# 嵌入注意力机制和通道重排的人脸表情识别研究

汪正权,朱成杰
安徽理工大学 电气与信息工程学院,安徽 淮南 232001

**摘 要:目的** 针对传统卷积神经网络在人脸表情识别过程中存在网络结构复杂、特征提取弱化更能反映情感状态的眼睛和嘴唇区域而造成的针对性不强、忽略了人脸表情中空间结构信息导致识别准确度不高等问题,在主流人脸表情识别方法的基础上提出了一种嵌入注意力机制和通道重排的人脸表情识别方法。**方法** 首先,将预处理的人脸图片传入空间注意力模块,通过增强空间维度信息,使得模型能够更好地关注图像中的关键区域;然后对获取到的空间注意力特征图的通道进行切分,使信息流通过不同的路径,再进行通道融合,从而增强了特征表达能力;其次,将得到的特征图传入到通道注意力模块,增强通道维度信息;最后,通过全局平均池化来进行网络预测。**结果** 所设计的网络仅以 1.9 M 的参数量在数据集 FER2013 和 CK+上分别达到 71.80%和 99.66%的识别准确度。**结论** 该方法相比许多传统经典算法有更好的识别效果,为提高人脸表情识别准确度提供了一种更有效的途径,具有很好的实用价值和应用前景。

**关键词:**人脸表情识别;残差网络;通道重排;注意力机制
**中图分类号:**TP391 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2025.0004.013

**Face Expression Recognition with Embedded Attention Mechanism and Channel Shuffle**

WANG Zhengquan, ZHU Chengjie
School of Electrical and Information Engineering, Anhui University of Science and Technology, Anhui Huainan 232001, China

**Abstract: Objective** Traditional convolutional neural networks face challenges in face expression recognition due to complex network structures and weakened feature extraction, particularly from critical regions like eyes and lips that reflect emotional states. This leads to issues such as weak specificity and neglect of spatial structural information in facial expressions, resulting in lower recognition accuracy. In response, this study proposes a face expression recognition method embedded with attention mechanisms and channel shuffle based on mainstream approaches in the field. **Methods** Firstly, preprocessed facial images are fed into a spatial attention module to enhance spatial dimensional information, allowing the model to better focus on key areas within the images. Secondly, the channels of the obtained spatial attention feature map are split to make the information flow pass through different paths, and then channel fusion is carried out to enhance the ability of feature expression. Subsequently, the obtained feature maps are introduced into the channel attention module to enhance the channel dimension information. Finally, global average pooling is applied for network prediction. **Results** The designed network achieves recognition accuracies of 71.80% on the FER2013 dataset and 99.66% on CK+, with only 1.9 M parameters. **Conclusion** This method outperforms many traditional classic algorithms, providing a more effective approach to improving face expression recognition accuracy. This method holds significant practical value and promising application prospects.

**Keywords:** face expression recognition; residual networks; channel shuffle; attention mechanism

**收稿日期:**2023-11-23 **修回日期:**2023-12-28 **文章编号:**1672-058X(2025)04-0102-07
**基金项目:**国家自然科学基金(62003001).
**作者简介:**汪正权(1997—),男,湖北仙桃人,硕士研究生,从事图像处理与分类算法研究.
**通信作者:**朱成杰(1979—),男,安徽淮北人,副教授,从事图像处理与分类算法研究. Email:15858277803@qq.com.
**引用格式:**汪正权,朱成杰. 嵌入注意力机制和通道重排的人脸表情识别研究[J]. 重庆工商大学学报(自然科学版),2025,42(4):102-108.
WANG Zhengquan, ZHU Chengjie. Face expression recognition with embedded attention mechanism and channel shuffle[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(4): 102-108.

## 1 引　言

人类情感的表达往往会伴随着人脸表情的变化。情感丰富的人类,在表达情感方面往往具有明显甚至夸张的肌肉动作,而情感细腻的人类,在表达自己情感方面往往比较含蓄,表现在人脸上的肌肉动作往往不易被人察觉。对于人脸表情的研究已经应用到人机交互,医疗卫生,社会互动和生理信号等方面[1-4]。

人脸表情识别的研究在 1 970 年开始兴起,研究方式通常包括传统人脸表情识别和基于深度卷积神经网络的人脸表情识别。传统的研究方式通常是手动提取特征,然而,由于数据集本身存在噪声、类内差异和样本不均衡等问题,人工提取特征的方法,往往要依赖于先验知识,其识别的效果不太理想。随着 VGGNet[5]、ResNet[6]、ShuffleNet[7]等兴起,也使得研究人员开始思考,卷积神经网络对人脸表情识别是否也具有相似的优越性能,高效的提取人脸表情特征。简腾飞等[8]利用数据增强和迁移学习来弥补数据集的不足,提出 Mixer Layer 网络结构,来对人脸表情进行识别。郭昕刚等[9]引入改进的残差模型,利用通道-空间注意力机制关注表情关键点的细微差别信息,同时引入联合损失函数增加类外距离,来对人脸表情进行识别。Arriaga 等[10]使用深度可分离卷积[11]和残差网络来对人脸表情进行识别,该方法在一定程度上提高了人脸表情识别的效率与准确率,同时缓解了梯度消失或爆炸问题。张鹏等[12]通过在 Inception[13]网络的基础上加入空洞卷积、多尺度注意力机制来对人脸表情进行识别,其结合了轻量级网络结构和注意力机制的优势,能够更好地利用数据,提高模型的泛化能力。同时,多尺度注意力机制可以关注到不同尺度的特征,使模型能更好地适应复杂的人脸表情变化。Minaee 等[14]在搭建网络过程中使用仿射变换来提取特征,并对人脸表情特征提取提供了一种可解释分析,这种方法通过引入仿射变换来提高特征提取的效果,同时提供可解释分析来增强模型的可靠性和可信度。此外,该方法专注于人脸表情特征提取,能够提取更有效的特征,从而提高人脸表情识别的准确性。

上述研究在人脸识别领域提供了一种新的思路和方法,为以后的研究提供了参考方向。但总的来说还存在一些局限性:网络结构复杂,调参和优化过程可能会比较复杂和耗时;特征提取针对性不强,在特征提取过程中,弱化了更能反映情感状态的眼睛和嘴唇区域;忽略了人脸表情图像中的空间结构信息,模型对人脸表情的解码能力不足。

针对目前卷积神经网络对人脸表情识别过程中网络结构复杂,参数计算量大,特征提取不充分等问题,设计了一种轻巧的网络结构,并在此网络结构嵌入了注意力机制和通道重排(Channel Shuffle, CS)技术[15]。传统深度可分离卷积将原始特征图分成几组后再分别进行组内卷积,得到的特征图只和对应组别输入有关系,学习到的特征有限,也很容易导致信息丢失。而通道重排将分组卷积后的结果再随机分组卷积,这样保证组与组之间信息得以交换,完美解决了信息丢失问题。注意力机制可以帮助模型关注人脸表情图像中的关键区域,提高关键特征权重,减小冗余信息对模型准确度的影响。这两种技术的结合可以使模型更加准确快速地解码人脸表情图像。所设计的网络能够以 1.9 M 的参数量在数据集 FER2013 和 CK+测试集上的准确度达到 71.80%和 99.66%。

## 2 网络系统设计与构建

将预处理后的人脸图片传入到空间注意力模块(Spatial Attention Module, SA)增强空间维度信息;再将特征通道等分切割,设置不同的卷积块进行特征提取,将提取后的特征进行拼接并进行通道融合;将特征信息传入到通道注意力模块(Channel Attention Module, CA)增强通道维度信号;最后通过全局平均池化(Global Average Pooling, GAP)来进行网络预测。网络模型结构如图 1 所示,其中卷积块 C 由 Conv2D+归一化层(Batch Normalization, BN)[16]+ReLu 激活函数组成。

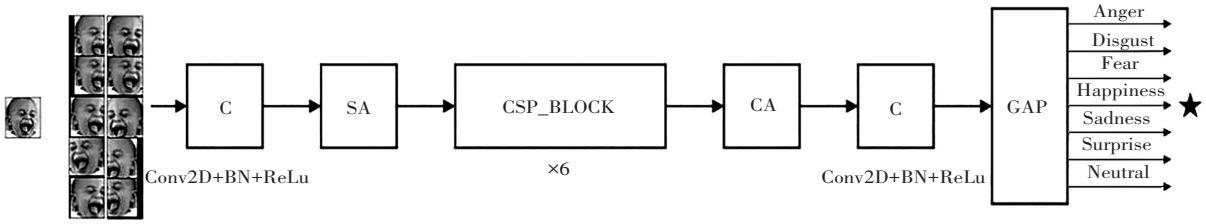


**图 1　网络结构**

**Fig. 1　Network structure**

### 2.1 通道和空间注意力

注意力机制关注视觉范围内重要的区域同时忽略视觉中不重要的内容,这与人能够从眼球观察到的视觉范围内,聚焦在自己所感兴趣部分的本能相似。Woo[17]等提出的一种卷积注意模块[16](Convolutional Block Attention Module, CBAM)是一种轻量级通用模块,能够嵌入任何卷积网络模型中,如 VGGNet、ResNet、ShuffleNet,能够以增加较小的参数量为代价,使权重参数自适应地调节,去关注

重要特征,从而有效地提升网络性能。

通道注意力模块是利用特征的通道关系生成通道注意图。其网络结构图如图 2 中 CA 所示,首先将输入特征图 $F_1 \in \mathbf{R}^{C\times H\times W}$ 在空间维度上进行压缩,利用全局平均池 Avg_Pool 和全局最大池化 Max_Pool 聚合特征图的空间信息,生成形状大小均为 $C\times1\times1$ 的平均池化特征 $F^c_{\text{avg}}$ 和最大池化特征 $F^c_{\text{max}}$;将两个池化特征传入到同一个共享网络,再将两个特征进行相加后经 Sigmoid 函数,生成通道注意力权重系数 $M_c \in \mathbf{R}^{C\times1\times1}$;最后将权重系数与输入特征图相乘,产生通道注意力特征图 $F_2 \in \mathbf{R}^{C\times H\times W}$。共享网络是由两个卷积核大小为 1×1 的 $W_0$ 和 $W_1$ 组成,这与 Woo 等在共享网络使用全连接层不同,卷积相对全连接层具有更好的跨通道信号获取能力,Sigmoid 函数常用 $\sigma$ 表示。

$$F^c_{\text{avg}} = \text{Avg\_Pool}(F_1) \tag{1}$$

$$F^c_{\text{max}} = \text{Max\_Pool}(F_1) \tag{2}$$

$$M_c = \sigma(W_1^{1\times1}(W_0^{1\times1}(F^c_{\text{avg}})) + W_1^{1\times1}(W_0^{1\times1}(F^c_{\text{max}}))) \tag{3}$$

$$F_2 = M_c \otimes F_1 \tag{4}$$

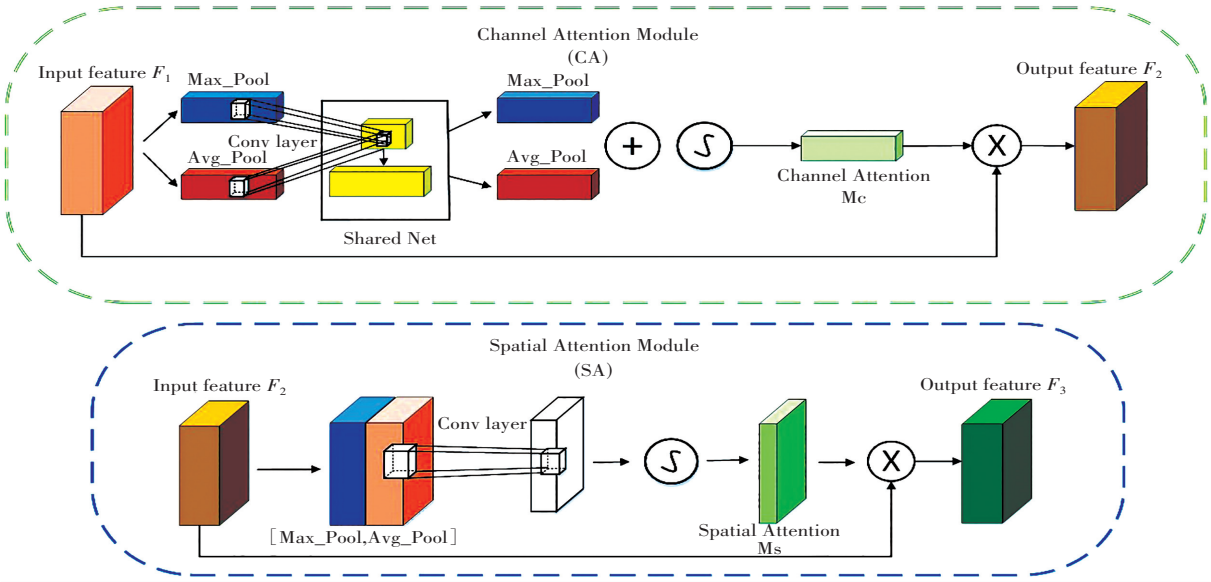


**图 2 通道注意力和空间注意力模块**

**Fig. 2 Channel attention and spatial attention module**

空间注意力模块是利用特征的空间关系生成的空间注意图。其网络结构图如图 2 中 SA 所示,首先将特征图沿着通道维度使用最大池化 Max_Pool 和平均池化 Avg_Pool 操作,生成形状大小均为 $1\times H\times W$ 的平均池化特征 $F^s_{\text{avg}}$ 和最大池化特征 $F^s_{\text{max}}$,并将其拼接成通道维度为 2 的特征图 $[F^s_{\text{max}};F^s_{\text{avg}}] \in \mathbf{R}^{2\times H\times W}$,所拼接的特征图可以有效地突出显示信息区域;之后再利用一个卷积核大小为 3×3 的卷积 $f$ 进行特征图融合;再经过 Sigmoid 操作,生成空间注意力权重系数 $M_s \in \mathbf{R}^{1\times H\times W}$;最后将权重系数与输入特征图相乘,产生空间注意力特征图 $F_3 \in \mathbf{R}^{C\times H\times W}$。

$$F^s_{\text{avg}} = \text{Avg\_Pool}(F_2) \tag{5}$$

$$F^s_{\text{max}} = \text{Max\_Pool}(F_2) \tag{6}$$

$$M_s = \sigma(f^{3\times3}([F^s_{\text{max}};F^s_{\text{avg}}])) \tag{7}$$

$$F_3 = M_s \otimes F_2 \tag{8}$$

CBAM 是结合了 SA 和 CA,来产生混合注意力特征图 $F_3 \in \mathbf{R}^{C\times H\times W}$

$$F_3 = \text{SA}(\text{CA}(F_1)) \tag{9}$$

从表 1 中不难看出,特征图的通道是不断增加,特征图的长和宽是不断减少,这也符合 VGGNet、ResNet 设计网络的一般规律,浅层网络具有低通道维度和高分辨率的特征,深层网络具有低分辨率和高通道维度的特征。利用空间注意力和通道注意力的特性,在 CSP_BLOCK 之前加入空间注意力模块,在 CSP_BLOCK 之后加入通道注意力模块。

**表 1 模型参数**

**Table 1 Model parameters**

| Input | Operator | Param # |
|---|---|---|
| [1,40,40] | $C$, $k=3$,$s=1$,$p=1$ | 88 |
| [8,40,40] | SA | 19 |
| [8,40,40] | CSP_BLOCK | 1,260 |
| [24,40,40] | CSP_BLOCK | 5,204 |
| [32,20,20] | CSP_BLOCK | 12,864 |
| [72,10,10] | CSP_BLOCK | 85,068 |
| [216,5,5] | CSP_BLOCK | 430,884 |
| [360,3,3] | CSP_BLOCK | 1,076,588 |
| [512,2,2] | CA | 262,144 |
| [512,2,2] | $C$, $k=3$,$s=1$,$p=1$ | 32,263 |
| [7,2,2] | AdaptiveAvgPool2d | 0 |
| [7,1,1] | Softmax | 0 |
| Total params: 1,906,382 | | |
| Forward/backward pass size (MB): 9.46 | | |

### 2.2　通道重排

针对堆叠多个深度可分离卷积会导致某个卷积的输出只来自于部分的输入通道,不利于不同组通道之间信息的传播,从而削弱了网络表征能力的问题,旷世科技提出的 ShuffleNet 中的通道重排 CS 使深度可分离卷积从不同组中获取输入数据,能够很好地解决这一问题,具体操作如图 3 所示。将输入通道数 $n$ 等分成 group 组,每组所含通道数为 $n//\text{group}$;将输入通道数形状改为(group, $n//\text{group}$);将得到的新的通道进行转置,再进行平坦操作即可完成通道重排。

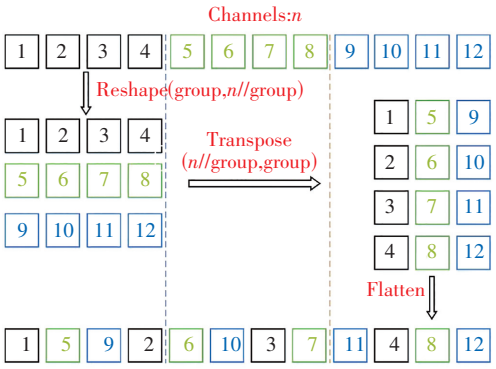


图 3　通道重排

Fig. 3　Channel shuffle

### 2.3　CSP_BLOCK

本研究所设计的基本网络单元 CSP_BLOCK 模块如图 4 所示。其中 G_Conv2D 为深度可分离卷积,($K$, $S$)中 $K$ 表示卷积核的大小,$S$ 表示卷积核的步长。通过分割通道维度从而分割梯度流,使梯度流通过不同的网络路径进行传播,从而达到增强梯度表现,将梯度的信息流从头到尾集成到特征图上,在减少了乘加运算量的同时又保证准确率;通道重排是为了解决堆叠多个深度可分离卷积而导致网络性能下降的问题,在实验过程中,在两个深度可分离卷积之间使用通道重排技术,且通道组数设置为 4 能达到较好的识别准确度;残差连接是为了解决“网络退化”问题,即随着网络深度的增加,准确度出现饱和现象,然后开始快速下降,导致网络难以训练,同时也针对训练过程中网络出现梯度消失和梯度爆炸的问题,残差连接的提出可以很好地解决这些问题,并且在一定程度上可以实现增加网络的深度来达到提高训练准确度的效果。另外在减少特征图维度方面,对不同的分支采用了不同的池化操作,最大池化可以获取特征图局部信息,可以更好地保留纹理上的特征;平均池化可以保留整体数据特征,能凸出背景特征,这点与 Woo 提出的全局池化相似,同时池化的大小不小于步长,能保证所有特征图的数据都被利用到,不会遗失特征图的信息。

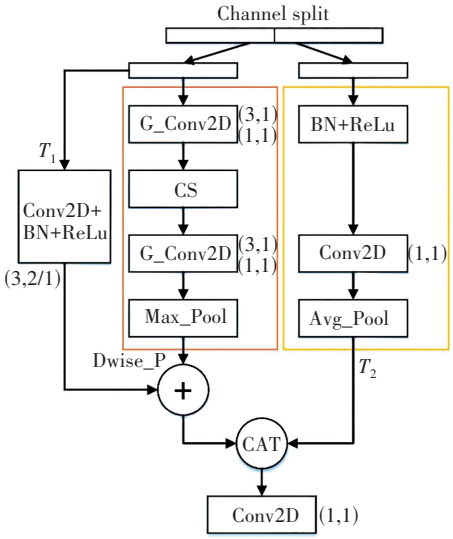


图 4　CSP_BLOCK

Fig. 4　CSP_BLOCK

首先将特征图 $X$ 沿着通道维度等分成两部分子特征图 $X_1$ 和 $X_2$,分别进行不同的卷积操作,来提取不同通道维度的特征信息;然后,将 $X_1$ 通过含有两个深度可分离卷积,一个通道重排和一个最大池化,这一卷积操作称为 Dwise_P($X_1$),其中深度可分解卷积中深度卷积的卷积核大小为 3×3,步长为 1,逐点卷积的卷积核大小为 1×1,步长为 1,并且两个深度可分离卷积都不改变特征图的分辨率,通道重排组数设置为 4,最大池化的大小为 3×3,步长为 2,在第一个 CSP_BLOCK 中不进行池化操作;将 $X_1$ 经过卷积块 $T_1$,即 Conv2D+BN+ReLu,再与 Dwise_P($X_1$)逐元素相加,其中 Conv2D 的卷积核大小为 3×3,在第一个 CSP_BLOCK 中步长为 1,其余部分步长为 2;将 $X_2$ 经过卷积块 $T_2$,即 BN+ReLu+Conv2D+Avg_Pool,其中 Conv2D 的卷积核大小为 1×1,步长为 1,平均池化的大小为 3×3,步长为 2,在第一个 CSP_BLOCK 中不进行池化操作;将两个子特征图 $X_1$ 和 $X_2$ 经过一系列卷积操作之后再进行通道拼接;最后再经过卷积核大小为 1×1 的卷积 $f$ 进行跨通道信息融合,生成最终的特征图 $F$。

$$[X_1;X_2]=X \tag{10}$$

$$F=f^{1\times1}([\text{Dwise\_P}(X_1)+T_1(X_1);T_2(X_2)]) \tag{11}$$

## 3　仿真实验与结果分析

在本次研究过程中,使用受研究人员所关注的且具有挑战的公开数据集 FER2013 和 CK+进行仿真实验。FER2013 数据集共有 35 886 张灰色人脸表情图片,其中训练集(Training)有 28 709 张,公有测试集(PublicTest)和私有测试集(PrivateTest)各有 3 589 张。数据集 FER2013 共有 7 种人脸表情,分别是 Anger、Fear、Sadness、Neutral、Happiness、Surprise、Disgust。CK+数据集包含 123 名参与者的 593 个图片序列,其中 327

个图片序列带有表情标签。本实验选取其中 7 种基本表情(Anger、Contempt、Disgust、Fear、Happiness、Sadness、Surprise),提取表情序列最后 3 帧,共 981 张图像,将数据集以 7∶3 的划分法式分为训练集和测试集。

以 FER2013 数据集为例,FER2013 的图片分辨率是 48×48 pixels,并且都是灰色图片。各样本类别展示,如图 5 所示。数据集识别难度较大,除了样本不均衡之外还有遮挡(眼睛被墨镜遮挡)、光照、背景和图片本身的分辨率低等因素,除此之外,数据集标签也存在标错的可能,人类对数据集 FER2013 的也仅有 65.5%[18] 的识别准确度。

**图 5 FER2013 数据集示例**

**Fig. 5 Example of FER2013 dataset**

### 3.1 数据集预处理

在将输入图片传入到网络训练之前对样本数据做以下增强:旋转、平移、仿射、中心和四角剪切、长宽比缩放等。使用数据增强来弥补数据集 FER2013 中样本不足和样本不均衡问题,剪裁大小为 40×40 pixels 并进行镜像操作,其增强效果如图 6 所示,其中最后一张为样本原图。

**图 6 数据增强**

**Fig. 6 Data augmentation**

### 3.2 实验环境

本研究所设计的网络采用 PyTorch 深度学习框架,显卡为 NVIDIA GeForce RTX 3060。

### 3.3 实验步骤

在具体实验过程中,在训练集和公有测试集上训练网络参数,在私有测试集上测试网络模型性能。采用了 Adam 优化器,交叉熵损失,迭代次数为 200。初始学习率为 0.01,权重衰减为 $10^{-6}$,采用余弦退火函数(Cosine Annealing LR)作为学习率的更新方式。

### 3.4 性能评价指标

为了验证网络设计的合理性,评价所设计的网络是否对人脸表情识别具有高效性,采用准确度作为评价标准。

### 3.5 消融实验

ShuffleNet 中的通道数对模型精度的影响是复杂的,增加通道数通常可以提供更多的特征信息,有助于提高模型的表达能力,从而提高精度。然而,如果通道数过多可能会导致模型过拟合或增加计算负担,从而降低精度。为了验证通道重排和注意力机制对测试集准确度的影响,在相同的环境和超参数下,对 FER2013 进行如下实验:

a) 将未嵌入注意力机制和通道重排的网络框架作为基础框架,记为 Base;

b) 在 Base 基础上加入通道重排,设置通道组数 $g$,分别取 1,2,4 和 8 记为 Base($g$);

c) 在 Base 基础上加入 SA、CA,记为 SA+Base+CA;

d) 在准确度最高的 Base($g$)基础上加入 SA、CA,记为 SA+Base($g$)+CA;

从表 2 中可以看出,对比 a)、b)组实验,通过增加通道重排技术可以增加网络的识别准确度,设置一个合适的通道组数能够进一步提升测试集的准确度,将通道组数设置为 4 时,可以提高 0.48%的识别准确度;对比 a)、c)组实验,在 CSP_BLOCK 模块前加入空间注意力模块和 CSP_BLOCK 模块后加入通道注意力模块能够提高 1.15%的识别准确度;对比 b)、d)和 c)、d)组实验,在通道重排技术($g$=4)的基础上加入注意力机制和在注意力机制上加入通道重排(技术 $g$=4)分别提升 1.5%和 0.83%。使用通道重排技术可以使深度可分离卷积不再单一依赖输入的部分通道,注意力机制能够有效地提取更多的细节特征信息,从而能够提高网络的测试准确度,实验结果也表明使用注意力机制和通道重排技术可以进一步增强网络性能,提高识别准确度。

**表 2 消融实验**

**Table 2 Ablation experiment**

| 模 型 | 准确度/% |
|---|---|
| Base | 69.82 |
| Base($g$=1) | 70.27 |
| Base($g$=2) | 69.23 |
| Base($g$=4) | 70.30 |
| Base($g$=8) | 69.57 |
| SA+Base+CA | 70.97 |
| SA+Base($g$=4)+CA | 71.80 |

### 3.6 实验结果分析

从图 7 的混淆矩阵,可以计算出本模型在数据集 FER2013 上的准确度为 71.80%。各类别的准确度依次分别为 67%、61%、54 %、89%、61%、84%、69%,所设计的网络对 Anger、Happiness、Surprise 和 Neutral 等表情表现出较高的识别度,Anger、Happiness、Surprise 等表情表现的面部运动幅度较大,Neutral 表情几乎没有面

部动作,网络模型更容易提取特征,更能准确地识别,而对 Disgust、Fear 和 Sadness 细微面部表情识别效果较差,将 Disgust 识别成 Anger,将 Fear 识别成 Sadness,这些识别度不高的表情在实际生活中也有可能被“曲解”。人类是最富有情感的生物,能够在某一时刻呈现出不同的表情特征,并且 Fear 和 Sadness 都有拉开嘴唇和紧张前额的特征,而 Anger 和 Disgust 则具有相同的眉毛特征、狭窄以及皱起的嘴角特征,这 3 种表情之间具有一定相似性,容易发生错误分类。

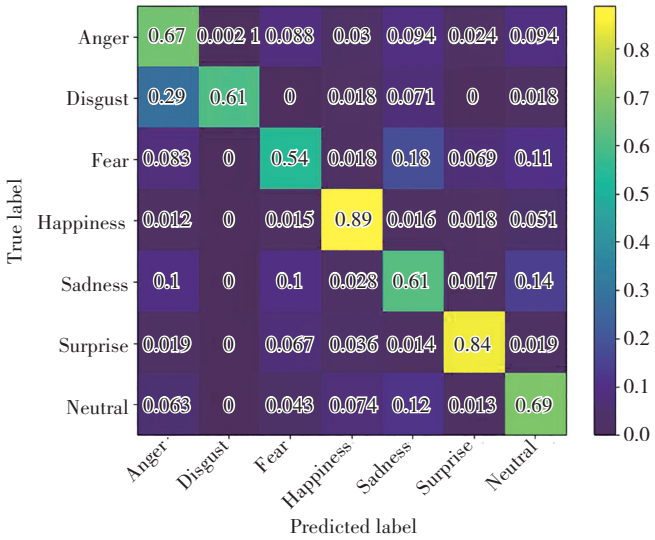


**图 7　FER2013 数据集上识别混淆矩阵**

**Fig. 7　Identify confusion matrix on FER2013 dataset**

从图 8 中不难看出,所设计的网络对 CK+数据集具有较高的识别性能,对 Anger、Contempt、Disgust、Fear、Happiness、Sadness 等表情的识别准确度达到 100%,对 Surprise 表情的识别准确度为 99%,整体的识别准确度为 99.66%,高于现有技术的准确度。也表示所设计的网络对人脸表情具有高效的解码性能。

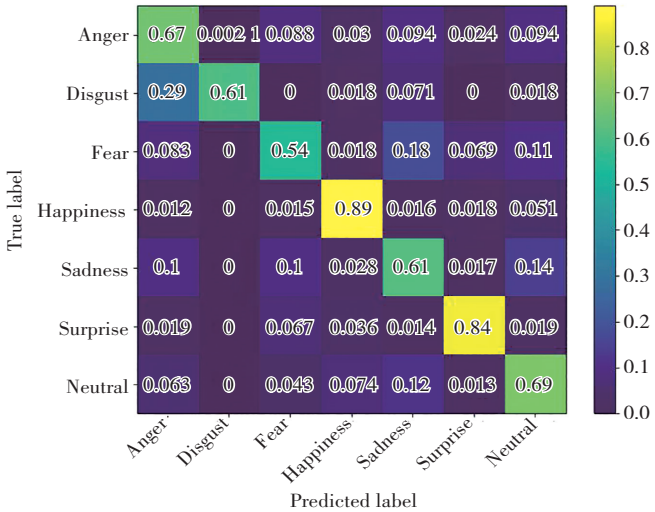


**图 8　CK+数据集上识别的混淆矩阵**

**Fig. 8　Identify confusion matrix on CK+ dataset**

为了验证本次实验所设计网络的优越性,将所设置的网络与其他在数据集 FER2013 和 CK+开展表情识别实验的深度卷积网络进行比较,结果如表 3 所示。文献[8]所使用的 Mixer Layer 网络采用无卷积结构,一定程度上降低计算复杂度,且通过迁移学习来弥补数据集不足问题。但其直接将像数值作为输入,这可能导致模型在面对不同大小或光照条件下的人脸图像时表现不稳定。文献[9]采用 ResNet-50 结合通道空间注意力机制和联合损失函数在特征提取以及模型鲁棒性方面表现优异,所付出的代价是网络结构复杂、参数计算量大、泛化能力差。从文献[11]可以看出深度可分离卷积可以显著减少模型的参数量,但在进行分组卷积时,每组卷积都独立运算,组与组之间特征信息相互独立,这会导致模型在提取人脸特征时失去一些空间信息,即使文献[10]在文献[11]的基础上结合残差网络提高一定的准确度,但也没考虑分组卷积带来的特征信息丢失问题。文献[12,14]中的 Inception 网络结构使用多个 1 * 1 的卷积核,增加感受野,保留更丰富的特征信息,但不可避免的增加参数量,过多的参数量可能导致过拟合问题。本研究以深度可分离卷积为基础来减少模型参数量,通过通道重排技术解决不同卷积组相互独立所带来的信息丢失问题,结合注意力机制,最终以更小的参数量实现更高的识别准确度,这也说明该方法的优越性。

**表 3　不同方法对 FER2013、CK+数据集上的准确度**

**Table 3　Accuracy of different methods on FER2013 and CK+ datasets**

| 算法 | FER2013/% | CK+/% | 参数量 |
|---|---|---|---|
| 文献[8] | 63.03 | 98.71 | 11,372,271 |
| 文献[9] | 63.91 | 97.98 | 2,464,032 |
| 文献[10] | 66.40 | 93.2 | 3,962,328 |
| 文献[11] | 66.00 | — | 1,825,465 |
| 文献[12] | 68.80 | 96.04 | 5,935,271 |
| 文献[14] | 70.02 | — | 7,151,502 |
| 本文 | 71.80 | 99.66 | 1,906,382 |

注:‘—’表示相应的方法未对数据集进行测试。

## 4　结　论

本文针对人脸表情识别提出了一种嵌入注意力机制和通道重排的卷积神经网络。将预处理的人脸图片传入空间注意力机制,增强空间维度信息;对获取到的空间注意力特征图的通道进行切分,使信息流通过不同的路径,再进行通道融合;最后将得到的特征图传入到通道注意力机制,增强通道维度信息。所设计的网

络仅以1.9 M的参数量就能在数据集FER2013和CK+上分别达到71.80%和99.66%的识别准确度，结果表明该方法相比许多传统经典算法有更好的识别效果。

所设计的网络是针对整个人脸来提取特征，再进行分类预测，人脸表情往往是集中表现在一些关键点，如眉毛、眼睛、鼻子和嘴巴等方面，而不是整张人脸。下一阶段将会先提取人脸的关键点，然后再进行网络特征提取，来实现对人脸表情的识别。后续也会在本文的基础上对网络结构进行进一步的改进，实现用更少的参数量、更合理的网络结构来达到更高的准确度。

**参考文献**(References):


[1] DHALL A, GOECKE R, JOSHI J, et al. Emotion recognition in the wild challenge 2013[C]//Proceedings of the 15th ACM on International Conference on Multimodal Interaction. New York: ACM, 2013: 509-516.

[2] ABDAT F, MAAOUI C, PRUSKI A. Human-computer interaction using emotion recognition from facial expression[C]//Proceedings of the UK Sim 5th European Symposium on Computer Modeling and Simulation. 2011: 196-201.

[3] FASEL B, LUETTIN J. Automatic facial expression analysis: A survey[J]. Pattern Recognition, 2003, 36(1): 259-275.

[4] SARIYANIDI E, GUNES H, CAVALLARO A. Automatic analysis of facial affect: A survey of registration, representation, and recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(6): 1113-1133.

[5] HE J, LI S, SHEN J, et al. Facial expression recognition based on VGGNet convolutional neural network[C]//Proceedings of the Chinese Automation Congress. 2018: 4146-4151.

[6] 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.

[7] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 6848-6856.

[8] 简腾飞, 王佳, 曹少中, 等. 基于Mixer Layer的人脸表情识别[J]. 计算机系统应用, 2022, 31(7): 128-134.
JIAN Teng-fei, WANG Jia, CAO Shao-zhong, et al. Facial expression recognition based on mixer layer[J]. Computer Systems and Applications, 2022, 31(7): 128-134.

[9] 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]. 吉林大学学报(工学版), 2024, 54(8): 2319-2328.
GUO Xin-gang, CHENG Chao, SHEN Zi-qi. Face expression recognition based on attention mechanism of convolution network[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(8): 2319-2328.

[10] HUO H, YU Y, LIU Z. Facial expression recognition based on improved depthwise separable convolutional network[J]. Multimedia Tools and Applications, 2023, 82(12): 18635-18652.

[11] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 4510-4520.

[12] 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.
ZHANG Peng, KONG Wei-wei, TENG Jin-bao. Facial expression recognition based on multi-scale feature attention mechanism[J]. Computer Engineering and Applications, 2022, 58(1): 182-189.

[13] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.

[14] MINAEE S, MINAEI M, ABDOLRASHIDI A. Deep-emotion: Facial expression recognition using attentional convolutional network[J]. Sensors, 2021, 21(9): 3046.

[15] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks[EB/OL]. 2015: arXiv: 1506.02025. http://arxiv.org/abs/1506.02025.pdf.

[16] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37. 2015: 448-456.

[17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the Computer Vision-ECCV 2018: 15th European Conference. Munich, Germany, 2018: 3-19.

[18] GOODFELLOW I J, ERHAN D, CARRIER P L, et al. Challenges in representation learning: A report on three machine learning contests[C]//Neural Information Processing. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013: 117-124.


**责任编辑:陈　芳**