

# 基于注意力拆分卷积残差网络的表情识别

陈加敏<sup>1</sup>, 徐杨<sup>1,2\*</sup>

<sup>1</sup>贵州大学大数据与信息工程学院, 贵州 贵阳 550025;

<sup>2</sup>贵阳铝镁设计研究院有限公司, 贵州 贵阳 550009

**摘要** 人脸表情识别是神经网络应用于模式识别上一项极具挑战性的任务,而表情识别过程中特征提取尤为重要。因此,提出了一种注意力拆分卷积残差网络来增强特征表现。该网络以 ResNet18 为骨干网络,用 Coordinate Attention Split Convolution Block(CASCBlock)替换 ResNet18 中的 basic block。CASCBlock 首先使用两个拆分卷积将特征在通道维度先拆分后融合降低冗余特征表现;然后在第 2 个拆分卷积后融入坐标注意力机制;最后构建一个全连接分类器进行表情识别。将所提方法在 FER2013 和 RAF-DB 数据集上进行了实验,实验结果表明,所提方法在 FER2013 和 RAF-DB 数据集上识别准确率相较于 ResNet18 分别提高了 2.897 个百分点和 2.575 个百分点,且模型的参数量下降了 60% 左右。

**关键词** 机器视觉; 拆分卷积; 残差网络; 特征融合; 表情识别; 注意力机制

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP202259.1815009

## Expression Recognition Based on Attention-Split Convolutional Residual Network

Chen Jiamin<sup>1</sup>, Xu Yang<sup>1,2\*</sup>

<sup>1</sup>College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, Guizhou, China;

<sup>2</sup>Guiyang Aluminum-Magnesium Design and Research Institute Co. Ltd., Guiyang 550009, Guizhou, China

**Abstract** Facial expression recognition is a challenging task for neural network applied with pattern recognition. Moreover, feature extraction is particularly important in the process of facial expression recognition. In this paper, a attention-split convolutional residual network was proposed to enhance the feature expression. This network used ResNet18 as the backbone network and replaced the basic block in ResNet18 with the coordinate attention-split convolutional block (CASCBlock), which is also proposed here. In the CASCBlock, two split convolutions were initially used to split and then fuse the features in the channel dimension to reduce redundant feature representations. Then, the coordinate attention mechanism was incorporated after the second split convolution. Finally, a fully connected classifier was developed for expression recognition. The proposed method was tested on the FER2013 and RAF-DB datasets, and the experimental results showed that the recognition accuracy of the proposed method on FER2013 and RAF-DB datasets is 2.897 percentage points and 2.575 percentage points higher than that of ResNet18, and the number of model parameters decreased by ~60% compared with ResNet18.

**Key words** machine vision; split convolution; residual network; feature fusion; facial expression recognition; attention mechanism

## 1 引言

随着深度学习的广泛应用,人脸表情识别进展显著。传统的表情识别主要应用机器学习方法人为设计特征,然后用分类算法判定表情,特征提取与分类过程相互独立,效率较低,且特征受人为因素的影响较大,

易丢失表情特征,从而降低表情识别准确率。

深度学习的方法将特征提取与分类结合在一起,自动学习特征,识别准确率也得到显著改善。因此,目前大多数研究者都基于深度学习的方法进行表情识别研究。尹鹏博等<sup>[1]</sup>采用分解卷积对模型进行参数降维,同时嵌入卷积注意力机制进行人脸表情识别;姚丽

收稿日期: 2021-06-11; 修回日期: 2021-07-19; 录用日期: 2021-08-10

基金项目: 贵州省科技计划项目(黔科合支撑[2021]一般 176)

通信作者: \*xuy@gzu.edu.cn xx

莎等<sup>[2]</sup>提出了一种采用卷积神经网络学习局部特征的人脸表情识别方法;杨旭等<sup>[3]</sup>在 AlexNet<sup>[4]</sup>中引入多尺度卷积的同时将低层次特征与高层次特征进行跨层连接,实现特征融合;吴慧华等<sup>[5]</sup>提出了一种基于余弦距离损失函数的表情识别方法,该方法减小了类内特征差异。目前这些方法都取得了较好的效果,但也面临着一些问题:1)网络复杂度随着层数的增加而增加;2)网络参数降低后导致准确率下降。

自 AlexNet<sup>[4]</sup>面世以来,深度卷积神经网络主导了图像分类,研究趋势从手工特征转移到网络体系结构。NIN<sup>[6]</sup>首先使用全局平均池化层来代替全连接层,并用  $1 \times 1$  卷积层来学习特征图通道的非线性组合,这是第 1 类特征图注意力机制;ResNet<sup>[7]</sup>中,跨层连接的引入减轻了深度神经网络中消失梯度的难度,并允许网络学习更深层的特征表示,ResNet 成为较为成功的卷积神经网络(CNN)架构之一,被广泛应用于计算机视觉领域;同时,一些新的卷积滤波器,如分组卷积(GWC)<sup>[8]</sup>、深度卷积(DWC)<sup>[9]</sup>和点卷积(PWC)<sup>[4]</sup>被广泛用于模型设计;ResNeXt<sup>[10]</sup>就与 GWC 和 PWC 在所有聚合转换中共享相同的拓扑结构,这减少了通道间连接的冗余,并引入了一个称为基数的新维度;为了进一步降低连接密度,Exception<sup>[11]</sup>和 MobileNet<sup>[12]</sup>先后使用 DWC 进行空间信息提取和 PWC 进行信道信息融合;ShuffleNet<sup>[13]</sup>在  $1 \times 1$  卷积上采用 GWC,然后采用信道洗牌操作;OctConv<sup>[14]</sup>与信道冗余正交,探索了特征图空间维度上的冗余;HetConv<sup>[15]</sup>设计了异质卷积滤波器,单个滤波器中同时存在  $3 \times 3$  核和  $1 \times 1$  核,然后,在特定层的滤波器中,这两个异质核以移位的方式排列;SKNet<sup>[16]</sup>引入了一种基于多尺度输入信息的动态核选择机制,实现了自适应感受野;OctConv、HetConv 和 SKNet 将  $3 \times 3$  内核应用于所有输入通道,然而许多特征映射出现了很大的模式相似性,这表明普通卷积基本上是重复的;GhostNet<sup>[17]</sup>首先利用普通卷积产生一些内在特征,然后利用线性运算进行特征增强;与 GhostNet 不同的是,SPConv<sup>[18]</sup>保证模型能够抓住每一个原始特征,而不是由内在特征生成的特征,

从而获得更好的性能。SPConv 将输入在通道维度按比例分成两个主要部分,一部分是代表性的应用(采用  $3 \times 3$  组卷积提供内在信息),另一部分是冗余的应用(采用  $1 \times 1$  点卷积补充微小的隐藏细节)。相比于普通卷积,SPConv 能够更好提取特征且能降低模型复杂度。

在计算机视觉中,常采用注意力机制来提高模型的效率,通道注意力可以提高模型的性能,但通道注意力会忽略对生成空间注意力图非常重要的位置信息。坐标注意力(CA)模型<sup>[19]</sup>将位置信息嵌入通道注意力中来提高模型的性能。

受以上启发,本文以 ResNet18 为骨干网络,用 SPConv 替换 ResNet18 中 basic block 的  $3 \times 3$  卷积;同时,为增强特征表现,在第 2 个 SPConv 后嵌入坐标注意力。并将所提方法应用于 FER2013、RAF-DB 数据集进行实验,实验结果表明,所提方法在 FER2013、RAF-DB 数据集上的识别准确率相比于 ResNet18 分别提高了 2.897 个百分点和 2.575 个百分点。

## 2 基本原理

### 2.1 组卷积和点卷积

组卷积就是先对输入的特征图进行分组,然后对每个组分别卷积。若输入的特征  $F \in \mathbf{R}^{C \times H \times W}$ ,输出的特征图数量为  $N$ ,分组为  $G$ ,则每组的输入 feature map 数量为  $C/G$ ,每组的输出 feature map 数量为  $N/G$ ,每个卷积核的尺寸为  $(C/G) \times K \times K$ ,卷积核的总数仍为  $N$  个,每组的卷积核数量为  $N/G$ ,卷积核只与其同组的输入 map 进行卷积,卷积核的总参数量为  $N \times (C/G) \times K \times K$ 。即,总参数量减少为原来的  $1/G$ 。组卷积如 [图 1(a)] 所示,图中以  $C=8, N=4, G=2$  为例。

点卷积的运算与普通卷积较为相似,卷积核的尺寸为  $1 \times M \times M, M$  为上一层输出特征的通道数,所以点卷积就是将上一层的输出特征在通道维度进行加权组合生成新的特征图。有几个卷积核就有几个输出特征图,如 [图 1(b)] 所示。可以通过点卷积更改特征

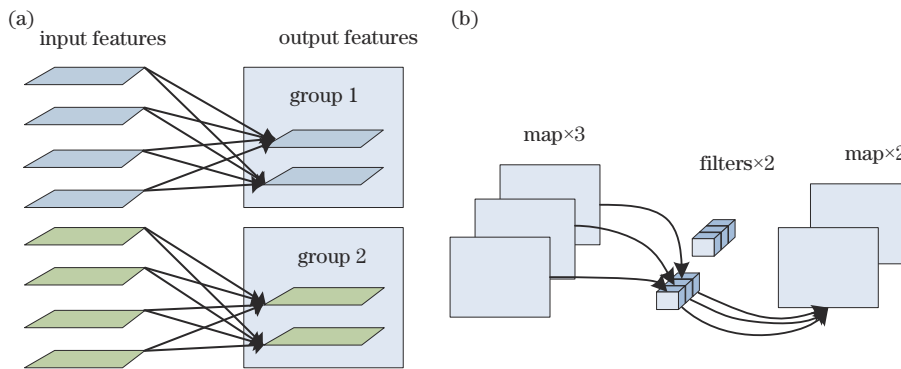


图 1 卷积。(a)组卷积;(b)点卷积

Fig. 1 Convolution. (a) Group convolution; (b) point convolution

图的数量及减少上一步生成的特征图在空间维度的加权组合。

## 2.2 SPConv

常规卷积、GhostConv、OctConv、HetConv 均在所有输入通道上执行  $k \times k$  卷积, 导致特征冗余问题<sup>[19]</sup>。

而 SPConv 将输入通道  $c_{in}$  按比例  $a$  分成两个部分, 分别是代表性的应用  $ac_{in}$  和冗余的应用  $(1-a)c_{in}$ 。将所有输入通道分成两个主要部分后, 代表部分之间还存在冗余, 在代表性信道上采用  $3 \times 3$  组卷积来进一步减少冗余, 如图 2 的中间部分所示。

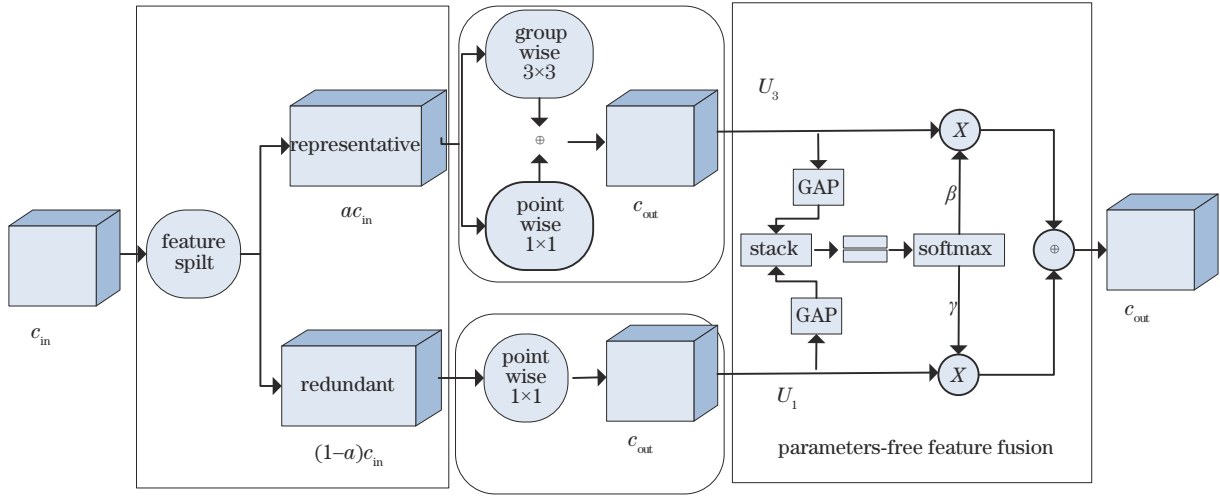


图 2 拆分卷积

Fig. 2 Split convolution

将组卷积视为具有稀疏块对角卷积核的普通卷积, 其中每个块对应于信道的分区, 并且分区之间没有连接<sup>[20]</sup>, 这意味着在组卷积之后, SPConv 进一步减少了代表部分之间的冗余, 同时也切断了通道之间的连接。为弥补这种信息丢失, 在所有代表性的信道中添加  $1 \times 1$  点卷积, 与常用的分组卷积和点卷积不同, SPConv 在同一代表信道上进行 GWC 和 PWC, 在冗余部分使用  $1 \times 1$  的点卷积。经过代表部分和冗余部分后可得到两个特征, 由于这两个特征源于不同的输入通道, 本实验组使用全局平均池来生成通道统计信息  $s_1, s_3 \in \mathbf{R}^c$  用于全局信息嵌入。  $U$  为图 2 中代表部分和冗余部分的输出特征,  $s$  的第  $c$  个元素为

$$s_{kc} = F_{\text{gap}}(U_{kc}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_{kc}(i, j), k \in [1, 3]. \quad (1)$$

如图 2 右侧所示, 将得到的  $s_1, s_3$  叠加在一起, 然后通过通道间的软注意操作生成特征重要向量  $\beta, \gamma$ , 其第  $c$  个元素分别为。

$$\begin{cases} \beta_c = \frac{\exp(s_{3c})}{\exp(s_{3c}) + \exp(s_{1c})} \\ \gamma_c = 1 - \beta_c \end{cases} \quad (2)$$

最终输出特征  $Y$  由代表部分特征  $U_3$  和冗余部分  $U_1$  分别与  $\beta$  和  $\gamma$  对应相乘相加而得, 即

$$Y = \beta U_3 + \gamma U_1. \quad (3)$$

## 2.3 坐标注意力

坐标注意力将位置信息嵌入通道中, 使网络能够在

大范围内进行注意。为了改善二维全局共享造成的信息丢失问题, 在通道维度将特征分解为两个并行的一维特征, 并使用两个一维全局池化操作分别将垂直方向和水平方向的输入特征聚合为两个独立的方向感知特征图。然后, 这两个嵌入特定方向信息的特征图分别被编码为两个注意力图, 每个注意力图都捕获了输入特征图沿着一个空间方向的长程依赖。因此, 位置信息就被保存在生成的注意力图里了, 两个注意力图接着被乘到输入特征图上来增强特征图的表示能力, 以强调感兴趣的表示。

坐标注意力模块如图 3 所示, 对于特征  $X$ , 首先使用尺寸为  $(H, 1)$  和  $(1, W)$  的池化核沿着水平坐标方向和垂直坐标方向对每个通道进行编码。因此, 高度为  $h$  的第  $c$  个通道的输出为

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i). \quad (4)$$

宽度为  $w$  的第  $c$  个通道的输出为

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w). \quad (5)$$

这两个变换沿着两个空间方向进行特征聚合, 返回一对方向感知注意力图。这两种变换能使注意力模块沿着一个空间方向捕捉到长期的相关性, 并在另一个空间方向上保留精确位置信息, 这有助于网络更准确地定位感兴趣的目标。将式 (4) 和式 (5) 生成的两个特征图连接起来, 并送入一个  $1 \times 1$  的卷积变换函数  $F_1$  中。

$$f = \delta \{ F_1 [z^h, z^w] \}, \quad (6)$$

式中:  $[\cdot, \cdot]$  表示沿着空间维度的拼接操作;  $\delta$  为非线性激活函数;  $f \in \mathbf{R}^{\frac{C}{r} \times (H+W)}$  为水平方向和垂直方向

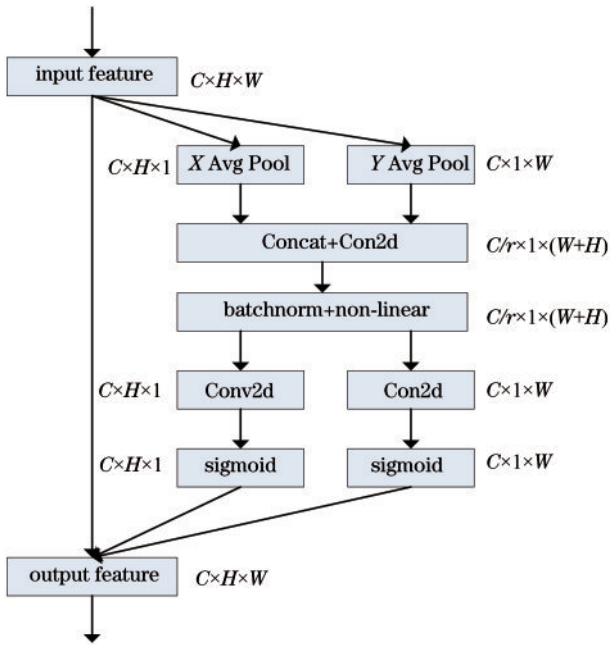


图 3 坐标注意力

Fig. 3 Coordinate attention

对空间信息进行编码的中间特征映射,  $r$  为控制块的减少比。然后沿着空间维把  $f$  分解成两个单独的张量  $f^h \in \mathbf{R}^{C/r \times H}$  和  $f^w \in \mathbf{R}^{C/r \times W}$ , 接着使用两个  $1 \times 1$  卷积变换  $F_h$  和  $F_w$  将  $f^h$  和  $f^w$  分别变换为相同通道数的张量。

$$g^h = \sigma[F_h(f^h)], \quad (7)$$

$$g^w = \sigma[F_w(f^w)], \quad (8)$$

式中:  $\sigma$  为 Sigmoid 函数, 为降低模型的复杂性, 通常采用适当的减少比来减少  $f$  的信道数。以输出的  $g^h$  和  $g^w$  作为注意力权值, 最后坐标注意力模块的输出为

$$y_c(i, j) = x_c(i, j) \times g^h(i) \times g^w(j), \quad (9)$$

式中:  $y_c(i, j)$  为输出;  $x_c(i, j)$  为输入;  $g^h(i)$ 、 $g^w(j)$  为注意力权值。

### 3 所提方法

#### 3.1 注意力拆分卷积残差网络

受文献[7, 18-19]的启发, 本实验组在 ResNet18 的基础上对 basic block 进行改进, 融合拆分卷积和坐标注意力机制, 提出了 Coordinate Attention Split Convolution Bloc(CASCBlock), 该模块由两个拆分卷积和一个坐标注意力组成, 如图 4 所示。

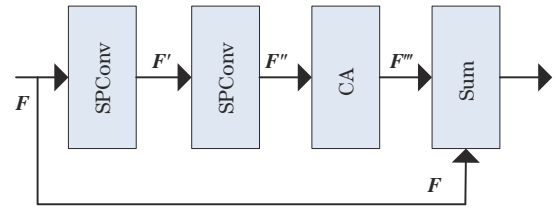


图 4 CASCBlock

Fig. 4 CASCBlock

拆分卷积将输入的特征  $F$  在通道维度按 0.5 的比例拆分为代表部分和冗余部分, 在代表部分使用分组为 2 的  $3 \times 3$  组卷积和  $1 \times 1$  的点卷积, 并将所得特征相加。在冗余部分使用  $1 \times 1$  的点卷积, 然后将两部分特征通过全局平均池化生成统计信息, 最后使用软注意力分配权重, 分配方式如式(2)所示, 通过式(3)得到经过一个拆分卷积后的特征  $F'$ ; 为增强特征表现, 采用两个拆分卷积, 经过两个卷积后的特征为  $F''$ ; 同时, 为增加具有判别性的局部特征, 在第 2 个拆分卷积后添加坐标注意力, 从而得到特征  $F'''$ 。为解决训练过程中梯度消失的问题, 借用残差结构中跨层连接的方式, 将特征  $F'''$  和特征  $F$  相加后输入下一层网络结构。

所提模型的整体框架如图 5 所示。首先使用一个  $7 \times 7$  卷积提取人脸图像特征信息, 然后将模型分为 4 个 layer, 每个 layer 由两个 CASCBlock 组成。

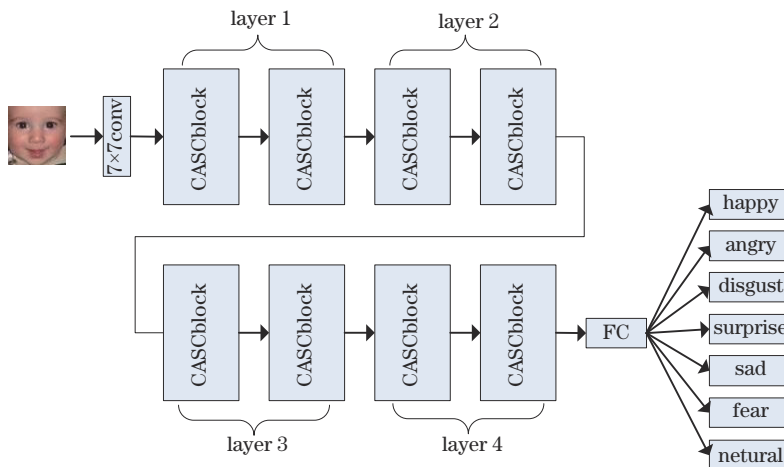


图 5 注意力拆分卷积残差网络

Fig. 5 Channel attention resolution convolutional residuals network



CASCBlock能够提升每个layer的特征表现力,最终达到提升表情识别准确率的效果。每个layer的详细参数如表1所示。

表1 注意力拆分卷积残差网络的参数

Table 1 Parameters of channel attention resolution convolutional residuals network

Layer	Output	Proposed network
	44×44	7×7, 64, s=1 3×3 maxpool, s=1
1	44×44	$\begin{bmatrix} \text{SPConv, 64} \\ \text{SPConv, 64} \\ \text{CA} \end{bmatrix} \times 2$
2	22×22	$\begin{bmatrix} \text{SPConv, 128} \\ \text{SPConv, 128} \\ \text{CA} \end{bmatrix} \times 2$
3	11×11	$\begin{bmatrix} \text{SPConv, 256} \\ \text{SPConv, 256} \\ \text{CA} \end{bmatrix} \times 2$
4	6×6	$\begin{bmatrix} \text{SPConv, 512} \\ \text{SPConv, 512} \\ \text{CA} \end{bmatrix} \times 2$
	1×1	Global avgpool 7-d fc

## 4 实验与分析

对注意力拆分卷积残差网络进行了对比实验。

表2 模型的性能

Table 2 Model performance

Model name	Number of parameters	Model size /Mbit	Accuracy in FER2013 /%	Accuracy in RAF-DB /%
ResNet18	11180103	42.7	69.769	81.845
SPResNet18	4188487	16.2	72.444	83.638
SPResNet18_CA	4263223	16.5	72.666	84.420

从表2可以看出:FER2013和RAF-DB数据集在初始的ResNet18模型上表情识别准确率分别为69.769%和81.845%;对ResNet18进行改进后,表情识别准确率在FER2013和RAF-DB数据集上分别增加了2.675个百分点和1.793个百分点,表明拆分卷积将特征在通道维度拆分为代表部分和冗余部分模型的有效性;融入坐标注意力机制后,识别准确率相比于ResNet18分别提高了2.897个百分点和2.575个百分点,表明注意力机制可以使网络关注具有判别性的表情特征,提高表情识别的准确率。所提模型在不同数据集上的测试准确率均得到了提升,表明了所提方法的泛化性;同时,所提模型的参数量和大小相比于ResNet18均降低60%左右。即,所提表情识别模型相比于ResNet18在提高准确率的同时降低了参数量和模型大小。

在Ubuntu20.04.1 LTS系统下使用python3.8+PyTorch1.7,显卡为NVIDIA GeForce GTX 3090。在FER2013和RAF-DB公开人脸表情数据集上对所提模型进行实验,并与目前的方法进行了对比分析。

### 4.1 数据库与实验设置

为测试CASCBlock的效果,选用FER2013数据集、RAF-DB数据集进行实验。FER2013数据集共有28709张图片,7种表情分类标签,测试集有3859张图片。RAF-DB数据集共有29672张图片,每张图片经过40人标注,实验使用其中经过Face++定位的人脸单标签类,含有12271张图片的训练集和3068张图片的测试集。参数设定如下:输入网络模型的图片尺寸为48×48;在图像处理阶段采用随机裁剪和水平翻转对数据进行扩充,将输入图片随机裁剪为44×44;训练过程中动量为0.9,初始学习率为0.1,迭代50次时将学习率设定为每增加5次下降0.9,总迭代次数为300;在FER2013中,批处理大小为128,RAF-DB中批处理大小为16。

### 4.2 实验结果

实验结果如表2所示,其中ResNet18为骨干网络的实验准确率,SPResNet18为将ResNet18中的3×3卷积替换为拆分卷积的模型,SPResNet18\_CA为将ResNet18中的basic block替换为所提CASCBlock的模型。

### 4.3 对比分析

为评估模型的效果,将所提模型(SPResNet18\_CA)与ResNet18在FER2013和RAF-DB数据集进行了对比分析,结果如图6、图7所示。从图6、图7可以看出:所提模型在FER2013和RAF-DB各类表情上的预测结果均有提高;在表情识别过程中Happy、Surprise、Neutral的结果偏高,因为这几类表情在我们的日常生活中比较容易辨别,特征明显。

为更好显示所提模型训练过程中的准确率变换趋势和收敛趋势,所提模型与ResNet18的准确率对比曲线如图8所示。从图8可以看出,所提模型在前期的训练过程波动较大,后期准确率曲线逐渐变得稳定。

为进一步说明所提方法的有效性,将其与近年来多个模型在FER2013和RAF-DB上进行了对比,如表3、表4所示。从表3可以看出,所提方法的准

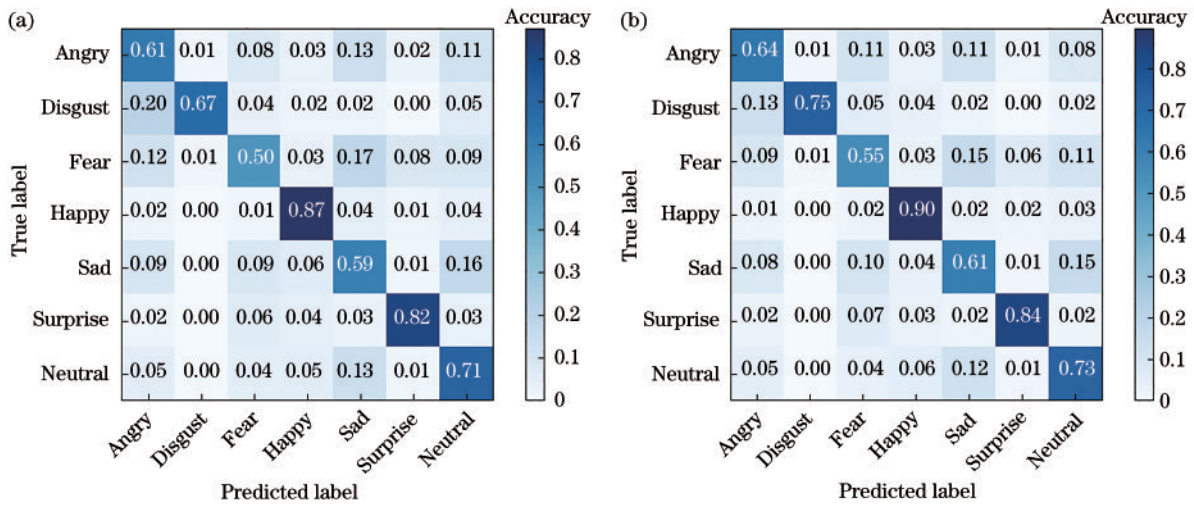


图 6 FER2013 实验结果。(a)ResNet18;(b)SPResNet18\_CA  
Fig. 6 Results of FER2013 experiment. (a) ResNet18; (b) SPResNet18\_CA

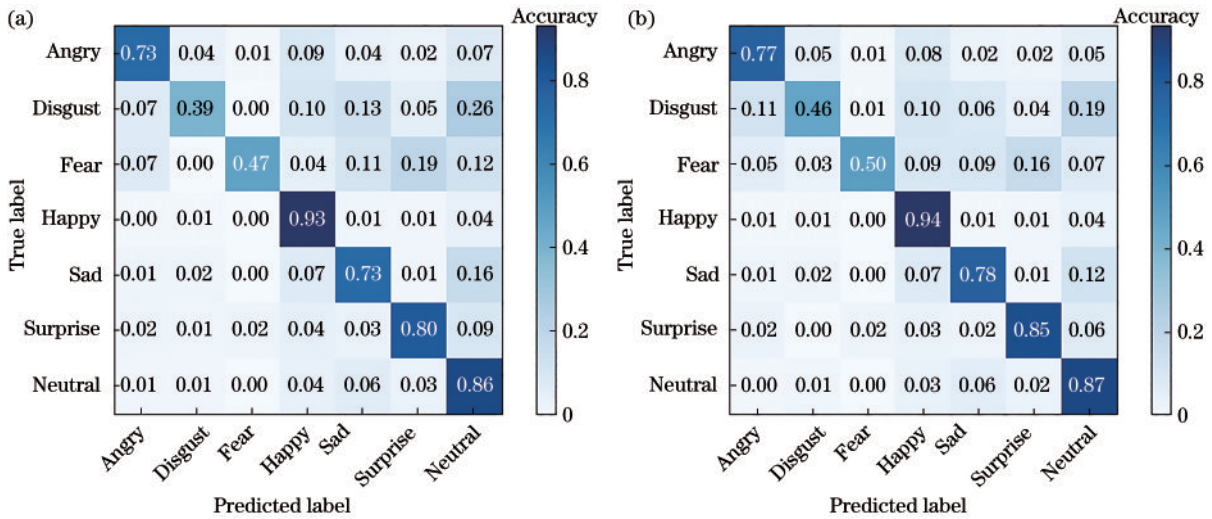


图 7 RAF-DB 实验结果(a)ResNet18;(b)SPResNet18\_CA  
Fig. 7 Results of RAF-DB experiment. (a) ResNet18; (b) SPResNet18\_CA

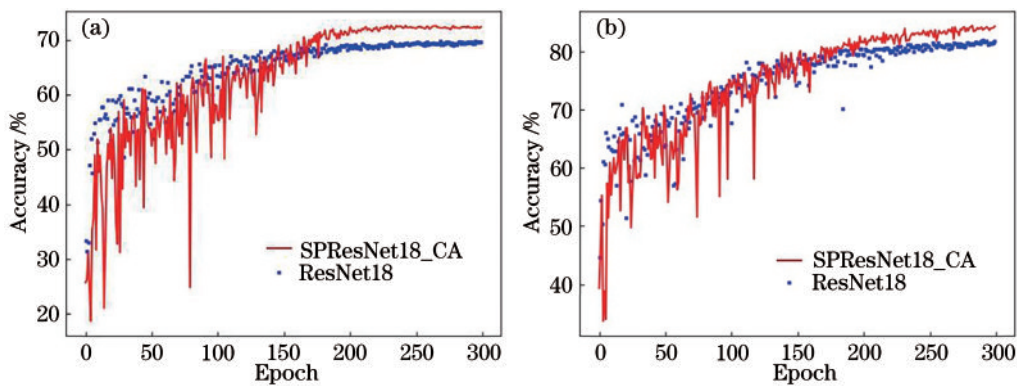


图 8 准确率曲线对比图。(a)FER2013;(b)RAF-DB  
Fig. 8 Accuracy curve comparison chart. (a) FER2013; (b) RAF-DB

准确率在 FER2013 上相比于 SHCNN 提高了 3.5 个百分点, 相比于 Zhou 等<sup>[23]</sup> 的方法提高了 1.7 个百分点。从表 4 可以看出, 所提方法在 RAF-DB 上相比于 DLP-CNN 提高了 10.22 个百分点, 相比于

pACNN 提高了 1.37 个百分点。从表 3、表 4 可以看出, 所提方法在 FER2013 和 RAF-DB 上的识别准确率均有一定的提升, 表明所提方法具有一定的竞争性。

表 3 在 FER2013 上的对比实验

Table 3 Comparative experiment on FER2013

Data	NetWork	Accuracy / %
FER2013	SHCNN <sup>[1]</sup>	69.1
	LCMA <sup>[21]</sup>	69.6
	LIANG <sup>[22]</sup>	70.3
	Zhou <sup>[23]</sup>	70.9
	<b>SPResNet18_CA</b>	<b>72.6</b>

表 4 在 RAF-DB 上的对比实验

Table 4 Comparative experiment on RAF-DB

Data	NetWork	Accuracy / %
RAF-DB	DLP-CNN <sup>[24]</sup>	74.20
	EAU-Net <sup>[25]</sup>	81.83
	DeepExp3D <sup>[26]</sup>	82.06
	pACNN <sup>[27]</sup>	83.05
	<b>SPResNet18_CA</b>	<b>84.42</b>

## 5 结 论

从特征提取维度提出了一种基于注意力拆分卷积残差网络的表情识别方法。该网络由 CACSBlock 组成,使用拆分卷积降低冗余特征的表现,使用注意力模块来标记特征权重,CACSBlock 中的拆分卷积增强了人脸表情的全局特征表现,注意力关注具有判别性的局部特征,通过全局和局部的增强来达到整体人脸表情识别特征的增强。从对比实验可以看出,所提方法在 FER2013、RAF-DB 数据集上的识别准确率均有提高。所提方法在特征提取维度有一定的提升效果。接下来将继续在特征提取维度深入地研究表情识别,并进一步将多特征融合方法应用于表情识别。

### 参 考 文 献

- [1] 尹鹏博,潘伟民,张海军. 基于卷积注意力的轻量级人脸表情识别方法[J]. 激光与光电子学进展, 2021, 58(12): 1210023.  
Yin P B, Pan W M, Zhang H J. Lightweight facial expression recognition method based on convolutional attention[J]. Laser & Optoelectronics Progress, 2021, 58(12): 1210023.
- [2] 姚丽莎,徐国明,赵凤. 基于卷积神经网络局部特征融合的人脸表情识别[J]. 激光与光电子学进展, 2020, 57(4): 041513.  
Yao L S, Xu G M, Zhao F. Facial expression recognition based on local feature fusion of convolutional neural network[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041513.
- [3] 杨旭,尚振宏. 基于改进 AlexNet 的人脸表情识别[J]. 激光与光电子学进展, 2020, 57(14): 243-250.  
Yang X, Shang Z H. Facial expression recognition based on improved AlexNet[J]. Laser & Optoelectronics Progress, 2020, 57(14): 243-250.

- [4] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [5] 吴慧华,苏寒松,刘高华,等. 基于余弦距离损失函数的人脸表情识别算法[J]. 激光与光电子学进展, 2019, 56(24): 241502.  
Wu H H, Su H S, Liu G H, et al. Facial expression recognition algorithm based on cosine distance loss function[J]. Laser & Optoelectronics Progress, 2019, 56(24): 241502.
- [6] Lin M, Chen Q, Yan S C. Network in network[EB/OL]. (2013-12-16)[2021-05-04]. <https://arxiv.org/abs/1312.4400>.
- [7] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [8] Xie X K, Zhou Y, Kung S Y. Exploring highly efficient compact neural networks for image classification[C]//2020 IEEE International Conference on Image Processing (ICIP), October 25-28, 2020, Abu Dhabi, United Arab Emirates. New York: IEEE Press, 2020: 2930-2934.
- [9] Huang W L, Wang J J, Xin X M, et al. Channel-modulated multibranch convolutional neural network[C]//2020 Chinese Automation Congress (CAC), November 6-8, 2020, Shanghai, China. New York: IEEE Press, 2020: 1854-1859.
- [10] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5987-5995.
- [11] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1800-1807.
- [12] Howard A G, Zhu M, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile applications[EB/OL]. (2017-04-17)[2021-02-04]. <https://arxiv.org/abs/1704.04861>.
- [13] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6848-6856.
- [14] Chen Y P, Fan H Q, Xu B, et al. Drop an octave: reducing spatial redundancy in convolutional neural networks with octave convolution[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3434-3443.
- [15] Singh P, Verma V K, Rai P, et al. HetConv: heterogeneous kernel-based convolutions for deep CNNs

- [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 4830-4839.
- [16] Li X, Wang W H, Hu X L, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 510-519.
- [17] Han K, Wang Y, Tian Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1577-1586.
- [18] Zhang Q L, Jiang Z Q, Lu Q S, et al. Split to be slim: an overlooked redundancy in vanilla convolution[EB/OL]. (2020-06-22)[2021-05-06]. <https://arxiv.org/abs/2006.12085>.
- [19] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.
- [20] Zhang T, Qi G J, Xiao B, et al. Interleaved group convolutions[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4383-4392.
- [21] Miao S, Xu H Y, Han Z Q, et al. Recognizing facial expressions using a shallow convolutional neural network[J]. IEEE Access, 2019, 7: 78000-78011.
- [22] 梁华刚, 雷毅雄. 增强可分离卷积通道特征的表情识别研究[J]. 计算机工程与应用, 2022, 58(2): 184-192.
- Liang H G, Lei Y X. Expression recognition with separable convolution channel enhancement features[J]. Computer Engineering and Applications, 2022, 58(2): 184-192.
- [23] Zhou J C, Jia X, Shen L L, et al. Improved softmax loss for deep learning-based face and expression recognition[J]. Cognitive Computation and Systems, 2019, 1(4): 97-102.
- [24] Li S, Deng W H. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition[J]. IEEE Transactions on Image Processing, 2019, 28(1): 356-370.
- [25] Deng J, Pang G Y, Zhang Z Y, et al. cGAN based facial expression recognition for human-robot interaction[J]. IEEE Access, 2019, 7: 9848-9859.
- [26] Koujan M R, Alharbawee L, Giannakakis G, et al. Real-time facial expression recognition "in the wild" by disentangling 3D expression from identity[C]//2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, November 16-20, 2020, Buenos Aires, Argentina. New York: IEEE Press, 2020: 24-31.
- [27] Zhou X Z, Jin K, Shang Y Y, et al. Visually interpretable representation learning for depression recognition from facial images[J]. IEEE Transactions on Affective Computing, 2020, 11(3): 542-552.