

基于卷积神经网络和池化算法的表情识别研究

陈航,邱晓晖

(南京邮电大学 通信与信息工程学院,江苏 南京 210003)

摘要:卷积神经网络(CNN)能够通过神经网络自主学习提取图像中的特征,并且具有局部响应、权值共享等优点,在人脸表情识别中获得了广泛的应用。池化算法是CNN的核心技术之一,通过对卷积层的特征进行聚合统计,池化算法可以减少CNN的特征维度,提高特征表征能力,但是目前常用的池化算法还存在提取特征单一,缺乏灵活性的情况。为了克服现有池化算法的不足,根据深度学习可采用BP算法自主调节参数的特性,提出一种改进的自适应池化算法。该算法在训练过程中能够根据损失函数,不断更新池化域的参数,最终使表情预测值和真实结果值之间的差值达到最小。基于CK+人脸表情数据库的实验结果表明,与现有池化算法相比,提出的自适应池化算法能有效提高表情识别准确率。

关键词:卷积神经网络;池化算法;人脸表情识别;深度学习;特征提取

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2019)01-0061-05

doi:10.3969/j.issn.1673-629X.2019.01.013

Research on Emotional Recognition Based on Convolution Neural Network and Pooling Algorithm

CHEN Hang, QIU Xiao-hui

(School of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Convolution neural network (CNN) can extract the features of the image by neural network autonomy learning with the advantages of local response and weight sharing, and has been widely used in facial expression recognition. The pooling algorithm is one of the core technologies of CNN. By aggregating the features of the convolution layer, the pooling algorithm can reduce the feature dimension of the CNN and improve the capability of feature representation. However, the commonly used pooling algorithms also have the disadvantage of single extraction feature and lack of flexibility. For this, we propose a modified adaptive pooling algorithm based on the characteristics of depth self-tuning parameters which can be adjusted by BP algorithm. According to the loss function, updating the parameters of the pooled domain eventually minimizes the difference between the predictive value of the expression and the true result value. The experiment based on CK+ face expression database shows that compared with the existing pooling algorithm, the proposed adaptive pooling algorithm can effectively improve the accuracy of facial expression recognition.

Key words: convolutional neural network; pooling algorithm; facial expression recognition; deep learning; feature extraction

0 引言

面部表情是人类用来表达个人情绪和心理的主要方式之一,在人际间的相互交往中发挥着重要的作用,而且往往比语言、动作更加生动真实,在虚拟现实、医学、心理学等领域有着广阔的应用前景。在表情识别中特征提取是影响表情识别的重要步骤,传统的表情识别方法有主元分析法(PCA)^[1]、Gabor小波法^[2]和局部二值模式(LBP)^[3-4]等,这些方法都是通过某种人工精心设计的算法提取具有特定分类信息的特征,

但是这种方法并不能完全提取图像中的所有信息,或者这些提取的特征都有特定的使用领域,往往不能适应复杂的现实环境。在2006年G E Hinton^[5]提出深度学习的概念以后,这个方向就成为了克服传统方法的突破点,并且在计算机视觉和模式识别等领域已经取得了显著的成就,其中卷积神经网络^[6-7]由于其局部响应、权值共享等优点,在表情识别方面获得了大量的应用。它不同于传统的依靠人工设计的特征提取方式,而是构建多层的深度神经网络,让机器自主地从样

收稿日期:2018-02-04

修回日期:2018-06-05

网络出版时间:2018-11-15

基金项目:江苏省自然科学基金(BK2011789);东南大学毫米波国家重点实验室开放课题(K201318)

作者简介:陈航(1993-),男,硕士研究生,研究方向为图像处理与模式识别;邱晓晖,博士,教授,研究方向为信号与信息处理与模式识别。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20181114.1554.002.html>

本数据中提取信息,并通过每层的网络参数逐层抽象、提取特征信息,从而获得更具有推广性和表征能力的特征。

从 2012 年, Krizhevsky 等^[8]首次将 CNN 应用于 ILSVRC 中,并在 ILSVRC-2012 挑战赛中他们所训练的深度卷积网络在图像分类和目标定位任务中都取得第一名,并打破了图像分类的世界记录。凭借在图像分类中的突出表现,越来越多的研究人员把 CNN 运用到更为细致的分类中,例如人脸识别^[9-10]、表情识别^[11-12]等等,并取得了一定的成果。

除输入和输出层以外,卷积神经网络主要由卷积层、池化层、全连接层构成,其中池化层的作用是对卷积层的输出特征进行聚合统计,目的就是提高特征的表征能力、减少特征维度。此外,池化层通过池化算法还可以有效避免网络在训练过程中因参数过多、模型过于复杂产生的过拟合现象。但是传统的池化算法并不能适用于所有情况,最大值池化算法未能考虑池化域中特征值之间的关系,而均值池化算法在特征信息较为集中的情况下会减弱特征信息,随机池化算法虽然考虑了特征值之间的关系但是传给下一层神经网络

时,并未包含所有的特征信息。文中在现有池化算法的基础上,结合人工神经网络的特点提出一种自适应池化算法,并应用于表情识别的卷积神经网络中,以提高表情识别的准确率。

1 卷积神经网络

卷积神经网络是一种结构上由输入层、隐层、输出层构成的深度神经网络,如图 1 所示,其隐层主要由卷积层、池化层、全连接层组成,卷积层主要通过卷积操作实现局部特征响应,然后使用相同的卷积核扫描整幅图像,提取整幅图像的特征,实现权值共享。一般每个卷积层都会对应多个不同的卷积核,每个卷积核提取出的图像特征称之为特征图谱,具体计算如下:

$$y_j^l = \theta \left(\sum_{i=1}^{N_j^{l-1}} \omega_j \otimes x_i^{l-1} + b_j^l \right), j = 1, 2, \dots, M \quad (1)$$

其中, y_j^l 表示第 j 个特征图谱; ω_j 表示对应的卷积核; x_i^{l-1} 表示上层第 i 个特征图谱作为当前的输入; b_j^l 表示偏置量; N_j^{l-1} 表示每个特征图谱的特征数量; M 表示每个卷积层的特征图谱数; $\theta()$ 为激活函数,常见的有 tanh、sigmoid、ReLU 等。

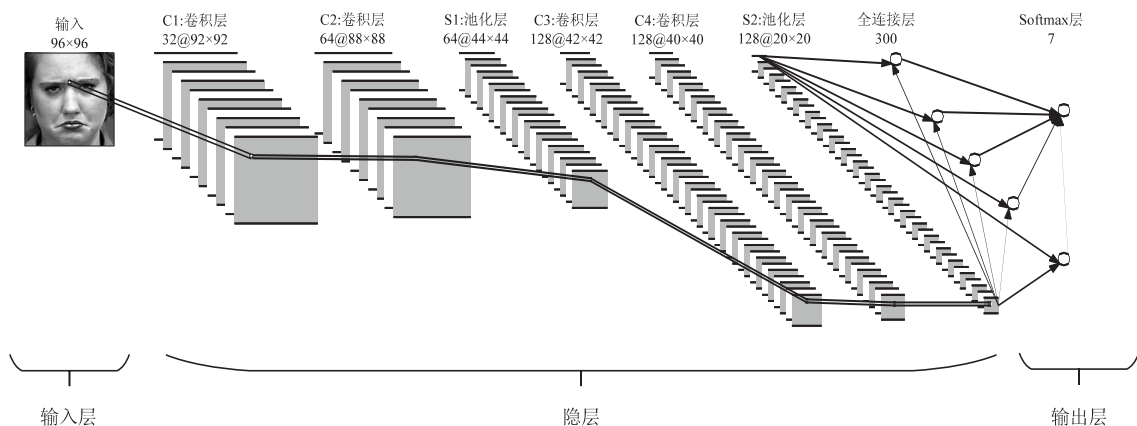


图 1 卷积神经网络结构

在卷积神经网络中卷积层的输出是图像的特征,有时在卷积层之后会紧跟着池化层,而池化层的主要作用就是对特征进行聚合统计,目的就是提高特征的表征能力、减少特征维度,而参数过多、模型过于复杂也是产生过拟合现象的重要原因之一,因此池化算法还可以有效避免训练过程中的过拟合现象。例如,在获得上一层卷积层的输出之后特征维度为 $n * n * m$ (m 为特征图谱的数量, n 为特征图谱的大小),可以采用 $2 * 2$ 的池化窗口进行池化操作,每个 $2 * 2$ 的窗口区域得到一个池化后的特征值,然后对每个特征图谱都进行池化操作,最后得到的池化层的输出维度为 $(n/2) * (n/2) * m$,和池化之前相比特征维度只有原来的 $1/4$ 。一般常用的池化算法有随机池化、最大值池化、平均池化等。

隐层的最后部分一般为全连接层,相当于传统神经网络中多层感知机中的隐层,即前一层的每个神经元都与后一层的每个神经元相连。具体公式为:

$$h_{w,b}(x) = \theta(W^T x) = \theta \left(\sum_{i=1}^3 W_i x_i + b \right) \quad (2)$$

其中, $h_{w,b}(x)$ 表示全连接层的输出; x_i 表示上一层神经元的输出即全连接层的输入; W_i 表示神经元之间连接的权值; b 表示偏置量, $\theta()$ 表示激活函数。

2 基于卷积神经网络的常用池化算法

在卷积神经网络中最常用的池化算法有最大值池化、平均池化和随机池化等。很容易理解,最大值池化算法就是对池化域中的特征值取最大值即可,而平均池化算法就是对池化域中的特征值取均值。最大值池

化和平均池化的表达式分别为:

$$S_{i,j} = \max_{i=1,j=1}^c (x_{ij}) + b \tag{3}$$

$$S_{i,j} = \frac{1}{c^2} (\sum_{i=1}^c \sum_{j=1}^c x_{ij}) + b \tag{4}$$

其中, c 为池化域的大小和步长,池化操作后的特征图为矩阵 S 。

2013 年,Zeiler 提出随机池化操作,并得到广泛的应用。和最大值池化、平均池化不一样的是,同样地输入特征图谱通过随机池化操作后得到的特征输出并不一定是完全一样的,减少了训练过程中可能出现的过拟合现象。在随机池化操作中池化所输出的值是根据池化区域特征值的概率分布来选取的,特征值大被选中的概率也大,但并不能确定一定会选取最大值。在训练过程中的操作步骤为:设给定某个池化域的输入特征值为 x ,首先计算每个特征被选中的概率值 P ,按照每个特征的概率值分布在 0-1 中,然后取 0-1 的随机数 r ,则 r 处于哪个特征值的概率区间内则选取此特征为池化后的值 S 如图 2 所示。

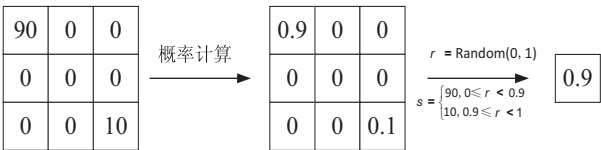


图 2 随机池化过程

而在使用网络模型对表情进行分类预测时,则采用概率加权的方式,即把池化域的特征值和特征值对应的概率值加权作为池化后的特征值。

3 改进的自适应池化算法

3.1 常用池化算法的缺陷

池化算法可以看作卷积运算,不同池化算法的区别仅是选取的卷积核不同,即:

$$S = \sum_{i=1,j=1}^c p_{ij}x_{ij} \tag{5}$$

其中, p_{ij} 为卷积核的参数,如果是平均池化算法则取 $p_{ij} = \frac{1}{c^2}$,如果是最大值池化则最大特征值的参数取 1,其余皆为 0。

假设有两种不同情况的池化域,如图 3 所示。图 3(a)中 v 阴影部分表示不同的特征值,其余特征值为 0,可以很明显地看出图 3(a)中的特征信息集中在 v 处,所以如果此情况使用平均池化算法则会明显地减弱特征信息。同理,在图 3(b)中 v_1, v_2, v_3 分别代表不同的特征值,如果此情况采用最大值池化算法直接选取最大值 v_3 ,而不顾 v_1, v_2 特征所包含的信息,也不管 v_1, v_2, v_3 之间是否有关联,则明显会丢失很多重要的特征信息。所以使用经典的

最大值池化算法对池化域中特征的聚合表示,会对全局的特征表示有一定的弱化。

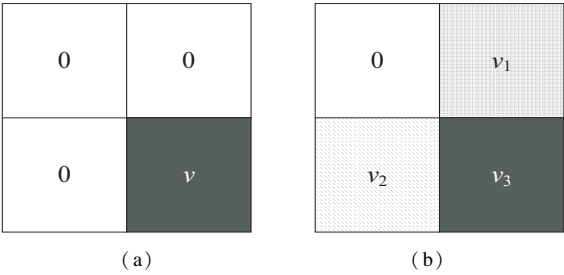


图 3 不同池化域

而采用随机池化算法,可以避免图 3(a)的情况,而且也考虑了图 3(b)中 v_1, v_2, v_3 之间的关系,通过计算概率的方式选取最终的特征值,但是最终的特征值并没有包含 v_1, v_2, v_3 中所有的特征信息,而在池化层中通过池化得到的特征值还是需要传给下一层神经网络继续进行运算,所以随机池化算法在传给下一层的过程中还是会丢失一部分重要的信息,并没有完全解决图 3(b)中出现的

3.2 改进的自适应池化算法原理

通过以上分析可以知道,传统池化算法中的卷积核参数不仅在卷积神经网络训练过程中取值固定,而且也不能根据具体样本及训练过程变化情况选取合适参数,存在特征信息提取模糊,甚至丢失等缺陷。

人工神经网络的最大特点就是能够模仿人脑自主学习的功能,人脑学习的过程是通过人的阅读、听等吸收外界信息的过程,而人工神经网络的学习过程靠的是在样本集上的训练过程,即寻找复杂函数关系式的过程。当神经网络所定义的函数式满足训练样本集时,则此网络模型训练完成,相反如果不满足训练样本集,则重新调整网络模型的参数,直到所定义的函数关系式能够满足训练样本集。

因此针对传统池化算法的缺陷,利用神经网络能够基于大量训练集、通过监督训练不断修正参数的特点,得到适用于不同场景的、自适应动态修正的卷积核参数。

3.3 自适应池化算法实现步骤

提出一种自适应改进池化算法。首先初始化池化参数,也就是卷积核的参数 p ,训练过程中通过卷积操作得到池化值,然后在训练迭代过程中使用梯度下降法更新参数 p ,通过不断的迭代学习不断的优化参数,最终实现自适应池化操作。

文中的自适应池化算法主要分为三个阶段。

(1)输入人脸图像灰度矩阵,通过卷积神经网络的映射关系输出对人脸表情类别的预测值。

(2)根据预测值和真实的结果值,通过损失函数计算出二者差值。一般在卷积神经网络的训练过程中

常用的损失函数有 SVMLoss 和 SoftmaxLoss,文中采用后者,具体表达式为:

$$\text{Loss} = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] + \frac{\lambda}{2} \sum_{i=1}^m \sum_{j=1}^n \theta_{ij}^2 \tag{6}$$

其中, y 为最终的预测值; θ 为神经元之间的权值; $1\{\}$ 为示性函数,当 $\{\}$ 内的表达式为真则取 1,否则取 0,表达式第二项为权重衰减项。

在每次训练迭代的过程中,根据 Loss 值调整网络模型参数达到损失函数的最小值,也就是神经网络的预测值能够最接近实际的结果。

(3)根据损失函数通过梯度下降法更新池化参数 p ,目的是找到最优的池化参数,即损失函数的极小值。具体公式为:

$$p_{n+1} = p_n - \gamma_n \cdot \nabla f(p_n) \tag{7}$$

其中, γ 为步长或学习率; $f()$ 为损失函数。

这样通过不断的迭代更新逐步收敛到损失函数的极值,也就得到了池化算法的参数值。

4 池化算法性能的仿真分析

4.1 卷积神经网络结构设计

根据表情识别的具体情况设计了一种卷积神经网络结构,其中整个网络模型由输入层、隐层、输出层组成,其中隐层包括 4 层卷积层、2 层池化层、1 层全连接层共 7 层。首先数据层的输入数据为 96 * 96 的人脸灰度矩阵,前 2 层卷积层 (C_1 和 C_2) 都是使用步长为 1,大小为 5 * 5 的卷积核,后两层卷积层 (C_3 和 C_4) 的卷积核步长为 1,大小为 3 * 3,两层池化层 (S_1 和 S_2) 的池化窗口大小为 2 * 2,步长也为 2。从第一层卷积层开始每层神经网络输出的特征图谱数量分别为 32、64、64、128、128、128,然后还包括拥有 300 个神经元的全连接层。最后的输出层为 Softmax 层,包含 7 个神经元,分别对应 7 种表情。

4.2 基于 CK+人脸表情数据集的仿真数据选取

在 CK+^[13]人脸表情数据集上进行实验。首先在 CK+数据集中每个样本的不同表情有一个渐变过程,为了保证数据的有效性,对每个样本的不同表情最多只选取表情达到峰值的三张图像。最终在整个 CK+数据集中一共选取了 981 张表情图像,其中高兴(354 张)、惊讶(249 张)、愤怒(135 张)、厌恶(177 张)、悲伤(84 张)、蔑视(54 张)、恐惧(75 张)。然后对选取后的数据集按照 1 : 5 的比例,将数据分为训练集和测试集,其中为了加强实验数据的有效性,在区分训练集和测试集时,必须确保同一个样本的表情不会同时出

现在训练集和测试集中。

4.3 不同池化算法的对比仿真结果

为了验证提出的动态自适应池化算法的有效性,在相同卷积神经网络结构中采用不同的池化算法,在训练集和数据集上分别进行实验并比较,其结果如表 1 所示。

表 1 不同池化策略和表情识别率

池化策略	识别率/%
最大值池化	86.53
平均池化	82.6
随机池化	83.17
动态自适应池化	84.13

表 1 中自适应池化算法的初始化参数为 $1/c^2$ (c 为池化域的大小),可以知道在训练开始时,自适应池化算法本质上就是平均池化算法,原本平均池化算法在三种经典的池化策略中准确率最低只有 82.6%,但是自适应池化算法经过不停的参数优化,最终卷积神经网络模型的预测结果不仅要高于原本的平均池化策略,而且还略高于随机池化策略(83.17%),仅次于最大值池化策略(86.53%)。经过训练后不同池化域的参数如图 4 所示。

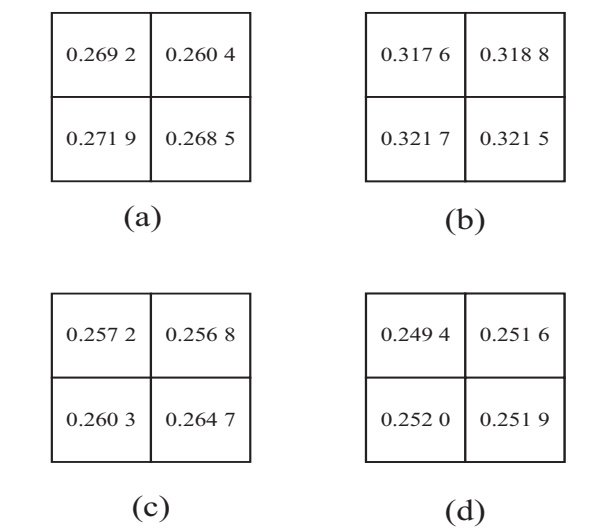


图 4 训练后不同池化域的参数

由图 4 可知,经过训练后的池化域参数并不是固定的 $1/c^2$,而是根据训练过程中的损失函数,不断优化得到,不同的池化区域参数都不相同。可以看出,文中提出的动态自适应池化对于卷积神经网络在表情识别中是切实有效的,并且当初始化条件等价于平均池化时,可以明显看出自适应池化算法的优越性。

当使用自适应池化算法时,在池化域大小为 2 * 2 的情况下,分别使用不同的初始化参数进行多次实验,其结果如表 2 所示。

表 2 不同初始池化参数和表情识别率

初始池化参数	识别率/%
0.25	84.13
0.5	84.61
0.8	90.38
1	94.23

在表 2 中可以明显看出,当初始化池化参数取 1 时,表情识别的最终识别率达到最高值为 94.23%,远远超出表 1 中使用最大值池化操作的 86.53%。结合表 1 和表 2 可以知道,提出的基于动态自适应池化的卷积神经网络在表情识别中,相对于使用传统的池化算法具有更好的表情识别率。

5 结束语

在卷积神经网络中池化操作的主要作用有两个:降低特征维度;聚合上层特征,避免过拟合现象。而传统池化方法中最大值池化和随机池化在聚合上层特征过程中存在丢失信息严重的情况,平均池化操作则会弱化特征信息,缺乏灵活性。针对传统池化策略存在的缺点,提出一种自适应池化算法,不仅能够完全考虑到池化域中所有的特征信息,并且能够根据损失函数不断优化池化参数。基于自适应池化算法构建了卷积神经网络并用于表情识别,在 CK+数据集上的实验证明,使用动态自适应算法相比较传统的池化算法,在表情识别中具有明显的优势,能够显著提高表情识别率。

参考文献:

[1] MORIZET N, AMIEL F, HAMED I D, et al. A comparative implementation of PCA face recognition algorithm [C]//IEEE international conference on electronics, circuits and systems. Marrakech, Morocco:IEEE,2008:865-868.

[2] LIU Chengjun, WECHSLER H. Gabor feature based classification using the enhanced fisher linear discriminant model for

face recognition [J]. IEEE Transactions on Image Processing,2002,11(4):467-476.

[3] AHONEN T, HADID A, PIETIKÄINEN M. Face recognition with local binary patterns [C]//European conference on computer vision. [s. l.]:[s. n.],2004:469-481.

[4] 陆 华. 基于局部二值模式的人脸识别和表情识别研究 [D]. 济南:山东大学,2014.

[5] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006,313(5786):504-507.

[6] 许 可. 卷积神经网络在图像识别上的应用的研究 [D]. 杭州:浙江大学,2012.

[7] 丰晓霞. 基于深度学习的图像识别算法研究 [D]. 太原:太原理工大学,2015.

[8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//International conference on neural information processing systems. Red Hook, NY:Curran Associates, 2012:1097-1105.

[9] HU Guosheng, YANG Yongxin, YI Dong, et al. When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition [C]//Proceedings of the IEEE international conference on computer vision. [s. l.]:IEEE,2015:142-150.

[10] 卢官明,何嘉利,闫静杰,等. 一种用于人脸表情识别的卷积神经网络 [J]. 南京邮电大学学报:自然科学版,2016,36(1):16-22.

[11] 刘 旷. 基于卷积网络集成的面部表情识别方法 [D]. 杭州:浙江大学,2016.

[12] LIU Mengyi, LI Shaoxin, SHAN Shiguang, et al. Au-inspired deep networks for facial expression feature learning [J]. Neurocomputing,2015,159:126-136.

[13] LUCEY P, COHN J F, KANADE T, et al. The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression [C]//IEEE computer society conference on computer vision and pattern recognition. San Francisco, CA, USA:IEEE,2010:94-101.