

文章编号: 2095-2163(2022)10-0068-07

中图分类号: TP391

文献标志码: A

# 基于特征提取和胶囊网络的人脸表情识别

黄小刚, 黄润才, 王桂江, 马诗语

(上海工程技术大学 电子电气工程学院, 上海 201620)

**摘要:** 人脸表情识别是近年来非常火热的一个研究领域,随着深度学习的发展,越来越多的深度学习方法用于表情识别中。针对胶囊神经网络(CapsNet)更关注的是图像高层空间信息、低层空间特征提取不全面的问题,提出了特征提取与胶囊网络结合的人脸表情识别算法。本文先使用局部二值模式(LBP)算子提取图像纹理特征,与胶囊网络结合形成多通道输入胶囊网络。为了进一步加强低层空间特征提取,在提取纹理特征后加入了深度残差网络(ResNet),与胶囊网络结合形成多通道输入增强胶囊网络。为了验证多通道输入胶囊网络和多通道输入增强胶囊网络的性能,本文在公开表情数据集CK+和RAF-DB分别进行了对照实验,得到了99.69%,82.02%准确率,优于其它的表情识别算法。

**关键词:** 人脸表情识别; 特征提取; 深度残差网络; 胶囊网络

## Facial expression recognition based on feature extraction and capsule network

HUANG Xiaogang, HUANG Runcai, WANG Guijiang, MA Shiyu

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

**[Abstract]** Facial expression recognition is a very hot research field in recent years. With the development of deep learning, more and more deep learning methods are used in expression recognition. Aiming at the problem that capsule neural network (CapsNet) pays more attention to the high-level spatial information of the image, but the low-level spatial feature extraction is not comprehensive, a facial expression recognition algorithm combining feature extraction and capsule network is proposed. Firstly, local binary pattern (LBP) operator is used in this paper to extract image texture features, and combined with capsule network, multi-channel input capsule network is formed. After that, in order to further strengthen low-level spatial feature extraction, depth residual network(ResNet) is added after extracting texture features. Finally, combined with capsule network, multi-channel input enhanced capsule network is formed. Based on the above, this paper carries out control experiments on the public expression data sets CK+ and RAF-DB respectively, which is used to verify the performance of the multi-channel input capsule network and the multi-channel input enhanced capsule network. The simulation achieves 99.69% and 82.02% accuracy, respectively and the results proves that the proposed method in this paper is better than other expression recognition algorithms.

**[Key words]** facial expression recognition; feature extraction; depth residual network; capsule network

## 0 引言

随着计算机技术的发展,人脸检测技术已经广泛应用于生活各领域中,如日常锁定和支付等,用来提高安全性。面部表情图像的获取正不断增多,如何进行表情情感分析、实现人机情感交互也就随即成为了当下的研究热点。

表情识别的过程可分为3个部分:图像采集与图像预处理、表情的特征提取、表情的分类。其中,人脸表情识别过程中又以表情特征提取为重点。在传统的人脸表情识别当中,有许多优秀的特征提取

算法:局部二值模式(LBP)<sup>[1]</sup>、方向梯度直方图(HOG)<sup>[2]</sup>、Gabor小波变换<sup>[3]</sup>、活动外观模型(AAM)<sup>[4]</sup>等。其中,局部二值模式由于其特征提取方式简洁明了,提取效果较好,受到众多学者青睐,因此衍生出了许多变体:CS-LBP<sup>[5]</sup>、CLBP<sup>[6]</sup>、CS-LOP<sup>[7]</sup>等。在特征提取当中,文献[8]使用LBP的变体,能更好地处理像素边缘的特征。文献[9]提取了LBP纹理特征以及用HOG算法提取眼睛、眉毛区域、连同嘴部区域的边缘信息,并对这3个区域进行不同权值的融合。文献[10]使用了一种活动外观模型与Gabor小波变换融合的方法。上述特征提取方法都具有一定的局限性,识别的稳定性也欠

**作者简介:** 黄小刚(1998-),男,硕士研究生,主要研究方向:机器学习、图像处理、深度学习;黄润才(1966-),男,博士,副教授,主要研究方向:智能计算、计算机网络与应用;王桂江(1990-),男,硕士研究生,主要研究方向:文本定位、文本识别、情感分析;马诗语(1997-),女,硕士研究生,主要研究方向:自然语言处理、知识图谱。

**通讯作者:** 黄润才 Email:hrc@sues.edu.cn

收稿日期: 2021-12-25

佳, 研究学界逐渐掀起了对深度学习的研究热潮。

近年来硬件技术取得了较大的突破, 深度学习也随之得到迅猛发展, 许多学者开始把深度学习方法应用于表情识别当中<sup>[11-12]</sup>。文献[13]使用了改进的 AlexNet 网络进行表情识别, 文献[14]构造表情局部特征融合的卷积网络模型。2017 年, Sara 等人<sup>[15]</sup>提出了胶囊网络, 使用胶囊单元存储信息, 并使用独特的动态路由机制传递传输胶囊信息, 该网络不仅能够检测到特征, 还能检测到特征的空间、大小、位置等信息, 具有可观发展前景。文献[16]验证了胶囊网络相比卷积神经网络在表情识别上具有更强的鲁棒性。文献[17]使用了胶囊网络和卷积网络的结合, 同时加入大量的卷积操作和层注意力机制, 实验结果表明该网络稳健性强, 但是也存在着网络结构冗余、网络收敛很慢、训练时间很长的问题。

本文把传统的特征提取方法与深度学习胶囊网络相结合, 提取了图像的局部二值模式特征, 与原图通道合并, 形成双通道输入, 送入胶囊网络, 形成一种多通道输入胶囊网络。残卷积网络(ResNet)<sup>[18]</sup>具有收敛速度快, 能避免梯度弥散的优点, 本文又在双通道输入的基础上添加残差网络形成多通道输入增强胶囊网络模型。通过局部二值模式、ResNet 提取复杂低层特征, 送入胶囊网络分类, 在 CK+<sup>[19]</sup>数据集和 RAF-DB<sup>[20]</sup>数据集上的实验结果表明 2 种特征增强能充分发挥胶囊网络的效果, 获取了更具表达能力的特征。

## 1 基本原理

### 1.1 局部二值模式

纹理信息对于图像的模式分析非常重要, 局部二值模式(LBP)是一种简单而高效的纹理描述方法。局部二值模式是能够获得中心像素和相邻像素之间差异的二进制模式。具体通过用中心值对每个像素的  $3 \times 3$  邻域进行阈值化并将结果作为二进制数来标记图像的像素, 中心点  $c$  像素的 LBP 编码如式(1)、(2)所示:

$$LBP(c) = \sum_{i=0}^{N-1} s(g_i - g_c) * 2^i \quad (1)$$

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2)$$

其中,  $g_c$  表示中心像素的灰度值;  $N$  是涉及邻域像素的总数;  $g_i (i = 0, 1, \dots, N - 1)$  是以  $c$  为中心的邻域像素的灰度值。中心像素及邻域半径为 1 的

$3 \times 3$  的像素如图 1 所示。

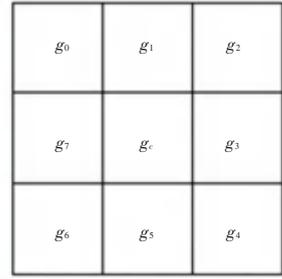


图 1  $3 \times 3$  像素点

Fig. 1 Pixels of  $3 * 3$

### 1.2 胶囊网络

胶囊网络(CapsNet)以胶囊作为基本神经元存储信息, 不同于卷积神经网络标量输入、标量输出的传递形式, 胶囊网络是一个向量输入、向量输出的形式。胶囊是一个向量, 具有多个值, 包含更多图像实体的信息, 不仅可以表示是否存在特征, 还可以表示特征之间的关系, 输出综合这些信息, 使用向量的模长表示实体存在的概率。用于 Mnist 手写数字识别的胶囊网络结构如图 2 所示。

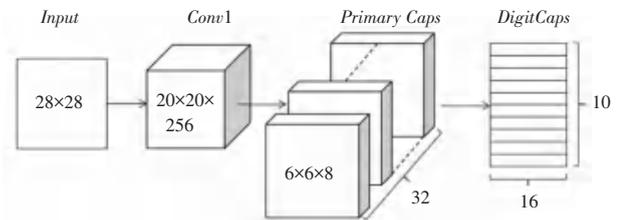


图 2 胶囊网络结构

Fig. 2 Capsule network structures

胶囊网络由输入层( Input)、卷积层( Conv1)、主胶囊层( PrimaryCaps)、数字胶囊层( DigitCaps)组成。输入层即输入  $28 \times 28$  大小的灰度图。Conv1 就是普通的卷积层, 使用 256 个步长为 1 的  $9 \times 9$  卷积核, ReLU 为激活函数, 用于提取图像的低层特征, 得到  $20 \times 20 \times 256$  的特征图。

主胶囊层开始使用胶囊单元存储信息, 该过程可以看作是普通卷积层的纵向扩展, 使用了 8 组, 32 个步长为 2 的  $9 \times 9$  卷积核, 把特征图进行三维拼接, 得到  $6 * 6 * 32$ , 即 1 024 个胶囊, 每个胶囊是一个长度为 8 的向量。对胶囊进行归一化, 使用非线性 squashing 压缩函数, 对胶囊进行压缩, 确保长度在  $0 \sim 1$  之间, 压缩函数如式(7)所示:

$$v_j = \text{squash}(s_j) = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (3)$$

其中,  $s_j$  是胶囊  $j$  的输入,  $v_j$  是输出。

数字胶囊层的输出是 10 个维数为 16 的向量,

每个向量的模长代表该类的预测概率。*DigitCaps* 层采用的是动态路由机制,进行3次动态路由迭代。动态路由的过程原理如图3所示。参数更新公式的数学表述可写为:

$$\hat{\mathbf{u}}_{ij} = \mathbf{W}_{ij} \mathbf{u}_{ij} \quad (4)$$

$$\mathbf{s}_j = \sum_i c_{ij} \hat{\mathbf{u}}_{ji} \quad (5)$$

$$\mathbf{v}_j = \text{squash}(\mathbf{s}_j) \quad (6)$$

$$b_{ij} = b_{ij} + \hat{\mathbf{u}}_{ij} \cdot \mathbf{v}_j \quad (7)$$

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_j \exp(b_{ij})} \quad (8)$$

其中,  $i \in (1, 2, \dots, 1024)$ ,  $j \in (1, 2, \dots, 10)$ , 这是由于主胶囊层共有1024个胶囊,数字胶囊层输出类别有10类;  $\mathbf{u}_{ij}$  是主胶囊层输出的1024个向量;  $\mathbf{W}_{ij}$  是权重矩阵;  $\hat{\mathbf{u}}_{ji}$  是预测向量,表示当前向量条件下预测为  $j$  类别的概率。式(5)中,  $\mathbf{s}_j$  是上一层预测向量  $\hat{\mathbf{u}}_{ji}$  乘以胶囊间耦合系数  $c_{ij}$  的加权和;式(6)是 *squash* 压缩函数,式(7)、式(8)用来迭代更新耦合系数  $b_{ij}, c_{ij}$ 。

胶囊网络采用的是边际损失函数,如式(9)所示:

$$L_c = T_c \max(0, m^+ - \|\mathbf{v}_c\|)^2 + \lambda(1 - T_c) \max(0, \|\mathbf{v}_c\| - m^-)^2 \quad (9)$$

其中,  $T_c$  表示  $c$  类是否存在:存在为1,不存在

为0;  $m^+$  取值为0.9,惩罚假阳性,当  $c$  类存在预测不存在时会使损失函数很大;  $m^-$  取值为0.1,惩罚假阴性,当  $c$  类不存在预测存在时会使损失函数很大;  $\lambda$  取值为0.5,调整假阴性的权重。

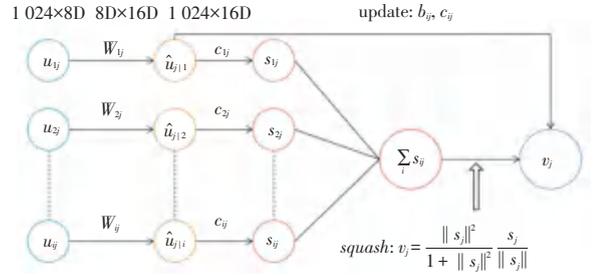


图3 动态路由过程

Fig. 3 Dynamic routing process

## 2 实验框架

### 2.1 整体框架

本文提出的基于特征提取与胶囊网络的表情识别算法实现流程如图4所示。图4中,首先对图片进行预处理操作,使用Harr级联器进行人脸检测,获取人脸部分图片,尺寸归一化为48x48、像素归一化为0~1之间。对图片进行数据增强,数据增强操作包括图片的旋转、平移、错切获取更多的训练样本。提取图像LBP纹理特征,提取图像ResNet18特征,送入胶囊网络分类训练,验证模型准确率。

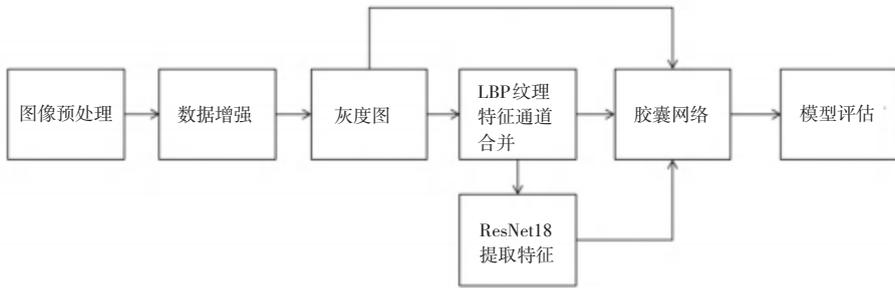


图4 识别算法流程图

Fig. 4 Flow chart of recognition algorithms

### 2.2 多通道输入胶囊网络

多通道输入胶囊网络使用LBP算子提取图像的纹理特征,灰度图和其纹理特征如图5所示。将纹理特征图与灰度图进行通道合并的多通道输入胶囊网络结构如图6所示。该网络共进行2组实验。第一组实验输入只使用灰度图作为对照实验,用于验证是否提取纹理特征的多通道输入胶囊网络相比单一的胶囊网络具有更好的识别效果,第二组实验输入为灰度图-LBP特征图。2组实验输入分别送

入胶囊网络的Conv1卷积层,得到的40x40x32的特征图,接着输入到主胶囊层中,特征图从标量变为矢量,得到特征长度为8的1024(16\*16\*4)个向量,最后送入到数字胶囊层,通过动态路由算法迭代3次得到预测结果为16x8的向量,8表示表情的类别(RAF-DB数据集为7类表情,预测结果为16x7,这里的16表示向量的长度,包含了预测该类别的特征信息)。网络各层详细编码参数见表1。



(a) 原图 (b) LBP 特征

图 5 特征图

Fig. 5 Feature map

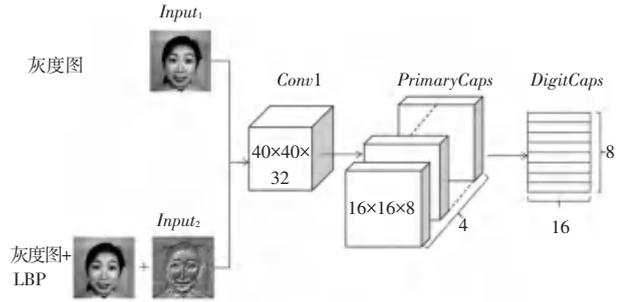


图 6 多通道输入胶囊网络

Fig. 6 Multi-channel input capsule network

表 1 特征提取的胶囊网络结构参数

Tab. 1 Structural parameters of capsule network based on feature extraction

Layer	Input shape	Output shape	Kernel structure	Activation
Conv1	$Input_1: 48 \times 48 \times 1$ $Input_2: 48 \times 48 \times 2$	$40 \times 40 \times 32$	$9 \times 9$ conv, 32	ReLU
PrimaryCaps	$40 \times 40 \times 32$	$1\ 024 \times 8$	8 groups $9 \times 9$ conv, 4, /2	ReLU, Squash
DigitCaps	$1\ 024 \times 8$	$16 \times 8 (16 \times 7)$	Dynamic routing	Squash

从图 6 和表 1 可以看出,该网络 2 组实验使用 2 组不同的特征作为输入进行训练,由于原图尺寸为  $48 \times 48$ ,对主胶囊层的卷积核个数进行了调整,使得输出仍为 1 024 个维度为 8 的向量,主胶囊层和数字胶囊层的胶囊维数并没有改变,数字胶囊层输出尺寸  $16 \times 8$  为 CK+数据集,  $16 \times 7$  为 RAF-DB 数据集输出,后续不再做重复说明。

### 2.3 多通道输入增强胶囊网络

多通道输入增强胶囊网络结构如图 7 所示。在输入图像为灰度图和 LBP 特征图基础上,再使用泛化能力强的 ResNet18 网络对胶囊网络进行增强。

ResNet18 网络由 1 个 Stem 层,4 个残差模块 BasicBlock 组成,该结构使用了大量的  $3 \times 3$  的卷积核,为了满足主胶囊 1 024 个 8 维向量的输出,对网络卷积核个数步长进行了相应的调整改进。输入图像经过 ResNet18 得到  $6 \times 6 \times 256$  特征图,送入主胶囊层,ResNet18 网络可以提取丰富特征,主胶囊层对特征进行整理,把标量变为向量,得到 1 024 个维度为 8 的向量。主胶囊层的胶囊送入数字胶囊层通过动态路由迭代 3 次得到最终分类结果,数字胶囊层的胶囊维数为 16。网络各层详细编码参数见表 2。

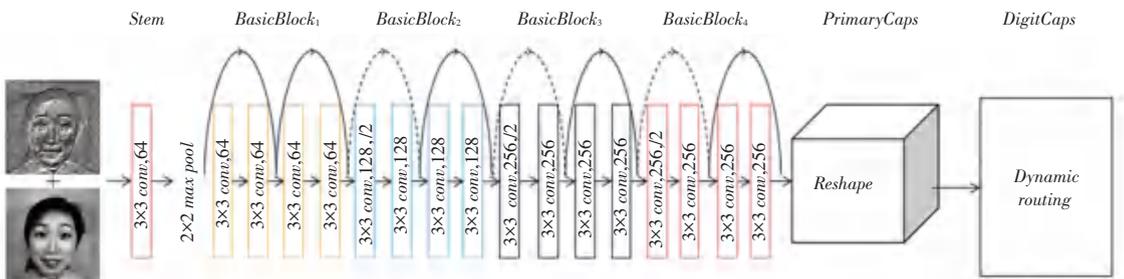


图 7 多通道输入增强胶囊网络

Fig. 7 Multi-channel input enhanced capsule network

表2 增强胶囊网络结构参数

Tab. 2 Enhanced capsule network structure parameters

Layer	Input shape	Output shape	Kernel structure	Activation
<i>Stem</i>	$48 \times 48 \times 2$	$48 \times 48 \times 1$	$3 \times 3$ conv, 64 $2 \times 2$ max pool	<i>ReLU</i>
<i>BasicBlock<sub>1</sub></i>	$48 \times 48 \times 1$	$48 \times 48 \times 64$	$3 \times 3$ conv, 64 $3 \times 3$ conv, 64 $3 \times 3$ conv, 64 $3 \times 3$ conv, 64	<i>ReLU</i>
<i>BasicBlock<sub>2</sub></i>	$48 \times 48 \times 64$	$24 \times 24 \times 128$	$3 \times 3$ conv, 128,/2 $3 \times 3$ conv, 128 $3 \times 3$ conv, 128 $3 \times 3$ conv, 128	<i>ReLU</i>
<i>BasicBlock<sub>3</sub></i>	$24 \times 24 \times 128$	$12 \times 12 \times 256$	$3 \times 3$ conv, 256,/2 $3 \times 3$ conv, 256 $3 \times 3$ conv, 256 $3 \times 3$ conv, 256	<i>ReLU</i>
<i>BasicBlock<sub>4</sub></i>	$12 \times 12 \times 256$	$6 \times 6 \times 256$	$3 \times 3$ conv, 256,/2 $3 \times 3$ conv, 256 $3 \times 3$ conv, 256 $3 \times 3$ conv, 256	<i>ReLU</i>
<i>PrimaryCaps</i>	$6 \times 6 \times 256$	$1\ 152 \times 8$	<i>Reshape</i>	<i>ReLU, Squash</i>
<i>DigitCaps</i>	$1\ 024 \times 8$	$16 \times 8(16 \times 7)$	<i>Dynamic routing</i>	<i>Squash</i>

### 3 实验结果与分析

实验程序运行在 Window10 操作系统上,使用 GeForce GTX3060 GPU,仿真环境使用 Tensorflow2.6 的深度学习框架。

本文采用 CK+和 RAF-DB 数据集进行实验。CK+数据集是实验环境下表情序列,在表情序列中选取了 1 085 张基本图片,这些图片中有 7 类基本表情,分别是:生气、厌恶、害怕、开心、伤心、惊讶、轻蔑。由于样本较少,添加了对应的中立表情,总共 8 种表情。对这些图片进行数据增强得到更多的数据集,其中 70% 的图片用于训练,30% 的图片用于测试。RAF-DB 数据集是自然环境下的表情图片,包含 7 类基本表情,训练集图片有 12 271 张,测试集图片有 3 068 张。

2 个数据集分别进行 3 组实验。第一组实验为基准实验,直接用灰度图送入胶囊网络进行训练,称为 CapsNet。第二组实验为纹理特征的多通道输入

胶囊网络实验,称为 LBP-CapsNet。第三组实验是基于纹理特征和深度残差网络的多通道输入增强胶囊网络,称为 ResNet-CapsNet。

实验中,使用 Adam 优化器进行梯度下降,总共迭代 200 次,批大小为 30。前 80 次迭代学习率为 0.01,80~160 次迭代学习率为 0.005,160 次后迭代学习率为 0.001,CK+数据集 3 组对照实验的验证集训练曲线如图 8 所示,RAF-DB 数据集 3 组对照实验的验证集训练曲线如图 9 所示。

根据图 8、图 9 可以看出,相比于只使用胶囊网络 CapsNet 实验,多通道输入胶囊网络 LBP-CapsNet 损失值较小,准确率相比于 CapsNet 实验有所提高。加入 ResNet 网络提取特征后的多通道输入增强型胶囊网络 ResNet-CapsNet 的初始损失值较大、准确率较低,这是由于添加 ResNet 网络,使网络结构变得更为复杂,随着网络不断训练迭代,学习的特征越来越多,网络损失值变小,准确率明显高于 CapsNet。

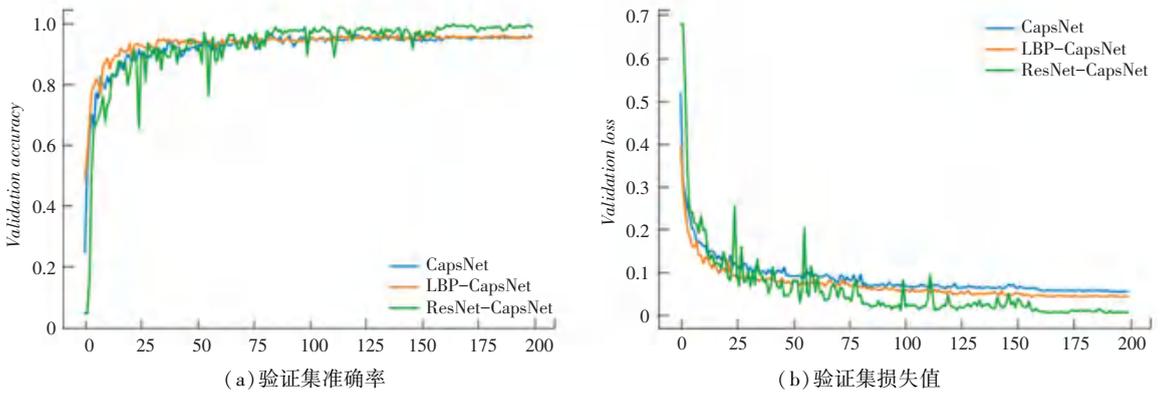


图 8 CK+训练曲线

Fig. 8 CK + training curve

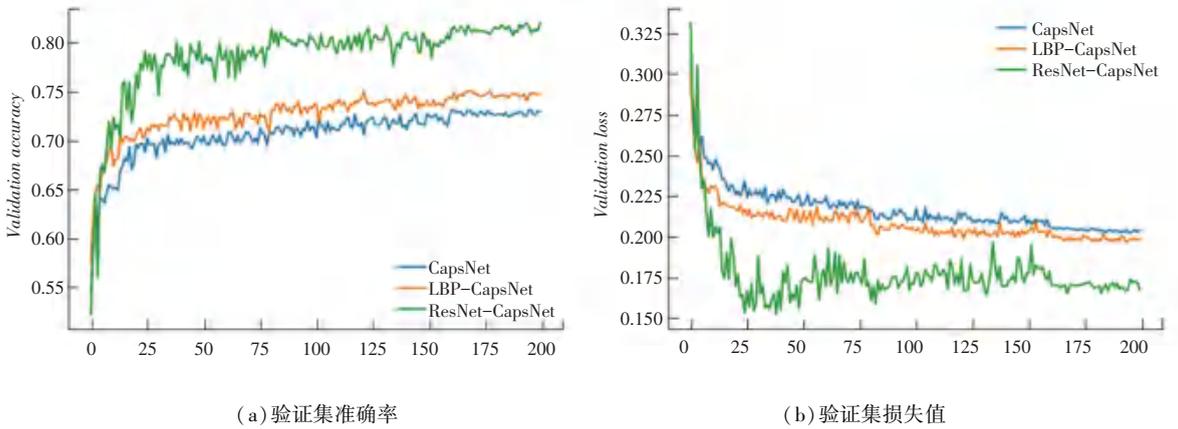


图 9 RAF-DB 训练曲线

Fig. 9 RAF-DB training curve

CK+和 RAF-DB 数据集的 3 组实验最终准确率见表 3。本文提出的多通道输入胶囊网络和多通道输入增强胶囊网络方法相比于只使用胶囊网络在 2 个数据集上准确率有明显提高,其中多通道输入增强胶囊网络在 CK+和 RDF-DB 数据集准确率分别达到 99.69%、82.02%。表 4 和表 5 分别列举了其他表情识别算法在 CK+和 RAF-DB 数据集上的准确率。分析可知,本文算法在表情识别方面具有很好的表现,准确率有明显的提升。

表 3 本文方法准确率比较

Tab. 3 Recognition rate comparison of the proposed method

Data set	Experiments	Input feature	Network	Accuracy/ %
CK+	CapsNet	\	CapsNet	96.61
	LBP-CapsNet	LBP	CapsNet	<b>97.23</b>
	ResNet-CapsNet	LBP	ResNet+CapsNet	<b>99.69</b>
RAF-DB	CapsNet	\	CapsNet	73.25
	LBP-CapsNet	LBP	CapsNet	<b>75.09</b>
	ResNet-CapsNet	LBP	ResNet+CapsNet	<b>82.02</b>

表 4 CK+的不同算法的准确率比较

Tab. 4 Recognition rate comparison of different algorithms on CK+

Reference	Method	Accuracy/ %
Ref. [6]	GF+CLBP+RF	92.80
Ref. [13]	AlexNet	97.46
Ref. [14]	CNN+SVM	94.56
Ref. [21]	Gabor+CapsNet	98.43
The proposed method	LBP+ResNet+CapsNet	<b>99.69</b>

表 5 RAF-DB 的不同算法的准确率比较

Tab. 5 Recognition rate comparison of different algorithms on RAF-DB

Reference	Method	Accuracy/ %
Ref. [22]	Boosting-POOF	73.19
Ref. [20]	DLP-CNN	74.20
Ref. [23]	Occams Razor	80.00
Ref. [24]	SAP	81.81
The proposed method	LBP+ResNet+CapsNet	<b>82.02</b>

## 4 结束语

本文提出了多通道输入胶囊网络和多通道输入增强胶囊网络表情识别方法,2种方法的本质都是通过充分提取低层特征,增强了胶囊网络提取特征的能力,得到了能够充分表示表情的特征的信息。在CK+和RAF-DB数据集上,本文的方法与其他论文方法对比可知,准确率有明显提高。本文算法只提取了图像的纹理特征作为低层特征,而几何特征也是人脸表情的重要特征,后续可以进一步加入几何特征进行实验。胶囊网络仍有较大发展潜力,许多细分方向仍有待深入挖掘,比如胶囊维数、动态路由迭代次数、squash函数等等对网络的影响。结合纹理特征、几何特征、深度学习这些方法,使用更多复杂场景的人脸图片进行验证以提高识别率是下一步的研究方向。

## 参考文献

- [1] GHIMIRE D, PARK S H, KIM M J. Facial expression recognition using local region specific dense optical flow and LBP features [M]//BLANC - TALON J, PENNE R, PHILIPS W, et al. Advanced Concepts for Intelligent Vision Systems. ACIVS 2017. Lecture Notes in Computer Science. Cham:Springer, 2017,10617: 28-38.
- [2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05). San Diego, CA, USA:IEEE, 2005,1:886-893.
- [3] ZHANG Z, LYONS M, SCHUSTER M, et al. Comparison between geometry - based and Gabor wavelets - based facial expression recognition using multi - layer perceptron [C]// Proceedings of the 3<sup>rd</sup>. International Conference on Face & Gesture Recognition (FG'98). Washington, DC, USA:IEEE, 1998:1-7.
- [4] ZHOU Huiling, LAM K M, HE Xiangjian. Shape - appearance - correlated active appearance model - ScienceDirect[J]. Pattern Recognition, 2016, 56(C):88-99.
- [5] HEIKKILÄ M, PIETIKÄINEN M, SCHMID C. Description of interest regions with local binary patterns[J]. Pattern Recognition, 2009, 42(3):425-436.
- [6] 王从澳,黄润才,孙延标,杨彬,等. 融合 CLBP 与几何显著特征的人脸情感识别[J]. 智能计算机与应用,2020,10(05):52-55.
- [7] HU Min, ZHENG Yaqin, YANG Chunjian, et al. Facial expression recognition using fusion features based on center - symmetric local octonary pattern [J]. IEEE Access, 2019, 7: 29882-29890.
- [8] 贾锋,王高,师钰璋,等. 基于改进的 LBP 及 KNN 算法的表情识别[J]. 国外电子测量技术,2020, 39(08):40-44.
- [9] 贾磊. 基于 LBP 和 HOG 特征融合的人脸表情识别算法研究 [D]. 太原:中北大学, 2019.
- [10] LILIANA D Y, BASARUDDIN C, WIDYANTO M R. Mix emotion recognition from facial expression using SVM - CRF sequence classifier [C]//Proceedings of the International Conference on Algorithms, Computing and Systems(ICACS '17). Jeju Island Republic of Korea:ACM, 2017:27-31.
- [11] VEMULAPALLI R, AGARWALA A. A compact embedding for facial expression similarity[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019:5676-5685.
- [12] JIA Xiuyi, ZHENG Xiang, LI Weiwei, et al. Facial emotion distribution learning by exploiting low - rank label correlations locally [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA:IEEE, 2019:9833-9842.
- [13] 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(09):1005-1012.
- [14] 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(04):338-345.
- [15] SARA S, FROSST N, HINTON G E. Hinton. Dynamic routing between capsules [C] // Proceedings of the 31<sup>st</sup> International Conference on Neural Information Processing Systems (NIPS 2017). Long Beach, CA, USA:ACM,2017:3859-3869.
- [16] LI Donghui, ZHAO Xingcong, YUAN Guangjie, et al. Robustness comparison between the capsule network and the convolutional network for facial expression recognition [J]. Applied Intelligence, 2021, 51:2269-2278.
- [17] CAO Shan, YAO Yuqian, AN Gaoyun. E2 - capsule neural networks for facial expression recognition using AU - aware attention[J]. IET Image Processing, 2020, 14(11):2417-2424.
- [18] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[J]. arXiv preprint arXiv: 1512.03385,2015.
- [19] LUCEY P, COHN J F, KANADE T, et al. The extended Cohn - Kanade dataset (CK+) : A complete dataset for action unit and emotion - specified expression [C]// Computer Vision & Pattern Recognition Workshops. San Francisco, USA: IEEE, 2010:94 - 101.
- [20] LI Shan, DENG Weihong, DU Junping. Reliable crowdsourcing and deep locality - preserving learning for expression recognition in the wild[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, Hawaii: IEEE, 2017: 2852-2861.
- [21] HOSSEINI S, CHO N I. GF - CapsNet: Using Gabor jet and capsule networks for facial age, gender, and expression recognition [C]//2019 14<sup>th</sup> IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). Lille, France: IEEE, 2019: 1-8.
- [22] LIU Zhiwen, LI Shan, DENG Weihong. Boosting - POOF: Boosting part based one vs one feature for facial expression recognition in the wild[C]// IEEE International Conference on Automatic Face & Gesture Recognition. Washington, DC, USA: IEEE, 2017:967-972.
- [23] VIELZEUF V, KERVADEC C, PATEUX S, et al. An Occam's Razor view on learning audiovisual emotion recognition with small training sets [C]// 20<sup>th</sup> ACM International Conference on Multimodal Interaction (ICMI). Boulder, Colorado: ACM, 2018: 589-593.
- [24] LI Huihui, WEN Guihua. Sample awareness - based personalized facial expression recognition[J]. Applied Intelligence, 2019, 49: 2956-2969.