计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(1):221−227 [doi: 10.15888/j.cnki.csa.007706]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

结合 C3D 与光流法的微表情自动识别①

何景琳[1], 梁正友[1,2], 孙　宇[1], 刘德志[1]

[1](广西大学 计算机与电子信息学院, 南宁 530004)
[2](广西多媒体通信与网络技术重点实验室, 南宁 530004)
通讯作者: 梁正友, E-mail: zhyliang@gxu.edu.cn

摘　要: 由于微表情动作幅度小且持续时间短, 使其识别难度大. 针对此问题, 提出一个结合三维卷积神经网络 (3D Convolutional neural network, C3D) 和光流法的微表情识别方法. 所提出的方法先用光流法从微表情视频中提取出包含动态特征的光流图像系列, 然后将得到的光流图像系列与原始灰度图像序列一起输入到 C3D 网络, 由 C3D 进一步提取微表情在时域和空域上的特征. 在开放数据集 CASMEⅡ上进行了模拟实验, 实验表明本文所提出的方法对微表情的识别准确率达到 67.53%, 优于现有方法.

关键词: 微表情识别; 光流法; 三维卷积神经网络; 数据增强; 深度学习

引用格式: 何景琳,梁正友,孙宇,刘德志.结合 C3D 与光流法的微表情自动识别.计算机系统应用,2021,30(1):221–227. http://www.c-s-a.org.cn/1003-3254/7706.html

Automatic Recognition of Microexpression Based on C3D and Optical Flow

HE Jing-Lin[1], LIANG Zheng-You[1,2], SUN Yu[1], LIU De-Zhi[1]

[1](School of Computer and Electronics Information, Guangxi University, Nanning 530004, China)
[2](Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning 530004, China)

Abstract: It is difficult to recognize microexpression because of its small range and short duration. To solve this problem, a micro expression recognition method based on 3D Convolutional neural network (C3D) and optical flow method is proposed. We first extract a series of optical flow images with dynamic features from the microexpression video by optical flow method, then input the obtained series of optical flow images with the original gray-scale image sequences into the C3D network, and then extract the features of micro expression in the time and space domain by C3D. Simulation experiments on the open data set CASMEⅡ show that the recognition accuracy of the proposed method is 67.53%, which is better than the existing methods.

Key words: micro expression recognition; optical flow method; 3-Dimensional Convolutional neural network (C3D); data augmentation; deep learning

1　引言

微表情是指人们尽最大努力抑制真实表情的视频片段. 微表情被内心真实情绪激发所产生, 难以抑制或伪造, 所以可以更准确地反映人内心的真情实感, 能够作为测谎的重要依据, 在心理治疗、刑事审问和国家安全等领域有广阔的应用前景.

微表情的持续时间短, 仅为 1/25 ~1/3 s[1], 并且动作幅度小. 尽管能够通过人力进行识别, 但识别的准确率并不高, 经过培训的人员识别准确率不超过 50%[2]. 因此, 近年来大量研究人员提出了利用计算机视觉和机器学习算法进行微表情自动识别.

最先用于微表情识别的是基于手工描述特征的识

① 基金项目: 国家自然科学基金 (61763002)
Foundation item: National Natural Science Foundation of China (61763002)
收稿时间: 2020-05-01; 修改时间: 2020-05-27; 采用时间: 2020-06-05; csa 在线出版时间: 2020-12-31

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方法来描述自发式微表情的特征并进行识别. Wang 等[4] 提出的 LBP-SIP 方法是在 LBP-TOP 的基础上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, 这两种方法都减低了数据的冗余度. Huang 等[6] 提出了时空全局部量化模式 (STCLQP) 的微表情分析方法, 该方法在进行微表情识别时考虑了更多信息, 如信号、大小和方向因素. Huang 等[7] 提出时空域局部二值模式整体映射 STLBPIP 的方法, 在不同图像上取得水平和垂直的整体映射, 保留人脸图像的形状属性, 并使用 LBP 提取在水平和垂直方向映射上的外观和动作特征. 在 CASME Ⅱ数据集上取得 59.26% 的识别率. He 等[8] 提出了一种多任务中间 (MMFL) 特征学习, 它通过学习一组特定类的特征映射来增强提取的低级特征的辨别能力, 并使用两种加权方案, 提高了微表情识别率. Xu 等[9] 提出了用人脸动态映射 (FDM) 来描述微表情, 通过人脸标注定位方法, 对没有任何预处理的微表情进行粗对齐和人脸图像裁剪, 然后在 FDM 特征提取之前采用基于像素级别的对齐方法. 通过分类以及多种评估方法, 在微表情数据集 SMIC、CASME 和 CASME Ⅱ上的准确率分别达到 71.43%, 42.02% 和 41.96%. Liu 等[10] 在时间空间局部纹理描述符 (SLTD) 方法的基础上, 提出一个简单并有效的特征描述符——主要方向性平均光流 (MDMO) 特征, 它运用了光流估计的方法来计算人脸局部感兴趣区域 (ROIs) 的微小运动, 36 个 ROIs 仅仅需要用长度为 72 的 MDMO 特征向量表示. 另外, 他们还提出了一个光流驱动的方法来对齐微表情视频的所有帧. Liong 等[11] 在光流法的基础上提出 Bi-WOOF, 对开始帧 (onset) 和高峰帧 (apex frame) 的动作信息进行加权, 在 CASMEⅡ数据集上取得 59.26% 的识别率.

近年来, 深度学习技术在识别方面获得巨大成功, 已经广泛应用于多个领域, 如行为识别[12]、自然语言处理[13]、语音识别[14] 等方面, 也逐步被应用到微表情自动识别当中. Patel 等[15] 提出了先用深度学习模型提取出微表情的深度特征, 然后使用特征选择的方法对深度特征进行选择, 减少了特征的冗余度. Kim 等[16] 通过结合时域和空域不同维度信息的提取方法进行微表情识别, 其中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[17] 提出了一种双时间尺度卷积神经网络 (DTSCNN) 用于自发微表情识别. DTSCNN 是一种双流网络, 不同的 DTSCNN 流用于适应不同帧速率的微表情视频. 每个 DSTCNN 流由独立的浅网络组成, 以避免过度拟合问题. 同时, 还为 DTSCNN 网络提供光流序列, 以确保浅网络可以进一步获得更好的性能. Khor 等[18] 提出了一个增强的长期递归卷积网络 (ELRCN), 首先使用光流法对微表情视频序列进行预处理, 以扩大输入数据的空间维度, 然后通过 CNN 模块将每个微表情帧编码成特征向量, 然后通过将特征向量通过一个长-短期记忆 (LSTM) 模块对微表情进行预测.

尽管微表情的自动识别取得了令人印象深刻进展, 但由于微表情动作微小和持续时间短, 使得其识别准确率还不高, 有进一步提高的空间. 利用深度学习技术进行微表情识别是一种趋势. C3D[19] 是一种深度学习技术, 能够同时提取视频的时域和空域信息, 较好地表示人类活动的特性, 在行为识别、场景识别、视频相似度分析等领域得到了成功的应用. 而光流法[20,21] 对视域中的物体运动检测有非常好的效果, 已被应用到微表情自动识别中. 为充分利用 C3D 和光流法的优点, 本文提出一种结合 C3D 与光流法的微表情自动识别方法, 通过结合 C3D 和光流法技术, 能有效提取微表情的时空特征; 同时, 我们还针对微表情数据规模小、容易过拟合等问题, 采用数据增强的方法增加样本的数量, 以满足深度学习网络的要求. 我们在 CASME Ⅱ[22] 上进行了实验, 实验表明所提出的方法比现有方法有更高的识别准确率, 微表情的识别准确率达到 67.53%.

2 C3D、光流法和微表情数据集

本节介绍本文用到的两个主要技术 C3D 和光流法, 以及实验中使用的微表情数据集.

2.1 C3D

C3D 是 Ji 等[23] 提出的一种在时域和空域上的三维卷积神经网络. 使用 C3D 可以同时对外观和运动信息进行建模, 在时间和空间的特征学习、行为识别、动作相似度等各种视频分析任务上都优于 2D 卷积网络[19,23].

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, 对于第 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_{lj}^{xyz}, 公式为:

$$a_{lj}^{xyz}=f\left[b_{lj}+\sum_{n}\sum_{s=0}^{S_l-1}\sum_{t=0}^{T_l-1}\sum_{r=0}^{R_l-1}w_{lj}^{str}a_{l-1}^{(x+s)(y+t)(z+r)}\right] \quad (1)$$

其中, b_{lj}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, w_{lj}^{str}为连接上层特征映射的三维卷积核 (s,t,r) 的权重, $f(x)$ 表示激活函数.

2.2 光流法

光流法常常用于视频的动作特征提取上, 使用光流法能够很好地捕捉相邻帧的动作信息. 光流法计算在时域上前一帧与当前帧之间的图像序列中像素的变化, 得到相邻帧之间的运动信息. 在微表情自动识别中, 光流法被用于提取微表情的时域特征[10,11,18]、增大输入数据的空间维度[16] 等, 有效提高了识别率.

光流法的目的是找到图像中每个像素的速度矢量. 本文使用的 Farneback 算法[24] 是一种密集光流方法, 用于计算帧中每个点的全局密集光流. 产生的光流是与原始图像大小相等的分别表示运动方向和亮度的双通道图像. Farneback 算法的原理是运用多项式展开的方法来估计相邻两帧图像中物体的运动, 这个运动即估计物体的位移场. 多项式展开指的是, 对每个像素的领域使用一个多项式来近似建模. 本文只对二次多项式展开变换, 对位置为 x 的每一个像素, 利用二次多项式构造一个局部信号模型, 表示为:

$$I(x)\sim x^{\mathrm{T}}Ax+b^{\mathrm{T}}x+c \quad (2)$$

其中, A 是对称矩阵, b 是向量, c 是标量. 这些系数使用加权的最小二乘法拟合领域中信号的值.

以下是在理想的情形下位移的估计过程. 对于第一个图像, 构造一个局部信号, 考虑的二次多项式:

$$I_1(x)=x^{\mathrm{T}}A_1x+b_1{}^{\mathrm{T}}x+c_1 \quad (3)$$

在经历一个全局的位移 d(不随空间变化, 恒定的方向和大小), 在第 2 幅图像上, 构造一个新的信号:

$$\begin{aligned}I_2(x)&=I_1(x-d)\\&=(x-d)^{\mathrm{T}}A_1(x-d)+b_1^{\mathrm{T}}(x-d)+c_1\\&=x^{\mathrm{T}}A_1x+(b_1-2A_1d)^{\mathrm{T}}x+d^{\mathrm{T}}A_1d-b_1^{\mathrm{T}}d+c_1\\&=x^{\mathrm{T}}A_2x+b_2^{\mathrm{T}}x+c_2\end{aligned} \quad (4)$$

然后, 通过观察, 可以对应得到:

$$A_2=A_1 \quad (5)$$

$$b_2=b_1-2A_1d \quad (6)$$

$$c_2=d^{\mathrm{T}}A_1d-b_1^{\mathrm{T}}d+c_1 \quad (7)$$

根据式 (6), A_1是非奇异矩阵的情况下, 可以计算全局位移 d:

$$2A_1d=-(b_2-b_1) \quad (8)$$

$$d=-\frac{1}{2}A_1^{-1}(b_2-b_1) \quad (9)$$

值得注意的是, 在任何维度下, 以上公式均是成立的.

而在实际情况中, 我们使用一个不随空间变化的位移 d, 使用单一个多项式拟合函数来研究两个图像的关系, 是不切实际的. 因此我们定义第 1 个图像随空间变化的参数$A_1(x)$, $b_1(x)$和c_1以及第 2 幅图像参数$A_2(x)$, $b_2(x)$和c_1, 由式 (5) 可以得到:

$$A(x)=\frac{A_1(x)+A_2(x)}{2} \quad (10)$$

再根据式 (6), 令:

$$\Delta b(x)=-\frac{1}{2}(b_2(x)-b_1(x)) \quad (11)$$

得到最主要的约束:

$$A(x)d(x)=\Delta b(x) \quad (12)$$

其中, $d(x)$说明已经使用一个随空间位置发生变化的位移场来代替方程 4 中恒定大小和方向的全局位移 d.

随后进行邻域估计. 假设位移场仅缓慢变化, 从而可以集成每个像素的邻域上的信息. 因此, 我们试图找到$d(x)$满足式 (12), 并且尽可能地超出邻域 P 的 x, 邻域估计表示为:

$$\sum_{\Delta x\in P}w(\Delta x)\|A(x+\Delta x)d(x)-\Delta b(x+\Delta x)\|^2 \quad (13)$$

其中, $w(\Delta x)$是邻域中的点的权重函数.

2.3 微表情数据集

目前自发的微表情数据集较少, 仅有的 3 个数据集分别是 SMIC[25], CASME[26], CASME Ⅱ[22]. 本文全部实验采用 CASME Ⅱ数据集.

CASME Ⅱ是中国科学院心理研究所收集的 CASME 数据库的升级版本. CASME Ⅱ包含由 200 fps 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 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、压抑等.

在我们的实验中, 由于恐惧和悲伤两个类的样本量分别是 2 个和 7 个, 不足以进行特征学习的训练. 因

此我们将它们排除在实验之外, 即其余的 246 个样本用于实验. 在我们的实验中使用了 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 个其它样本).

3 本文提出的方法

本文所提出的方法流程主要分为预处理, 特征提取和分类 3 个步骤, 过程如图 1: 首先经过预处理得到标准化的微表情视频. 为了捕获人脸表情动态信息, 通过光流法逐帧计算得到两通道的包含动态信息的特征序列. 然后将原始图像的灰度图作为一个通道的特征, 与光流的两个通道组合成 3 个通道的特征序列, 将这样的三通道的特征序列输入到 C3D, 由 C3D 的卷积层、池化层和完全连接层自动提取高级特征. 最后, 由 C3D 的最后一层全连接层计算出每类的预测概率, 实现对微表情的分类.

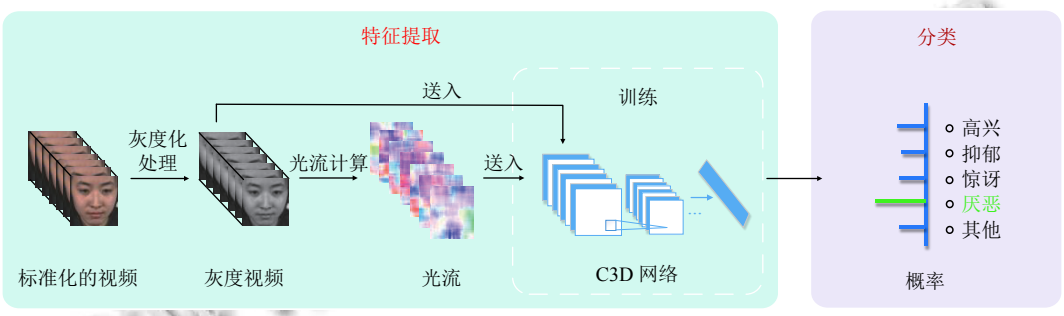


图 1 本文所提出的微表情识别方法流程

3.1 微表情视频预处理

3.1.1 人脸检测和裁剪

由于原始帧中含有无关的背景, 因此需要定位边界框以移除背景, 并从原始图像中保留面部区域. 首先, 通过 OpenCV 中的 Dlib 库实现的 68 点检测算法[27] 检测微表情视频片段第一帧中的面部区域, 并根据第一帧剪裁该视频片段的其余帧. 此外, 为了准确地在原始帧中裁剪面部区域, 从 68 点中选出特定周围点 (即最左侧点、最顶部点、最右侧点和最下侧点) 以形成面部区域周围的边界框, 如图 2 所示, 通过裁剪周围点得到最贴合脸部的人脸区域.

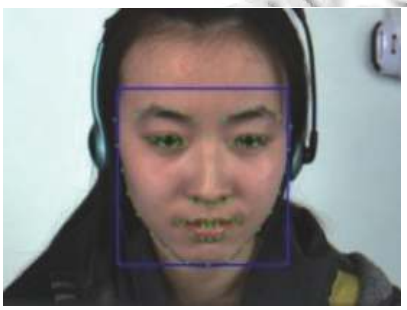


图 2 使用 68 点人脸检测算法

3.1.2 标准化处理

由于帧序列的时域和空域大小不统一, 得到的裁剪帧序列需要进行标准化操作. 具体来说, 在时域上使用时间插值模型 (TIM)[28] 统一了微表情视频的帧长. 例如, 在时域上通过 TIM 的方法统一帧数为 96, 在空域上使用平面线性插值的方法将每一帧的平面尺寸统一为 96×96. 经过这样的时域和空域上的大小尺寸调整, 每个样本大小尺寸统一为 96×96×96×3 (3 为 RGB 通道), 如图 3 所示.

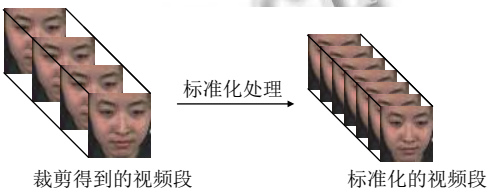


图 3 微表情视频的标准化处理

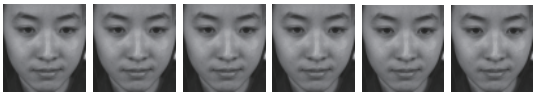
3.1.3 灰度化处理

微表情视频的灰度化处理. 由 RGB 三通道的图片序列经过灰度化得到一通道的灰度图片序列. 如图 4 所示, 图中上方为原始微表情 RGB 图片序列, 下方是对应灰度化的图片序列.

3.1.4 数据增强

深度学习常常需要大量的数据进行学习, 在样本量较少的时候, 一般采用数据增强策略; 仿射变换就是其中之一. 仿射变换一般包括平移、翻转、旋转、缩放. 这些方法对于卷积神经网络提取特征具有不变性[29], 并且广泛用于各个领域的深度学习[17,29].

(a) 原始图



(b) 灰度

图 4 微表情的原始图及灰度图序列

本文用仿射变换进行数据增强. 所的方法包括平移、翻转和旋转 3 类. 具体地, 对于水平翻转、垂直翻转; 旋转角度 90°、旋转角度 180°、使用旋转角度 270°, 单种方式有 5 种, 同类仿射变换不混合的混合方式有 6 种, 水平翻转加垂直翻转本身及和旋转三个角度混合有 4 种, 共 15 种仿射变换.

3.2 光流特征提取

光流特征提取是对微表情视频进行光流估计, 提取微表情视频的低级特征. 对于微表情识别的视频序列, 用光流法计算面部区域的微小移动, 计算工具使用 OpenCV 库[24], 得到与原始图像大小相等的双通道图像, 双通道分别表示强度和方向. 为了更直观地可视化光流, 可以使用 Munsell 颜色系统[30] 将强度和方向矩阵转换为可视化图像, 使用该颜色系统的微表情光流分布如图 5 所示. 得到的两个通道的光流估计序列视为低级特征, 和一个通道的灰度图像序列合并为 3 个通道图像序列, 共同输入 C3D 的输入层中, C3D 将自动提取出时域和空域上的特征并进行最后的分类.

(a) 裁剪的图像

(b) 光流图

图 5 使用 Munsell 颜色系统表示的微表情光流分布

3.3 C3D 网络结构设计

在深度学习中, 一个好的神经网络结构在训练过程中起着重要作用, 良好的结构模型可以有效地提取特征. 因此, C3D 结构上的参数需要适当地设置, 包括层数、每层的核种类等. 本文所提出的 C3D 结构如表 1 所示. 其中, 网络结构的卷积层核是 (3×3×3). 表 1 中: (1) Conv, Pool 和 F 分别是卷积层, 最大池化层和完全连接层的缩写. (2) 第一层 Conv1 的输入大小为 (96×96×95×3), 其中 95 表示一个视频在时域上的大小, (96×96) 表示一个视频在空域上的大小, 3 表示输入视频的通道, 包括两个光流通道和一个上一帧灰度帧. (3) 最后一层 F2 用于分类, 本文中的微表情类数为 5, 因此输出大小为 (5×1). 这个网络结构设计借鉴了经典的深度学习网络 VGG[31](Visual Geometry Group) 的优点, VGG 在图像处理方面表现出良好的性能. 它具有 3 个优点: (1) VGG 使用多个小卷积核而不使用过多大卷积核, 一方面可以在获得相同大小的特征映射的情况下捕获到更多的空间上下文信息, 但使用较小的卷积核时, 使用的参数和计算量较少. 另一方面, 因为使用更小的核, 意味着要使用更多的滤波器, 即能够使用更多的激活函数, 从而使卷积网络学习到更具区分性的映射函数. (2) 层组的设计. 两个卷积层后面紧接着一个池化层, 其中前两个卷积层更注重局部特征, 适合于需要提取微小局部信息的微表情识别. (3) 第 2 组层组比第 1 组层组多一个卷积层, 这意味着可以在第 1 组层组的基础上进一步细化特征的学习.

表 1 C3D 网络结构参数

网络层	输入大小	核(kernel)	stride	Pad	本层输出大小
Input	96×96×95×3	–	–	–	96×96×95×3
Conv1	96×96×95×3	3×3×3	1	0	94×94×94×16
Pool1	94×94×94×16	2×2×2	2	0	47×47×47×16
Conv2a	47×47×47×16	7×7×7	1	0	41×41×41×32
Conv2b	41×41×41×32	7×7×7	1	0	35×35×35×32
Pool2	35×35×35×32	2×2×2	2	0	17×17×17×32
Conv3a	17×17×17×32	3×3×3	1	0	15×15×15×64
Conv3b	15×15×15×64	3×3×3	1	0	13×13×13×64
Pool3	13×13×13×64	2×2×2	2	0	6×6×6×64
F1	96×96×95×3	–	–	–	128×1
F2	96×96×95×3	–	–	–	5×1

4 实验

4.1 实验参数说明

网络模型的学习过程由 Keras 编码实现. 参数是经反复试验决定的. 本文中, 初始学习率设置为 0.01. 对于 C3D, 训练 epoch 设定为 160. 对于损失函数, 使用均方误差 (Mean Square Error, MSE), 均方误差损失函数是使用最广泛的函数, 并且在大部分情况下, 均方误差有着不错的性能, 因此被用作损失函数的基本衡量指标. 实验的主要硬件设备是两块 NVIDIA Titan X GPU, 编程语言使用 Python.

4.2 实验结果和分析

在本节中, 我们通过评估本文提出的方法在 CASME

II 数据集的分类准确率, 并与其他现有方法进行了比较, 包括现有的手工描述特征方法和深度学习方法.

4.2.1 和现有方法的对比

由于留一受试者交叉验证方法 (LOSO) 能防止学习过程中的主体偏差[15]; 因此, 我们的实验采用 LOSO 交叉验证法. 在此情景下, 我们将本文所提出方法和其他现有的方法进行比较, 包括手工描述特征方法和深度学习方法.

所提出方法的识别准确率比较如表 2 所示. 如表中所示, 所提出的方法优于其他现有方法. 与手工描述特征的方法相比, 深度学习方法通过调整参数和权重, 能够自动学习特征并在训练期间优化模型. 深度学习方法尽管更依赖于训练样本的数量, 但这个问题可以通过数据增强来解决, 通过逐层学习样本, 获取到深层次的特征. 如表 2 所示的基于深度学习的方法总体比手工描述特征方法表现更好. 特别地, 本文所提出的方法结果比手工描述特征方法中的最佳方法高约 4%, 这表明本文所提出的方法作为一种深度学习方法, 能够自动提取特征, 省去了人工寻找特征的步骤, 也提高了识别准确率.

表 2 本文提出的方法与现有方法的微表情识别准确率比较

方法类别	方法名称	准确率 (%)
手工描述特征方法	LBP-TOP[22]	63.41
	STLBP-IP[7]	59.51
	STCLQP[6]	58.39
	MDMO[10]	52.12
	FDM[9]	41.96
	MMFL[8]	59.81
	Bi-WOOF[11]	59.26
深度学习方法	SDF[15]	47.30
	文献[16]提出的方法	60.98
	C3D(本文)	61.34
	本文所提出的方法	67.53

4.2.2 光流法对微表情识别的影响分析

本文所提出的方法是通过计算光流获取低级特征, 在时间维度上提取相邻帧上的强度和方向的特征, 以便捕获更多的动态信息, 然后对 C3D 进行训练, 提取高级特征, 实现微表情的自动识别.

从表 2 可以看到, 本文所提出的方法比 C3D 的识别准确率高了 6.19%. 即光流法贡献了 6.19% 的识别正确率. 原因是微表情视频是一个动态时域上出现动作变化的视频, 光流法能通过计算出微表情的微小运动的大小和方向, 逐帧地提取微表情的动态特征, 捕捉到更多的动作信息, 从而提高识别准确率.

5 结束语

本文提出了结合 C3D 与光流法的微表情自动识别方法, 通过光流法逐帧提取微表情的动态信息, 得到的光流序列和原始灰度序列输入 C3D 网络, 通过 C3D 提取时域和空域上的特征, 同时捕捉微表情的动态信息. 实验中, 为了满足大量的深度学习数据训练需要, 采用数据增强策略, 扩大了微表情数据规模, 防止深度学习网络容易过拟合. 在开放的微表情数据集 CASME II 上进行了模拟实验, 实验表明所提出的方法提高了微表情识别准确率, 准确率达到 67.53%.

参考文献

1 Corneanu CA, Simón MO, Cohn JF, *et al*. Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(8): 1548 –1568. [doi: 10.1109/TPAMI.2016.2515606]

2 Russell TA, Chu E, Phillips ML. A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool. British Journal of Clinical Psychology, 2006, 45(4): 579–583. [doi: 10.1348/014466505X90866]

3 Pfister T, Li XB, Zhao GY, *et al*. Recognising spontaneous facial micro-expressions. Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain. 2011. 1449–1456.

4 Wang YD, See J, Phan RCW, *et al*. LBP with six intersection points: Reducing redundant information in LBP-TOP for micro-expression recognition. Proceedings of the 12th Asian Conference on Computer Vision. Singapore. 2015. 525–537.

5 Wang YD, See J, Phan RCW, *et al*. Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition. PLoS One, 2015, 10(5): e0124674. [doi: 10.1371/journal.pone.0124674]

6 Huang XH, Zhao GY, Hong XP, *et al*. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. Neurocomputing, 2016, 175: 564–578. [doi: 10.1016/j.neucom.2015.10.096]

7 Huang XH, Wang SJ, Zhao GY, *et al*. Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection. Proceedings of the IEEE International Conference on Computer Vision Workshop. Santiago, Chile. 2015. 1–9.

8 He JC, Hu JF, Lu X, *et al*. Multi-task mid-level feature

learning for micro-expression recognition. Pattern Recognition, 2017, 66: 44–52. [doi: 10.1016/j.patcog.2016.11.029]
9 Xu F, Zhang JP, Wang JZ. Microexpression identification and categorization using a facial dynamics map. IEEE Transactions on Affective Computing, 2017, 8(2): 254–267. [doi: 10.1109/TAFFC.2016.2518162]
10 Liu YJ, Zhang JK, Yan WJ, *et al*. A main directional mean optical flow feature for spontaneous micro-expression recognition. IEEE Transactions on Affective Computing, 2016, 7(4): 299–310. [doi: 10.1109/TAFFC.2015.2485205]
11 Liong ST, See J, Wong K, *et al*. Less is more: Micro-expression recognition from video using apex frame. Signal Processing: Image Communication, 2018, 62: 82–92. [doi: 10.1016/j.image.2017.11.006]
12 Huang WB, Fan LJ, Harandi M, *et al*. Toward efficient action recognition: Principal backpropagation for training two-stream networks. IEEE Transactions on Image Processing, 2019, 28(4): 1773–1782. [doi: 10.1109/TIP.2018.2877936]
13 Young T, Hazarika D, Poria S, *et al*. Recent trends in deep learning based natural language processing. IEEE Computational Intelligence Magazine, 2018, 13(3): 55–75. [doi: 10.1109/MCI.2018.2840738]
14 Xiong W, Wu L, Alleva F, *et al*. The Microsoft 2017 conversational speech recognition system. Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. Calgary, AB, Canada. 2018. 5934–5938.
15 Patel D, Hong XP, Zhao GY. Selective deep features for micro-expression recognition. Proceedings of 2016 23rd International Conference on Pattern Recognition. Cancun, Mexico. 2016. 2258–2263.
16 Kim DH, Baddar WJ, Ro YM. Micro-expression recognition with expression-state constrained spatio-temporal feature representations. Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam, the Netherlands. 2016. 382–386.
17 Peng M, Wang CY, Chen T, *et al*. Dual temporal scale convolutional neural network for micro-expression recognition. Frontiers in Psychology, 2017, 8: 1745. [doi: 10.3389/fpsyg.2017.01745]
18 Khor HQ, See J, Phan RCW, *et al*. Enriched long-term recurrent convolutional network for facial micro-expression recognition. Proceedings of 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition. Xi'an, China. 2018. 667–674.
19 Tran D, Bourdev L, Fergus R, *et al*. Learning spatiotemporal features with 3D convolutional networks. Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 4489–4497.
20 Barron JL, Fleet DJ, Beauchemin SS. Performance of optical flow techniques. International Journal of Computer Vision, 1994, 12(1): 43–77. [doi: 10.1007/BF01420984]
21 Negahdaripour S. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(9): 961–979. [doi: 10.1109/34.713362]
22 Yan WJ, Li XB, Wang SJ, *et al*. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PLoS One, 2014, 9(1): e86041. [doi: 10.1371/journal.pone.0086041]
23 Ji SW, Xu W, Yang M, *et al*. 3D convolutional neural networks for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221–231. [doi: 10.1109/TPAMI.2012.59]
24 Farnebäck G. Two-frame motion estimation based on polynomial expansion. Proceedings of the 13th Scandinavian Conference on Image Analysis. Halmstad, Sweden. 2003. 363–370.
25 Li XB, Pfister T, Huang XH, *et al*. A spontaneous micro-expression database: Inducement, collection and baseline. Proceedings of 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). Shanghai, China. 2013. 1–6.
26 Yan WJ, Wu Q, Liu YJ, *et al*. CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. Proceedings of 2013 10th IEEE International Conference and Workshops on Automatic face and Gesture Recognition (FG). Shanghai, China. 2013. 1–7.
27 Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 1867–1874.
28 Zhou ZH, Zhao GY, Pietikäinen M. Towards a practical lipreading system. Proceedings of CVPR 2011. Providence, RI, USA. 2011. 137–144.
29 Zhong Z, Zheng L, Zheng ZD, *et al*. CamStyle: A novel data augmentation method for person re-identification. IEEE Transactions on Image Processing, 2019, 28(3): 1176–1190. [doi: 10.1109/TIP.2018.2874313]
30 Gargi U, Kasturi R, Strayer SH. Performance characterization of video-shot-change detection methods. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(1): 1–13. [doi: 10.1109/76.825852]
31 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. The 3rd International Conference on Learning Representations. San Diego, CA, USA. 2015. 1–5.