文章编号：1002-1582(2020)06-0712-09

# 利用双流卷积神经网络的人脸表情识别方法

翟海庆[1,3]*，刘丹[1,3]，刘峻[1,2]
(1. 河南工学院 计算机科学与技术学院，河南 新 乡 453003)
(2. 武汉理工大学 计算机科学与技术学院，湖北 武汉 430063)
(3. 河南省生产制造物联大数据工程技术研究中心，河南 新乡 453003)

**摘　要**：近年来，人脸表情识别(FER)方法已经取得比较好的识别准确度，但实际环境中由于姿态、遮挡、光照等因素，会对其检测准确度有不小的减弱效果。针对这些问题，提出了一种新的基于双流卷积神经网络(CNN)的FER算法。从外观和几何特征差异两方面入手，建立双流CNN，基于外观特征的网络是提取预处理后图像的局部方向模式(LDP)特征作为该网络的输入，而基于几何特征的网络主要是基于动作单元(AUs)标志点的坐标变化，AUs标志点主要是标志面部做表情时运动肌肉的位置。此外，利用了一种自动编码器技术生成具有中性情绪的面部图像的技术。算法在CK+和JAFFE数据集上进行了验证，检测准确度分别为98.81%和96.05%，与其他最新方法比较均显示出更好的效果。

**关　键　词**：人脸表情识别；双流卷积神经网络；LDP特征；几何特征；深度学习
**中图分类号**：TP394.1；TH691.9　　**文献标识码**：A
DOI:10.13741/j.cnki.11-1879/o4.2020.06.014

# A facial expression recognition method based on dual-stream convolutional neural network

ZHAI Haiqing[1,3]*, LIU Dan[1,3], LIU Jun[1,2]
(1. School of Computer Science and Technology, Henan Institute of Technology, Xinxiang453003, China)
(2. School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430063, China)
(3. Big Data Engineering Research Center of Henan for Production and Manufacturing IoTs, Xinxiang 453003, China)

**Abstract**: In recent years, the research of facial expression recognition has achieved good recognition accuracy, but in the actual environment, due to the influence of posture, occlusion, light and other factors, its detection accuracy has a little weakening effect. To solve these problems, a new FER system based on the dual-stream convolutional networks is proposed. A dual stream CNN from two aspects of appearance and geometric characteristics is established . The network based on appearance features is to extract the local derivative pattern features of the preprocessed image as the input, whereas the geometric feature-based network learns the coordinate change of action units' landmark, which is a muscle that moves mainly when making facial expressions. In addition, a technique to generate facial images with neutral emotion using the autoencoder technique is be proposed. By this technique, the dynamic facial features between the neutral and emotional can be extracted images without sequence data. The detection accuracy of the algorithm is 98.81% and 96.05% respectively on CK+ and Jaffe datasets, which shows better results compared with other latest methods.

**Key words**: facial expression recognition; Dual-stream convolutional neural network; LDP features; geometric characteristics; deep learning

**收稿日期**：2020-06-28；**收到修改稿日期**：2020-09-01
**基金项目**：国家自然科学基金项目(61802116)；河南省科技厅科技计划项目(192102210248)
**作者简介**：翟海庆(1979—)，男，讲师，硕士，从事人工智能、图像处理等研究。
刘丹(1978—)，女，副教授，硕士，从事人工智能、数据分析等研究。
* **通讯作者**：zhaihaiqingtwo@163.com

## 0　引　言

传统技术领域中，基于感知的通信技术在人类交互中起主要作用。由于人工智能技术的进步，人工智能机器人可以在现实生活与人类进行互动，但是，为了更准确地与人类的互动，需要安装接收人类感官发出信息的装置。在与人类进行信息交互时，大部分信息由眼睛接收，因此，视觉信息十分重要。在利用人与机器之间交互的人工智能机器人中，人脸提供了众多可以了解用户当前状态的重要信息。因此，在之前的一段时间中，很多专家学者前赴后继，在面部表情识别（Facial Expression Recognition，FER）领域进行了广泛的研究。

近年来，随着相关数据的不断增加和深度学习的不断发展，一种能够在各种环境下准确识别面部表情的技术得到了积极深入的研究。基于普遍性、相似性、生理学和进化特性，FER 研究中的情绪可以分为六类：快乐、悲伤、恐惧、厌恶、惊讶和愤怒。此外，加上中性情绪，情绪可以分为七类。

实现基于表情识别的 FERs 需要四个步骤。首先，需要一个人脸检测步骤来定位人脸。典型的算法包括级联形状回归[1]、Haarcascade[2]等。第二步涉及面部配准，通过该步骤获得主要特征点，以识别面部旋转或肌肉运动。由于存在各种角度的光照和旋转的可能，检测后的人脸在识别准确度方面偏低。因此，有必要通过获取标志点来改善识别准确度，标志点是做面部表情时主要肌肉运动的位置，一般将其称为动作单元（Action Units，AUs），主要位置包括眉毛、眼睛、鼻子和嘴。检测标志点的一个典型算法是主动外观模型（Active Appearance Models，AAM）[3]。第三，在特征提取步骤中，通过获取特征点的运动或位置信息，提取出能够识别面部表情的特征。最后，基于获得的特征，选取合适的分类器，例如支持向量机（Support Vector Machine，SVM）和隐马尔可夫模型（hidden Markov Model，HMM）[4,5]。

近年来，由于机器视觉的发展和硬件技术的进步，诞生了许多基于深度学习的人脸表情识别算法。例如，文献[6]利用多任务卷积神经网络（Multi-Task Convolutional Neural Network，MTCNN）模型进行人脸检测后，结合 Inception 的思想提出一种新的卷积神经网络模型，使用 1×1 卷积核对特征维数进行缩减，增加并平衡网络深度和宽度的同时不增加额外的计算负担，更加精准的对人脸特征进行提取。文献[7]在局部方向强度模式（Local Directional Strength Pattern，LDSP）特征的基础上加入局部方向直方图模式（Local Directional Rank Histogram Pattern，LDRHP）特征，然后进行核主成分分析和广义判别分析，得到更为鲁棒的特征。最后，采用卷积神经网络对人脸特征进行训练，实现了人脸表情识别。在文献[8]中，研究人员使用了双重网络的 FER 方法，该方法融合了人脸的整体特征和侧重于人脸标志的部分特征。在文献[9]中，研究人员设计了一种基于端到端的低质人脸表情识别方法，并利用生成对抗网络修补破损的区域，通过分类器建立分类约束判别损失，以达到更好的分类效果。文献[6]、[8]、[9]只是训练出一个新的神经网络，并没有具体分析新的神经网络达到较好效果的原因，可解释性差。文献[7]只利用了外观特征，有一定的局限性。

尽管前人已经进行了许多研究，但由于光照、遮挡等各种环境变化的影响以及个体特征的差异，大部分文献只从外观特征或几何特征一种特征入手[10,11]，人脸表情识别率仍然有待进一步提高。因此，本文提出了一种有效的人脸表情识别方法——基于双流 CNN，将外观特征和几何特征进行融合，还考虑了第二高情感（Top-2）预测结果的误差，从而获得更精确的结果。

## 1　提出的算法

所提出的算法如图 1 所示。利用双流 CNN 来提升辨识准确度，该结构可以对最容易发生错误的结果（Top-2 错误情绪）重新分类。第一个 CNN 利用 LDP 特征对 AUs 进行训练，其中 LDP 特征是人脸研究领域中典型的特征提取技术。第二个网络提取每个区域标志的几何变化，并对所有的六种情绪进行训练。基于外观和几何差异这两个特征，本文提出了一种加权函数的混合算法。

### 1.1　预处理

在对人脸进行表情识别这一过程之前，首先要定位到人脸的位置，并将人脸位置区域划分出来。因此，人脸和非人脸部分必须通过人脸检测过程进行分离。只有保留情感识别的重要部分，才能防止由于脸部周围环境的变化而导致的准确性下降等问题。受文献[12]的启发，采用 HOG 算法确定人脸边界坐标的检测器，利用该检测器分割出面部。利用线性 SVM 对由滑动窗口组成的正（含一个目标）和负（不含目标）样本进行 HOG 特征训练来识别人脸区域。裁剪后的面部图像通常是从额头中部到下巴，从最左边的脸到最右边的脸。

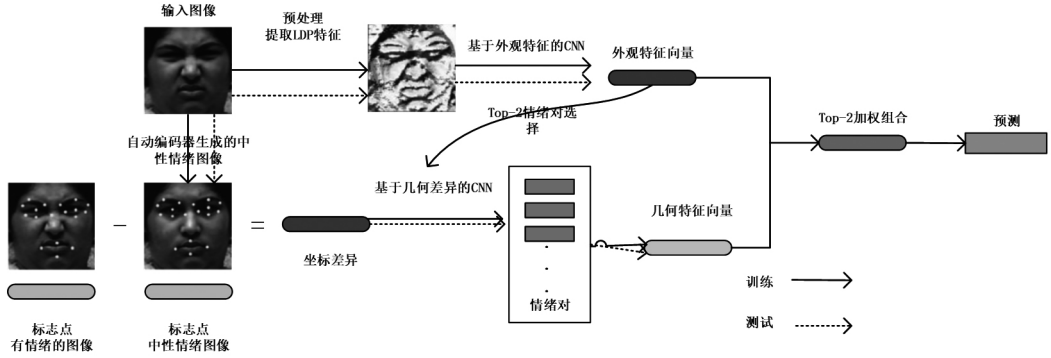



图 1 本文算法流程

面部区域被裁剪后，在创建输入特征之前需进行模糊处理以去除噪声，去噪后的 LDP 图像是基于外观特征的网络的输入。如果从未经处理的面部图像中提取特征，则可能导致性能下降。另外，由于 AUs 在情绪变化中起着重要作用，其重要性高于其他人脸部位，因此在预处理后将其转换为 LDP 图像。

图像预处理结果如图 2 所示。利用传统高斯滤波技术后，图像边缘会模糊，利用双边滤波对图像进行预处理[13]，如图 2(a)，相比图 2(b)中没有进行预处理的图像，可以有效解决边缘模糊的问题。

(a) 进行双边滤波预处理的图像 (b) 没有预处理的图像

图 2 预处理的图像

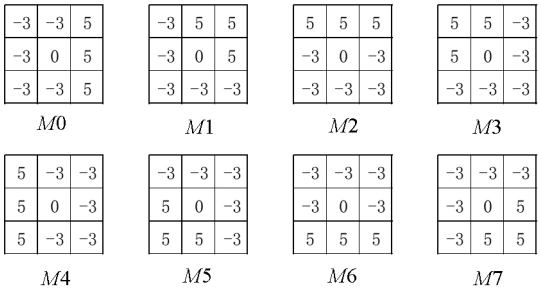
### 1.2 基于外观特征的网络

一般可以通过几何特征和纹理特征来提取图像的局部特征信息，典型的算法有：LBP、LDP 等。2010 年，Jabid 等提出 LDP 算法，已有较多应用[14]，其与 LBP 算法较为相似，保留了各项优点的同时对其缺点也进行了改进。例如一致性光照变化对 LBP 算法影响不大，但如果附加了随机噪声或者光照变化是非一致性的图像，并不能很好的提取特征，而 LDP 算法正是改进了这些地方。所以本文使用 LDP 算法提取外观特征。

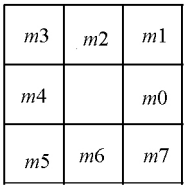
对于某个中心像素点，该算子不是直接比较中心像素点和邻域像素点的灰度差，而是通过将其与 8 个 Kirsch 模板进行卷积计算，得到 8 个方向的边缘响应值，然后把绝对值较大的前 $K$ 个边缘响应值所对应的二进制位设置为 1，剩余 8-$K$ 个二进制位设置为 0。Kirsch 算子以及将图像经过该算子计算得出的 8 个方向的边缘响应值模板如图 3 所示。该算子具体计算方法为

$$\mathrm{LDP}(x,y)=\sum_{i=0}^{7}S(m_i-m_k)2^i$$
$$S(m_i-m_k)=\begin{cases}1, & m_i\geqslant m_k\\ 0, & m_i<m_k\end{cases} \tag{1}$$

式中，$m_k$ 表示第 $k$ 个最大的边缘响应值，在计算中一般令 $k=3$。




(a) 8 个方向的 Kirsch 模板




(b) 8 个方向的边缘响应值

图 3 8 个方向的 Kirsch 算子和各个方向上的边缘响应值

CNN 已经广泛地应用于计算机视觉领域。CNN 一般都会有卷积层和池化层这两大核心部分，这两大部分通过对图像不断的进行卷积和池化运算，可以提取到图像中不同尺度的抽象特征。CNN 可以将人脸的重要信息保留，同时提取出有用的人脸表情特征。

如图 4 所示，基于外观特征网络的输入是大小为 128×128 的图像，并且总共经历三遍卷积层和池化层。第一层卷积使用 5×5 大小的核执行卷积运算。接下来，在第一个池化层中，执行 max-pooling，即在 2×2 像素块中只保留卷积值最大的像素。考虑到 3×3 大小的卷积核在 64×64 像素输入图像中广泛应用，实验确定卷积核大小为 5×5，以适应

128×128输入图像的大小。结果，在上一步中获得的 64 个 128×128 尺寸的图像将更改为 64 个 64×64 尺寸的图像，这表示尺寸减小了一半。

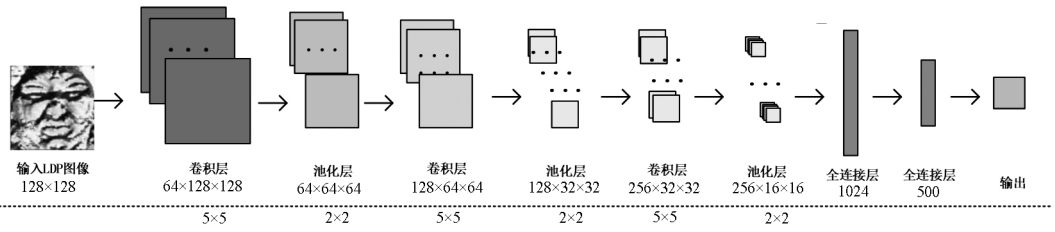



图 4　本文基于外观特征的 CNN

以类似的方式，将卷积和池化操作重复 3 次。经过最后一个池化层得到了 256 张 16×16 尺寸的图像。卷积和池化完成后，这些值将平铺并通过全连接层（隐藏层）传递。第一个全连接层有 1024 个节点，而第二个全连接层有 500 个节点。

在所提出的网络中，在全连接层之间使用 dropout 操作。当网络训练时，它会随机关闭神经元节点，这可以防止网络过拟合。ReLU（Rectified Linear Unites）函数可以通过卷积运算将特征映射的量化值转换为非线性值，使用该函数作为卷积层和全连接层之间的激活函数。在网络的最后，使用 softmax 函数将六种情绪提取为连续值。六种情绪的 softmax 提取结果表示为

$$s_i = \frac{e^{a_i}}{\sum_{k=0}^{n-1} e^{a_k}} \tag{2}$$

式中，$n$ 对应于需要分类的情绪数量；$s_i$ 是第 $i$ 类的 softmax 函数值。这个值是第 $i$ 类情感得分的 e 的指数值除以 $a_k$ 的 e 的指数之和，即所有类别的值。整个卷积过程结束之后，利用交叉熵损失函数减小网络输出与真实值误差，更新网络参数。损失函数表示为

$$\text{Lost} = -\sum_{j=0}^{n-1} y_j \log(s_j) \tag{3}$$

式中，$y_j$ 是正确分类向量的第 $j$ 个元素。使用交叉熵函数，可以通过负对数似然法获得交叉熵，从而灵活地响应模型的各种概率分布。另外，寻找梯度的过程也相对简单。

若要对六个类别（即 $n$）进行分类，如果第一个元素正确，则 $y=[1,0,0,0,0,0]$，$y_1=1$，其他元素为 0。$s_j$ 是 softmax 函数的输出值。此外，使用随机梯度下降（Stochastic Gradient Descent，SGD）作为优化器以计算交叉熵损失。利用该模型提取的 softmax 结果中情感值最高的两种结果，与几何特征的结果相融合，用于更精确的情感检测。因此，这两种最高情感的信息被传送到基于几何特征的网络中。

### 1.3　基于几何特征的网络

如果仅使用外观特征网络，当遇到光照，外围环境变化等因素时，会大大降低结果的识别准确度。所以在此使用基于几何特征的 CNN 网络来捕获情绪标志点的位置及其运动。通过检测标志点的移动将获得的部分元素特征添加到总体特征中去，这样特征信息的提取会更全面。最后通过 softmax 层，挑选出 Top-2 的情绪值，进行加权计算，将两个网络的结果融合成最终结果对情绪进行分类。

需要利用人的中性面部图像求取与情绪图像的差异，但是实际的系统中没有足够的中性面部图像，因此，采用自动编码器生成中性图像数据，这个基于几何特征的网络可通过生成的中性面部图像和情绪图像之间的坐标差来进行训练。

此网络是根据 VGG19 网络构建的，主要有编码和解码两个过程。编码时，将带有情绪的面部图像输入到 VGG19 的 pool3 层，然后执行 1 个卷积层和 maxpooling 层，最后将特征通过 4096 个节点的全连接层；解码时，与编码过程相反，进行上采样和卷积过程，得到和输入大小一样的图像。自动编码器缩小了输入和输出之间的差异，误差函数为

$$\text{error} = \text{MSE}(G_n, X_n) \tag{4}$$

式中，$G_n$ 是生成的中性面部图像；$X_n$ 是现有的图像，MSE 的计算公式表示为

$$\frac{1}{m}\sum_{i=0}^{n-1}(g_i - x_i)^2 \tag{5}$$

式中，包含生成的中性图像的第 $i$ 个像素 $g_i$ 和真实图像的第 $i$ 个像素 $x_i$，$m$ 是图像总的像素数。

#### 1.3.1　AAM 几何特征提取

对于提取人脸特征点，本文使用 AAM 模型。该模型算法主要分为两个步骤，第一是建立模型，其次为模型匹配。采用统计分析方法建立 AAM 的先验模型，包含人脸的形状信息和纹理信息。模型分为 4 个步骤建立：(1)标注训练图片；(2)形状对齐；(3)形状和纹理建模；(4)建立混合模型。在图像标

注过程中，手动标注的图像质量直接影响模型的结果，因此，每张图像均由多人完成标注。每个对象选取 35 张图像，原图和手动标注图像如 5 所示。

(a) 原图 (b) 特征点图

图 5 原始图像及标定特征点后图像

人脸的形状信息是由图像中特征点的横纵坐标组成的向量来表示的，标注完即可得到训练集中的所有样本数量。由于人脸位置、尺寸会对图像特征提取造成一定的影响，在统计分析之前，需用普鲁克分析(Procrustes Analysis，PA)[15]对人脸进行归一化处理，对于归一化后的图像进行主成分分析，原理如下

$$\boldsymbol{s}=\boldsymbol{s}_0+\sum_{i=1}^{n}b_i\boldsymbol{s}_i \tag{6}$$

式中，$s_0$ 是基础形状向量；$s_i$ 是训练集图像形状空间的正交基；$b_i$ 表示图像形状特征在 $s_i$ 方向上的投影坐标。这里不同的 $b_i$ 表示不同形状的人脸。

人脸的纹理信息是由与形状无关的图像中面部区域的像素的灰度值来表示的，所以考虑到需要去除光照带来的影响，对于纹理向量也要进行相应的处理，所以纹理模型可表示为

$$\boldsymbol{A}(x)=\boldsymbol{A}_0(x)+\sum_{i=1}^{n}a_i\boldsymbol{A}_i(x) \tag{7}$$

式中，$\boldsymbol{A}_0(x)$ 是基础纹理向量；$A_i(x)$ 是训练图像形状空间的正交基；$a_i$ 表示纹理在 $A_i(x)$ 方向上的投影坐标，这里不同的 $a_i$ 表示不同纹理的人脸。

形状特征和纹理特征建立完成后，它们就包含了完整的人脸信息，但由于两个模型中的参数存在相关性，所以用 AAM 算法去除它们之间的相关性，原理表示为

$$\boldsymbol{s}=\boldsymbol{s}_0+\boldsymbol{Q}_s\times c \tag{8}$$

$$\boldsymbol{A}(x)=\boldsymbol{A}_0(x)+\boldsymbol{Q}_g\times c \tag{9}$$

式中，训练集中形状子空间的投影矩阵和纹理子空间的投影矩阵分别是 $\boldsymbol{Q}_s$，$\boldsymbol{Q}_g$；$c$ 同时控制纹理和形状特征的变化，变换 $c$ 可以形成不同的人脸图片，是两个特征的联合参数。

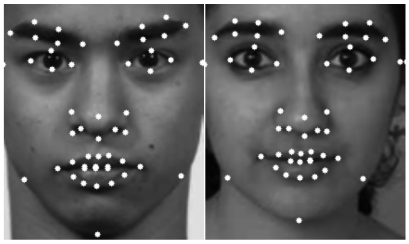
第二步是匹配标定，这部分主要是定义一个能量函数来计算两者匹配过程产生的误差，定义为

$$\sum_{x\in S_0}[A(x)-I(W(x,p))]^2 \tag{10}$$

式中，$A(x)$ 代表 $S_0$ 对应的模型实例中像素 $x$ 的纹理值；$W(x,p)$ 是 $S_0$ 对应实例中像素 $x$ 在输入图像中的对应位置；$I(W(x,p))$ 代表输入图像中像素 $W(x,p)$ 的纹理值。将上式泰勒展开

$$\sum_{x}\left[I(W(x,p)-A_0(W(x,0))-\nabla A_0\frac{\partial W}{\partial P}\Delta P\right]^2 \tag{11}$$

普鲁克分析中用反转组合拟合算法对上式进行迭代求解，可求得最小的外观系数 $c$。平均表观特征图像和输入图像的角色被调换，以保持 Hessian 矩阵在迭代中不变，这样整体算法的复杂度就会大大降低了。



(a) 实例 1 (b) 实例 2

图 6 提取特征点实例

图 6 中两张人脸图像是模型拟合的结果，白色的点是提取的特征点，所以可用 AAM 模型对每一个人脸数据集进行特征点提取，进而得到几何特征。

1.3.2 几何特征 CNN 训练

获得标志点后，计算中性面部图像和情绪图像之间 AU 坐标(比如眉毛，嘴唇，眼睛)之间的差异。

结果总共有 36 个长度的 $x$ 和 $y$ 坐标，接着将其做差，得到

$$\boldsymbol{V}_e=[x_e^0-x_n^0,\ y_e^0-y_n^0,\cdots,x_e^{35}-x_n^{35},\ y_e^{35}-y_n^{35}] \tag{12}$$

式中，$\boldsymbol{V}_e$ 是情绪 $e$ 的几何向量，每个 $x_e$ 和 $y_e$ 坐标是按从左到右，从上到下的顺序排列的，眼睛的边缘和嘴唇坐标按顺时针从上、右、下和左构成，$\boldsymbol{V}_e$ 是通过从如下的 CNN 网络训练得来的。

根据表 1，在基于外观特征的 CNN 的 top-2 结果上进行加权，在第一个卷积层中，使用 64 个 3x1 大小的卷积核进行卷积运算，第二个卷积层也是采用同样的方式，接着使用 2x1 的内核进行池化运算，经过卷积池化之后，将其展平与大小为 500 个节点的全连接层相连，最后获得这两个类别的 softmax 结果。网络中使用的激活函数是 ReLU，使用 dropout 来防止过拟合，同样的这里的损失函数也是使用交叉熵，存储每对模型，且从之前获得的外观模型

的 softmax 的结果选择和 top-2 相对应的模型，通过这种方式，进而确定最后的加权函数。

表 1　几何特征 CNN 的结构

| | 尺寸 | 核大小 |
|---|---|---|
| 输入：标志点坐标差异 | 36×1 | |
| 卷积层 1_1 | 64×36×1 | 3×1 |
| 卷积层 1_2 | 64×36×1 | 3×1 |
| 池化层 1 | 64×18×1 | 2×1 |
| 卷积层 2_1 | 128×18×1 | 3×1 |
| 卷积层 2_2 | 128×18×1 | 3×1 |
| 卷积层 2_3 | 128×18×1 | 3×1 |
| 池化层 2 | 128×9×1 | 2×1 |
| 卷积层 3_1 | 256×9×1 | 3×1 |
| 卷积层 3_2 | 256×9×1 | 3×1 |
| 卷积层 3_3 | 256×9×1 | 3×1 |
| 池化层 3 | 256×4×1 | 2×1 |
| 全连接层 | | |
| 输出 | | |

### 1.4　特征加权融合及分类

对外观特征的 top-2 情绪结果和基于几何特征的网络进行加权计算，如此，可有效提高结果的准确率。外观特征网络 CNN 中选择的 6 个类别中有 2 个类别含有 top-2 softmax 值，因此构建如下的加权函数

$$C_l = \alpha A_l + (1-\alpha) G_l \tag{13}$$

式中，$\alpha$ 是 0 到 1 之间的数；$A_l$ 是与第 $l$ 个情绪类别对应的基于外观特征网络的 softmax 结果；$G_l$ 是基于几何特征网络的 softmax 结果，softmax 表示为

$$A_{\text{top\_}l} = \frac{\alpha_{\text{top\_}l}}{\sum_{i=0}^{j} \alpha_{\text{top\_}i}} \tag{14}$$

式中，$a$ 是不同类别的 softmax 值；top_l 是具有第 l 个最高 softmax 值的类别编号；$j$ 的最大值为 1，由此得到的结果和几何结果 $G_l$ 融合。为确定最终的面部表情，使用组合 $G_l$ 将得到的两个特征所占的比例重新缩放，表示为

$$R_l = (a_{\text{TOP}_0} + a_{\text{TOP}_1}) \times C_l \tag{15}$$

式中，$\boldsymbol{R}_l$ 是具有第 $l$ 个最高 softmax 值的类别是基于外观特征中 top-2 所占比例的组合，即 $(a_{\text{TOP}_0} + a_{\text{TOP}_1})$，因此融合两个网络获得的 softmax 向量，预测的情绪如下

$$\boldsymbol{F} = [V_0, V_1, \cdots, V_5] \tag{16}$$

$$\text{pred}_e = \underset{i}{\text{argmax}}\, v_i \tag{17}$$

所获得的 $\boldsymbol{R}_l$ 用作与 0 和 5 个 softmax 值相对应的情绪类别的值 $\boldsymbol{V}_i$，因此 softmax 函数值向量由 6 个分量构成，如上式 $\boldsymbol{F}$，最后情绪是由式(17)决定的，从 0 到 5 对应的情绪分别为“快乐”，“悲伤”，“惊喜”，“愤怒”，“反感”，“恐惧”。

## 2　实验结果和分析

在这部分，通过实验验证所提方法，通过与现有算法的识别准确度比较，证明了所提方法的优越性。接着介绍实验数据集、实验的硬件和参数设置，最后给出实验结果并对其性能进行了分析。

### 2.1　数据集

(1) 扩展的 COHN-KANADE(CK+)

CK+是一个面部表情数据集，它具有从中性情绪到极端情绪的序列帧并且其标记完整。共有 123 名受试者参与，593 个图像序列，其中 327 个被标记为 7 种普遍情绪(愤怒、蔑视、厌恶、恐惧、快乐、悲伤和惊讶)[8]。该数据集被各种与面部表情识别相关的算法所采用。因此，适用于对最新技术的评价。

本文以愤怒、厌恶、恐惧、快乐、悲伤、惊讶六种情绪作为实验数据。每个序列中的最后三帧被用作极限情绪帧，并且每个情感大约有 80～120 个图像，因此总共使用了 927 个图像。准确度验证方法使用 10 倍交叉验证方法，以便与其他算法进行比较。它将数据集分成 10 组，其中 9 组用于学习，最后一组用于验证。利用这些结果对人脸表情识别算法的准确度求取平均值。图 7 显示了正面面部图像的 CK+数据集示例。

图 7　扩展的 Cohn-Kanade(CK +)数据集

(2) 日本女性面部表情(JAFFE)

JAFFE 总共包含 7 种面部表情共 213 幅图像，分别有快乐、愤怒、悲伤、恐惧、厌恶、惊讶和中性表情，是一个由 10 位日本女性的灰度正面表情图像组成的数据集[16]。将其数据增强至 915 个以进行学习和验证，包括旋转、翻转和加噪声。旋转方向分别为顺时针 5 度和逆时针 5 度，或将方差为 0.01、均值为零的高斯噪声添加到原始图像中。

为了将该方法与最新算法进行比较，采用与 CK+数据集相同的方法：10 倍交叉验证方法进行了验证，并通过对结果求平均来测量识别准确度。图 8 显示了 JAFFE 数据集的示例。

图 8 JAFFE 数据集

### 2.2 实验环境

本文实验设备 CPU 是 i7-8700，3.2GHz 频率，GTX1080 显卡，基于 Windows 系统。本文方法是一种基于深度学习的算法。因此，使用 TensorFlow 和 Keras 对本文算法进行建模。它还基于 python 语言，该语言针对深入学习建模和相关库进行了优化。通过训练以及利用 10 倍验证方法对每个数据集的实验结果（包括准确性）进行验证。在每个网络的训练过程中，每个数据集迭代 30 次，学习率为 0.01，采用 SGD 作为优化算法。

### 2.3 实验结果

#### 2.3.1 融合特征的权重确定

为了改善两个网络的误差，设计了一种融合两个网络的算法。将基于外观特征的网络的 softmax 结果与基于几何特征的网络的结果相结合，将预测结果计算为最高值。式(13)中 $\alpha$ 的值是指基于外观特征的网络的贡献度。它的实际值介于 0 和 1 之间。实验数据总共用了 2000 个，每 200 个数据为一组，共 10 组，按照 9∶1 的比例进行训练数据和测试数据的划分。实验结果如图 9 所示。

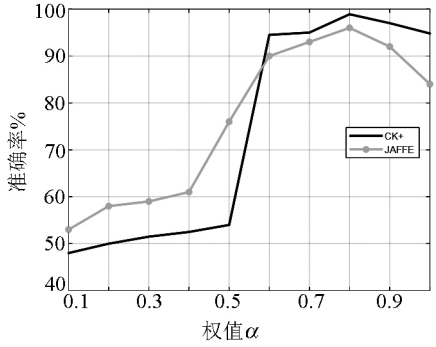



图 9 不同权值下的准确度

根据图中显示，无论是 CK＋数据集还是 JAFFE 数据集，都是权值 $\alpha$ 在 0.8 时展现了最优的准确度，所以本文采用权值为 0.8 来结合基于外观和基于几何特征这两个网络。

#### 2.3.2 定量验证

在这一部分中，将针对每个数据集验证所提出的算法是否正确识别面部表情。使用 10 倍交叉验证法对每个数据集的准确性进行测量。使用0.8为权值结合两个网络进行实验，并给出最后结果的混淆矩阵。在 CK＋数据集上的混淆矩阵如 2 表所示。

表 2 在 CK ＋数据集上提出的方法的混淆矩阵

| | 愤怒 | 反感 | 恐惧 | 快乐 | 悲伤 | 惊喜 |
|---|---|---|---|---|---|---|
| 愤怒 | **96.6** | 0.0 | 0.8 | 0.9 | 1.7 | 0.0 |
| 反感 | 0.4 | **98.8** | 0.0 | 0.0 | 0.8 | 0.0 |
| 恐惧 | 0.0 | 0.0 | **100.0** | 0.0 | 0.0 | 0.0 |
| 快乐 | 0.0 | 0.1 | 0.5 | **99.4** | 0.0 | 0.0 |
| 悲伤 | 1.4 | 0.0 | 0.0 | 0.0 | **98.6** | 0.0 |
| 惊喜 | 0.5 | 0.0 | 0.0 | 0.0 | 0.1 | **99.4** |

在 CK＋数据集中，通过将 927 个数据划分为 10 组来测量准确度。在 CK＋数据集中，生气的准确性为 96.6%，恐惧的准确性最高。愤怒的情绪常常被错误地预测为悲伤的情绪。这是因为在这两种情绪中，面部图像中人嘴巴的末端总是朝下的。

将现有算法与本文算法在 CK＋数据集上的实验结果进行了比较，如表 3 所示，其中都是使用了 10 倍交叉验证法。文献[8]提出了一种基于深度综合多块聚集双 CNN(DCMA-CNNS)方法的人脸表情识别算法。与提出的方法相似之处在于，都使用了双流网络。该方法针对 LBP 图像的静态特征，而不是动态变化的特征，CK＋数据集的准确率为 93.46%。文献[7]还关注一帧中每个活动块的外观，而不是动态特性。该方法在 CK＋数据集中的准确率为94.09%。文献[10]针对细致的表情变化特点，基于特征点计算角度几何特征，在支持向量机上采用不同的核函数对表情分类和识别，识别率达 95.16%。在传统方法中还与 LBP 和 HOG 特征融合检测[16]进行了比较，该文献利用 LBP 提取脸部区域纹理特征，HOG 提取眼眉区域和嘴区域等边缘形状信息，将这 3 个区域赋予不同的权值再做融合。将融合的特征信息送入 SVM 分类器得出最终检测结果。

表 3 在 CK ＋数据集上几种方法的结果比较

| 方法 | 准确度/% |
|---|---|
| DCMA-CNNs [8] | 93.46 |
| LDSP-LDRHP-CNN [7] | 94.09 |
| 角度几何特征[10] | 95.16 |
| 改进 GAN[11] | 94.73 |
| LBP-HOG [16] | 94.70 |
| 提出的算法 | 98.81 |

在 JAFFE 数据集中，依然测量了 10 倍交叉验证的准确性，并计算了平均值。此数据集包含具有相对较高噪声信息量的黑白图像。因此，结果的准确性略低于 CK＋数据集。该算法的混淆矩阵结果如表 4 所示。

表 4　JAFFE 数据集中提出的方法的混淆矩阵

| | 愤怒 | 反感 | 恐惧 | 快乐 | 悲伤 | 惊喜 |
|---|---|---|---|---|---|---|
| 愤怒 | **98.7** | 1.1 | 0.0 | 0.0 | 0.2 | 0.0 |
| 反感 I | 1.2 | **97.6** | 0.9 | 0.0 | 0.3 | 0.0 |
| 恐惧 | 0.0 | 1.7 | **95.7** | 0.3 | 1.8 | 0.5 |
| 快乐 | 0.0 | 0.2 | 0.2 | **99.5** | 0.0 | 0.1 |
| 悲伤 | 2.3 | 0.7 | 7.8 | 1.6 | **87.6** | 0.0 |
| 惊喜 | 0.0 | 0.0 | 1.5 | 1.3 | 0.0 | **97.2** |

在表 5 中，在 JAFFE 数据集上将本文算法与其他几种最新算法进行对比。Zhao 等[17]提出了一种基于 3DCNN 的面部表情识别方法。首先提取人脸表情的关键区域，对关键区域的特征进行提取并融合，用融合的特征进行分类，最终准确率为95.56％。另一个最新的算法是 Liu 等提出的[18]，它使用 LBP 和 HOG 特征以及 gamma 校正来提取显著区域的特征，该方法的识别准确度为 90.08％。张雪梅等[19]融合局部二值模式(LBP)和韦伯局部描述算子(WLD)这两种纹理特征，利用 SVM 进行人脸表情识别分类，得到95.77％这样比较好的效果。

表 5　JAFFE 数据集中几种方法的对比

| 方法 | 准确度/％ |
|---|---|
| 3D-CNN [17] | 95.56 |
| salient feature [18] | 90.08 |
| LBP-HOG [16] | 95.21 |
| 基于 Inception 思想[6] | 94.47 |
| 纹理特征融合[19] | 95.77 |
| 提出的算法 | 96.05 |

同时，为了验证算法的时间复杂度，从 CK＋数据集和 JAFFE 数据集中各随机抽取 100 张图片，用本文方法与其他传统方法进行时间与准确率的比较，这些传统方法通常具有维度低，实时性好的优点。下表所示即为平均检测准确率和时间消耗。

表 6　不同方法的时间、准确率比较

| 方法 | 平均准确度/％ | 时间消耗/ms |
|---|---|---|
| Gabor-ACI-LBP [20] | 92.5 | 4690 |
| DLBP [21] | 96.5 | 124 |
| LBP-HOG [16] | 95.0 | 234 |
| 基于 Inception 思想[6] | 97.0 | 1728 |
| 纹理特征融合[19] | 96.5 | 2321 |
| 提出的算法 | 97.5 | 1539 |

为了更直观的比较算法复杂度，将传统算法也在 GPU 上运行。虽然本文提出的算法在时间耗时方面，相比文献[16]和文献[21]提出的算法，不占优势，但准确率却是最高的。此外，相比文献[20]提出的算法，无论是识别准确度还是时间消耗都具有较大优势，主要是因为文献[20]提出的算法为了更高的准确率，其选取的特征的维度也较高，造成了很高的时间消耗。

## 3　结　论

针对人脸表情识别识别率低的问题，本文提出了一种基于外观和几何双流 CNN 的算法模型。运用先进的特征提取算法，首先，使用 LDP 算法提取外观特征，放入基于外观的 CNN 中训练，再用 AAM 提取几何特征，放入基于几何的 CNN 中训练。选取阈值，将两个网络的特征进行融和。实验结果表明，与其他几种相关算法相比，本文提出的算法平均准确率可以达到约 97.5％，比其他最新算法均具有更高的准确率。

参考文献：

[1] 桑高丽，王国滨，朱蓉，等. 基于级联形状回归的多视角人脸特征点定位[J]. 浙江大学学报：工学版，2019，53(7)：1374—1379.
Sang Gaoli，Wang Goubing，Zhu Rong，et al. Multi-view facial landmark location method based on cascade shape regression[J]. Journal of Zhejiang University：Engineering Science，2019，53(7)：1374—1379.

[2] 张海涛，李美霖，董帅含. 两层级联卷积神经网络的人脸检测[J]. 中国图象图形学报，2019，24(2)：203—214.
Zhang Haitao，Li Meilin，Dong Shuaihan. Two-layer cascaded convolutional neural network for face detection[J]. Journal of Image and Graphics，2019，24(2)：203—214.

[3] Tzimiropoulos G，Maja P. Fast algorithms for fitting active appearance models to unconstrained images[J]. International Journal of Computer Vision，2017，122(1)：17—33.

[4] 谭小慧，李昭伟，樊亚春. 基于多尺度细节增强的面部表情识别方法[J]. 电子与信息学报，2019，41(11)：2752—2759.
Tan Xiaohui，Li Zhaowei，Fan Yachun. Facial expression recognition method based on multi-scale detail enhancement[J]. Journal of Electronics & Information Technology，2019，41(11)：2752—2759.

[5] Narendra Kohli，Mayur Rahul，Rashi Agarwal. Facial expression recognition using moments invariants and modified hidden markov model[J]. International Journal of Applied Engineering Research，2018，13(8 Pt.4)：6081—6088.

[6] 王晓红，梁祐慈，麻祥才. 一种基于 Inception 思想的人脸表情分类深度学习算法研究[J]. 光学技术，2020，46(3)：347—353.
Wang Xiaohong，Liang Youci，Ma Xiangcai. Facial expression classification algorithm research based on ideology of Inception [J]. Optical Technique，2020，46(3)：347—353.

[7] Uddin M Z，Khaksar W，Torresen J. Facial expression recogni-

tion using salient features and convolutional neural network[J]. IEEE Access,2017,5(1):26146—26161.

[8] S Xie, H Hu. Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks[J]. IEEE Trans. Multimedia,2019,21(1):211—220.

[9] **刘全明，辛阳阳. 端到端的低质人脸图像表情识别[J]. 小型微型计算机系统**,2020,41(3):668—672.
Liu Quanming, Xin Yangyang. Face expression recognition based on end-to-end low-quality face images[J]. Journal of Chinese Computer Systems,2020,41(3):668—672.

[10] **吴珂，周梦莹，李高阳，等. 基于角度几何特征的人脸表情识别[J]. 计算机应用与软件**,2020,37(7):120—124.
Wu Ke, Zhou Mengying, Li Gaoyang, et al.Facial expression recognition based on geometrical features of angles[J]. Computer Applications and Software,2020,37(7):120—124.

[11] **李婷婷，胡玉龙，魏枫林. 基于 GAN 改进的人脸表情识别算法及应用[J].吉林大学学报:理学版**,2020,58(3):605—610.
Li Tingting, Hu Yulong, Wei Fenglin. Improved facial expression recognition algorithm based on gan and application[J]. Journal of Jilin University:Science Edition,2020,58(3):605—610.

[12] Chen J K, Chi Z R, Fu H. A new framework with multiple tasks for detecting and locating pain events in video[J]. Computer Vision and Image Understanding: CVIU, 2017, 155: 113—123.

[13] Ruturaj G Gavaskar, Kunal N Chaudhury. Fast adaptive bilateral filtering[J]. IEEE Transactions on Image Processing, 2019,28(2):779—790.

[14] **罗元，余朝靖，张毅，等. 基于改进的局部方向模式人脸表情识别算法[J]. 重庆大学学报**,2019,42(3):85—91.
Luo Yuan, Yu Chaojing, Zhang Yi, et al.Facial expression recognition algorithm based on improved local direction pattern [J]. Journal of Chongqing University:Natural Science Edition, 2019,42(3):85—91.

[15] **张娟. 稀疏正交普鲁克回归处理跨姿态人脸识别问题[J]. 计算机科学**,2017,44(2):302—305.
Zhang Juan. Sparse orthogonal procrustes problem based regression for face recognition with pose variations[J]. Computer Science,2017,44(2):302—305.

[16] **贾磊. 基于 LBP 和 HOG 特征融合的人脸表情识别算法研究[D]. 山西：中北大学**, 2019.
Jia Lei. Research on facial expression recognition algorithm based on LBP and HOG feature fusion[D]. Shanxi: North University of China,2019.

[17] Zhao J F, Mao X, Zhang J. Learning deep facial expression features from image and optical flow sequences using 3D CNN[J]. Visual Computer,2018,34(1):1461—1475.

[18] Liu Y, Li Y, Ma X, et al. Facial expression recognition with fusion features extracted from salient facial areas[J]. Sensors, 2017,17(4):712.

[19] **张雪梅，公维宾，邹建志，等.基于纹理特征融合的人脸表情识别[J]. 计算机技术与发展**,2020,30(3):57—61.
Zhang Xuemei, Gong Weibing, Wu Jianzhi, et al.Facial expression recognition based on texture feature fusion[J]. Computer Technology and Development,2020,30(3):57—61.

[20] Shi S, Liu J M, Si H Q, et al. Facial expression recognition based on Gabor features of salient patches and ACI-LBP[J]. Journal of intelligent & fuzzy systems: Applications in Engineering and Technology,2018,34(4):2551—2561.

[21] **黄丽雯，杨欢欢，王勃. 非对称方向性局部二值模式人脸表情识别[J]. 计算机工程与应用**,2018,54(23):189—194.
Huang Liwen, Yang Huanhuan, Wang Bo. Facial expression recognition based on asymmetric region-directional local binary pattern[J]. Computer Engineering and Applications,2018,54(23):189—194.