

# 基于特征融合的人脸表情识别算法研究

李波, 王坤侠

(安徽建筑大学 电子与信息工程学院, 安徽 合肥 230601)

**摘要:** 为了避免传统表情识别方法中复杂的特征手动提取过程, 同时保证能够提取到更多的表情特征, 文中提出一种融合卷积神经网络 (Convolutional Neural Network, CNN)、方向梯度直方图 (Histogram of Oriented Gradient, HOG) 以及人脸关键点定位 (facial landmark detection) 的人脸表情识别方法。该方法首先通过在图像预处理中使用多任务卷积神经网络 (Multi-task convolutional neural network, MTCNN) 对不同尺度输入图像进行人脸检测并得到人脸的关键点位置信息 (facial landmark)。然后根据 facial landmark 的位置信息计算出人脸表情图像的几何结构特征, 并且计算人脸表情图像局部区域的方向梯度直方图来构成 HOG 特征, 采用特征融合的方式将 facial landmark 和 HOG 特征做进一步的融合形成新的特征向量 LM\_HOG。最后将融合后的特征与经过 CNN 提取的全局特征再次融合输入到支持向量机 (Support Vector Machine, SVM) 和 Softmax 中进行表情识别。在 FER2013 和 Extended Cohn-Kanade (CK+) 人脸表情库上的实验结果表明, 将融合得到的 LM\_HOG 特征作为局部特征, 用以描述图像的局部性差异, CNN 提取的特征作为全局特征, 用以描述人脸表情图像的整体性差异, 融合后的特征能更好的提取图像细节特征, 平均识别率分别达到了 75.14% 和 97.86%, 具有优越的性能。

**关键词:** 人脸表情识别; 方向梯度直方图; 人脸特征点; 特征融合

中图分类号: TP393

文献标识码: A

文章编号: 2095-8382 (2021) 01-094-09

## Facial Expression Recognition Method Based on Feature Fusion

LI Bo, WANG Kunxia

(School of Electronics and Information Engineering, Anhui Jianzhu University, Hefei 230601, China)

**Abstract:** In order to avoid the complex manual feature extraction process in traditional facial expression recognition methods and to extract more facial features, this paper proposes a facial expression recognition method based on convolutional neural network, histogram of oriented gradient and facial landmark detection. First, the method uses multi-task convolutional neural network to perform face detection on input images of different scales and obtain landmark position information of the face in image preprocessing. Then, calculate the geometric structural features of the facial expression image according to the location information of the facial landmark, and calculate the histogram of oriented gradient of the local area of the facial expression image to form the HOG feature. The feature fusion method is used to make the facial landmark and the HOG feature forms a new feature vector LM\_HOG. Finally, the fused features and global features extracted by CNN are fused again and input into the SVM and Softmax for expression recognition. The experimental results on the FER2013 and Extended Cohn-Kanade (CK+) show that the LM\_HOG features obtained by the fusion are used as local features to describe the local differences of the images, and the features extracted by CNN are used as global features to describe the overall difference in facial expression images, the fused features can better extract the image details, and the average recognition rate has reached 75.14% and 97.86%, respectively, and has superior performance.

**Key words:** facial expression recognition; histogram of oriented gradient; facial landmark; feature fusion

收稿日期: 2020-09-07

基金项目: 安徽省自然科学基金面上项目 (1708085MF167)

作者简介: 李波 (1994-), 男, 硕士研究生, 研究方向为情感计算、人工智能。

面部表情作为人类语言的一部分<sup>[1]</sup>,是表达内在情感和意图的最重要方法之一。由于面部表情识别在人机交互,商业产品和医学研究中的广泛应用,在过去的几十年中,它已经在计算机视觉和模式识别领域得到了广泛的研究。Ekman 等<sup>[2]</sup>通过描绘不同的脸部肌肉动作与不同表情的对应关系,进而确定了六种不同的基本情感。Pantic 等<sup>[3]</sup>通过从面部表情图像的正面轮廓中提取 facial landmark 来获得 32 个肌肉运动单元,并大致完成对六种基本类型的情感识别。胡敏等<sup>[4]</sup>提出了一种 Gabor 梯度特征直方图的表情识别方法,该方法使用 Gabor 表示原始图像数据的局部特征,然后在相同比例的不同方向上进行特征的融合,并计算融合特征的直方图。面部表情的识别由多种特征组成,进行面部情感识别时,不相关的特征会影响情感识别的准确性,因此具有高影响因子的特征集的获取已成为面部表情识别的重要研究方向。

目前用于表情识别特征提取的方法主要有 Gabor 小波、局部二值模式(Local Binary Patterns, LBP)、方向梯度直方图(Histogram of Oriented Gradient, HOG)和主动外观模型(Active Appearance Model, AAM)等方法。Gabor 小波变换<sup>[5]</sup>通常用于提取面部表情图像的局部纹理特征,Log-Gabor<sup>[6-7]</sup>改进了原始的 Gabor 滤波器以提取与人类视觉系统识别相似的特征。HOG<sup>[8-9]</sup>方法通过使用 HOG 算子计算局部图像像素值变化的幅度信息和方向信息,并统计梯度方向密度来描述图像的局部边缘特征。该方法<sup>[10]</sup>计算复杂度低,提取效果好,更好地保持图像的几何不变性和光学不变性,是经典的特征提取方法之一。LBP<sup>[11]</sup>方法是一种有效的局部纹理描述算子,对图像的灰度变化以及旋转具有一定的鲁棒性。AAM<sup>[12]</sup>方法综合考虑全局形状和纹理信息,对人脸形状特征和纹理特征进行统计分析。胡敏等<sup>[13]</sup>分别对眉毛和嘴巴部分进行特征提取和表情分类,取得了很好的分类效果。赵等<sup>[16]</sup>提出了一种融合 LBP 特征和多方向 Gabor 直方图特征的方法,实验结果表明,与单独使用其中一种特征进行表情识别相比,融合特征可以获得更好的识别效果。文献[17]将人脸表情图像的几何特征和纹理特征进行融合,充分利用了纹理信息和人脸的结构特性,增强了因

个体差异而产生的不同表情特征的鲁棒性。与传统的通过分类器训练的人工特征提取方法不同, CNN 无需预处理即可直接处理图像并输出最终的情感分类结果,这是一个“端到端”的网络模型。Hinton 等<sup>[14]</sup>考虑到单个 CNN 的结构太小这一事实,在 2006 年提出了深度信念网络(Deep Belief Network, DBN)模型,通过模拟人类的视觉认知系统, DBN 逐层训练网络,以实现高维非线性数据的特征提取和分类。在 2013 年 Kaggle 举行的面部表情识别挑战赛中, Tang 和 Bergstra<sup>[15]</sup>使用了卷积神经网络,其准确率达到 69.7%,最终攀升至榜首。CNN 进行特征提取时,训练数据会隐式的进行学习,从而避免了显式的特征提取。然而, CNN 在进行人脸表情识别任务中,需要大量的训练数据来训练模型的参数,由于人脸表情库的数据量较小远远不能满足网络参数训练的要求,为了提高面部表情识别的准确率,一种方式是通过改进 CNN 网络模型,例如扩大卷积的宽度和加深网络模型的深度,另一种是将 CNN 与传统方法进行融合,弥补彼此的不足,提高特征的提取能力,从而达到更好的识别效果。Jung 等<sup>[18]</sup>提出了两步法进行表情识别,使用深度学习提取面部的外观特征和几何特征,之后,再次提出新的两步法识别模型<sup>[19]</sup>,首先使用类似 haar 的特征作为面部检测,再分别使用深度神经网络(Deep Neural Network, DNN)和卷积神经网络(CNN)两种类型的网络模型,最终结果表明, CNN 比 DNN 具有更好的准确性。赵等<sup>[20]</sup>提出了一种基于 DBN 和逻辑回归的面部表情识别方法,该方法首先检测面部,利用鼻子、眼睛和嘴巴的位置信息,以进行面部表情识别。

将基于 facial landmark 的几何特征用于面部表情识别也是一种已被应用的常用方法。Alexandra 等<sup>[21]</sup>通过描述面部表情然后使用多层感知器(Multilayer Perceptron, MLP)和支持向量机完成情感分类。Paul 等<sup>[22]</sup>使用了 facial landmark 和基于轨迹的几何特征,在 UvA-NEMO 和 CK 数据集上验证了该方法的有效性。本文综合利用人脸表情图像的局部特征和全局特征,提出了一种基于几何特征和纹理特征相融合的人脸表情识别方法。首先,使用基于三层级联体系结构的多任务卷积神经网络(MTCNN)从人脸面部表情图像中提取 facial

landmark, 从而利用 facial landmark 的位置信息来计算相关面部点之间的相对距离, 以捕获由表情变化而产生的特征向量。同时, 采用 HOG 方法来获取人脸图像的幅度信息和方向信息, 通过融合这两个特征创建 LM\_HOG 特征向量。最后, 将 LM\_HOG 特征与经过 CNN 网络模型获得的一维特征矢量进行拼接, 使用分类器进行人脸表情识别。为了验证所提出方法的有效性, 分别采用 SVM 和 Softmax 在 FER2013 和 CK+ 人脸数据集上进行人脸表情识别, 实验结果表明, 该方法在两个数据集上的识别率分别为 75.14% 和 97.86%, 验证了本文方法的优越性能。

### 1 本文算法

本文所提出的面部表情识别算法流程如图 1 所示, 主要分为图像预处理、特征提取以及表情识别三个模块。在特征提取阶段, 针对单一特征对人脸表情信息表征不全面的情况, 将特征提取网络分为两个通道, 通道一通过对输入的表情图像进行预处理后使用 CNN 模型提取人脸表情图像的全局特征, 对光照等物理因素的变化具有很好的鲁棒性; 通道二根据 facial landmark 的位置信息计算出人脸结构特征与 HOG 特征, 提取出面部图像的几何特征和纹理特征。表情识别模块分别采用了 SVM<sup>[23]</sup> 和 Softmax 分类器进行情感分类。

#### 1.1 预处理

为了更好地提取人脸表情特征, 需对输入的面部图像进行预处理。预处理可以提高图像质量, 消

除噪点, 为后面的特征提取和分类奠定良好的基础。本文对输入的面部图像进行预处理的流程为:

步骤一: 对图像进行人脸检测, 以达到减少图像中与人脸无关的背景信息对特征提取等产生影响的目的。由于人脸可能出现在图像中的任何位置, 本文使用一种固定大小的滑动窗口 (sliding window) 扫描整个输入图像以确定人脸位置。当使用滑动窗口进行图像扫描时, 可能会得到多个候选框, 此时采用非极大值抑制 (non-maximum suppression, NMS)<sup>[24]</sup> 方法对候选框进行合并以及消除重复数据。

步骤二: 根据人脸检测得到的候选框利用旋转矩阵进行旋转校正, 增强其旋转的鲁棒性, 然后裁剪出人脸面部区域。旋转矩阵定义如下:

$$(x', y') = (x, y) \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \quad (1)$$

其中  $(x, y)$  表示原始坐标;  $(x', y')$  表示旋转后的坐标;  $\theta$  表示旋转角度, 旋转角度可通过测量两个眼睛中心点向量与水平方向的夹角得到。

步骤三: 归一化操作, 将裁剪后的面部图像归一化为统一大小, 多维特征将具有相似的缩放比例。从而减少了特征提取的面积, 有助于梯度下降算法<sup>[25]</sup> 更快的收敛。并且, 若输入图像为 RGB 格式, 将其进行灰度化处理以降低数据维数, 在保持大部分信息的情况下减少了计算量, 同时便于后续的 facial landmark 定位以及 HOG 特征提取。

#### 1.2 人脸特征点定位

在进行图像预处理之后, 进行人脸特征点定位

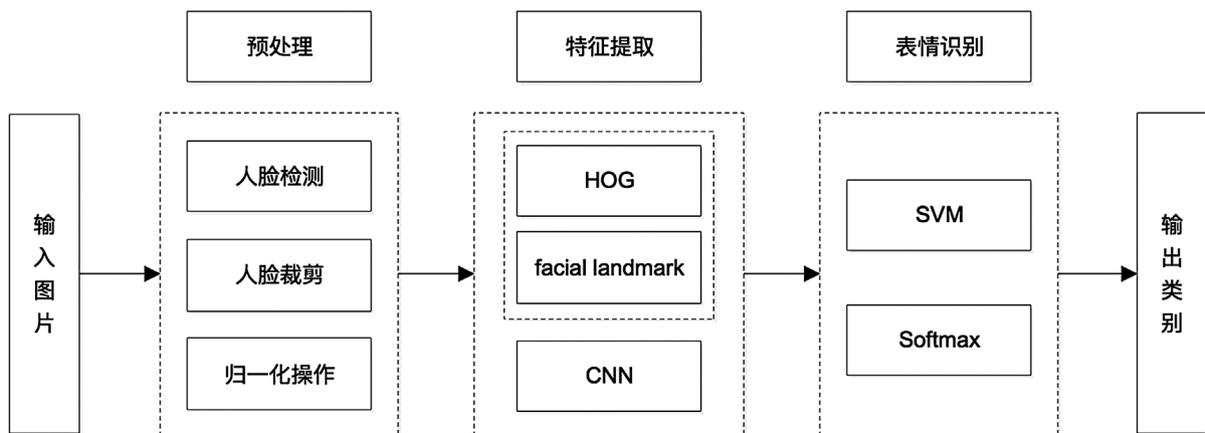


图 1 本文算法流程图

和特征提取。汤晓鸥等<sup>[26]</sup>提出了一种基于 DCNN 的多任务学习方法,并用于人脸特征点检测。该方法采用人脸特征点定位任务为主,其他多任务为辅的策略,最终的实验结果表明该方法在一些遮挡和大姿态人脸面部图像中取得了较好的性能。在文献<sup>[27]</sup>中,人脸特征点分为两个部分,第一部分用于预测面部图像特征的 51 个特征点,这部分被称为内部点;另一部分被用来预测面部轮廓的 17 个特征点,即轮廓点。其中,当对着两部分的特征点进行预测时,它们是完全分离的,并且分别对各个网络进行了训练和预测,从而获得相应的边界框。此外,在预测内部点时,使用四级 DCNN 模型分离面部特征,并且不共享相同的损失函数,从而避免了不平衡问题并实现了精确定位。由于直接使用人脸检测器检测到的人脸边界框通常包含太多不相关的背景,因此会干扰训练,故使用 CNN 模型预测人脸边界框时更有效。基于该方法,考虑到人脸特征点定位与姿态估计关系密切,二者使用相似的特征与内部表示,都需要对光照、遮挡等因素表现出鲁棒性,因此本文结合了新的级联深度神经网络(Deep Alignment Network, DAN)<sup>[28]</sup>,并且添加了特征点热图(feature heatmap)的方式对人脸特征点进行定位任务(如公式 2)。DAN 每个阶段包含三个输入和一个输出,输入分别是被矫正后的图片、关键点热图和由全连接生成的特征图,输出是面部表情的形状。其中,CONNECTION LAYER 的作用是将本阶段得到的输出进行一些列变换,生成下一阶段所需要的的三个输入,具体操作如图 2 所示。由于 CNN 的输入是变换后的图像,故得到的偏移量为相对偏移量,偏移后应通过逆变换将其还原到

原始空间,因此对姿态变换具有良好的适应性。

$$H(x,y) = \frac{1}{1 + \min_{s_i \in T_i(S_{i-1})} \|(x,y) - S_i\|} \quad (2)$$

关键点热度图的计算就是一个中心衰减,关键点处值最大,越远则值越小,其中  $T_i$  表示变换后的图像, $t$  是当前 stage 的序号, $T_i(S_{i-1})$  表示变换后的特征点, $H(x,y)$  表示变换后的热力图图像。

通过 DAN 获得特征点后,利用这些锚点坐标来进行一些相关距离的计算,如图 3 所示。在本文提出的方法中,将眼睛靠近鼻子最接近的点即图 2 中的特征点 39 和 42 作为固定点,鼻子的下端为特征点 33。从这些固定点出发计算相关距离。基本上,需要考虑与眉毛与运动有关的 8 个特征点和与嘴巴相关的 6 个特征点。但是,这些点之间的绝对距离不能直接使用,因为可能存在各种其他因素产生干扰,例如不同的脸部大小以及图片拍摄距离等,因此为了消除这些变化需要对距离进行归一化。为了归一化左眉的距离,将计算出来的距离除以特征点 39 和 21 之间的距离(如公式 3)。类似的,利用特征点 42 和 22 对右眉的距离进行归一化。同时,还可以使用特征点 33 和 51 之间的距离来归一化嘴巴的宽度、嘴巴的高度。对于每只眼睛,相对于嘴巴上唇的左上角和右上角将计算出两个距离,最终计算出六个值的特征向量。

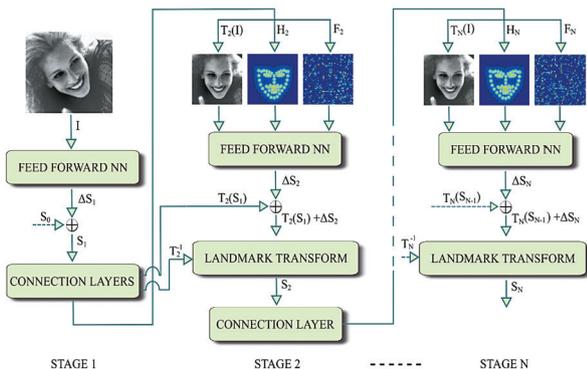


图 2 DAN 基本框架

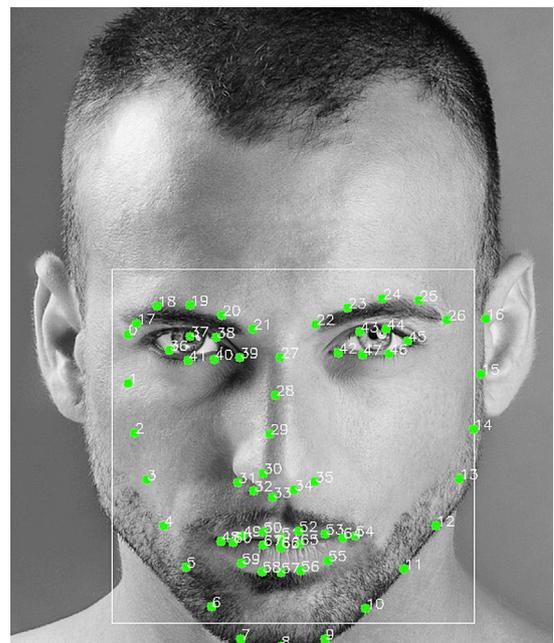


图 3 人脸特征点定位

$$eyebrow_{right} = \sum \frac{dis(E_{right}, i)}{normDis_{rightEyebrow}} \quad (3)$$

$$normDis_{rightEyebrow} = dis(E_{right1}, E_{right2}) \quad (4)$$

$$lip_{right} = \sum \frac{distance(c_1, i)}{normDis_{Lips}} \quad (5)$$

$$normDis_{Lips} = distance(C_1, C_2) \quad (6)$$

$$feature_{vector} = \langle mouth_{width}, mouth_{height}, eyebrow_{left}, eyebrow_{right}, lip_{left}, lip_{right} \rangle \quad (7)$$

### 1.3 HOG 特征提取

方向梯度直方图 (Histogram of Oriented Gradient, HOG) 最初由 Dalal 和 Triggs<sup>[29]</sup> 提出用于目标检测, 主要被用于统计图像的局部梯度特征。HOG 特征具有很多优点, 首先, 由于 HOG 是在图像的局部单元上计算得出的, 因此可以很好的保持图像的几何不变性和光学不变性; 其次, HOG 描述了边缘结构特性, 因此很好的描述图像的局部形状信息。HOG 特征提取具体过程主要分为以下几个步骤:

步骤一: 归一化处理, 利用 Gamma 校正对输入进行颜色归一化, 从而调节图像的对比度, 降低图像局部阴影和光照变化对结果造成影响, 消除一些噪声的干扰。

步骤二: 梯度计算, 通过梯度计算对图像进行边缘检测, 计算出人脸表情图像中像素点  $(x, y)$  的水平和垂直方向的梯度, 公式如下:

$$G_x(x, y) = I(x+1, y) - I(x-1, y) \quad (8)$$

$$G_y(x, y) = I(x, y+1) - I(x, y-1) \quad (9)$$

步骤三: 梯度幅值和角度计算, 使用前面梯度计算得到的梯度值  $G_x$  和  $G_y$ , 通过公式 (10) 和 (11) 计算梯度幅值  $G(x, y)$  和梯度方向  $\theta(x, y)$ 。

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (10)$$

$$\theta(x, y) = \arctan \frac{G_y(x, y)}{G_x(x, y)} \quad (11)$$

步骤四: C-HOG 计算 [29], 将人脸表情图像利用其维度信息平均分割成  $m \times m$  个像素大小的 Cell, 在获取 Cell 之后统计以 Cell 为单元的方向梯度直方图 C-HOG, 之后将梯度方向平均分为  $N$  个方向角度区域 (Bin), 用  $\lambda_k(x, y)$  表示在  $(x, y)$  位置处对第  $n$  个 Bin 的权值, 权值的取值计算如下:

$$\lambda_k(x, y) = \begin{cases} G(x, y), \theta(x, y) \in Bin_n \\ 0, \theta(x, y) \notin Bin_n \end{cases} \quad (12)$$

然后根据每个 Cell 的位置信息计算出方向梯度直方图向量  $[H_1^l, H_2^l, H_3^l, \dots, H_N^l]$ , 其中,

$$H_n^l = \sum_{(x, y) \in cell} \lambda_n(x, y), 1 \leq n \leq N。$$

步骤五: B-HOG 计算, 由步骤三分割的 Cell, 利用其相邻的  $m \times m$  个从而组成一个较大的单元块 (Block), 将块内所有 Cell 的 C-HOG 通过串联得到块的方向梯度直方图 (B-HOG), 然后, 采用  $L_2$  模标准化的方式对每个 B-HOG 进行标准化 (如公式 13)。

$$v \leftarrow v / \sqrt{|v| + \mu^2} \quad (13)$$

其中,  $\varepsilon$  是一个无穷小量, 为了避免分母为 0。

步骤六: HOG 特征提取, 将每个 B-HOG 以串联的形式拼接在一起, 顺序级联这些特征, 进而得到人脸表情图像的 HOG 特征。

在本文中, 将所提取到的 HOG 特征与通过人脸特征点定位获得的  $feature_{vector}$  进行向量间串联拼接融合, 形成融合后的 LM\_HOG 特征向量。

### 1.4 表情识别

表情识别是最后也是最重要的一步, 使用基于分类的识别策略执行。创建 LM\_HOG 特征向量后与经过 CNN 得到的特征进行融合输入到分类器中对面部表情进行分类识别。在本文中, 分别使用 SVM<sup>[22]</sup> 和 Softmax 进行表情识别。

## 2 实验

实验采用的环境为: 软件环境 Windows10 下的 TensorFlow 平台; 硬件环境: 处理器是 Intel Core i7-7700 CPU, 显卡是 GTX-1070;

### 2.1 表情数据库介绍

FER2013 是 Pierre 和 Aaron 通过 Google 图像搜索 API 自动收集的无约束的大规模数据集, 是 Kaggle 面部表情识别挑战赛中的公开数据集, 共包含 35887 张人脸表情图像, 28709 张图像作为训练集, 3589 张图像作为验证集, 3589 张图像作为测试集, 每张人脸图片的像素大小为  $48 \times 48$ , 所有的图像都进行相同的预处理操作。该数据集反映了不同环境下人脸表情呈现的真实状态, 每一类表情在

年龄、光照、面部姿态、表情程度等方面都有着显著的差异,具有一定的挑战性。

CK+ 数据集是在人脸表情数据库 Cohn-Kanade (CK) 基础上扩增而来的,是目前评估人脸表情识别方法最广泛的实验室控制条件下的数据集。该数据集包含来自 123 名受试者 593 个视频序列,其中共有 327 个样本,总共涵盖了七种基本表情。由于 CK+ 数据量较少,故此在实验中采用了十折交叉验证的方式进行,即加载全部表情图像然后将人脸表情数据数据均分为 10 份,每次训练时取其中 9 份进行训练,然后另一份进行测试,共进行 10 次实验,为了避免单次实验的影响,最后取实验的平均结果作为表情识别率<sup>[30]</sup>。

### 2.2 实验结果与分析

在对 CNN 网络模型的训练中,网络模型的详细训练参数见表 1。

表 1 网络模型训练参数

参数	参数值
学习率 (learning rate)	0.001 (每 5 个 epoch 衰减 90%)
批量大小 (batch size)	FER2013:128 CK+:16
迭代次数 (epoch)	50
丢失率 (dropout)	0.5

为了验证该方法融合特征的有效性,从人脸表情图像中提取了基于几何特征的 facial landmark, HOG 特征,以及融合后的 LM\_HOG 特征,并且在人脸表情图像预处理中,分别采用是否加入 sliding window 的操作进行实验效果对比,验证其有效性。表 2 为基于 CNN 融合模型结构在不同的特征组合下,分别在 FER2013 和 CK+ 两个数据集中进行实验得到的表情识别率。由表 2 中的实验结果数据可以看出,将基于 facial landmark 的几何特征与基于纹理的 HOG 特征进行融合可以更为有效的提高表情识别率,验证了本文方法的有效性。

表 2 不同特征组合方式识别率对比 (%)

特征提取方法	FER2013	CK+
CNN	58.72	72.68
CNN + facial landmark	72.54	90.25
CNN + HOG	71.86	93.48
CNN + LM_HOG	74.80	96.68
CNN + LM_HOG (sliding window)	75.14	97.86

表和表分别为在 FER2013 和 CK+ 上的表情识别混淆矩阵。从表 3 的混淆矩阵可以看出,该方法对“高兴”情感的识别率最高,而对于“恐惧”的识别率相对较低。这种情况出现的原因主要有两个:一个是数据集中的情感类别数量不均衡,例如有 7215 张“高兴”表情图像,而“恐惧”表情图像只有 436 张,统计发现每种类别的平均数量大约为 4101 张,而“恐惧”数量不足从而导致学习不够;其次,“恐惧”、“厌恶”、“伤心”这几类情感在情绪表达时有一定的相似之处,这使得它们有时难以区分,并且在原数据集中同样存在很多错误分类的情况,这使得训练难度更大;第三是因为模型对于面部夸张、运动幅度大的表情,提取更加明显且更易提取,而对于运动幅度小、面部变化不明显的表情,识别率较低。从表 4 的混淆矩阵可以看出,本文所采用的特征融合的方法对每种表情的识别率都能达到 95% 以上,表明本文多特征融合的方法可以对表情进行有效的区分。

表 3 基于 CNN + LM\_HOG 的表情识别率混淆矩阵 (FER2013)

%	生气	厌恶	恐惧	高兴	惊讶	伤心	中性
生气	68.72	6.10	8.46	1.84	2.24	8.32	4.32
厌恶	21.26	64.26	6.02	0	1.83	6.63	0
恐惧	7.02	0	58.24	2.25	5.07	17.40	10.02
高兴	1.01	0	1.02	89.86	1.96	0.92	5.23
惊讶	1.04	0	4.28	3.76	84.28	0.86	5.78
伤心	8.44	1.08	10.76	3.56	1.02	65.76	9.38
中性	7.22	0	5.07	6.85	11.06	1.18	68.62

表 4 基于 CNN + LM\_HOG 的表情识别率混淆矩阵 (CK+)

%	生气	厌恶	恐惧	高兴	惊讶	伤心	中性
生气	98.18	0.92	0	0	0.66	0.24	0
厌恶	0.31	97.92	0.40	0.32	0.19	0	0.86
恐惧	1.06	0.50	97.88	0	0	0.56	0
高兴	0	0.10	0	99.65	0	0.03	0.22
惊讶	0.12	0.38	0	0.36	98.84	0.18	0.12
伤心	0.36	1.74	0.67	0	0.35	96.60	0.28
中性	1.10	2.16	0.54	0	0	0.52	95.68

在 FER2013 和 CK+ 数据集中,将本文方法与目前主流的人脸表情识别方法进行了比较,包括 GLTP<sup>[31]</sup>、Gabor<sup>[32]</sup>、LDN<sup>[33]</sup>、CNN<sup>[34]</sup>、Geome-

tric feature + LBP + SVM<sup>[35]</sup>、HOG + DBN + Gabor + SAE<sup>[36]</sup>、LBP + MTSL<sup>[37]</sup>、Boosted DBN<sup>[38]</sup>、DNNRL<sup>[39]</sup>、T-DNN<sup>[39]</sup>、CPC<sup>[40]</sup>、RTCRelief-F<sup>[41]</sup> 等方法。实验结果如表 5,6 所示,可以看出,本文所提出的方法对表情具有更好的识别效果,识别率明显高于其他几种方法。

表 5 本文方法与目前主流的方法对比  
(FER2013)

方法名称	识别率 (%)
DNNRL <sup>[39]</sup>	70.60
T-DNN <sup>[39]</sup>	70.13
CPC <sup>[40]</sup>	71.35
RTCRelief-F <sup>[41]</sup>	73.36
本文方法	75.14

表 6 本文方法与目前主流的方法对比  
(CK+)

方法名称	准确率 (%)
GLTP <sup>[31]</sup>	90.65
Gabor <sup>[32]</sup>	91.4
LDN <sup>[33]</sup>	89.3
CNN <sup>[34]</sup>	93.68
Geometric feature+ LBP+SVM <sup>[35]</sup>	90.08
HOG + DBN + Gabor + SAE <sup>[36]</sup>	91.11
LBP+MTSL <sup>[37]</sup>	91.53
Boosted DBN <sup>[38]</sup>	96.70
本文方法	97.86

### 3 结论

本文提出一种基于卷积神经网络融合特征的人脸表情识别方法,结合 facial landmark 和 HOG 特征对人脸表情图像进行几何特征、局部梯度特征、纹理特征的提取,然后将两种特征进行融合形成 LM\_HOG 特征,最后将融合后的特征与经过 CNN 提取的全局特征再次融合输入到分类器中进行表情识别。在 FER2013 和 CK+ 数据集均取得了较好的识别效果。特征融合的方式能够将单一特征的优缺点进行有效互补,可以更充分地提高面部表情信息,因此能够取得较好的表情识别效果。在接下来的工作中,对于含有较大遮挡的面部表情识别任务将成为研究的重点,提高识别网络模型对于遮挡情况的鲁棒性。

### 参考文献:

- [1] 李勇,林小竹,蒋梦莹.基于跨连接 LeNet-5 网络的面部表情识别[J].自动化学报,2018,44(1):176-182.
- [2] Ekman P, Friesen W V. Constants across cultures in the face and emotion[J]. Journal of Personality and Social Psychology, 1971, 17(2):124-129.
- [3] Valstar M F, Gunes H, Pantic M. How to distinguish posed from spontaneous smiles using geometric features[C]// Proceedings of the ninth international conference on Multimodal interfaces - ICMI '07. November 12-15, 2007. Nagoya, Aichi, Japan. New York: ACM Press, 2007.
- [4] 胡敏,朱弘,王晓华,等.基于梯度 Gabor 直方图特征的表情识别方法[J].计算机辅助设计图形学学报,2013,25(12):1856-1861.
- [5] Zhang H H, Zhang B L, Huang W M, et al. Gabor wavelet associative memory for face recognition[J]. IEEE Transactions on Neural Networks, 2005, 16(1):275-278.
- [6] 刘元,吴小俊.基于 Log-Gabor 滤波与黎曼流形学习的图像识别算法[J].模式识别与人工智能,2015,28(10):946-952.
- [7] 苑玮琦,范永刚,柯丽.相位一致性和对数 Gabor 滤波器相结合的掌纹识别方法[J].光学学报,2010,30(1):147-152.
- [8] Chen J H, Takiguchi T, Ariki Y. Facial expression recognition with multithreaded cascade of rotation-invariant HOG[C]//2015 International Conference on Affective Computing and Intelligent Interaction (ACII). September 21-24, 2015, Xi'an, China. IEEE, 2015:636-642.
- [9] 刘斌,赵兴,胡春海,等.面向颜色深度图像手脸近距离遮挡的手势识别[J].激光与光电子学进展,2016,53(6):140-149.
- [10] 胡敏,滕文娣,王晓华,等.融合局部纹理和形状特征的人脸表情识别[J].电子与信息学报,2018,40(6):1338-1344.
- [11] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions[J]. Pattern Recognition, 1996, 29(1):51-59.
- [12] Proceedings third IEEE international conference on automatic face and gesture recognition[C]//Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. April 14-16, 1998, Nara, Japan. IEEE, 1998:i.
- [13] 胡敏,江河,王晓华,等.精确局部特征描述的表情识别[J].中国图象图形学报,2014,19(11):1613-1622.
- [14] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7):1489-1502.

- 1527–1554.
- [15] Yichuan Tang. Deep learning using linear support vector machines[EB/OL]. <https://www.doc88.com/p-0701519074532.html>
- [16] Zhao L, Wang Z C, Zhang G X. Facial expression recognition from video sequences based on spatial-temporal motion local binary pattern and Gabor multiorientation fusion histogram[J]. *Mathematical Problems in Engineering*, 2017, 2017: 1–12.
- [17] 江河. 人脸表情特征提取及层级分类研究[D]. 合肥: 合肥工业大学, 2015.
- [18] Jung H, Lee S, Yim J, et al. Joint fine-tuning in deep neural networks for facial expression recognition[C]//2015 IEEE International Conference on Computer Vision (ICCV). December 7–13, 2015, Santiago, Chile. IEEE, 2015: 2983–2991.
- [19] Jung H, Lee S, Park S, et al. Development of deep learning-based facial expression recognition system[C]//2015 21st Korea–Japan Joint Workshop on Frontiers of Computer Vision (FCV). January 28–30, 2015, Mokpo, South Korea. IEEE, 2015: 1–4.
- [20] Zhao X M, Shi X G, Zhang S Q. Facial expression recognition via deep learning[J]. *IETE Technical Review*, 2015, 32(5): 347–355.
- [21] Bandrabur A, Florea L, Florea C, et al. Emotion identification by facial landmarks dynamics analysis[C]//2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP). September 3–5, 2015, Cluj–Napoca, Romania. IEEE, 2015: 379–382.
- [22] Desrosiers P A, Daoudi M, Devanne M. Novel generative model for facial expressions based on statistical shape analysis of landmarks trajectories[C]//2016 23rd International Conference on Pattern Recognition (ICPR). December 4–8, 2016, Cancun, Mexico. IEEE, 2016: 961–966.
- [23] Joachims T. Making large-scale SVM learning practical[R]. Technical report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund, 1998.
- [24] 赵文清, 严海, 邵绪强. 改进的非极大值抑制算法的目标检测[J]. *中国图象图形学报*, 2018, 23(11): 1676–1685.
- [25] 常永虎, 李虎阳. 基于梯度的优化算法研究[J]. *现代计算机*, 2019(17): 3–8, 15.
- [26] Zhang Z P, Luo P, Loy C C, et al. Facial landmark detection by deep multi-task learning[C]//Computer Vision – ECCV (C)1994–2021 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net> 2014, 2014.
- [27] Zhou E J, Fan H Q, Cao Z M, et al. Extensive facial landmark localization with coarse-to-fine convolutional network cascade[C]//2013 IEEE International Conference on Computer Vision Workshops. December 2–8, 2013, Sydney, NSW, Australia. IEEE, 2013: 386–391.
- [28] Kowalski M, Naruniec J, Trzcinski T. Deep alignment network: a convolutional neural network for robust face alignment[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). July 21–26, 2017, Honolulu, HI, USA. IEEE, 2017: 2034–2043.
- [29] 岑瑶, 赵烜赫, 潘新, 等. 基于稀疏表示和 HOG 特征的掌纹识别方法[J]. *内蒙古农业大学学报(自然科学版)*, 2020, 41(1): 76–80.
- [30] 高理想, 高磊. 基于特征融合的人脸表情识别研究[J]. *软件工程*, 2019, 22(12): 24–27.
- [31] 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). June 20–25, 2005, San Diego, CA, USA. IEEE, 2005: 886–893.
- [32] Holder R P, Tapamo J R. Improved gradient local ternary patterns for facial expression recognition[J]. *EURASIP Journal on Image and Video Processing*, 2017, 2017(1): 1–15.
- [33] 龚安, 曾雷. 基于 Gabor 变换与改进 SLLE 的人脸表情识别[J]. *计算机系统应用*, 2017, 26(9): 210–214.
- [34] Ramirez Rivera A, Rojas Castillo J, Oksam Chae O. Local directional number pattern for face analysis: face and expression recognition[J]. *IEEE Transactions on Image Processing*, 2013, 22(5): 1740–1752.
- [35] Lopes A T, de Aguiar E, de Souza A F, et al. Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order[J]. *Pattern Recognition*, 2017, 61: 610–628.
- [36] Ghimire D, Jeong S, Yoon S, et al. Facial expression recognition based on region specific appearance and geometric features[C]//2015 Tenth International Conference on Digital Information Management (ICDIM). October 21–23, 2015, Jeju, South Korea. IEEE, 2015: 142–147.
- [37] Fathallah A, Abdi L, Douik A. Facial expression recognition via deep learning[C]//2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA). October 30 – November 3, 2017, Hammamet, Tunisia. IEEE, 2017: 745–750.
- [38] Zhong L, Liu Q S, Yang P, et al. Learning multiscale active facial patches for expression analysis[J]. *IEEE Transactions*

on Cybernetics, 2015, 45 (8) :1499-1510.

- [39] Liu P, Han S Z, Meng Z B, et al. Facial expression recognition via a boosted deep belief network[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. June 23-28, 2014, Columbus, OH, USA. IEEE, 2014:1805-1812.
- [40] Li W B, Ke C. Ensemble deep neural networks for domain-specific Image Recognition[C]//2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW).

July 11-15, 2016, Seattle, WA, USA. IEEE, 2016:1-4.

- [41] Chang T Y, Wen G H, Hu Y, et al. Facial expression recognition based on complexity perception classification algorithm[EB/OL]. 2018
- [42] Li D Y, Wen G H, Hou Z, et al. RTCRelief-F: an effective clustering and ordering-based ensemble pruning algorithm for facial expression recognition[J]. Knowledge and Information Systems, 2019, 59 (1) :219-250.

(上接第 93 页)

据,巧妙地表示突发因素对各行各业的影响,灵活的调整信贷策略。基于层次分析法建立多元回归模型对信贷风险量化分析,使所得影响因素的权重更加合理,决策更加准确。通过将贷款额度、信贷风险、年利率等以分等级的方式进行信贷决策,决策层次清晰明了。将信贷风险分等级,根据风险等级进而确定不同信贷决策,没有将风险等级与决策建立准确函数关系,但本文建立的模型通过检验均可行,证明方法可行。

#### 参考文献:

- [1] 周延鲸. 银行信贷风险量化研究 [J]. 系统工程, 1997, 15 (3) :28-31, 45.
- [2] 司守奎, 孙玺菁. 数学建模算法与应用 [M]. 北京: 国防工业出版社, 2014.
- [3] 田军, 张朋柱, 王刊良, 等. 基于德尔菲法的专家意见集成模型研究 [J]. 系统工程理论与实践, 2004, 24 (1) :57-62, 69.

- [4] 孙晓元, 解北京, 薛洪来, 等. 基于隶属度分析的回采面冲击地压微震能量阈值判定 [J]. 中国煤炭, 2016, 42 (1) :35-40.
- [5] Mo Qi Kong, Jian Xin Tang, Si Ming Yu. Financial risk assessment of an ocean shipping company based on the AHP[J]. Journal of Coastal Research, 2020, 106 (sp1) :481.
- [6] Smales L A. News sentiment and bank credit risk[J]. Journal of Empirical Finance, 2016, 38:37-61.
- [7] 李萍, 毛琼, 王新颖, 等. 基于多元线性回归分析的操作成本预测模型建立与应用 [J]. 石油规划设计, 2018, 29 (3) :33-37, 52.
- [8] 宋海萍. 基于灰色系统模型的矿区地表沉降预测研究 [J]. 矿山测量, 2020, 48 (4) :35-37, 56.
- [9] 张永娟, 魏夕原, 申伟平, 等. 层次分析法在长城钻探公司审核标准中的应用 [J]. 油气田环境保护, 2020, 30 (4) :65-69, 77.
- [10] 严长春, 生佳根, 於跃成, 等. 基于主题模型包含突发因素的推荐算法研究 [J]. 计算机与数字工程, 2020, 48 (6) :1304-1308, 1366.