

DOI:10.13718/j.cnki.xsxb.2019.07.013

一种深度自编码器面部表情识别新方法^①

何 明

重庆工业职业技术学院 建筑工程与艺术设计学院, 重庆 401120

摘要: 为解决面部表情特征维度高的问题, 该文提出了一种基于深度学习自编码器的表情识别新方法, 该方法利用深度自编码器在多层隐层上进行特征选择, 能够在较低维度上表示高维度的面部特征. 首先采用定向梯度直方图从面部表情的选定区域提取特征, 然后在多个层面上使用深度自编码器, 得到最优编码特征, 降低特征维度, 最后使用支持向量机模型对降维特征进行分类. 实验表明, 与其他现有特征选择和降维技术相比, 该文方法提取的特征优于其他特征, 并能够有效实现面部表情识别.

关键词: 定向梯度直方图; 深度自编码器; 支持向量机; 表情识别

中图分类号: TP391

文献标志码: A

文章编号: 1000-5471(2019)07-0081-06

面部表情识别可以反映人脸情绪, 现已经得到广泛研究^[1]. 面部表情识别由以下部分组成: 人脸检测(使用图像处理技术从整个图像中检测人脸)、特征提取(从检测到的脸部提取关键特征)和图像分类(使用机器学习模型对图像提取特征进行分类)^[2]. 面部表情是视觉上可观察到的非语言交流信号, 是表达和理解情绪的关键机制. 情绪识别领域的大部分研究都集中在对 6 种情绪(愤怒、恐惧、厌恶、喜悦、惊奇和悲伤)的分类上.

对于特征提取方法, 面部表情识别中广泛使用的特征描述技术有梯度方向直方图(histograms of oriented gradients, HOG)、韦伯局部描述等方法^[3-4], 而梯度方向直方图已被证明在面部表情识别中效果显著. 不过由于多视点特征的复杂性, 使得提取面部表情特征的维度通常很高, 因此诸如主成分分析、局部二元模式和非负矩阵分解等技术^[5-6]正被用于解决特征高维问题, 在较低维度中表示最相关的特征来代替高维特征.

由于机器学习技术的特性, 使其更适应于非线性变换的自动特征学习和建模, 人们通过使用机器学习算法来进行各种面部特征分类, 如在文献[7]中提出了深度网络面部表情识别方法, 深度网络是基于两个不同的模型, 一种模型是从图像序列中提取时间外观特征, 而另一种模型是从时间面部界标点提取时间几何特征, 这两种模型结合使用一种新的集成方法, 以提高面部表情识别的性能.

文献[8]中提出了一种从面部显著区域提取融合特征的简化算法框架来实现面部表情识别, 该算法将梯度方向直方图与本地二进制模式相结合提取显著区域特征, 并使用主成分分析方法减少提取的特征维数, 最后对减少的特征应用几个分类器, 获得较高的识别率.

自编码器模型已被广泛地用于分类和识别问题. 文献[9]中提出一种面部表情识别的新框架, 以高精度自动区分表情. 该方法将面部几何特征和外观特征组合而成的高维特征引入到面部表情识别中, 使用稀疏的自编码器网络进行特征学习和分类. 该方法表情识别精度达到 95.79%, 但是以计算复杂度为代价.

在研究了现有方法的基础上, 本文提出了一种深度堆叠自编码器的面部表情识别新方法. 该方法利用

① 收稿日期: 2018-05-16

基金项目: 重庆市社会科学规划项目(2017YBYS108).

作者简介: 何明(1980-), 女, 硕士, 讲师, 主要从事数字化艺术设计及信息化视觉传达研究.

自编码器网络特性,在较低维度上有效地表示高维度面部特征,同时能够检测到最相关的特征,最终使用支持向量机(SVM)对特征进行分类,大大提高了面部表情识别率.

1 基于深度自编码器面部表情的识别方法

本文方法中面部表情识别通过人脸检测、特征提取、深度自编码器特征降维和 SVM 特征分类 4 个步骤实现. 对于人脸检测算法,本文采用的是 Viola-Jones 人脸检测算法,检测并提取人脸区域,消除影响识别率的冗余区域,通过这种方式获得嘴巴和眼睛周围的脸部局部区域,这些部位代表了最有鉴别力的信息. 本文面部表情识别算法流程如图 1 所示.

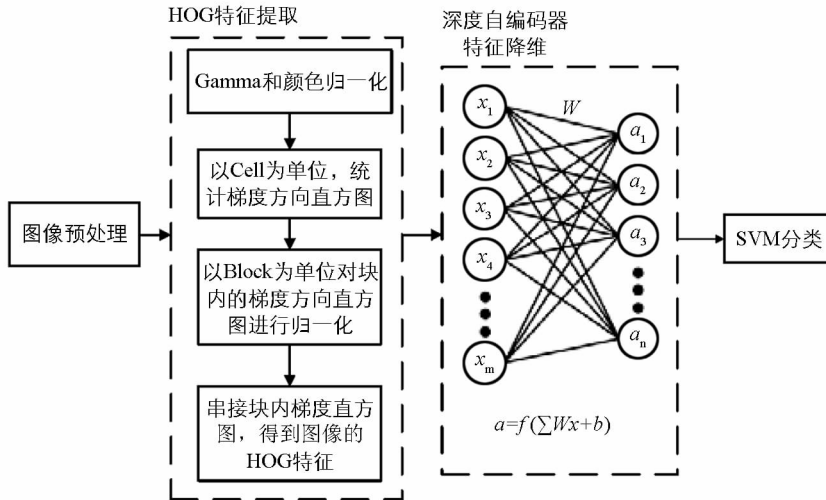


图 1 本文算法流程图

1.1 HOG 特征提取

定向梯度直方图(HOG)是一种广泛用于计算机视觉和图像处理的特征描述方法,除了对象的方向,HOG对于几何和光度转换是不变的.数据库中的图像具有不同表情和不同的眼睛、鼻子和嘴角的方向.本文表情识别方法将HOG应用于人脸图像特征向量提取,128×128尺寸的裁剪图像使用HOG得出尺寸为1×8100的特征向量,特征向量连接起来形成特征矩阵.对于输入图像像素点 (x,y) 在水平方向梯度 $G_x(x,y)$ 和垂直方向梯度 $G_y(x,y)$,则有

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

$$G_y(x,y) = I(x,y+1) - I(x,y-1) \quad (2)$$

其中, $I(\cdot)$ 表示像素点的位置,则像素点 (x,y) 的梯度幅值 $G(x,y)$ 和梯度方向 $\theta(x,y)$ 可以表示为

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \quad (3)$$

$$\theta(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \quad (4)$$

然后,使用高斯矩阵对梯度向量幅值做点乘,对像素块边缘进行暗化,建立梯度方向直方图,对直方图归一化以后得到HOG特征描述值.

1.2 深度自编码器特征降维

机器学习可以在较低维度上获得最相关特征的整体表示,由于提取的HOG特征具有较高的特征尺寸,通过使用不同的降维技术(如主成分分析法PCA、线性判别分析法LDA和非负矩阵分解法NMF)来降低特征向量的维数.本文采用深度自编码器进行特征降维.

自编码器是一种无监督的体系结构,在其输出处复制给定的输入,其采用输入特征向量 X ,并通过将原始输入数据从一个表示改变到另一个表示来学习代码字典.自编码器通过将目标值设置为输入来应用反向传播,自编码器的结构如图2所示.

长度为 m 的输入向量 X 被编码为长度为 n 的低维特征向量 a ,其中 $n < m$,然后被重构成长度为 m 相

似于 x 的向量 y . 如果自编码器输入相关的结构数据, 则网络将发现这些相关性, 在自编码器中较低维度向量 a 表示为

$$a = f(\sum Wx + b) \quad (5)$$

其中, W 是与输入单元和隐藏单元相关联的权向量, b 是与隐藏单元相关的偏置向量, a 是网络中隐藏层的输入, $f(\sum Wx + b)$ 表示非线性输入映射函数, 本文采用 sigmoid 函数作为映射函数, 表示为

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (6)$$

$$z_i = \sum_{j=1}^m W_{ij}x + b_i \quad (7)$$

其中, W_i 是第 i 个输入单元和隐藏单元相关联的权值向量, b_i 是与隐藏单元相关的偏置向量, 低维特征向量可以表示为 $a_i = f(z_i)$, 具体表达式见式(8), 即为深度自编码器数学描述.

$$a_i = f(\sum W_i a_{i-1} + b_i) \quad (8)$$

即使当隐藏单元数量很大时, 自编码器也可以通过对隐藏单元施加稀疏约束来发现数据的相关结构, 这种架构被称为稀疏自编码器. 稀疏自编码器的代价函数 $J(W, b)$ 为

$$J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \| h_{w,b}(x^{(i)}) - x^{(i)} \|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{L-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 + \beta \sum_{j=1}^{s_2} KL(\rho \| \hat{\rho}_j) \quad (9)$$

其中, 式(9)中第一部分公式尽量减小输入与输出的差异, 第二部分是避免过度拟合的权重衰减, $h_{w,b}(x)$ 是激活函数, W 和 b 分别表示权重矩阵和偏置矩阵, λ 是权重衰减参数, L 是自编码器网络层数, s_l 表示第 l^{th} 层的单元数, $W_{ji}^{(l)}$ 表示层 l 第 i 个单元和层 $l+1$ 第 j 个单元间的权重值, $b_i^{(l)}$ 表示层中 $l+1$ 第 i 个单元的偏置量, $\beta \sum_{j=1}^{s_2} KL(\rho \| \hat{\rho}_j)$ 表示稀疏惩罚项, β 为惩罚系数, ρ 是稀疏参数, KL 表示 Kullback-Leibler (KL) 散度, 其表达式为

$$KL(\rho \| \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \quad (10)$$

式中 ρ 被设置为接近于 0 的值. KL 散度是用于测量两个不同分布之间差异的标准函数.

本文中使用的 HOG 提取的特征输入到自编码器网络, 通过限制隐藏层中的隐藏单元将其编码在较低维度上. 另外, 通过实验来验证增加深度自编码器隐藏层数与表情识别率的关系, 为了保证自编码器获得编码特征的质量, 本文使用反向传播来微调网络参数, 并使用均方误差 (MSE) 作为损失函数.

1.3 支持向量机面部表情分类

支持向量机是二元和多类分类问题的强大工具, 最初 SVM 被设计用于二进制分类, 以最大化的余量分离二进制类数据, 但是对于现实世界的问题, 通常需要区分两个以上类别的数据. 因此, 两个有代表性的方案存在于支持向量机中, 即一对一和一对多类数据进行分类. 本文使用具有高斯核的 SVM 一对多分类方法.

在一对多方案中, SVM 构造 k 个独立二元分类器对 k 类数据进行分类, 第 m 个二元分类器通过使用第 m 类的数据作为正训练样本, 其余 $k-1$ 个数据为负训练样本. 在实验过程中, 类标签由二元分类器进行预测, 从而给出最大输出值, 对于训练数据 $x_i (i = 1, 2, \dots, N)$ 和对应标签 $y_i = \pm 1$ 的二元分类任务, 可以将判定函数表示为

$$f(x) = \text{sign}(w^T + b) \quad (11)$$

其中, $w^T + b = 0$ 表示分离超平面, w 是与分离超平面垂直的权向量, 并且 b 表示超平面的偏置向量,

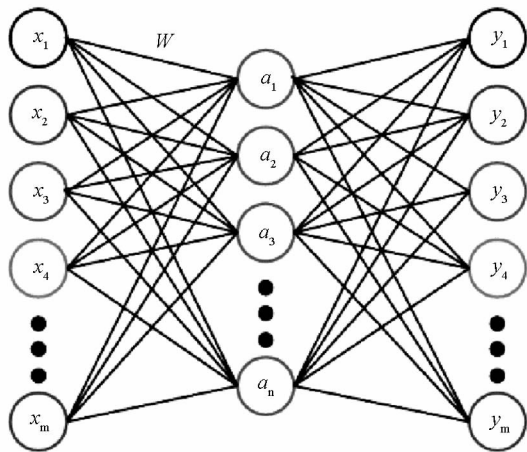


图 2 自编码器网络的体系结构

以下是超平面之间的区域,也称为边缘带.

$$\gamma = \frac{2}{|\omega|} \quad (12)$$

最后,选择 ω 和 b 的最优值作为优化问题. 其中,式(12)受到以下约束最大化为

$$y_i(\omega^T x_i + b) \geq 1 \quad \forall i \quad (13)$$

其中, x_i 为训练数据, y_i 为对应标签, ω 是与分离超平面垂直的权向量, b 表示超平面的偏置向量.

2 实验结果与分析

使用 CK+数据库数据来对本文面部表情识别方法进行评估. 该数据集包含来自 123 个主题的 593 个图像序列,其中 327 个被赋予 7 个人脸表情的标签,本文用了 6 个表达式(即愤怒、厌恶、恐惧、快乐、悲伤和惊奇),每个表达图像序列以中性表达开始,并以峰值表达(即愤怒)结束. 在本文实验中,使用每个表达式的 5 个峰值图像来合并表达式的时间信息.

图 3 显示了用于训练的 6 种情绪的图像序列. 本文使用 80% 的数据进行训练,而使用剩下 20% 数据进行测试. 在实验过程中,只使用每个表达式的一个峰值图像.

本次实验对本文深度自编码器和 PCA 获得不同长度的特征进行了对比实验,如表 1 中所示.



(a)愤怒、(b)厌恶、(c)恐惧、(d)快乐、(e)悲伤和(f)惊奇

图 3 各种情绪 5 个图像的样本序列

表 1 使用不同维度特征的表情识别准确度对比

特征数量	PCA(准确性)/%	本文方法(准确性)/%
10	88.34	97.80
20	90.29	98.10
40	96.01	98.50
60	96.11	99.60
80	96.44	98.10
100	94.17	98.40
200	94.82	97.84
300	95.15	96.28
400	95.46	96.98
500	95.45	96.98

由表 1 中数据可知,堆叠自编码器获得的编码特征优于 PCA 性能,采用自编码器进行降维处理,在 60 维的情况下识别率达到了 99.60%,PCA 的识别率在维度为 80 维时达到了 96.44%.

另外,本文还对自编码器网络中增加隐藏层与识别效果的关系做了实验,引入更多具有不同数量的隐藏层(即 500,400,300,200),编码特征数量范围为 5~100. 本文使用 5 个自编码器进行实验,对自编码器网络中隐含层数的数量与识别效果的关系进行研究(表 2).

表 2 不同特征维度在多个隐藏层上实现的识别率

特征数量	隐藏层 1	隐藏层 2	隐藏层 3	隐藏层 4	隐藏层 5
5	78.60%	89.00%	90.10%	93.40%	94.90%
10	78.40%	93.00%	94.90%	97.80%	97.10%
20	77.70%	95.20%	98.50%	98.10%	97.80%
40	96.20%	98.30%	98.90%	98.50%	98.10%
60	96.90%	98.70%	99.20%	99.60%	98.80%
80	97.60%	98.90%	98.40%	98.10%	97.30%
100	97.90%	98.50%	98.90%	98.40%	98.20%

表 2 显示了引入不同数量隐藏层的实验结果，由此可知隐藏层数量越多，不一定会提高精度，在每个特征数的隐层数达到一定数量后精度开始下降，例如当特征维度为 80 时，隐藏层增加到第二层的时候识别率最高。

另外，将本文方法和新的一些面部表情方法进行对比实验，具体对比结果见表 3。

表 3 本文方法与其他方法面部识别准确率对比

研究	识别方法	识别准确率/%
文献[10]	本地 Gabor 特征+Adaboost 分类器	95.10
文献[11]	配对功能	98.07
文献[12]	HOG+LDA+PCA	99.51
文献[13]	本地二进制模式(LBP)+HOG 与 PCA	98.30
文献[14]	CNN+Softmax	98.10
本文方法	HOG+深度自编码器	99.60

从表 3 中数据可以知，使用文献[12]中方法得到的最大识别准确率是 99.51%，本文方法得到 99.60% 的识别精度，高于现有的识别方法。

虽然本文提出的方法已经达到了较好的识别率，但是自编码器的时间复杂度线性地依赖于特征和隐藏层的数量，特征或隐藏层数量越多，训练模型所需的时间就越多。未来的工作将要考虑如何减少训练模型需要的时间，提高表情识别效率。

3 结 论

本文提出了一种深度自编码器的面部表情识别新方法，该方法使用 HOG 方法提取面部特征，利用深度自编码器特性，将高维度面部特征在较低维度上表示，使用 SVM 进行分类，实现面部表情识别，具有降低特征维度的性能。在 CK+数据集上的实验结果表明，使用自编码器的非线性降维比 PCA 的线性降维性能更有效；对自编码器网络中隐含层数数量与识别效果的关系进行研究，增加了网络的学习能力，为面部表情识别提供了更好的理论支撑。与现有表情识别方法相比，本文方法具有更好的识别准确度，高达 99.60%，证明了本文方法的有效性。

参考文献：

- [1] 徐琳琳, 张树美, 赵俊莉. 基于图像的面部表情识别方法综述 [J]. 计算机应用, 2017, 37(12): 3509-3516, 3546.
- [2] 杨雨浓, 房鼎益, 王 洪. 一种基于混合深度置信模型的面部表情识别方法 [J]. 西南大学学报(自然科学版), 2016, 38(6): 142-147.
- [3] KHAN S A, HUSSAIN A, USMAN M. Reliable Facial Expression Recognition for Multi-Scale Images Using Weber Local Binary Image Based Cosine Transform Features [J]. Multimedia Tools and Applications, 2018, 77(1): 1133-1165.
- [4] 何 俊, 何忠文, 蔡建峰, 等. 一种新的多角度人脸表情识别方法 [J]. 计算机应用研究, 2018, 35(1): 282-286.
- [5] 杨 勇, 蔡舒博. 一种基于两步降维和并行特征融合的表情识别方法 [J]. 重庆邮电大学学报(自然科学版), 2015, 27(3): 377-385.
- [6] SIDDIQI M H, ALI R, KHAN A M, et al. Human Facial Expression Recognition Using Stepwise Linear Discriminant Analysis and Hidden Conditional Random Fields [J]. IEEE Transactions on Image Processing, 2015, 24(4): 1386-1398.
- [7] JUNG H, LEE S, YIM J, et al. Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition [C]. Santiago; IEEE International Conference on Computer Vision (ICCV), 2015.
- [8] LIU Y P, LI Y B, MA X, et al. Facial Expression Recognition with Fusion Features Extracted from Salient Facial Areas [J]. Sensors, 2017, 17(4): 712-728.
- [9] ZENG N Y, ZHANG H, SONG B Y, et al. Facial Expression Recognition Via Learning Deep Sparse Autoencoders [J]. Neurocomputing, 2018, 273: 643-649.
- [10] XING Y X, LUO W. Facial Expression Recognition Using Local Gabor Features and Adaboost Classifiers [C]. Shanghai; 2016 International Conference on Progress in Informatics and Computing (PIC), 2016.
- [11] COSSETIN M J, NIEVOLA J C, KOERICH A L. Facial Expression Recognition Using a Pairwise Feature Selection and

Classification Approach [C]. Vancouver: 2016 International Joint Conference on Neural Networks (IJCNN), 2016.

- [12] KAR N B, BABU K S, JENA S K. Face Expression Recognition Using Histograms of Oriented Gradients with Reduced Features[M]//Advances in Intelligent Systems and Computing. Singapore: Springer Singapore, 2016.
- [13] LIU L, FIEGUTH P, GUO Y L, et al. Local Binary Features for Texture Classification: Taxonomy and Experimental Study [J]. Pattern Recognition, 2017, 62: 135-160.
- [14] 杨金鑫, 杨辉华, 李灵巧, 等. 结合卷积神经网络和超像素聚类的细胞图像分割方法 [J]. 计算机应用研究, 2018, 35(5): 1569-1572, 1577.

A New Method for Facial Expression Recognition Based on Deep Autoencoders

HE Ming

Institute of Construction Engineering and Art Design, Chongqing Industry Polytechnic College, Chongqing 401120, China

Abstract: To solve the problem of high facial expression feature dimensions, a new method for facial expression recognition based on deep autoencoders has been presented in this paper. In this method, autoencoders is used to perform feature selection on multi-layer hidden layers, which can be able to represent high-dimensional facial features in lower dimensions. Firstly, the histograms of oriented gradients is used to extract features from selected regions of facial expressions. And then deep autoencoders is used on multiple levels to obtain the best coding feature, reducing the feature dimension. Finally, the support vector machine model is used to classify the dimensionality reduction features. Experiments show that the features extracted from the deep autoencoder outperformed when compared to other feature selection and dimension reduction techniques, and can effectively realize facial expression recognition.

Key words: histograms of oriented gradients; deep autoencoders; support vector machine; expression recognition

责任编辑 夏 娟