

基于卷积神经网络的多人行为识别方法

龚安 费凡 郑君

(中国石油大学(华东)计算机与通信工程学院 山东 青岛 266580)

摘要 为了解决多人行为识别中人物角色多且难以区分、图片增加的特征维数难以表达和学习以及行为背景复杂且容易产生干扰等问题,提出了一种基于卷积神经网络的多人行为识别方法。考虑到多人行为识别的复杂性,选择较为容易的两人交互行为作为研究对象,对实验中需要的图像数据库进行了初步的收集与预处理;然后选用在特征提取中不受拍摄角度、光照强度影响的 Dense-sift 算法来对原始图像进行初步的特征提取。由于人体行为图片相对手写字数字图片更为复杂,因此为了使该网络能够很好地识别人体行为,针对该网络在其输入、网络层数、滤波器核数、学习率、输出等方面进行了修改。实验结果表明,提出的方法对拳击、拥抱、接吻 3 类交互行为的识别是有效的。

关键词 多人行为识别,卷积神经网络,Dense-sift 特征提取

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.02.053

Multi-person Behavior Recognition Method Based on Convolutional Neural Networks

GONG An FEI Fan ZHENG Jun

(College of Computer & Communication Engineering, China University of Petroleum, Qingdao, Shandong 266580, China)

Abstract In order to solve the problems in multi-person behavior recognition, for example, it is difficult to distinguish many characters, it is difficult to express and learn increased feature dimension of image, the behavior background is complex and it is easy to cause interference, this paper proposed a method of multiplayer behavior recognition based on convolutional neural network. At first, considering the complexity of multi-person behavior recognition, the simple two-person interactive behavior is chosen as the research object and the picture database is collected. Then, because multiplayer behavior recognition has complicated background and many features in the recognition progress, a method using the Dense-sift algorithm for feature pretreatment mode is proposed. Against the complexity of the multiplayer behavior recognition, this network makes various modifications, such as input dimensions which is expanded to include layer convolution, convolution kernel increasing, output reduction, etc. Experimental results show that the proposed method can recognize simple multi-person behavior recognition, such as boxing, hug and kissing effectively.

Keywords Multi-person behavior recognition, Convolutional neural network, Dense-sift feature extraction

1 引言

随着科技和硬件技术飞速发展,大量的摄像监控设备涌入人们的生活,从而导致图像、视频数据日益增加。目前对于这些数据的利用主要是通过人眼,这使得人们的工作量巨大且效率低下。在人机交互方面,人们的要求越来越高,简单地利用键盘、鼠标进行人机交互已经无法满足人们的需求。人体行为的识别则能够很好地改善这些问题。目前,针对人体行为的识别方法大多是针对单人行为,但是在实际应用中使用较多的却是多人行为识别。

多人行为识别通常包括两人间的行为识别以及 3 人或 3 人以上的群体行为识别。在两人间的交互行为中,如何区分两人的角色关系以及如何有效地提取关键特征一直是研究的

难点。Du 等人^[1]使用动态贝叶斯网络对两人行为进行建模,该方法能对运动细节进行多尺度的特征提取。韩磊等人^[2]将时空单词应用到两人行为识别中。在群体行为识别中,人物角色会有多种情况,其识别方法有所区别。对于开会这种人物角色不同的情况(如有人在听报告而有人在提问),Peng 等人^[3]针对开会过程中的 3 种行为提出了使用多层结构的动态贝叶斯网络进行多人行为辨识。对于游行这种人物角色相同的情况,Ryoo 等人^[4]使用量词对行为事件进行描述,并以此描述来对多人(人群)行为进行辨识分类。

近年来,深度学习以其在特征提取、学习上的优势受到了越来越多的关注。Ji 等人^[5]对二维图像中的 CNN 进行扩展,该方法是深度学习模型在行为识别领域中的一次很好的尝试。Karpathy 等人^[6]利用慢融合模型对不同图像帧进行融

到稿日期:2017-05-07 返修日期:2017-07-25

龚安(1971—),副教授,硕士生导师,CCF 会员,主要研究方向为大数据智能处理,E-mail:414625329@qq.com(通信作者);费凡(1993—),硕士生,主要研究方向为人体行为识别、自然语言处理。

合,构建了基于图像序列的 CNN 模型,提升了行为特征的表达能力。

卷积神经网络作为深度学习的一种,不仅在特征提取上具有优势,而且在进行训练时具有可直接输入图片、无需提前处理的特点,能够在数据输入前减少误差;而且其网络稀疏连接和权值共享的特点大大降低了需要训练的参数数量,简化了算法的训练。因此,本文采用卷积神经网络的方法对多人行为进行识别。由于缺乏相应的图片数据库,实验数据集均由人工在网络上采集,工作量巨大,因此本文仅建立了 3 种图片数据库并对其进行识别。

2 图像预处理

图像预处理是行为识别的第一步。预处理分为两个步骤:使用 Dense-sift^[7] 算法进行特征提取和对特征的数据格式进行处理。Dense-sift 在目标分类和场景分类中有重要的应用,它是一种稠密的 Sift^[8] 算法。对比两种算法,Sift 算法提取到的特征只是感兴趣区域或者是图像上若干个稳定的关键点的特征^[9];而 Dense-sift 算法提取的则是感兴趣的块中每一个位置的 Sift 特征,即每一个位置的 Sift 描述子,该算法与 Sift 最大的不同在于关键点的选取是稠密且同规格的。

图像预处理过程如下:

1)把原始图片划分为固定大小的矩形块

确定一个像素步长作为划分区域的标尺,利用步长 x 对原始图片(分辨率为 $m * n$)进行划分。为了提取边缘不稳定的噪声点,首先将边缘的 $x/2$ 个像素去除,再以 x 个像素为步长对图片进行分割处理,从而可以得到 $\frac{(m-x)}{x} * \frac{(n-x)}{x}$ 个大小为 $x * x$ 的矩形块。

(2)对划分后的小矩形块进行 Sift 特征提取

对划分后的每个小矩形块用 Sift 算法进行特征提取,得到若干个描述子。Sift 描述子是关键点邻域高斯图像梯度统计结果的一种表示,通过对关键点周围的图像区域进行分块,计算块内梯度直方图,来生成具有独特性的向量,这个向量是该区域图像信息的一种抽象,具有唯一性。David Lowe 提出当梯度方向直方图为 $4 * 4$ 维时,Sift 描述子具有最好的区分度^[10],因此本文以每个特征点为中心区域进行采样、合并、累加,最终得到若干个 $4 * 4 * 8$ 维的梯度信息,这些梯度信息即为每个特征点的特征向量。特征向量形成后,为了去除光照变化的影响,对其进行归一化处理,如果图像的对比度发生变化,则每个像素值都会乘以一个数值,这样可以消除归一化后对比度的影响;对于图像亮度的变化,为每个像素值加上一个数值,这对计算梯度是没有影响的,因此描述子对亮度的仿射变换是鲁棒的。由于某些图像亮度变化较大,可能会导致特征点周围像素的梯度变化过大,为了减少对亮度的依赖,在 128 维的特征向量中,滤除梯度大于 0.2 的梯度值(0.2 是实验得出的经验值),然后重新归一化,这样梯度大小的作用就会被削弱,而方向信息的作用会被强化。经过处理,最后得到

$\frac{(m-x)}{x} * \frac{(n-x)}{x}$ 个 128 维的描述子。

3)对描述子进行拆分组合处理

如图 1 所示,将 128 维的描述子中相同方向的维度信息提取出来,并按照其原有的位置关系组合在一起,从而形成一个 $4 * 4$ 维的中间信息。其矩阵表示如下:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

(1)

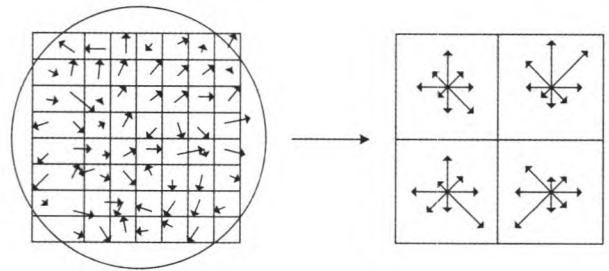


图 1 描述子拆分示意图

Fig. 1 Structure diagram of splitting feature descriptors

由于对描述子进行归一化处理后,每一个维度的数值都不大于 1,且描述子有方向的信息不会出现负值,因此有 $0 \leq \alpha_{ij} \leq 1$ 。将 $\frac{(m-x)}{x} * \frac{(n-x)}{x}$ 个描述子进行相同的拆分,可以得到 $\frac{(m-x)}{x} * \frac{(n-x)}{x}$ 个 $4 * 4$ 维的中间信息。接着将每一个小矩形块的 $4 * 4$ 维的中间信息放到该矩形块相对于原始图像相同的位置进行组合,进而得到一个 $[4 * \frac{(m-x)}{x}] * [4 * \frac{(n-x)}{x}]$ 维的描述信息。其矩阵表示如下:

$$\begin{bmatrix} a_{11} & \Lambda & a_{1N} \\ M & o & M \\ a_{M1} & \Lambda & a_{MM} \end{bmatrix}$$

(2)

这相当于是一张分辨率为 $[4 * \frac{(m-x)}{x}] * [4 * \frac{(n-x)}{x}]$

维的图像的灰度信息,正好可以作为卷积神经网络的输入。如果将每一个方向的特征图按照相同的方式进行拆分组合,那么就能得到 8 张图片的灰度信息,而这 8 张灰度图片将成为整个网络的输入。

3 基于神经网络的多人行为识别模型

3.1 LeNet-5 卷积神经网络

LeNet-5^[11] 是一种比较成熟的卷积神经网络,其准确性可以满足商业用途的要求,国外很多银行利用其识别支票上的手写数字。本文将该网络进行深度研究,并对其进行修改,使其能够应用于多人行为识别。该网络一共包括 7 层(不含输入层),分别为 C1 层、S2 层、C3 层、S4 层、C5 层、F6 层和输出层,其中 C1 层、C3 层、C5 层为卷积层,S2 层、S4 层为下采样层,F6 层是全连接层,如图 2 所示。

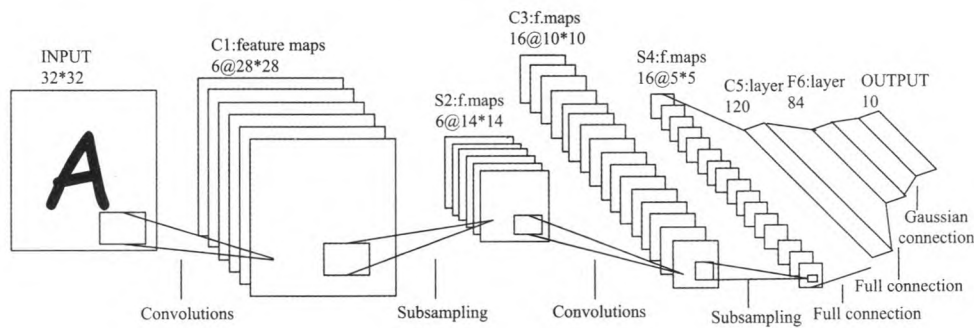


图 2 LeNet-5 的网络结构
Fig. 2 Structure of LeNet-5

3.2 优化卷积神经网络

由于 LeNet-5 网络输入的手写数字图片具有图片尺寸小、图片背景单一、识别目标简单等特点,将其用于多人行为的图片时,输入图片变得十分复杂,因此需要对该网络进行改进。具体改进如下:

1)修改输入。首先将输入图片的大小设置为 $[4 * \frac{(m-x)}{x}] * [4 * \frac{(n-x)}{x}]$ 维。这是因为分辨率为 $m * n$ 的原图片经过预处理后得到的输入图片的分辨率为 $[4 * \frac{(m-x)}{x}] * [4 * \frac{(n-x)}{x}]$ 维。其次为了使预处理后从 8 个不同方向得到的 8 个不同的特征图在训练过程中被充分利用,将原始的只能处理一个通道的灰度图像的网络更改为能够处理 8 个通道的网络。

2)修改网络层数及每一层的输入。由于需进行行为识别的图片较为复杂,如果只依靠原始网络的三次特征抽象,得到的抽象特征无法产生良好的识别效果,因此对网络层数结构进行修改。

对网络层数结构进行修改的具体方案为:在卷积层 C3 层与下采样层 S4 层之后分别添加一层卷积层和一层下采样层,将其称为 C3'与 S4'。

输入大小的改变以及网络层数的修改使得每一层的输入、输出都有所不同,因此需要对其进行修改:

①对于卷积层 C,当卷积邻域为 $a * a$,输入图片大小为 $m * n$ 时,输出图片的大小则为 $(m-a+1) * (n-a+1)$ 。由于原始网络的卷积邻域为 $5 * 5$,因此输出图片的大小将为 $(m-4) * (n-4)$ 。

②对于采样层 S,当采样邻域为 $a * a$,输入图片大小为 $m * n$ 时,输出图片大小则为 $(\frac{m}{a}) * (\frac{n}{a})$ 。由于原始网络的采样邻域为 $2 * 2$,且输入层为上一卷积层的输出,即 $(m-4) * (n-4)$,其输出图片大小为 $\frac{(m-4)}{2} * \frac{(n-4)}{2}$ 。

按照上述算法,可得到每一层的输入、输出参数如下。

- C1 层输入: $[4 * \frac{(m-x)}{x}] * [4 * \frac{(n-x)}{x}]$
- C1 层输出: $[4 * \frac{(m-x)}{x}-4] * [4 * \frac{(n-x)}{x}-4]$
- S2 层输入: $[4 * \frac{(m-x)}{x}-4] * [4 * \frac{(n-x)}{x}-4]$
- S2 层输出: $[4 * \frac{(m-x)}{x}-4] * [4 * \frac{(n-x)}{x}-4] \times \frac{1}{4}$

3)修改滤波器核数。滤波器的核数是每一层能获取到的特征的种类数,也是下一层可用于组合的特征数量。在特征学习中,特征的多样性能够提高学习效果。由于进行行为识别的图片较为复杂,其包含的信息、拥有的特征数相对较多,需要增加原始网络每层滤波核的数量,且由于改进后的网络中增加了一层卷积层,对其进行改进后,滤波核分别为 20,30 和 60。

4)修改学习率。在学习过程中梯度的随机因素称作学习率,学习率是深度学习中非常重要的参数之一,合适的学习率将会产生良好的识别效果。将学习率分别设置为 0.1,0.01,0.05,0.001,0.0001,0.0005 和 0.00001 进行实验,可以发现学习率为 0.001 时的学习效果最好,因此本文将学习率修改为 0.001。

5)修改输出数。输出数需要根据实验中待识别的行为种类进行重新设置。由于本文将识别 3 种不同行为,为了方便起见,用 0,1,2 来表示这 3 种行为。

3.3 算法的完整流程

本文提出的基于卷积神经网络的多人行行为识别方法的主要步骤如下:

- Step 1 对原始图片的分辨率进行处理并进行 Dense-sift 特征提取;
- Step 2 将 128 维特征向量在 8 个方向上进行拆分,将每个方向的特征向量进行组合,形成 8 个方向特征的灰度图像;
- Step 3 将所有由 8 通道组成的灰度图像的图片数据集输入到改进后的卷积神经网络中进行学习和训练,最后进行识别。

本文算法的流程图如图 3 所示。

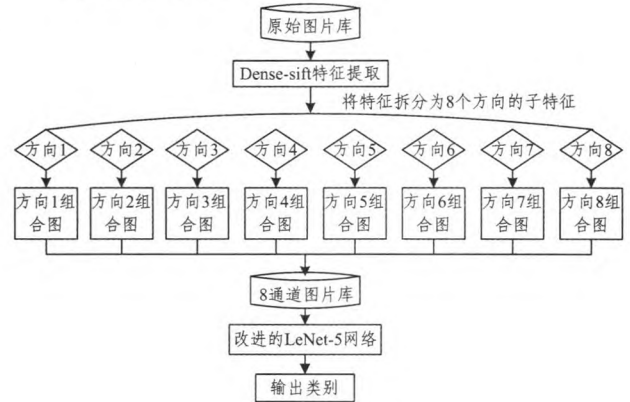


图 3 算法流程图
Fig. 3 Flow chart of the algorithm

4 实验分析

4.1 实验准备

现有的行为识别数据库大多是针对单人行为的视频数据,缺乏双人交互行为和多人行为的图片数据。

针对实验使用的网络模型,在网上进行人工查找和收集,最后形成了初步的图像数据库(每个类别 3000 张图片)。在进行实验之前,对该图像数据库进行以下 3 步的处理。

1)人工标记图片。为了确保每个类别的图片都是被均匀输入卷积神经网络的,本文对数据库中的图片以“随机数”+“_”+“类别”的命名规则来处理,其中 box 代表拳击,hug 代表拥抱,kiss 代表亲吻,结果如图 4—图 6 所示。

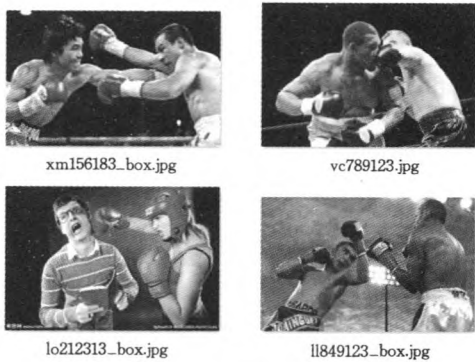


图 4 拳击图片集
Fig. 4 Boxing pictures



图 5 拥抱图片集
Fig. 5 Hug pictures



图 6 接吻图片集
Fig. 6 Kissing pictures

2)筛选图片。从图像库中筛选出背景、光照、角度具有代表性的拥抱、接吻、拳击的图片各 2200 张,将每个类别的 2000 张图片作为训练集,从剩下的 200 张图片中选出 100 张作为验证集,100 张作为测试集。为了得到更加适合卷积神经网络输入的图片,将图片统一设置为 720 * 480 像素大小;为了保留更多的图片信息,这里无需对图片进行灰度化处理。由于需要进行对比实验,同时准备一组经过灰度化处理的图片,其分辨率为 176 * 128。

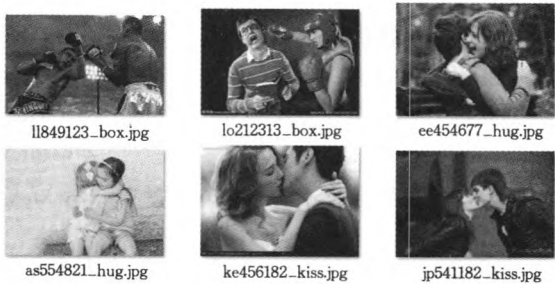


图 7 灰度化后的训练集
Fig. 7 Training set after graying

3)对图片进行特征提取并对特征进行拆分组合。将划分图片的步长设置为 20 个像素,这样一张分辨率为 720 * 480 大小的图片将被分为 35 * 23 张大小为 20 * 20 的小图片;用 Dense-sift 算法对其进行处理,可得到 35 * 23 个 4 * 4 * 8 维的描述子。对所有描述子按照第 2 节中的拆分组合方法进行处理,可以得到 8 个通道的灰度图像,图像分辨率为 (35 * 4) * (23 * 4) = 140 * 72。

如图 8 所示,中间的图片即是原始图像,而其他图片都是由该图进行 Dense-sift 特征提取并进行拆分组合后得到的图像。由图 8 可知,经过处理后得到的灰度图像可以反映原始图像在各个方向的轮廓信息,且经过处理后图像的信息量会减少,有利于卷积神经网络对特征的学习。



图 8 原始图像与 8 通道图像
Fig. 8 Original image and 8 channels image

4)确定网络结构。对于分辨率为 140 * 72 大小的输入图像,按照 3.2 节中的方法对每一层的参数进行修改确定,其结果如表 1 所列。

其中,卷积核个数分别为 20,30,60,即 C1 层、S2 层有 20 个 feature map,C3 层、S4 层有 30 个 feature map,C3'层、S4'层有 60 个 feature map。C5 层有 500 个 feature map。输出一共有 3 类,分别用 0,1,2 代表拥抱、接吻、拳击。学习率为 0.001。

表 1 卷积网络结构参数表(1)

Table 1 Parameter list of convolution network structure(1)

层	输入维数	输出维数
C1	140 * 72	136 * 68
C2	136 * 68	68 * 34
C3	68 * 34	64 * 30
S4	64 * 30	32 * 15
C3'	32 * 15	28 * 11
S4'	28 * 11	14 * 5
C5	15 * 5	500
F6	500	3

4.2 网络结构修改前、后的对比实验

本实验具体的环境配置如下:处理器为 Intel(F) Core(TM) i5-6500 CPU @ 3. 2GHz; 显卡为 NVIDIA GeForce GTX1050TI;内存(RAM)为 8. 00GB;操作系统为 64bit Windows10 企业版;编程平台为 Spyder, Matlab;开发语言为 Python。

1)原始网络

原始网络的层数为 7,卷积核数为 6 和 16,学习率为 0. 1。与 LeNet-5 网络不同的是,这里将输入图片大小改为 176 * 128。其他层输入、输出的网络结构参数如表 2 所列,其训练的结果如图 9 所示。

表 2 卷积网络结构参数表(2)

Table 2 Parameter list of convolution network structure(2)

输入层	维数	输出维数
C1	176 * 128	172 * 124
S2	172 * 124	86 * 62
C3	86 * 62	82 * 58
S4	82 * 58	41 * 29
C5	41 * 29	500
F6	500	3

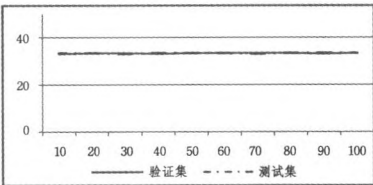


图 9 迭代次数与识别率的关系
Fig. 9 Relationship between iterations and recognition rate

由图 9 可知,在整个训练过程中正确率为 33%,且没有变化,这说明原始网络无法学习到有用的特征。

2)改进后的网络

输入为 8 个通道的分辨率为 140 * 72 的灰度图像,结果如图 10—图 14 所示。

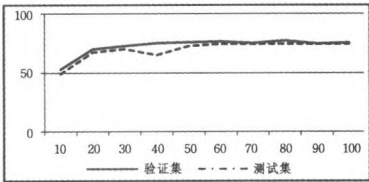


图 10 拳击图片的识别率与迭代次数的关系
Fig. 10 Relationship between recognition rate of boxing pictures and the number of iterations

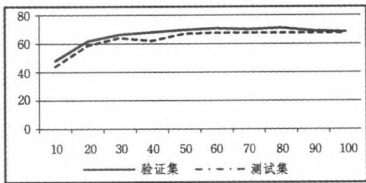


图 11 拥抱图片的识别率与迭代次数的关系
Fig. 11 Relationship between recognition rate of hug pictures and the number of iterations

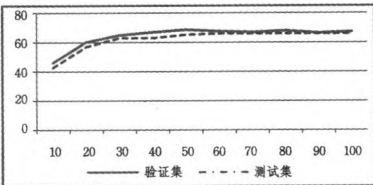


图 12 接吻图片的识别率与迭代次数的关系
Fig. 12 Relationship between recognition rate of kissing pictures and the number of iterations

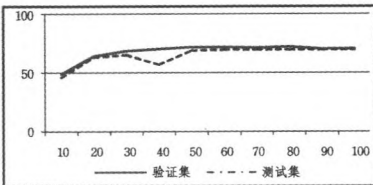


图 13 改进后迭代次数与识别率的关系
Fig. 13 Relationship between the number of iterations and recognition rate after improvement

由图 10—图 13 可知,这 3 种不同种类的图片的识别率在迭代 50 次左右都会趋于一个稳定的状态,其中拳击类图片的识别率最高,达到了 75%左右;拥抱类图片的识别率为 68%左右;接吻图片的识别率达到了 67%左右。由图 13 可知,改进后的网络对行为的识别率从 46%提升到 70%左右,相比于原始网络 33%的识别率有较大的提升;并且整个过程中识别率基本呈上升的趋势,在迭代 50 次之后,识别率基本稳定在 70%左右,这说明其不存在过训练的问题。因此,可以说明该网络结构能够对多人行为图像的特征进行一定的学习。图 14 为 3 种图片的识别率对比,由图可知,拳击图片的识别率高于拥抱和接吻图片的识别率,这是因为拥抱与接吻图片比较相似,它们之间存在相互的干扰,从而使识别率相对较低于拳击。

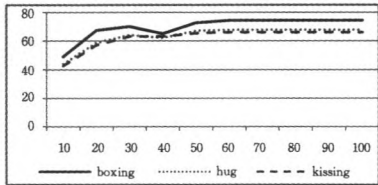


图 14 3 种图片识别率的对比
Fig. 14 Comparison of recognition rates of three pictures

3)结论

通过这两个实验可知,在多人行为识别中使用针对手写数字识别的原始网络时效果很差,这是因为原始网络结构过

于简单,对行为图片不能进行有效的特征提取与特征的学习,因此其识别率仅仅与随机猜测的概率相同。实验证明,使用本文提出的基于卷积神经网络的多人行为识别方法,能够有效地进行特征训练,将其识别率提升到 70%左右,其效果远远超过了原始网络结构。

4.3 学习率对实验结果的影响

1)学习率变化的情况

学习率的变化方式为:初始学习率为 0.01,在训练过程中,每训练 10 个 epoch 时,学习率缩小为原始学习率的 1/10。实验结果如图 15 所示。

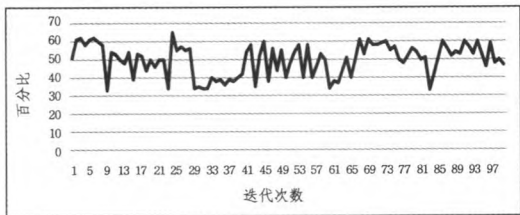


图 15 学习率变化时代次次数与识别率的关系
Fig. 15 Relationship between the number of iterations and recognition rate when learning rate changes

由图 15 可知当学习率变化时,识别率十分不稳定,且没有规律,在 33%~66%之间变动。

2)学习率固定不变的情况

将学习率分别设置为 0.1,0.05,0.01,0.001,0.0005,0.0001,0.00001 时,识别率的实验结果如图 16 和表 3 所示。

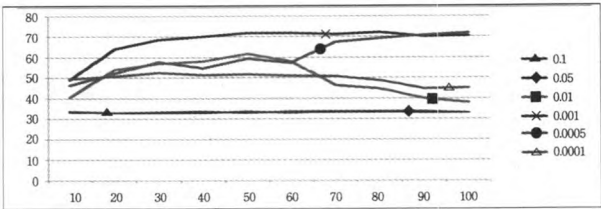


图 16 不同学习率下的识别率对比
Fig. 16 Comparison of recognition rates in different learning rates

由图 16 可知,当学习率为 0.1 和 0.05 时,识别率基本不变化,说明此时未能对特征进行有效的提取、学习;当学习率为 0.01 时,在 50 次迭代之前识别率逐步提升,但在 50 次迭代之后识别率出现了下降的趋势,说明在迭代 50 次后发生了过训练的现象。当学习率为 0.0001 时,其识别率最高仅能达到 51%左右,且之后出现了过训练现象;当学习率为 0.00001 时,其识别率虽然没有出现过训练的现象,但只是达到了 53%,因此都不予选择。当学习率为 0.001 和 0.0005 时,识别率均能逐步提高,最终都提高到 70%左右,并且没有出现下降的趋势,这说明这两个学习率都能使网络进行有效的特征提取和学习,且不会出现过训练的情况。在学习率为 0.001 时,仅迭代 40 次识别率就达到了 70%;而学习率为 0.0005 时,在迭代近 90 次时才能使识别率达到 70%。这说明学习率为 0.001 时,网络结构性能更为优秀,因此在 3.2 节的改进中选择 0.001 的学习率。

表 3 学习率与识别率的关系
Table 3 Relationship between learning rate and recognition rate

学习率	10	20	30	40	50	60	70	80	90	100
0.1	33.67	33.00	33.00	33.00	33.67	33.00	33.67	33.67	33.67	33.00
0.05	33.67	33.00	33.33	33.67	33.00	33.67	33.67	33.67	33.00	33.00
0.01	40.67	54.00	57.00	58.00	61.67	58.00	46.33	44.67	40.00	38.00
0.001	49.00	64.00	68.33	70.00	71.67	71.67	71.00	72.00	70.00	70.33
0.0005	46.33	52.00	57.77	54.77	59.33	57.33	67.33	69.00	70.66	71.66
0.0001	49.33	50.67	52.33	51.33	51.67	51.00	50.67	48.67	44.67	45.00
0.00001	47.33	47.33	49.00	49.67	50.33	51.00	50.67	51.33	52.33	53.00

3)结论

实验结果表明,变动的学习率不利于本网络学习特征,当学习率固定时,学习率的大小直接影响了网络的学习效果,且学习率过大或者过小都不利于网络特征的学习。因此通过实验的方式来寻找一个较为合理的学习率是比较合理的。

结束语 本文深入探究了目前人体行为识别领域主要的识别方法,针对人体行为中多人人体行为识别方法的不足,根据深度学习在图片识别方面的优势,提出了一种基于深度学习的多人人体行为的识别方法。实验结果表明,本文方法对多人人体行为识别是有效的。

参 考 文 献

[1] DU Y,CHEN F,XU W,et al. Activity recognition through multi-scale motion detail analysis[J]. Neurocomputing, 2008, 71 (16-18): 3561-3574.
[2] HAN L,LI J F,JIA Y D. Human Interaction Recognition Using Spatio-Temporal Words [J]. Chinese Journal of Computers, 2010,33(4): 776-784. (in Chinese)

韩磊,李君峰,贾云得. 基于时空单词的两人交互行为识别方法 [J]. 计算机学报, 2010,33(4): 776-784.
[3] PENG D,HUIJUN D,LIGEN D,et al. Group interaction analysis in dynamic context[J]. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society, 2009, 39(1): 275-282. .
[4] RYOO M S,AGGARWAL J K. Recognition of High-level Group Activities Based on Activities of Individual Members[C]//IEEE Workshop on Motion and Video Computing. IEEE Computer Society, 2008: 1-8.
[5] JI S,YANG M,YU K. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(1): 221-231.
[6] KARPATHY A,TODERICI G,SHETTY S,et al. Large-Scale Video Classification with Convolutional Neural Networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014: 1725-1732.

(下转第 321 页)

高,因此分割结果较好。由于使用本文算法进行图像分割所花费的分割时间少于传统的 AFWFCM 算法,说明本文所提算法实现过程较为简单,计算复杂度较低。

结束语 随着图像采集技术的提高,图像的数量越来越多,图像中包含的信息也越来越多。由于在图像采集过程中散焦图像的焦距调整不佳,导致图像的清晰度不高,难以对图像中的信息进行准确、完整的提取,因此需要对图像进行局部特征模糊分割。本文提出了一种利用免疫谱聚类算法的分割算法,主要研究内容如下:

1)采用分块方法对散焦图像局部特征进行模糊,比较模糊前后散焦图像的奇异值转变情况,依据奇异值变化对散焦图像进行标识,实现散焦图像的奇异值特征提取,确定散焦图像的分割目标。

2)利用谱聚类的方法对散焦图像中的像素点样本进行聚类,采用 Nyström 逼近方法降低计算的复杂度,对像素点相似性矩阵的特征向量进行计算,并利用免疫算法提高聚类效果,保证聚类结果的精确度,从而保证散焦图像的局部特征模糊分割结果。

参考文献

- [1] WANG W Z, LI N. A Segmentation Algorithm of Moving Target Image Anti Light Interference[J]. Bulletin of Science and Technology, 2015, 31(6): 166-168. (in Chinese)
王维哲,李娜. 一种去光照干扰的运动目标图像分割算法[J]. 科技通报, 2015, 31(6): 166-168.
- [2] ZHANG J, FAN H H. An Improved Image Segmentation Algorithm and Simulation Based on Fuzzy Clustering[J]. Computer Simulation, 2015, 32(4): 380-383. (in Chinese)
张杰,范洪辉. 一种改进的模糊聚类图像分割算法研究与仿真[J]. 计算机仿真, 2015, 32(4): 380-383.
- [3] DONG D D, ZHOU S G, FAN L, et al. Research of Remote Sensing Image Segmentation Based on Watershed and Alpha Expansion[J]. Science Technology and Engineering, 2015, 15(10): 204-209. (in Chinese)
董丹丹,周绍光,凡莉,等. 基于分水岭和 α 扩展的遥感影像分割方法研究[J]. 科学技术与工程, 2015, 15(10): 204-209.
- [4] ZHANG Y M, BA D K, XING K. A Method of Fuzzy Threshold for Adaptive Image Segmentation[J]. Computer Measurement & Control, 2016, 24(4): 126-128. (in Chinese)
张永梅,巴德凯,邢阔. 基于模糊阈值的自适应图像分割方法[J]. 计算机测量与控制, 2016, 24(4): 126-128.
- [5] GONG W W, GE Y R. Image segmentation by spectral clustering based on IRAM and semi-supervised[J]. Electronic Design Engineering, 2016, 24(17): 156-159. (in Chinese)
龚文文,葛玉荣. 基于 IRAM 和半监督的谱聚类图像分割[J]. 电子设计工程, 2016, 24(17): 156-159.
- [6] WANG T, JI Z X, SUN Q S. A Segmentation Algorithm Combined with Non-local Information and Graph Cut[J]. Journal of Computer-Aided Design & Computer Graphics, 2015, 27(5): 783-791. (in Chinese)
王涛,纪则轩,孙权森. 结合非局部信息与图割的图像分割算法[J]. 计算机辅助设计与图形学学报, 2015, 27(5): 783-791.
- [7] SHEN X J, PAN H, CHEN H P. Medical Image Segmentation Algorithm Based on One-Dimensional Otsu Multiple Threshold[J]. Journal of Jilin University (Science Edition), 2016, 54(2): 344-348. (in Chinese)
申铨京,潘红,陈海鹏. 基于一维 Otsu 的多阈值医学图像分割算法[J]. 吉林大学学报(理学版), 2016, 54(2): 344-348.
- [8] CHEN X, HE Z S, LI Y H. Improved color image segmentation of GrabCut algorithm based on SLICO[J]. Application Research of Computers, 2015, 32(10): 3191-3195. (in Chinese)
陈鑫,何中市,李英豪. 一种新的基于 SLICO 改进的 GrabCut 彩色图像分割算法[J]. 计算机应用研究, 2015, 32(10): 3191-3195.
- [9] WANG S H, DI L, LIANG J Z. Multi-dimensional fuzzy clustering image segmentation algorithm based on kernel metric and local information[J]. Journal of Computer Applications, 2015, 35(11): 3227-3231. (in Chinese)
王少华,狄岚,梁久祯. 基于核与局部信息的多维度模糊聚类图像分割算法[J]. 计算机应用, 2015, 35(11): 3227-3231.
- [10] HOU X F, WU C M. Fast Fuzzy Local Information C-means Clustering Segmentation Algorithm [J]. Computer Science, 2016, 43(10): 297-303. (in Chinese)
侯晓凡,吴成茂. 一种快速的模糊局部 C-均值聚类分割算法[J]. 计算机科学, 2016, 43(10): 297-303.
- [11] YANG M, SU Y K. Adaptive Algorithm Based on Fuzzy C-Means for Image Segmentation[J]. Journal of Chongqing University of Technology (Natural Science), 2015, 29(6): 94-99. (in Chinese)
杨漫,苏亚坤. 采用模糊 C-均值聚类的自适应图像分割算法[J]. 重庆理工大学学报(自然科学), 2015, 29(6): 94-99.

(上接第 311 页)

- [7] DAS A, PAL U, FERRER B M A, et al. Sclera recognition using dense-SIFT[C]//International Conference on Intelligent Systems Design and Applications. 2013: 74-79.
- [8] GAO C, ZHANG X, WANG Y L, et al. Automatic stitching approach of aerial image sequence based on SIFT features[J]. Computer Applications, 2007, 21(11): 2789-2792. (in Chinese)
高超,张鑫,王云丽,等. 一种基于 SIFT 特征的航拍图像序列自动拼接方法[J]. 计算机应用, 2007, 21(11): 2789-2792.
- [9] BROWN M, LOWE D G. Invariant Features from Interest Point Groups[C]//British Machine Vision Conference. 2002: 656-665.
- [10] LOWE D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [11] ZHAO Z H, YANG S P, MA Z Q. License Plate Character Recognition Based on Convolutional Neural Network LeNet-5 [J]. Journal of System Simulation, 2010, 22(3): 638-641. (in Chinese)
赵志宏,杨绍普,马增强. 基于卷积神经网络 LeNet-5 的车牌字符识别研究[J]. 系统仿真学报, 2010, 22(3): 638-641.