C\left(f^{3\times3}\left(MP(F)\right);f^{3\times3}\left(AP(F)\right)\right)\right)\right) \tag{6}$$

其中：f 表示卷积操作，3×3，1×1 表示卷积核大小，C 表示对两条支路得到的特征进行通道特征拼接，σ 表示 Sigmoid 激活函数。

把通道注意力支路和空间注意力支路并联结合，可以获得重要通道特征及特征与特征之间的空间关系，再经过 GELU 函数提高非线性的同时增加泛化力，最后将得到的特征图 F_{CSAM} 与输入特征图相乘得最终特征图 F_{final}。该过程可由公式 7，公式 8 表示：

$$F_{CSAM}=GELU\left(M_c(F)+M_S(F)\right) \tag{7}$$

$$F_{final}=F\times F_{CSAM} \tag{8}$$

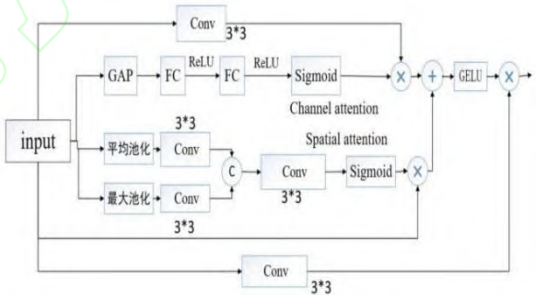

图 2　通道-空间注意力模块

Fig.2　Channel-spatial attention module

1.3　中间特征提取层

传统卷积网络得到更深层次的特征，需要利用大量的卷积层和池化层进行堆积，但在达到目的的同时也会增加网络的参数量，增加计算难度。本文利用 3 个改进型高阶残差模块增强网络的抗干扰能力，提高网络对表情区域的关注度，将原卷积层换成深度可分离卷积（Depthwise separable convolution, DS Conv）来加宽网络，减少参数量。该模型在两个改进型高阶残差模块的中间加入过渡层，目的是在不改变特征图尺寸的情况下改变维度，提高网络转换能力。

1.3.1　高阶残差模块

残差网络[16]在 IRSVRC 中取得不错的成绩后，到现在仍然使用广泛。残差网络不仅可以极大的缓解因增加模型层数而梯度消失的问题，还能使网络的抗干扰能力变得更强。本文网络模型引入改进型高阶残差模块进行网络训练。引入的模块如图 3 所示。

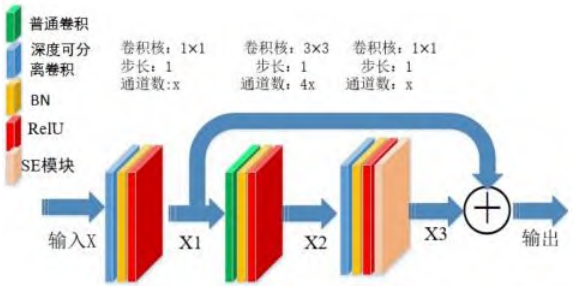

图 3 改进型高阶残差模块

Fig. 3 Improved high-order residual module

由图 3 看出，该模块是由 2 个 DS Conv、1 个常规卷积、3 个归一化层、3 个 ReLU 和 1 个压缩-激励模块（Squeeze-and - Excitation，SE）组成。利用 DS Conv 不仅可以改变通道数量，还可以减少冗余参数量。按照瓶颈层残差模块卷积核摆放模式，将残差模块的卷积核按照 1×1、3×3、1×1 的方式排列，利用 1×1 的卷积核对特征图升/降维的同时减少参数量，接着 SE 模块对输入通道加权，突出重要通道作用，图 3 的具体过程是：输入特征 X 分别经过 2 次块卷积层得到特征分别为 X_1，X_2，将 X_2 经过 1 次块卷积层和 1 次 SE 模块后得到特征 X_3，引入 Resnet 网络中残差莫模块的残差机制，将 X_1 与 X_3 建立直接的关联通道，即将经过第一个块卷积得到的一般特征与经过第三个块卷积和 SE 模块得到的高级特征级联相加，以提取更深层的特征。假设定义残差函数为F(x)，那么最后输出的结果H(x)由公式 9 表示为：

$$H(x)=F(x)+X_1=X_3+X_1 \tag{9}$$

其中：H(x) 表示的是经过整个残差模块后的总输出；F(x) 表示经过 3 次块卷积层和 1 次 SE 模块后得到的特征；X_1 表示的是在第 1 个块卷积层后的输出特征。X_3 表示的是经过 3 个块卷积层和 1 次 SE 模块后的输出特征。

本文网络在两个改进型高阶残差模块之间加入一个过渡层。过渡层是将经过 2 次卷积归一化后得到的特征再进行激活和平均池化。为避免过拟合，过渡层中再加入 1 次正则化操作。加入过渡层的目的是：借助过渡层的卷积核升维以扩大提取图像特征，并为下一个改进型高阶残差模块的输入做准备。

1.4 末端特征提取分类层

1.4.1 细化模块

实际应用中人脸表情特征之间存在共享特性，即在不同的类别之中其差异表现不明显，得到的结果最后导致人脸识别准确度低。因此，本文网络提出细化模块，进一步提取深度面部信息，具体结构如图 4 所示。

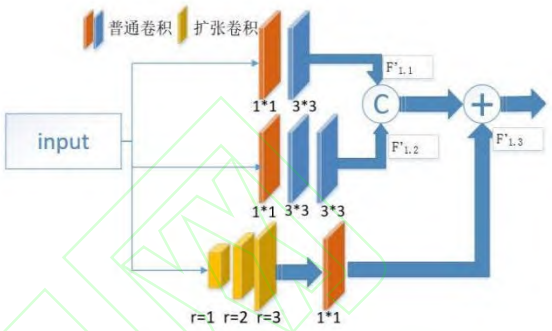

图 4 细化模块

Fig.4 Refinement module

输入特征 F′进入三条不同路径得到的特征分别为 $F'_{1.1}$、$F'_{1.2}$ 和 $F'_{1.3}$。$F'_{1.2}$ 在 $F'_{1.1}$ 的基础上多利用一次 3×3 卷积，在增加网络深度的同时更充分提取上下文特征信息。将提取到不同深度的多尺度特征 $F'_{1.1}$ 和 $F'_{1.2}$ 通道拼接融合得到 F'_1，保证得到的特征图 F'_1 与 F′维数相同。对 $F'_{1.3}$ 所在的支路使用具有多尺度感受野的金字塔卷积块，将膨胀率设置为锯齿状。不仅得到更大感受野，而且在有效构建空间内包含更多局部和全局上下文信息。接着将输出特征送入 1×1 卷积层，保证输出维度与 F'_1 相同。最后将 F'_1 与 $F'_{1.3}$ 特征图相加得最终特征 F'_{sum}。$F'_{1.1}$ 和 $F'_{1.2}$ 分别由公式 10，公式 11 所示，F'_1 由公式 12 所示，$F'_{1.3}$ 由公式 13 表示，F'_{sum} 由公式 14 表示。

$$F'_{1.1}=f^{3\times3}\left(f^{1\times1}\left(F'\right)\right) \tag{10}$$

$$F'_{1.2}=f^{3\times3}\left(f^{3\times3}\left(f^{1\times1}\left(F'\right)\right)\right) \tag{11}$$

$$F'_1=C\left(F'_{1.1};F'_{1.2}\right) \tag{12}$$

$$F'_{1.3}=f^{1\times1}(\sigma(W_{r=3}\sigma(W_{r=2}\sigma(W_{r=1}F')))) \tag{13}$$

$$F'_{sum}=F'_1+F'_{1.3} \tag{14}$$

其中：f 表示卷积操作，3×3，1×1 表示卷积核大小，C 表示将对两条支路得到的特征图通

道特征拼接，σ 表示 Sigmoid 激活函数。

1.4.2　联合损失函数

特征图经过网络处理后，一般会直接使用交叉熵函数进行计算，如公式 15 所示：

$$L_{loss}=-\sum_{i=1}^{m}\log\frac{e^{w_{y_i}^{T}x_i}}{\sum_{j=1}^{n}e^{w_j^{T}x_i}} \tag{15}$$

其中：x_i 表示第 i 个样本在进入 FC 层以前的输出，其属于第 y_i 类别，w_j 表示第 j 个 FC 层权重参数；m 表示一次训练中批量大小；n 表示类别数目。

为将同一类的表情更加紧凑，增大不同表情之间的差异，故在交叉熵损失的基础上添入中心损失，构成联合损失函数[10]，中心损失计算过程如公式 16 所示：

$$L_c=\frac{1}{2}\sum_{i=1}^{m}\| x_i-c_{y_i}\|_2^2 \tag{16}$$

其中：c_{y_i} 表示第 y_i 类的特征中心，当 y_i 类训练更新时，为避免新中心抖动太大，选择在更新值之中再加入系数，则 c_{y_i} 的更新值如公式 17 所示：

$$\triangle c_i=\beta\times\frac{\partial L_c}{\partial x_i} \tag{17}$$

其中：β 表示类别中心更新系数。最后，总的损失函数如公式 18 所示：

$$L=\lambda L_c+L_{loss} \tag{18}$$

其中：λ 表示中心损失系数，用于控制损失函数所占比重。

2 实验结果与分析

2.1 选取数据集

本文选取 FER2013 和 CK+[17]这两种经典的人脸表情数据集。FER2013 一共有 35886 张图片，其中训练集占了 28708 张图片，验证和测试集各自占了 3589 张图片，共有 7 种表情。而 CK+数据集的样本来自不同的国家，民族和性别，是较完善的公开数据集。该数据集同样将表情分为 7 类。图 5 是两种数据集中不同类别的部分样本图像。

图 5 两种数据集的部分图片

Fig.5　A partial picture of both datasets

2.2 实验环境

本文所有训练测试都是在如下环境下进行：编程语言是 python3.7，操作系统是 64 位的 Ubuntu 18.04.5，其 CPU 是 i7-10700，深度学习框架则是 TensorFlow 2.1.0。

2.3 数据增强

本文在训练测试时采用 10 折交叉验证法，即将数据集的样本总共分成 10 份,其中 9 份作为训练样本，1 份作为测试样本，反复进行 5 次训练，最后取平均数成为最后结果。

2.4 消融实验

为验证本文设计模型中每个模块的有效性，分别在 FER2013 数据集和 CK+数据集上进行消融实验，实验结果如表 1 所示。具体内容是：以不含任何改进网络模块的基础卷积网络 VGG 网络为基线，将改进型高阶残差模块（Improved high-order residual，IHOR）、通道-空间注意力模块（Channel-spatial attention module, CSAM）、细化模块（Refinement module，RM）和联合损失函数（Joint loss function，JLF）添加到该基线中，形成本文提出的结构。为验证输入尺寸是否会弱化人脸细微区域，在本文网络模型都存在的情况下，将经过预处理的样本尺寸放大到 64×64，与本文样本对比实验，观察表情识别情况。CBAM 是将空间注意力机制(Spatial attention, SA)和通道注意力机制(Channel attention, CA)相结合，JLF 是将交叉熵损失(CrossEntropy loss, CEL)和中心损失(Center loss, CL)相结合。将 CBAM 拆分成 SA 和 CA，JLF 拆分成 CEL 和 CL 分别在两种数据集上分别进行实验分析。具体数据都如表 2 所示。其中，实验过程均采用相同训练设置。

由表 1 可以看出：当仅通过 IHOR 模块进行实验时，识别率在 FER2013 和 CK+上分别优于基线网络 1.21%和 1.17%；当仅通过 CSAM 时，识别率在 FER2013 和 CK+上分别优于基线网络 1.29%和 1.41%；当仅通过 RM 时，识别率在 FER2013 和 CK+上分别优于基线网络 0.71%和 0.50%；当仅通过 JLF 时，识别率在 FER2013 和 CK+上分别优于基线网络 1.48%和 0.96% ；当同

时通过 IHOR 和 CSAM 时，识别率在 FER2013 和 CK+上分别低于本文网络的 1.51%和 0.66%；当同时通过 IHOR、CSAM、RM 时，识别率在 FER2013 和 CK+上分别低于本文网络的 0.97%和 0.44%。由此可看出各模块在网络中的必要性。对两种不同尺寸样本分别实验可以看出：本文样本尺寸没有弱化人脸细微变化。

表 1　在 FER2013 和 CK+上的消融实验结果

Table 1 Ablation experiment results on FER2013 and CK+

预处理	IHOR	CSAM	RM	JLF	FRR2013 准确率/%	CK+ 准确率/%
48×48	×	×	×	×	60.83	95.31
48×48	√	×	×	×	62.04	96.48
48×48	×	√	×	×	62.12	96.72
48×48	√	√	×	×	62.38	97.32
48×48	×	×	√	×	61.54	95.81
48×48	√	√	√	×	62.94	97.54
48×48	×	×	×	√	62.31	96.27
64×64	√	√	√	√	62.44	96.64
48×48	√	√	√	√	63.91	97.98

由表 2 可以看出：在 IHOR、RM、JLF 不变的前提下对 CSAM 进行拆分；在 IHOR、CSAM、RM 不变的前提下对 JLF 进行拆分。其中，当网络仅存在 CA 时，表情识别率在 FER2013 和 CK+数据集上分别低于存在 CBAM 识别率的 1.34%和 1.54%；当网络仅存在 SA 时，表情识别率在 FER2013 和 CK+数据集上分别低于存在 CBAM 识别率的 1.04%和 1.17%。当仅存在 CL 时，表情识别率在 FER2013 和 CK+数据集上分别低于存在 JLF 识别率的 0.80%和 0.48%；当仅存在 CEL 时，表情识别率 FER2013 和 CK+数据集上分别低于存在 CBAM 识别率的 1.16%和 0.91%。由此可见：将 CA 和 SA 并联后的表情识别效果优于独立使用结果，将 CL 和 CEL 结合后得到的表情识别效果优于独立使用结果。

表 2 在 FER2013 和 CK+上验证联合模块的实验结果

Table 2 The experimental results of the joint module were verified on FER2013 and CK+

CA	SA	CL	CEL	FER2013 准确率/%	CK+ 准确率/%
√	×	√	√	62.57	96.44
×	√	√	√	62.87	96.81
√	√	√	×	63.21	97.50
√	√	×	√	62.95	97.07
√	√	√	√	63.91	97.98

2.5 实验验证与分析

本文网络模型在训练参数更新时使用的优化器是随机梯度下降法，并且在随机梯度下降法的基础上又增加了衰减和动量，两个数据集各训练了 200 次，具体准确率变化分别如图 5，图 6 所示。需要说明：因为 FER2013 数据集样本中会放置一些有遮挡物的图片、漫画图片、不存在人脸的图片、表情不明的图片等无法进行表情识别，部分图像如图 6 所示，例：图 6 中的第一张图像（从左往右）在数据集中是放在生气类别中，但是我可能归于伤心类别中。这种类似的图像在数据集中还有很多，所以该数据集在进行训练时准确率并不是很高。

图 6　无法识别的样本图像

Fig 6　Unrecognition sample images

图 7 是本文网络在 FER2013 数据集上得到的实验结果，其中蓝色曲线代表着训练集识别精度，黄色曲线则表示测试集识别精度。由图 7 可以看出：随着训练次数的不断增加，测试集识别精度从第 4 次到第 63 次识别精度增长迅速，从第 64 次第 68 次有一次较大波动，在此之后，识别精度变得平稳起来，训练集识别精度在第 1 次到前 63 次增长迅速，之后便增长缓慢，其中训练从第 20 次到第 117 次，测试集识别精度略高于训练集识别精度，之后便逐渐相似起来。

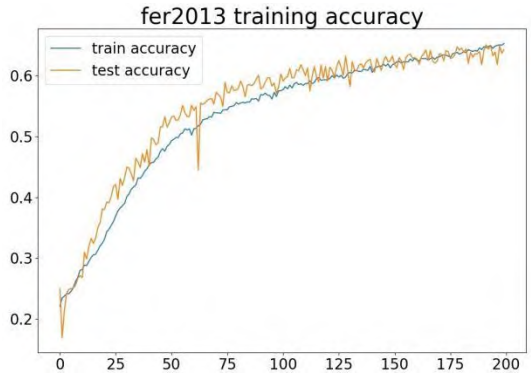

图 7　FER2013 数据集实验结果

Fig.7　The results of the FER2013 dataset

图 8 为本文网络在 CK+数据集训练后得到的最终实验结果，由图 8 可以看出：训练集识别精度在 1 到 200 次训练过程中一直在上升状态，其

中在前 34 次训练过程中迅速增长，在此之后，识别精度的增长就逐渐平稳。而测试集识别精度在前 24 次训练过程中识别率有一次下降过程，第 25 次到第 45 次训练过程中识别率增长较快，训练次数到 46 次之后识别率波动减小至平稳。

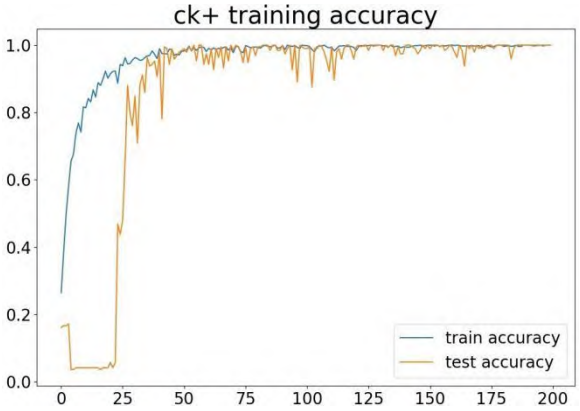

图 8　CK+数据集实验结果

Fig.8　The results of the CK+ dataset

为验证本文网络在人脸表情识别上是否更有效。分别选用了多种性能良好的网络相比较。本文采用 FER2013 数据集进行验证的结果如表 3 所示。结果表明：本文模型在 FER2013 数据集上的识别率达到 63.91%，比 VGG16 网络高 3.51%识别率，比 CNN 网络高 1.83%识别率。同样本文网络模型比文献[11]识别率高 1.26%，比文献[22]识别率高 6.78%，比文献[23]识别率高 0.35%。由表 3 可知本文网络模型在 FER2013 数据集上得到的识别率皆高于表 3 中其他的网络模型，表示本文网络模型在人脸表情识别方面的可行性。

表 3　在 FER2013 数据集上不同方法比较

Table 3 Comparison of different methods on FER2013 dataset

识别方法	准确率/%
Zhang et al[11]	62.65
Hu et al[22]	57.13
Liu et al[23]	63.56
VGG16	60.40
CNN	62.08
本文网络	63.91

表 4 为本文网络模型和部分卷积网络进行参数量的比较。由表 4 可以看出本文网络模型的参数量比 VGG16 低 10.78M，比 Resnet50 低 44.39M，比文献[18]低 11.46M，比文献[24]低 4.06M，比文献[25]低 15.02M，比文献[26]低 17.16M，由此可以说明本文模型面对基础网络和近年改进卷积网络是存在很大优势。

表 4　模型与网络的对比

Table 4　Comparison of model and network

识别方法	参数量(/M)
VGG16	22.12
Resnet50	55.73
Xception[18]	22.80
Chen et al[24]	15.40
Zhao et al[25]	26.36
Ma[26]	28.50
本文模型	11.34

表 5 为本文模型在 FER2013 数据集识别结果的混淆矩阵。由表 5 所示：生气可以被正确分类，惊讶相对于其他表情来说识别率较低，仅有 81%的概率可以被正确识别，有 12%的概率错误识别成生气，7%的概率错误识别成悲伤。剩下的表情混淆率较小，厌恶有 93%的概率可以被正确识别，其中有 7%的概率错误识别成惊讶；恐惧有 96%的概率可以被正确识别，其中有 4%的概率错误识别成高兴；高兴有 89%的概率可以被正确识别，其中有 11%的概率错误识别成悲伤；中性有 93%的概率可以被正确识别，其中有 7%的概率错误识别成高兴；悲伤有 88%的概率可以被正确识别，其中有 5%的概率错误识别成恐惧和 7%的概率错误识别成中性。

表 5　FER2013 识别结果混淆矩阵

Table 5 FER2013 identification results obfuscation matrix

真实标签	预测标签						
	生气	厌恶	恐惧	高兴	中性	悲伤	惊讶
生气	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00
厌恶	0.00	**0.93**	0.00	0.00	0.00	0.00	0.07
恐惧	0.00	0.00	**0.96**	0.04	0.00	0.00	0.00
高兴	0.00	0.00	0.00	**0.89**	0.00	0.11	0.00
中性	0.00	0.00	0.00	0.07	**0.93**	0.00	0.00
悲伤	0.00	0.00	0.05	0.00	0.07	**0.88**	0.00
惊讶	0.12	0.00	0.00	0.00	0.00	0.07	**0.81**

表 6 表示的是使用不同的网络在 CK+数据集上得到的结果。本文网络模型的识别率为 97.98%，对比之下，Resnet50 网络比其低 2.72%识别率，文献[18]比其低 3.48%识别率。同样，本文网络模型比文献[7]识别率高 2.42%，比文献[8]

识别率高 0.42%，比文献[19]识别率高 4.13%，比文献[20]识别率高 1.98%，文献[21]识别率高 0.52%。由表 6 可以看出本文网络模型在人脸表情识别方面存在可行性。

表 6 CK+数据集在不同识别方法准确率比较

Table 6 Comparison of accuracy of CK data sets in different identification methods

识别方法	准确率/%
Xie et al[7]	95.56
He et al [8]	97.56
Gao et al[19]	93.85
Liu et al[20]	96.00
Shi et al[21]	97.46
Resnet 50	95.26
Xception[18]	94.50
本文网络	97.98

表 7 是本文模型在 CK+数据集识别结果的混淆矩阵。由表 7 所示：厌恶，悲伤和惊讶可以正确分类，而恐惧相对于其他表情来说比较低。仅有 79%的概率可以正确识别，其中 10%的概率错误识别成厌恶，5%的概率错误识别成悲伤，6%的概率错误识别成惊讶。其他表情混淆率较小，生气有 85%的概率可以正确识别，其中 15%的概率错误识别成恐惧；高兴有 81%的概率可以正确识别，其中 6%的概率错误识别成恐惧，6%的概率错误识别成中性，6%的概率错误识别成悲伤；中性有 94%的概率可以正确识别，其中 6%的概率错误识别成厌恶。

表 7 CK+识别结果混淆矩阵

Table 7 CK+ identification results obfuscation matrix

真实标签	预测标签						
	生气	厌恶	恐惧	高兴	中性	悲伤	惊讶
生气	**0.85**	0.00	0.15	0.00	0.00	0.00	0.00
厌恶	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00
恐惧	0.00	0.10	**0.79**	0.00	0.00	0.05	0.06
高兴	0.00	0.00	0.06	**0.81**	0.06	0.06	0.00
中性	0.00	0.00	0.00	0.00	**0.94**	0.00	0.06
悲伤	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00
惊讶	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**

3. 结束语

针对人脸表情识别方面，本文提出对传统卷积网络进行改进的方法。在普通卷积网络中引入改进型高阶残差模块，在减少不必要参数的同时加强对表情区域的关注，有助于提高模型对表情的识别能力；同时加入通道-空间注意力模块，对网络提取出的人脸表情区域实现不同维度和位置上的权重分配，专注于模型对人脸表情关键点中细微差别特征信息；最后加入细化模块对提取出来的特征进一步细致化处理并且利用联合损失函数增加类与类之间的距离，减少类内距离进一步减小表情混淆率，从而提高表情识别的正确率。而且，在卷积层后加入 BN 层缓解梯度消失问题。本文在 FER2013 和 CK+数据集上进行实验验证，证明网络的可操作性；通过与其他常用卷积网络和部分研究者们的改进网络相比较，进一步得出本文网络的可行性。接下来会增加生活中的视频数据集，提高方法的实用性。

参考文献

[1] 方明，陈文强. 结合残差网络及目标掩码的人脸微表情识别[J]. 吉林大学学报(工学报)，2021,51(01) : 303-313.
Fang Ming, Chen Wenqiang. Facial micro-expression recognition based on residual error network and object mask [J]. Journal of Jilin University Engineering, 2021, 51(01) : 303-313.

[2] 叶继华，祝锦泰，江爱文，等. 人脸表情识别综述[J]. 数据采集和处理, 2020, 35 (01): 21 - 34.
YE Jihua, ZHU Jintai, JIANG Aiwen, et al. A review on facial expression recognition [J]. Data acquisition and processing, 2020, 35(01): 21 - 34.

[3] 李珊，邓伟洪. 深度人脸表情识别研究进展[J]. 中国图象图形学报, 2020, 25 (11) : 2306 - 2320.
LI Shan, DENG Weihong. Progress in deep facial expression recognition [J]. Chinese Journal of image and graphic , 2020 , 25 (11) : 2306 - 2320.

[4] Xianye Ben, Yi Ren, Junping Zhang, et al. Video-based Facial Micro-Expression Analysis: A Survey of Datasets, Features and Algorithms. EEE Transactions on Pattern Analysis and Machine Intelligence, DOI: 10.1109/TPAMI.2021.3067464.

[5] 贲晛烨，杨明强，张鹏，等.微表情自动识别综述[J]. 计算机辅助设计与图形学学报，2014，26(09) : 1385-1395.
BEN Xianyu, YANG Mingqiang, ZHANG Peng, et al. Survey on automatic micro expression recognition methods[J]. Journal of Computer-Aided Design and Computer Graphics, 2014,26(09) : 1385-1395.

[6] 王志良，陈锋军，薛为民. 人脸表情识别方法综述[J]. 计算机应用与软件，2003，(12) : 63-66.

WANG Zhiliang, CHEN Fengjun, XUE Weiming. A survey of facial expression recognition [J]. Computer Applications and Software, 2003,(12) : 63-66.

[7] 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(04) : 64-69.
XIE Yincheng,LI Xi,LI Tian, et al. Expression recognition based on improved RESNET and loss function[J]. 2022, 37(04) :64-69.

[8] 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，(05) : 107-111+117.
HE Chao, HOU Ming. Facial expression recognition base on improved convolutional neural network[J]. Information Technology, 2022，(05) : 107-111+117.

[9] 崔子越,皮家甜,陈勇,等.结合改进 VGGNet 和 Focal Loss 的人脸表情识别[J]. 计算机工程与应用，2021，57(19): 171-178.
CUI Ziyue, PI Jiatian, CHEN Yong, et al. Facial expression recognition combined with improved VGGNet and Focal Loss[J]. Computer Engineering and Applications,，2021，57(19): 171-178.

[10] 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.
Kang Jie, Li Jiawei, Yang Sili. Facial Expression Recognition Based on Convolution Neural Network with Domain Adaption[J]. Computer Engineering, 2019, 45(12):201-206.

[11] 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(01)：148-153.
ZHANG Bo, Lan Yanting, Li Dawei, et al. Face Expression Recognition Based On Convolution Network Channel Attention [J]. Radio Engineering, 2022,52(01): 148-153.

[12] Jiang D H, Hu Y Z, Dai L, et al. Facial Expression Recognition Based on Attention Mechanism [J]. Scientific Programming,2021(1) :1-5.

[13] 余久方，李中科，陈涛. 基于分离混合注意力机制的人脸表情识别[J]. 电讯技术，2021.
YU Jiufang, LI Zhongke, Chen Tao. Facial expression recognition based on separate hybird attention mechanism. Telecommunication Engineering, 2021.

[14] 梁华刚, 王亚茹, 张志伟. 基于 Res-Bi-LSTM 的人脸表情识别[J]. 计算机工程与应用，2020,56(13): 204-209.
LIANG Huagang, WANG Yaru, ZHANG Zhiwei. Facial expression recognition based on Res-Bi-LSTM [J]. Computer Engineering and Applications, 2020, 56(13): 204-209.

[15] Sanghyun W, Jongchan P, Joonyoung L, et al.CBAM: Convolutional Block Attention Module [C]// European Conference on Computer Vision. Munich: Springer, 2018: 3-19.

[16] He K，Zhang X，Ren S，et al. Deep residual learning for image recognition [C]// proceedings of the IEEE Computer Vision and Pattern Recognition. Las Vegas: IEEE，2016: 770 - 778.

[17] Lucey P, Cohn JF, Kanade T, et al. The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset For Action Unit And Emotion-Specified Expression [C] // Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Regression-Workshops. San Francisco, 2010: 94-101.

[18] CHOLLET F.Xception: deep learning with depth wise separable convolutions[C]//IEEE Conference on Computer Vision and Pattern Recognition. Hawaii : IEEE,2017:1800-1807.

[19] 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.
GAO Tao，SHAO Qian，ZHANG Yanan，et al. Research on facial expression recognition based on depth residual network [J]. Electronic design engineering, 2020, 28 (23) : 101 - 104.

[20] 刘尚旺，刘承伟，张爱丽. 基于深度可分离卷积神经网络的实时人脸表情和性别分类[J]. 计算机应用，2020, 40 (4) : 990 - 995.
Liu Shangwang, Liu Chengwei, Zhang Aili. Real-time facial expression and gender recognition based on depthwise separable convolutional neural network [J]. 2020, 40(4): 990-995.

[21] 石翠萍，谭聪，左江，等.基于改进 AlexNet 卷积神经网络的人脸表情识别[J]. 电讯技术，2020，60 (09) : 1005 - 1012.
SHI Cuiping，TAN Cong，ZUO Jiang，et al. Facial expression recognition based on improved AlexNet convolutional neural network [J]. Telecommunications Technology, 2020, 60(09): 1005 - 1012.

[22] Hu J，Shen L，Sun G. Squeeze-and-Excitation Networks [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018 : 7132-7141.

[23] 刘全明，辛阳阳. 端到端的低质人脸图像表情识别[J]. 小型微型计算机系统，2020，41(03) : 668-672.
LIU Quanming, XIN Yangyang. End-to-end low-quality face image expression recognition[J]. Small microcomputer system, 2020,41(03) :668-672.

[24] 程学军，邢萧飞. 利用改进型 VGG 标签学习的表情识别方法[J]. 计算机工程与设计. 2022, 43(04) :1134-1144.

CHEN Xuejun, XING Xiaofei. An expression recognition method based on improved VGG tag learning [J]. Computer Engineering and Applications. 2022, 43(04) :1134-1144.

[25] 赵家琦，周颖玥，王欣宇,等. 采用支路辅助学习的人脸表情识别[J]. 计算机工程与应用，2022.
ZHAO Jiaqi, ZHOU Yingyue, WANG Xinyu, et al. Facial expression recognition using branch-assisted learning [J]. Computer Engineering and Applications. 2022.

[26] 马金峰. 基于密集连接卷积结构的人脸表情识别研究[J].电脑与电信. 2021，(04):1-5.
MA Jinfeng. Research on facial expression recognition based on dense convolution structure [J]. Computers and telecommunications. 2021, (04): 1-5.