

一种用于微表情自动识别的三维卷积神经网络进化方法



梁正友 何景琳 孙宇

广西大学计算机与电子信息学院 南宁 530004

摘要 由于微表情持续时间短、动作幅度小,因此微表情自动识别一直是一个具有挑战性的问题。针对上述问题,提出一种用于微表情识别的三维卷积神经网络进化(Three-Dimensional Convolutional Neural Network Evolution, C3DEvol)方法。该方法使用能有效提取动态信息的三维卷积神经网络(Three-Dimensional Convolutional Neural Network, C3D)来提取微表情在时域和空域上的特征;同时使用具有全局搜索和优化能力的遗传算法对C3D的网络结构进行优化,以获取最优的C3D网络结构和避免局部优化。利用CASME2数据集在带有两块NVIDIA Titan X GPU的工作站上开展了实验,结果表明C3DEvol微表情自动识别的准确率达到63.71%,优于现有的微表情自动识别方法。

关键词: 微表情识别;遗传算法;三维卷积神经网络;特征提取;网络结构优化

中图分类号 TP391

Three-dimensional Convolutional Neural Network Evolution Method for Facial Micro-expression Auto-recognition

LIANG Zheng-you, HE Jing-lin and SUN Yu

School of Computer and Electronics Information, Guangxi University, Nanning 530004, China

Abstract Due to the short duration of micro-expressions and the small amplitude of motion, the automatic recognition of micro-expressions is still a challenging problem. Aiming at the problems, this paper proposes a Three-Dimensional Convolutional Neural Network Evolution (C3DEvol) method for micro-expression recognition. In the C3DEvol, three-dimensional Convolutional Neural Network (C3D) which can extract dynamic information effectively is used to extract micro-expression features in time domain and space domain. At the same time, the genetic algorithm with the capabilities of global search and optimization is used to optimize the network structure of C3D in order to obtain the optimal network structure and avoid local optimization. Experiments are performed on a workstation with two NVIDIA Titan X GPUs using the CASME2 dataset. Experiments show that the accuracy of C3DEvol micro-expression automatic recognition reaches 63.71%, which is better than the existing micro-expression automatic recognition method.

Keywords Micro-expression recognition, Genetic algorithm, Three-dimensional convolutional neural network, Feature extraction, Network structure optimization

近年来,微表情识别逐渐成为研究热点。与普通表情相比,自发式的微表情更能够反映人的真实情绪。微表情指人们尽最大努力抑制真实表情的视频片段,它的持续时间短,仅为 $1/25\sim 1/3\text{ s}$ ^[1],并且动作幅度小。尽管能够通过人力对微表情进行识别,但识别的准确率并不高,经过培训的人员对其进行识别的准确率也不低于50%^[2]。因此,有必要使用先进的计算机视觉和机器学习技术来提高识别率。

最先用于微表情识别的是手工描述特征的识别方法。Pfister等^[3]提出LBP-TOP手工描述特征识别方法,通过在3个正交平面中组合局部二进制模式来对自发式微表情进行手动特征识别。Wang等^[4]提出的LBP-SIP方法基于LBP-TOP,在所有相邻点中选取6个点,而LBP-MOP方法^[5]沿3

个正交平面仅提取了3个平均图像,两种方法都降低了数据的冗余度。另外,Huang等^[6]考虑了更多的信息,如信号、大小和方向因素,以及时空全局量化模式。Liu等^[7]提出了一种基于时空空间局部纹理描述符的主方向平均光流法,识别更加简单有效。

近年来,深度学习技术在识别方面获得了巨大的成功,已经被广泛应用于多个领域,如行为识别^[8]、自然语言处理^[9]、语音识别^[10]等,也逐步被应用到微表情自动识别中。Tran等^[11]提出了C3D深度学习技术,能够同时提取视频的时域信息和空域信息,较好地表示了人类活动的特性,在行为识别、场景识别、视频相似度分析等领域开展了实验。Kim等^[12]提出了结合时域和空域不同维度信息的提取方法,其中

到稿日期:2019-06-30 返修日期:2019-10-17 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61763002)

This work was supported by the National Natural Science Foundation of China(61763002).

通信作者:梁正友(zhlyiang@gxu.edu.cn)

空间维度信息通过搭建 CNN 提取帧序列的 5 个不同状态信息获得,时域信息通过 LSTM 网络获得。Peng 等^[13]采用两种时间维度差异的卷积神经网络对不同数据集的微表情进行识别。

深度学习技术尽管在微表情自动识别的应用中取得了一些成果,但有待进一步研究。深度学习网络模型的结构参数复杂,人工选取合适的参数工程量大,而且神经网络的层数及参数目前还没有严格的理论指导,这导致不容易设计出一个好的网络模型,且设计出的深度网络存在训练时间过长且识别效果不佳等问题。因此,寻找适合的网络结构参数优化技术对于微表情识别的深度学习网络进行优化,有利于节约深度学习网络的时间花销并提高其准确率。

用进化算法优化深度学习网络模型参数是一种提高深度学习性能的有效方法。进化算法(Evolutionary Algorithms, EAs)^[14]是一类模拟自然界遗传进化规律的仿生学算法,主要包括遗传算法(Genetic Algorithm, GA)^[15]、遗传规划(Genetic Programming, GP)^[16]、进化策略(Evolutionary Strategies, ES)^[17]和进化规划(Evolutionary Programming, EP)^[18]4种典型方法。进化算法的产生借鉴了大自然中生物的进化理论,它主要包括基因编码、种群初始化、交叉变异算子、经营保留机制等基本操作。目前已有不少文献基于二维卷积神经网络(2D Convolution Neural Network, 2DCNN)提出了进化算法,并将其应用于不同的领域。Chen 等^[19]提出了一种深度信任神经网络的进化算法,提高了传统网络在时间复杂度和精度方面的性能。Ijjina 等^[20]通过遗传算法和梯度下降算法优化 2DCNN,并将其应用于人体动作识别;Oullette 等^[21]通过遗传算法优化 2DCNN 的卷积核种类,以避免陷入常规 2DCNN 中的局部最小值,并将得到的网络应用于纹理检测;Rikhtegar 等^[22]通过遗传算法优化 2DCNN 网络,并用其来识别带有姿势和光照变化的图片。上述研究的共同点在于使用遗传算法优化 2DCNN 的参数值,以获得最好的应用效果。但目前为止,未发现针对三维卷积神经网络进行进化优化的研究。

本文提出了一种用于微表情识别的三维卷积神经网络进化方法 C3DEvol,该方法利用遗传算法实现全局搜索及优化,训练评估得出最佳的参数组合结果,再通过最佳参数组合的网络结构在维度上提取特征,实现微表情的自动识别。实验结果表明,C3DEvol 提高了微表情自动识别的准确率,达到了 63.71%。

1 C3D

C3D 是 Ji 等^[23]提出的一种基于时域和空域的深度学习网络。与二维卷积神经网络相比,C3D 更能提取出时域的特征,可用于行为识别、场景识别、视频相似度分析等领域。

C3D 的优点在于,采用三维卷积核对上一层网络中的特征映射进行卷积操作,可以一次性提取时域特征,即可以捕捉到多个帧的动作信息。具体地,第 l 层网络的第 j 个特征映射上的像素点 (x, y, z) 的特征值可以记作 (a_{ij}^{xyz}) ,如式(1)所示:

$$a_{ij}^{xyz} = f[b_{ij} + \sum_n \sum_{s=0}^{S_l-1} \sum_{t=0}^{T_l-1} \sum_{r=0}^{R_l-1} \omega_{ij}^{srt} a_{i-1}^{(x+s)(y+t)(z+r)}] \quad (1)$$

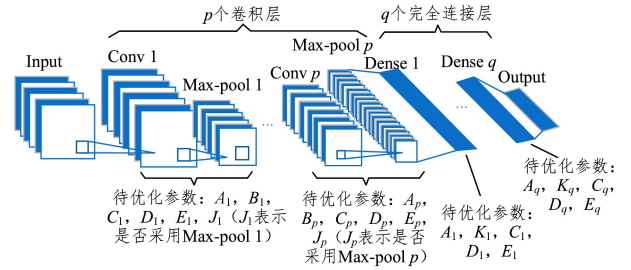
其中, b_{ij} 为特征映射的偏置值, n 为连接当前特征映射的第 $(l-1)$ 层网络的特征映射集,而 S_l 和 T_l 分别是三维卷积核的高度和宽度, R_l 是三维卷积核的时域维度大小, ω_{ij}^{srt} 为连接上层特征映射的三维卷积核位置 (s, t, r) 的权重, $f(x)$ 表示激活函数。

2 C3DEvol 方法

C3DEvol 方法是将 C3D 参数组看作进化算法中的个体,每个个体代表一种参数选择策略,适应度较高的个体进入下一代种群;对每一个体进行评估,直到最后一代评估结束,筛选出最优的个体。

2.1 基本介绍

用于微表情识别的 C3D 网络如图 1 所示,由输入层 Input、 p 个卷积层 Conv(附带池化层 Max-pool)、 q 个完全连接层 Dense 和输出层 Output 组成,其中 p 和 q 的值可变。



注: Conv 1 表示第 1 层卷积层, Conv p 表示第 p 层卷积层; Dense 1 表示第 1 层完全连接层; Dense q 表示第 q 层完全连接层; Output 为最后的分类层

图 1 待优化的 C3D 网络结构

Fig. 1 C3D network structure to be optimized

使用 C3DEvol 进行微表情的识别过程是:首先对微表情进行前期的预处理;然后在每一代中使用进化的 C3D 网络对微表情进行特征提取和分类,并对得到的准确率进行适应度值评估;经过多次这样的迭代,筛选出最佳的网络结构,使用该网络进行微表情识别可以得到最佳的微表情识别率。

C3DEvol 的目标是用遗传算法不断优化 C3D 网络参数,以获得最佳的网络结构。待优化的 C3D 网络结构包括卷积层、完全连接层(包括最后的分类层)和优化器。待优化的 C3D 网络结构参数及标识符设置如表 1 所列。

表 1 待优化 C3D 网络结构的参数组设置

Table 1 Parameter group settings for C3D network structure to be optimized

参数名称(参数标识符)	
卷积层	该层是否有效(A_n)、卷积核种类(B_n)、是否采用批标准化(C_n)、激活函数(D_n)、dropout 比例(E_n)、是否采用最大池化(J_n)
完全连接层	该层是否有效(A_m)、该层神经网络节点数(K_m)、是否采用批标准化(C_m)、激活函数(D_m)、dropout 比例(E_m)
优化器	优化器(L)

注:卷积层中,待优化参数 A_n 表示第 n 层是否有效; B_n 表示第 n 层卷积核(Kernel)种类; C_n 表示第 n 层是否采用批标准化(Batch Normalization); D_n 表示第 n 层激活函数(Activation); J_n 表示第 n 层是否采用最大池化(Max-Pooling),其中 $n \in [1, p]$ 。同理,完全连接层中待优化参数为 $(A_m, K_m, C_m, D_m, E_m)$, m 表示第 m 层完全连接层, $m \in [1, q]$

待优化的 C3D 网络结构由表 1 所列的参数组成,通过组合以上参数,得到一个新的 C3D 网络结构,用于后续对微表

情进行训练和评估。

2.2 C3D 网络结构参数组编码

C3D 网络结构参数组编码是对 C3D 网络结构的网络层数、每一层的节点数、每层的参数及优化器进行编码,具体参数组编码如表 2 所列。

表 2 C3D 网络结构参数组编码

Table 2 C3D network structure parameter group coding

参数标识符	参数编码
A_n 和 A_m	1 或 0
B_n	$2^i, i \in [3, \log_2 M_1]$ 且 $i \in \mathbb{N}$
C_n 和 C_m	1 或 0
D_n 和 D_m	1 或 0
E_n 和 E_m	$\frac{i}{20}, i \in [0, 10]$ 且 $i \in \mathbb{N}$
J_n	1 或 0
K_m	$2^i, i \in [4, \log_2 M_2]$ 且 $i \in \mathbb{N}$
L	$i, i \in [0, 3]$ 且 $i \in \mathbb{N}$

注:第一列中带下标的 $A-E, J-L$ 为表 2 中对应的参数标识符, M_1 为最大卷积核种类数, M_2 为最大完全连接层数, \mathbb{N} 为正整数集合

如表 2 所列,卷积层(附带池化层)的层数和完全连接层的层数均可变, $A_n=1$ 表示第 n 层卷积层有效(参与深度学习计算),此时 B_n, C_n, D_n, E_n 和 J_n 等描述本层网络结构的参数取值范围见表 2; $A_n=0$ 表示第 n 层卷积层无效(不参与深度学习计算),此时 B_n, C_n, D_n, E_n 和 J_n 等参数无意义;同理, $A_m=1$ 表示当前完全连接层有效。激活函数种类 D_n 包括 relu 和 sigmoid, 分别编码为 1 和 0; 优化器 L 的种类包括 adam, rmsprop, adagrad, adadelta, 分别编码为 0, 1, 2, 3。

2.3 算法设计

用于寻找最优 C3D 网络结构的 C3DEvol 的整个流程如图 2 所示。首先,初始化 C3D 参数组并对微表情进行预处理;其次,通过使用初始参数组的 C3D 结构对微表情进行训练,并根据训练结果求出个体的适应度值进行评估,对种群采取不同的选择、交叉、变异等操作,在不断地迭代中筛选出适应度值最高的个体,即得到最佳的 C3D 参数组及分类结果。

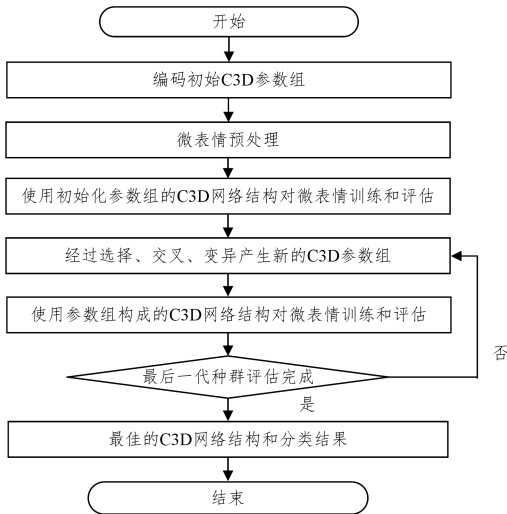


图 2 基于 C3DEvol 的微表情自动识别总流程

Fig. 2 C3DEvol based micro-expression auto-recognition process

首先初始化种群,然后根据评估产生新一代种群,新种群通过对种群中的个体进行选择、交叉、变异得到,直到最后一代种群评估完成。

(1)初始化种群。将表 1 编码后的参数组视为个体,多组这样的个体组成一个种群。第 g 代种群可以表示为:

$$H_g = \{H_{i,g}, H_{i+1,g}, \dots, H_{N,g}\} \quad (2)$$

其中, $H_{i,g}$ 为 g 代种群中第 i 个参数组个体, N 为种群大小。种群中的每个参数组个体可以表示为:

$$H_{i,g} = \{h_{i,1,g}, h_{i,2,g}, \dots, h_{i,D,g}\} \quad (3)$$

其中, $h_{i,D,g}$ 表示第 g 代种群的第 i 个参数组个体元素, D 表示该个体的维度,即待优化参数组的维度。

(2)评估种群。为了获取最佳的网络模型参数组合,对于每个个体,通过 C3D 训练得到该个体的识别准确率,再根据准确率设计适应度函数。适应度函数为:

$$F(a) = \begin{cases} \frac{a - \min(A)}{a}, & a \in A \text{ 且 } a \neq 0 \\ 0, & a = 0 \end{cases} \quad (4)$$

其中, a 为个体准确率, A 为个体所在种群中所有个体的准确率。

(3)选择操作。求出种群中每个参数组个体的适应度值,对这些适应度值由高到低进行排序,保留前 5% 的最优个体。保留的个体不参加后续的交叉操作,直接进入新一代种群的个体。

$$H_{i,g+1} = H_{i,g}, \text{ if } F(H_{i,g}) \geq F(B_{i,g}) \quad (5)$$

其中, $F(B_{i,g})$ 表示临界适应度值。前 5% 的最优个体会进入新一代种群。

(4)交叉操作。95% 的个体来自原种群交叉。对于两个待交叉个体,根据随机数产生交叉位置,根据交叉位置对两个待交叉个体进行交叉操作,交叉后进入新一代种群。进入新一代种群的个体表示为:

$$H_{i,g+1} = \{h_{i,1,g}, h_{i,2,g}, \dots, h_{i,D-n,g}, h_{j,D-n,g}, \dots, h_{j,D-1,g}, h_{j,D,g}\} \quad (6)$$

其中, j 为另一交叉个体的元素, n 为交叉的元素位置, D 表示该个体的元素个数,即待优化参数组的维度。交叉后的个体将进入新一代种群。

(5)变异操作。对所有进入新一代的个体进行变异操作,随着代数的增加,变异元素个数会逐渐增加。第 g 代有 n 个个体变异元素, n 的公式为:

$$n = \max\left(3, \left\lfloor \frac{g}{4} \right\rfloor\right) \quad (7)$$

由式(7)可以看出, n 最低取 3,元素变异范围见表 2。变异后的个体形成最终的新一代种群。

(6)新生成的种群继续循环步骤(2)一步(5),若代数 g 超过所给的最大代数或结果达到要求,则停止搜索,并输出结果。

本文实验中产生的每一代种群的个体代表一种参数选择策略,将每一种参数组解码出新的 C3D 结构参数,进而得到新的 C3D 结构。通过对该结构进行训练和分类求出结果,根据结果求出适应度值,将适应度值较高的个体送入下一代种群中。经过不断地迭代,在达到搜索停止条件后,得到适应度值最高的个体,即搜索出的最优参数选择策略。

3 实验

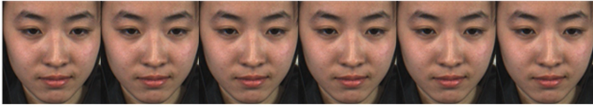
3.1 微表情数据集

目前自发的微表情数据集较少,仅有 3 个,分别是

CASME1^[24], CASME2^[25] 和 SMIC^[26]。本文实验采用中国科学院心理研究所的最新数据集 CASME2。

CASME2 是中国科学院心理研究所收集的 CASME 数据库的升级版。CASME2 包含由 200fps 相机记录的 26 个受试者的 255 个微表情视频序列。获得的微表情样本由 AU 编码, 包括 3 个部分: 起始、顶点和结束。微表情数据集可以分为 7 类: 高兴、惊讶、恐惧、悲伤、厌恶、压抑、其他。

实验中, 由于两个类的样本量不足以进行特征学习的训练, 因此不使用样本过少的视频序列类别(即恐惧和悲伤)来进行识别, 将其余的 246 个样本用于实验。本实验使用了 5 个类(包括 32 个高兴样本、63 个厌恶样本、25 个惊讶样本、27 个压抑样本和 99 个其他样本)。图 3 给出了第 17 位受试者的部分帧序列样本。



注: 第 17 位受试者(视频编号为 EP05_02)的 53—58 帧序列

图 3 CASME2 数据集的部分帧序列

Fig. 3 Partial frames sequence cropped by CASME2 dataset

3.2 实验过程

(1) 前期的预处理。微表情识别的预处理流程包括人脸检测、裁剪以及统一帧长处理环节。如图 4 所示, 在微表情预处理阶段, 首先检测人脸区域, 对每个视频序列的第一帧采用 Haar 人脸检测器^[27]进行人脸检测, 第一帧识别得到的人脸区域矩形识别框适用于剩余帧, 将整个视频序列对应的人脸区域裁剪下来。之所以将第一帧应用到剩余帧, 是因为微表情的实验是在受控环境下进行的, 头部的摆动可忽略不计, 而本文选取的是有效的表情段(即从 onset 帧到 offset 帧), 平均时间不足 $1/3$ s^[1], 对微表情识别的影响可忽略不计。接着进行规范化操作。由于帧序列的时域和空域大小不统一, 得到的裁剪帧序列可以进行规范化操作, 在时域上使用时间插值模型(Temporal Interpolation Model, TIM)^[28]进行统一帧长处理。具体来说, 通过平面线性差值的方法将每一帧的平面尺寸统一为 64×64 , 在时间序列上通过 TIM 的方法统一帧数为 63。经过这样的时域和空域的尺寸调整, 每个样本的尺寸被统一为 $64 \times 64 \times 63 \times 3$ (3 为 RGB 通道)。

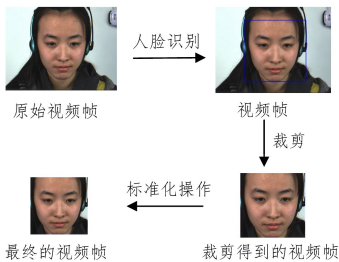


图 4 微表情预处理过程

Fig. 4 Micro-expression pre-processing process

(2) 验证方法和指标说明。本文实验均采用留一受试者交叉验证方法, 即测试集的视频序列不参与训练。具体地, CASME2 数据集的受试人数为 26, 将其中 1 人对应的视频序列用作测试集, 将其他人的视频序列用作训练集, 如此循环,

就能得到 26 个测试集的预测结果(识别准确率), 最终的结果为 26 个预测结果的平均值。

(3) 参数及配置说明。网络结构采用 C3DEvol 算法中所组合的参数。除用于优化的可变参数外, 其余公共参数有: 最大卷积层数为 6, 最大完全连接层(不包括最后的分类层)为 2, 最大卷积核种类为 64, 最大完全连接节点数为 512, 种群代数为 5, 种群大小为 5, 三维卷积核大小为 $(3 \times 3 \times 3)$, 池化层大小为 $(2 \times 2 \times 2)$, epoch 设定为 30。C3DEvol 算法用 Keras 实现。实验的主要硬件设备是两块 NVIDIA Titan X GPU。编程语言使用 python。

3.3 实验结果和分析

3.3.1 C3DEvol 方法选取的参数分布及最佳网络结构

本实验设定种群代数为 5, 种群大小为 5, 即经过 25 次训练。图 5 和图 6 分别给出了这 25 次训练中选取的激活函数概率分布和优化器概率分布。其中, 选取的激活函数概率分布是指在种群迭代训练中, C3DEvol 算法选取每个激活函数的概率大小, 从而构成整个激活函数的选取概率分布。选取的激活函数概率分布的具体计算公式为:

$$F(x) = \frac{x}{T} \quad (8)$$

其中, x 表示选取某个激活函数的训练次数; T 表示种群迭代训练次数, 这里由于设定种群代数为 5, 种群大小为 5, 因此 T 为 25。

同理, 选取的优化器概率分布是指在种群迭代训练中, C3DEvol 算法选取每个优化器的概率大小, 从而构成整个优化器的选取概率分布。选取的优化器概率分布的具体计算公式为:

$$F(y) = \frac{y}{T} \quad (9)$$

其中, y 表示选取某个激活函数的训练次数; T 表示种群迭代训练次数, 这里取值为 25。

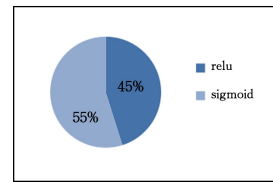


图 5 选取的激活函数概率分布

Fig. 5 Selected activation function distribution

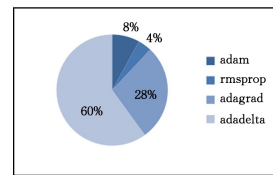


图 6 选取的优化器分布

Fig. 6 Selected optimizer distribution

从选取的激活函数来看, 大多选取 sigmoid 为激活函数, 这是因为在层数较少的情况下, 采用 sigmoid 有利于前向传播。在微表情的识别上, 由于网络层数不多, 更倾向于选择 sigmoid 激活函数。

从选取的优化器来看, C3DEvol 算法的搜索决策更倾向

于选择 adadelta,而更少地选择 adam。原因如下:1) adadelta 能够自适应调整学习率,解决在训练微表情的过程中分母累积的和越来越大而学习率变得极其小的问题,因此在学习的最后阶段,能够维持网络的更新能力;2) adam 比较适合用于大规模的数据及参数的场景,而本文的识别对象微表情本身的样本量有限,特征多样性不够,不适合运用 adam 优化器。实验中得到的最佳分类结果正是运用了 adadelta 优化器,这进一步说明了该优化器适合用于微表情识别,并且 C3DEvol 算法的搜索决策是正确的。

根据表 2 所列的参数编码,通过对每一代进行训练和评估来求出适应度值,并经过选择、交叉、变异决定进入下一代的个体,在不断优化中得到最优参数组,从而得到最佳网络结构。根据实验最终得到的最优的参数组合,形成的 C3D 网络结构具有 2 个卷积层($A_1=1, B_1=8, C_1=1, D_1=1, E_1=0.35, J_1=1; A_2=1, B_2=32, C_2=1, D_2=1, E_2=0.35, J_2=1$)和 1 个完全连接层($A_3=1, K_3=64, C_3=0, D_3=0, E_3=0.45$),优化器 L 为 3。实验使用 3.2 节所述的硬件进行 GPU 加速,每一个体的训练时间为 32 min;一次实验需要对 25 个个体进行训练,总的训练时间为 800 min。

3.3.2 与其他方法的对比

目前的微表情识别研究主要使用留一受试者交叉验证法。微表情识别使用的手工描述特征方法和深度学习方法的比较如表 3 所列。表 3 中的对比方法均是在 CASME2 数据集上进行的,并且都采用的是留一受试者的验证方法。

表 3 微表情识别方法的准确率比较

Table 3 Comparison of accuracy of micro-expression recognition

methods		
	方法	准确率/%
手工描述 特征方法	LBP-TOP ^[3]	63.41
	STLBP-IP ^[4]	59.51
	STCLQP ^[6]	58.39
	MDMO ^[7]	52.12
	Bi-WOOF ^[29]	59.26
	FDM ^[30]	41.96
	MMFL ^[31]	59.81
深度学习 方法	SDF ^[32]	47.30
	MOST ^[12]	60.98
	C3D(本文)	42.08
	C3DEvol(本文)	63.71

本文提出的方法获得了最高的准确率,为 63.71%。从表 3 可以看到,微表情识别领域更多地采用了手工描述特征方法,但识别准确率仍然偏低,并且手工识别方法需要人工提取特征,工程量大。相比手工识别方法,优化的 C3DEvol 能够自动地提取出特征,在像素层面上通过时域和空域多个维度进行卷积和池化操作来实现最后的分类,可操作性强,更加自动化。

没有使用优化方法的 C3DEvol,即表格中的 C3D,其识别效果较差,这是由于在优化参数前,模型中的参数是初始化的,是没有经过选择、交叉和变异的参数;而在经过数次迭代之后,各项参数都得到了很好的优化,得到的识别准确率有了进一步的提升,达到了 63.71%。

3.3.3 与二维卷积神经网络进化算法的对比

为了探索 C3DEvol 对网络的优化能力,本文用同种条件

下的二维卷积神经网络进化算法(C2DEvol)对微表情开展了实验。具体地,由于二维卷积神经网络的输入比三维的要小一个维度,因此在每一次迭代实验中需要将三维卷积神经网络的输入大小 $64 \times 64 \times 63 \times 3$ 转换为二维卷积神经网络的输入大小 $64 \times 64 \times 63$,即去掉颜色通道,改为灰度图像输入;对应的卷积核也改为 3×3 ,池化层大小为 2×2 ,其他参数不变。实验结果如图 7 所示。

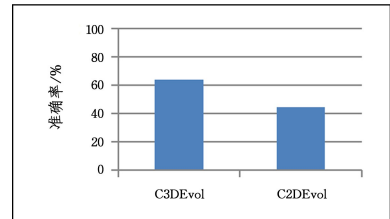


图 7 C3DEvol 和 C2DEvol 筛选出的最佳微表情识别准确率的比较

Fig. 7 Comparison of micro-expression recognition accuracy between C3DEvol and C2DEvol

由图 7 可以看出,在同样的迭代训练次数下,相比 C2DEvol, C3DEvol 筛选出的最佳识别准确率更好,这说明了三维卷积神经网络更适合于微表情识别。C2DEvol 在微表情识别上的效果较差,这是由于仅仅通过简单的堆叠帧的方式,不能有效提取时域上的特征信息,也就是无法捕获到更多微表情的微小动态变化细节。因此,尽管通过迭代训练能够筛选到更好的网络结构,但微表情识别效率还是不高。而 C3DEvol 正好解决了这个问题,能够利用进化算法筛选出最佳的三维卷积神经网络结构,进而在不断优化网络结构的过程中提高微表情识别率。

结束语 本文利用遗传算法能够全局搜索及优化的特点,提出了 C3DEvol 的微表情自动识别方法,通过对深度网络参数的编码、选择、交叉、变异等操作,筛选出最佳的网络参数组,将最优的网络参数组构成新的 C3D 网络,实现微表情的自动识别。与传统的手工方法以及现有的深度学习方法相比,所提方法明显提高了识别准确率,达到了 63.71%。

尽管本文方法能够提供一种优化策略,但微表情本身的样本量不够丰富,样本的有效特征提取受限,今后的工作应考虑增加样本的数量和多样性。另外,在优化策略的改进上,可以探索出更好的策略,以更快取得最佳的参数组合,进一步提高识别准确率。

参考文献

- [1] CORNEANU C, OLIU M, COHN J F, et al. Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-related Applications [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(8): 1548-1568.
- [2] RUSSELL T A, CHU E, PHILLIPS M L. A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool [J]. British Journal of Clinical Psychology, 2006, 45(4): 579-583.
- [3] PFISTER T, LI X, ZHAO G, et al. Recognising spontaneous facial micro-expressions[C]// 2011 IEEE International Conference

- on Computer Vision (ICCV). IEEE, 2011; 1449-1456.
- [4] WANG Y, SEE J, PHAN R C W, et al. Lbp with six intersection points; Reducing redundant information in lbp-top for micro-expression recognition [C] // Asian Conference on Computer Vision. Cham: Springer, 2014; 525-537.
- [5] WANG Y, SEE J, PHAN C W, et al. Efficient Spatio-Temporal Local Binary Patterns for Spontaneous Facial Micro-Expression Recognition [J]. Plos One, 2015, 10(5): 1-20.
- [6] HUANG X, ZHAO G, HONG X, et al. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns [J]. Neurocomputing, 2016, 175: 564-578.
- [7] LIU Y J, ZHANG J K, YAN W J, et al. A main directional mean optical flow feature for spontaneous micro-expression recognition [J]. IEEE Transactions on Affective Computing, 2016, 7(4): 299-310.
- [8] HUANG W, FAN L, HARANDI M, et al. Toward Efficient Action Recognition: Principal Backpropagation for Training Two-Stream Networks [J]. IEEE Transactions on Image Processing, 2019, 28(4): 1773-1782.
- [9] YOUNG T, HAZARIKA D, PORIA S, et al. Recent trends in deep learning based natural language processing [J]. IEEE Computational Intelligence Magazine, 2018, 13(3): 55-75.
- [10] XIONG W, WU L, ALLEVA F, et al. The Microsoft 2017 conversational speech recognition system [C] // 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018; 5934-5938.
- [11] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3d convolutional networks [C] // Proceedings of the IEEE International Conference on Computer Vision. 2015; 4489-4497.
- [12] KIM D H, BADDAR W J, JANG J, et al. Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition [J]. IEEE Transactions on Affective Computing, 2017, 10(2): 223-236.
- [13] PENG M, WANG C, CHEN T, et al. Dual temporal scale convolutional neural network for micro-expression recognition [J]. Frontiers in Psychology, 2017, 8: 1-12.
- [14] PETKE J, HARALDSSON S O, HARMAN M, et al. Genetic Improvement of Software; A Comprehensive Survey [J]. IEEE Transactions on Evolutionary Computation, 2018, 22(3): 415-432.
- [15] KIM Y H, YOON Y, GEEM Z W. A comparison study of harmony search and genetic algorithm for the max-cut problem [J]. Swarm and Evolutionary Computation, 2019, 44: 130-135.
- [16] METEVIER B, SAINI A K, SPECTOR L. Lexicase Selection Beyond Genetic Programming [M] // Genetic Programming Theory and Practice XVI. Cham: Springer, 2019; 123-136.
- [17] NGUYEN S, ZHANG M, JOHNSTON M, et al. Genetic Programming for Job Shop Scheduling [M] // Evolutionary and Swarm Intelligence Algorithms. Cham: Springer, 2019; 143-167.
- [18] SHANMUGAPRIYA K, MALAR R M S M. An Effective Technique to Track Objects with the Aid of Rough Set Theory and Evolutionary Programming [J]. Journal of Intelligent Systems, 2019, 28(1): 1-13.
- [19] CHEN Z, XIA J, BAI J, et al. Feature extraction algorithm based on evolutionary deep learning [J]. Computer Science, 2015, 42(11): 288-292.
- [20] IJJINA E P, CHALAVADI K M. Human action recognition using genetic algorithms and convolutional neural networks [J]. Pattern Recognition, 2016, 59: 199-212.
- [21] OULLETTE R, BROWNE M, HIRASAWA K. Genetic algorithm optimization of a convolutional neural network for autonomous crack detection [C] // Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No. 04TH8753). IEEE, 2004; 516-521.
- [22] RIKHTEGAR A, POOYAN M, MANZURI-SHALMANI M T. Genetic algorithm-optimised structure of convolutional neural network for face recognition applications [J]. IET Computer Vision, 2016, 10(6): 559-566.
- [23] JI S, XU W, YANG M, et al. 3D convolutional neural networks for human action recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221-231.
- [24] YAN W J, WU Q, LIU Y J, et al. CASME database; a dataset of spontaneous micro-expressions collected from neutralized faces [C] // 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). IEEE, 2013: 1-7.
- [25] YAN W J, LI X, WANG S J, et al. CASME II: An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation [J]. Plos One, 2014, 9(1): 1-8.
- [26] LI X, PFISTER T, HUANG X, et al. A spontaneous micro-expression database; Inducement, collection and baseline [C] // 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). IEEE, 2013; 1-6.
- [27] VIODA P. Rapid object detection using a boosted cascade of simple features [C] // Proc. IEEE CVPR 2001. 2001; 905-910.
- [28] ZHOU Z, ZHAO G, PIETIKÄINEN M. Towards a practical lip-reading system [C] // 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2011; 137-144.
- [29] LIONG S T, SEE J, WONG K S, et al. Less is more; Micro-expression recognition from video using apex frame [J]. Signal Processing: Image Communication, 2018, 62: 82-92.
- [30] XU F, ZHANG J, WANG J Z. Micro-expression identification and categorization using a facial dynamics map [J]. IEEE Transactions on Affective Computing, 2017, 8(2): 254-267.
- [31] HE J, HU J F, LU X, et al. Multi-task mid-level feature learning for micro-expression recognition [J]. Pattern Recognition, 2017, 66: 44-52.
- [32] PATEL D, HONG X, ZHAO G. Selective deep features for micro-expression recognition [C] // 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016; 2258-2263.



LIANG Zheng-you, born in 1968, Ph.D., professor, is a member of China Computer Federation. His main research interests include computer vision, wireless sensor networks, parallel distributed computing and artificial intelligence.