

一种基于生成对抗网络的行为数据集扩展方法

牛斌,吴鹏,马利,刘景巍

(辽宁大学信息学院,辽宁沈阳110036)

摘要:深度学习作为人工神经网络的分支,在图像识别领域有广泛的应用,但其数据集的不足导致模型学习不够完善。通过对深度学习的数据规模要求进行分析,针对人体行为识别中的应用,发现人体数据集的采集工作是一个极具耗时耗力的工程,很难满足目前深度学习网络的需求。为了解决这一难题,提出了一种依靠原有的小规模数据集产生大量可靠数据集的半监督深度学习模型。通过将循环神经网络和生成式对抗网络相结合的方法使循环神经网络学习到数据的序列关系和特征,使生成式对抗网络产生合理数据进而扩展人体行为数据集。依靠该网络结构,可以很好地分析出采集数据的特征,并且依据这些特征可以生成大量的合理的数据,后经过数据处理等工作,形成可用于模型训练的可靠数据集,缓解了深度学习工作中数据集紧缺的问题。

关键词:数据生成;深度学习;循环神经网络;生成式对抗网络

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2019)07-0043-06

doi:10.3969/j.issn.1673-629X.2019.07.009

A Behavior Data Set Extension Method Based on Generative Adversarial Network

NIU Bin, WU Peng, MA Li, LIU Jing-wei

(School of Information, Liaoning University, Shenyang 110036, China)

Abstract: As a branch of artificial neural network, deep learning has a wide range of applications in the field of image recognition. The lack of data sets leads to incomplete model learning. Through the analysis of the data size requirements of deep learning, it is found that the collection of human data sets is a very time-consuming and labor-intensive project for the application of human behavior recognition. It is difficult to meet the needs of the current deep learning network. To solve this problem, we propose a semi-supervised deep learning model that relies on the original small-scale data set to generate a large number of reliable data sets. By combining the cyclic neural network and the generative confrontation network, the cyclic neural network learns the sequence relationship and characteristics of the data, so that the generation-oriented network generates reasonable data and then expands the human behavior data set. Relying on this network structure, the characteristics of the collected data can be well analyzed, and a large amount of reasonable data can be generated according to these features, and then processed through data processing to form a reliable data set that can be used for model training, thereby alleviating the shortage of data sets in deep learning work.

Key words: data generation; deep learning; recurrent neural networks; generative adversarial network

0 引言

随着信息技术的发展,深度学习和人体行为识别逐渐得到了重视,并在各个领域有着广泛的应用,如安防监控、运动分析、医学辅助诊断和智能人机交互等。如何快速准确地对人体行为进行识别已成为极具挑战性的问题^[1]。早期人体行为数据采集十分困难,包括传感器性能低下,成本高昂,场景和背景的干扰等。近

年来,由于高清摄像技术和传感器技术的发展^[2],使得人体行为识别更加高效准确。人体行为识别技术得到了有力支撑。随着人体行为数据提取技术的快速发展,人们依旧对行为识别的实验结果不满意,难以满足工业要求。

对行为识别的速度和准确性的追求从来没有停止过。尤其近年来,由于深度学习的飞速发展,人们开始

收稿日期:2018-07-29

修回日期:2018-11-20

网络出版时间:2019-03-21

基金项目:2017年辽宁省博士科研启动基金指导计划项目(20170520276)

作者简介:牛斌(1963-),男,硕士,教授,研究方向为嵌入式系统设计、图像处理、模式识别;吴鹏(1991-),男,硕士,研究方向为图像处理、深度学习。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190321.0909.030.html>

探索通过深度学习和行为识别相结合^[3-5],以不断提高人体行为识别的准确性和高效性。在 2006 年 Hinton^[6]提出了深度学习的概念,为解决深层结构相关的优化难题带来希望,随后又提出多层自动编码器深层结构,从此深度学习得到了迅猛发展。由于深度网络可以无监督地从数据中学习特征,而这种学习方式也符合人类感知世界的机理,因此当训练样本足够多时,通过深度网络学习到的特征往往具有一定的语义特征,并且更适合目标和行为的识别。

深度学习的首要条件就是对于数据规模的要求。如果深度学习用于行为识别领域,则通常情况下需要大量的行为动作数据序列。而对于数据集的要求,又让深度学习框架的可信性受到了限制。在当前“大数据”时代,大量的数据获取相比原来更容易,计算资源和能力的大大提升,各种新模型、新理论如雨后春笋般涌现,验证周期大大缩短。但是各方面的努力依旧难以满足工业要求。

所以,数据集的可靠扩展对于深度学习和行为识别具有重要意义。目前,使用深度学习的方法进行行为识别依然充满挑战。行为识别的应用^[7]背景相当广泛,主要集中在人机交互、虚拟现实、自动驾驶等应用中,推动着人类社会的发展。人工智能时代的开启必然会在很大程度上改变世界,无论是交通、医疗、购物,还是军事等方面^[8]。

1 相关工作

在行为识别研究初期,有些论文使用一些结构简单的行为数据库^[9-11]。行为数据库(比如 KTH 和 WEIZANN)的主要出发点是集中在一些基本简单的行为上,比如行走、下蹲等。它们都是在可控场景(背景相对静止)的情况下进行拍摄,行为类别和样本数量比较少。很多经典行为识别算法都使用这些数据集进行验证,都得到了预期的结果。

这些简单的数据集是行为识别的起步数据集,为计算机视觉的发展做出了巨大贡献。直至今日,这些数据集仍然应用广泛。随着传感器技术的发展和行为识别研究的不断深入,数据集也在不断发生变化,主要包括以下几个方面:

(1)从单一视角到多视角的改变。比如,法国国家信息与自动化研究所和中国科学院自动化研究所等国内外很多研究机构都发布了多视角的行为数据,如 IXMAS Action^[12]数据库、CASIA 行为数据库^[13]。多视角相对于单视角,优点包括拥有者包含的数据量更加全面,可以从各个角度对人体进行观测等。但缺点是数据处理更加复杂,消耗时间变长等。

(2)背景和场景更加复杂。静止背景的数据集虽

然处理相对简单,但是难以满足实际需求。使用该数据集训练得到的模型识别成功率低,不具有通用性。因此,Hollywood^[14],UCF YouTube^[15]和 UCF 101^[16]数据集应运而生,这些数据集都是取自电影或者真实录像,具有通用性和可信性等特点,其场景更具有真实性和复杂性。

(3)数据规模更加庞大。虽然当前行为识别数据库的规模还远远不及图像识别数据库,但其行为类别和样本数量也在迅速增加。当今正处于一个“大数据”的时代,大量的数据如洪水般充斥在生活中。比如 UCF 101 行为识别数据库就具有规模更大、类别更多的特点。

(4)形式更加多样化。随着人体传感器的发展,行为识别数据集也从单一视频的形式转换成各种容易处理的样式。比如 2010 年微软公司推出的 Kinect 传感器,用于 Xbox360 游戏机的输入设备。该设备将人体骨骼抽象成三维坐标,使行为识别工作预处理部分变得简单、容易理解。MSR Action3D 和 MSR Daily Activity 3D 数据集都是使用该设备进行采集。行为数据采集是行为识别工作的重要部分,对于识别工作的结果有着决定性的影响。目前人们对该工作的努力一直没有停止,尤其是多人物的行为识别也取得了不少研究成果。

2 RGAN 扩展数据集方法

2.1 RGAN 网络结构

生成式对抗网络^[17]是一种半监督学习算法,可以将一些随机数据经过神经网络的复杂运算,生成具有和原始的骨骼数据相似的数据。一个优秀的生成式对抗网络可以很好地模仿原始骨骼数据的特征,它的输出的准确性直接影响后续行为识别相关研究的结果,对后续工作具有巨大的意义。

生成式对抗网络在各领域应用广泛^[18-20],但是在行为数据集生成领域的应用还是十分稀少。文中通过分析行为数据的特点,总结出行为数据和序列数据并无太大区别,所以生成式对抗网络同样适合于对行为数据集的扩展。对于生成行为数据集,需要一种特殊的网络结构,从而可以生成令人满意的行为数据。

经过对生成器和鉴定器的选择,文中提出的 RGAN 结构如图 1 所示。

由图 1 可知,鉴定器和生成器除了采用非传统循环神经网络^[21](RNN)和卷积神经网络^[22](CNN),还需要和全连接层(FC)相连接。其中生成器主要包括 3 层 RNN,用于生成具有时间相关性的特征数据。每层 RNN 包括 $3n$ 个 RNN 节点,其中 n 代表原始数据中每一帧的骨骼点总数。不同的行为采集设备可能会有

不同的骨骼点数量,所以依据不同的设备,RNN 的节点数据随之变化。紧接着为 2 层全连接层,每层全连接层依旧含有 $3n$ 个节点。

文中使用全连接层进一步生成动作数据,通过该

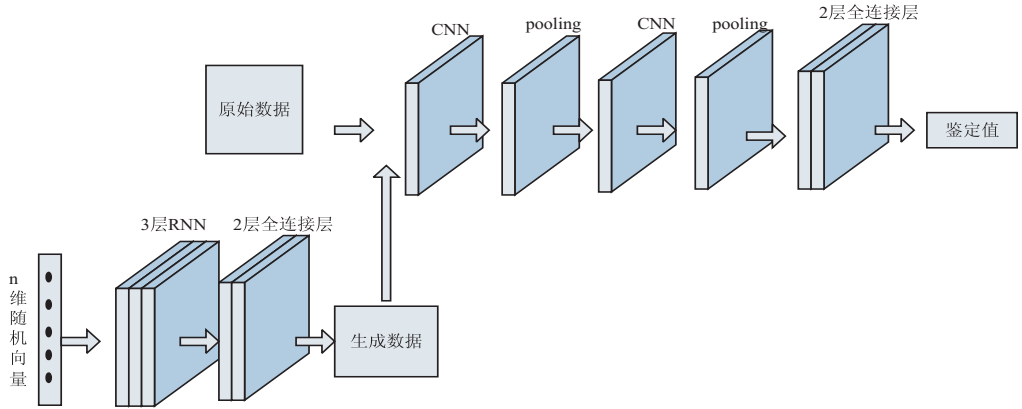


图1 RGAN 结构

鉴别器主要由 CNN、池化层和 FC 组成,每层 CNN 包括一个 3×3 卷积核,池化层采用最大池化,第一层全连接层的节点数量为 13,第二层全连接层的节点数量为 1。CNN 的主要任务是提取矩阵的形状特征。由于动作序列本质上与图片无异,所以训练后的 CNN 可以很好地分析出生成数据和原始数据的差别。3 层全连接层的任务是将提取的特征换算成鉴定值,从而分辨出数据的真实性。

2.2 RGAN 网络的训练

在经过对数据集的预处理之后,还需要对生成对抗网络的各个参数进行初始化,一般通过将全部参数赋值为零或者符合正态分布的参数进行赋值。当所有的初始化工作完成后,可以开始模型训练。

GAN 的训练过程是生成器和鉴别器交替训练的过程^[17]。在训练生成器的过程中,需要保证鉴别器的参数不变。在训练鉴别器的过程中,需要保证生成器的参数不变。以上过程为一次迭代,每一次迭代都包括两次的正向传播和两次反向传播。

在训练生成器的过程中:

保证鉴别器参数不变。首先从数据集中选择一条数据,即抽取一条动作数据,一般以矩阵的形式存储,矩阵的行数为该动作的帧数,设为 m ,矩阵的列数为 n ,表示在该帧下所有骨骼点的三维空间坐标维度的总数,所以该数据可以表示为 $E_{m \times n}$:

$$E_{m \times n} = F(U) \quad (1)$$

其中, U 为原始的骨骼数据集; F 为从原始的骨骼数据集中选取的某个动作数据。

在选取某个动作之后,一般需要获取其帧数 m 和骨骼点的空间坐标的数量 n 。为了保证生成的数据和原始的骨骼数据尽可能相似,需要设置生成的循环

全连接层,可以产生和原始数据同维度并具有时间相关性的数据。再通过训练,即可产生符合需求的动作数据序列。

迭代次数为 m ,即保证生成的数据矩阵的行数和列数与原始的骨骼数据相同。

之后,需要由系统生成一个随机向量设为 V ,将随机向量输入到生成器,经过 m 次迭代,按照前文所述的算法,生成一个和原始数据 R 同维度的矩阵,称之为伪动作序列 V 。

可用式 2 表示:

$$Z = G(V) \quad (2)$$

再将生成动作序列 Z 送入鉴别器中,生成一个特征值设为 A :

$$A = D(Z) \quad (3)$$

最后,依据误差公式 4,调整生成器参数:

$$\text{Lg}(Z) = - \sum \log(D(G(Z))) \quad (4)$$

使用优化器 Adam 对生成器的所有参数进行优化。在训练鉴别器的过程中,保证生成器参数不变,同样在真实动作序列中选取一条数据,如式 1,在选取某个动作之后,仍然需要获取其帧数 m 和骨骼点的空间坐标的数量 n 。然后同样设置生成的循环迭代次数为 m 。

按照上文所述,获得随机变量 V ,按照式 2 经过生成器之后,仍然获得伪行为数据矩阵。再输入到鉴别器中,按照式 3 获得特征值 A_1 :

$$A_1 = D(Z) \quad (5)$$

与训练生成器不同的是,还要将原始的骨骼数据 R 输入到鉴别器中,得到特征值 A_2 :

$$A_2 = D(R) \quad (6)$$

最后,依据误差公式 7,调整鉴别器参数:

$$\text{Ld}(Z, R) = - \sum (\log A_1 + \log(1 - A_2)) \quad (7)$$

最后使用优化器 Adam 对鉴别器的所有参数进行优化。

不断地迭代两个过程,当鉴定器无法对生成器产生的数据鉴定时,说明训练完成。

当训练完成后,需要将生成器从 RGAN 系统中提取出来,即需要的数据集扩展工具。不断地输入随机变量,即可大量产生数据集,从而扩充原本数量不足的行为数据集。

本算法的伪代码如下:

算法:train RGAN

Input:random value,real data

Output:estimation value,generator

1:while D can distinguish fade data

//保证鉴定器不能鉴定生成的数据

2:input random value to G

//训练鉴定器

3:G output fade data

4:input fade data and real data to D

5:D output a estimation value

6:update the parameter of D

7:input random value to G

//训练生成器

8:G output fade data

9:input fade data to D

11:D output a estimation value

12:update the parameter of G

13:end while

3 实验

文中实验使用的计算机主要包括两颗英特尔至强 CPU,8 G 三星 DDR3 内存和 8 G 显存 GTX1080 显卡加速器。

3.1 RGAN 算法评价标准

由于行为数据集的存储为基本数字序列,则对某一动作序列可以抽象为一个矩阵,再通过转换方法可以抽象成一个向量。所以一个动作序列可以通过抽象的方法变成一个矩阵或者向量。如果想要评测两个动作之间的相似度,只需要通过数学等方法获得两个矩阵或者向量的相似度即可。

目前采用两种评测方法:

(1) 欧氏距离评测方法。

对于矩阵的相似度,可以采用欧氏距离的方法:设两个矩阵为 A 和 B ,其中的元素分别为 a 和 b ,则欧氏距离可以表示为:

$$L = \sum_{A,B} |a - b| \quad (8)$$

该方法的主要思想是将数值的差相加,目的是获取数据的差异性,而忽略其他方面的影响。优点是直观、简单;缺点是计算量较大,尤其是 1 范式计算还存在开方的计算。

(2) 余弦相似度评测方法。

余弦相似度是衡量两个向量相关性的方法,其定义为:

$$\cos(x, y) = \frac{x \cdot y}{|x| \cdot |y|} \quad (9)$$

计算结果为 -1 至 1 的小数, -1 表示两个向量方向相反, 0 表示两个向量为垂直方向, 1 表示两个向量方向相同。

该方法从向量的方向角度上描述了向量的相似性。如果两个向量的方向在方向上相差较大,则通过该方法可以体现出来。

文中将采用两者相结合的方法,从两个角度来衡量 RGAN 产生的数据集,同时还要和其他半监督算法,无监督算法进行比较。

总结以上公式,定义评测公式为:

$$\alpha = \frac{L}{2 + \cos(x, y)} \quad (10)$$

3.2 实验结果及分析

对 RGAN 算法进行实验,并采用上述评价标准来衡量算法优劣,进行客观分析。

3.2.1 RGAN 系统关键参数变化

在理想状态下,当 RGAN 训练结束后,生成器和鉴定器会有很强的生成能力和鉴定能力。两者形成一种对抗的状态,当某个 loss 函数值上升时,另一个会下降,最后趋于平衡。

文中使用 tensorboard 收集实验数据。实验结果如图 2 所示。

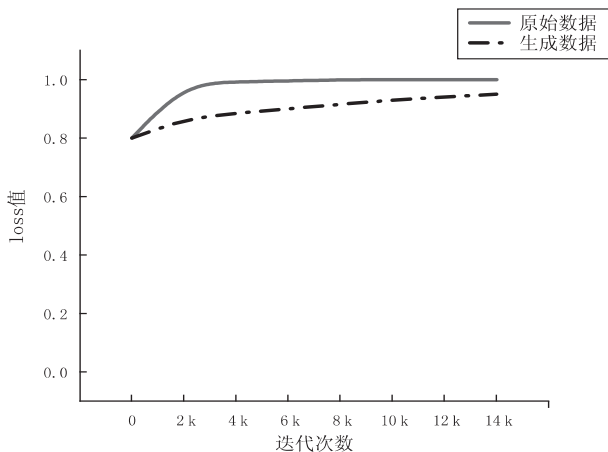


图 2 鉴定器鉴定原始数据与生成数据的结果

鉴定器对于原始数据的鉴定结果如图 2 所示,对应式 6,可以看出,根据迭代自出的增多,鉴定原始数据的结果收敛至 1,说明它的鉴定能力在增强。

鉴定器对于生成器的生成结果如图 2 所示,对应式 5,可以看出,鉴定器对于生成器产生的结果逐渐增加,接近于 1,即生成器产生的数据和原始数据越来越相似。

鉴定器和生成器的 loss 函数图如图 3 所示。其中鉴定器的 loss 函数在训练过程中的变化情况对应式 7;生成器的 loss 函数在训练过程中的变化情况对应式 4。

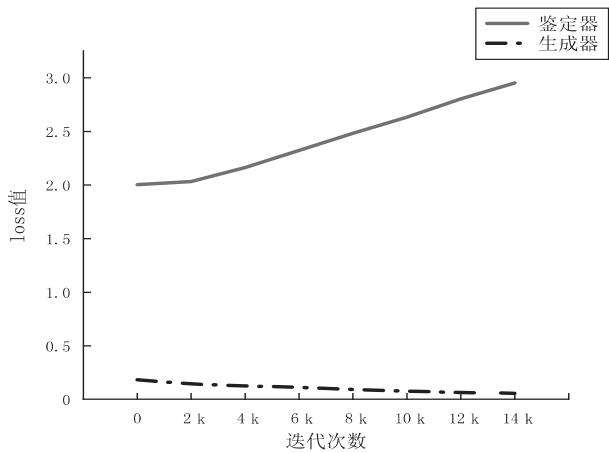


图3 鉴定器 loss 与生成器 loss 的计算结果

可以看出,两者的趋势相反,完全体现了一种对抗思想,并且最后分别收敛。以上参数的变化反映了RGAN的训练过程,在上文所述硬件条件下,整个训练过程总共消耗时间为6小时左右。

3.2.2 结果及分析

为了评价文中算法的优势,本节将使用几种常见的半监督学习算法与文中算法进行比较。

传统生成式对抗网络(GAN),即将RNN模块删除后的GAN;将原始数据输入到传统GAN后,使用GAN生成一定规模的数据,然后计算数据集的评价值,对应式9。

(1)在UTKinct数据集进行实验,结果如图4所示。

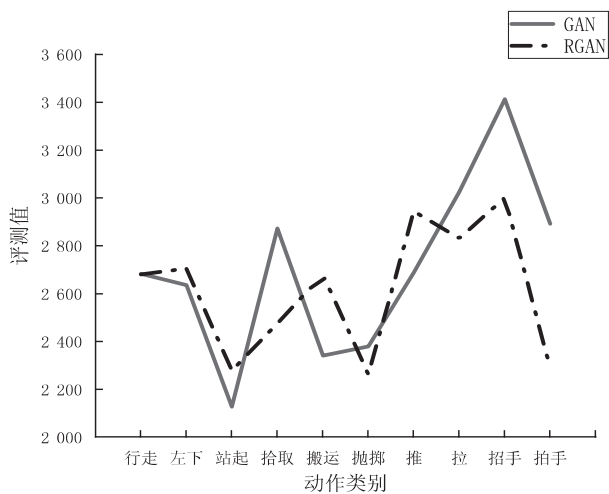


图4 UT-Kinect 实验结果

随机选取了若干动作进行实验。在UT-Kinect数据集的实验中可以得出结论, RGAN产生的数据和原始数据的相似度更高,也就是RNN在RGAN的系统中得到了预期的效果。

文中提出的RGAN算法评价值的平均值虽然较小,平均值达到2 613. 355 272,但是由于相关领域相似实验和评价标准较少, RGAN生成的数据集和原始数据集相似度是否已经具有合理性,是否已经很好地生成数据,确实还需要更多对比实验。但是实验结果比传统GAN的评价值小,说明该算法在一定程度上对GAN在行为识别领域得到了改进。这是由于RGAN可以学习到数据序列之间的依赖关系。实验结果表明, RGAN在一定程度上可以很好地生成人体行为UT-Kinect相似数据集。

(2)在MSR Action 3D数据集上的实验结果如图5所示。

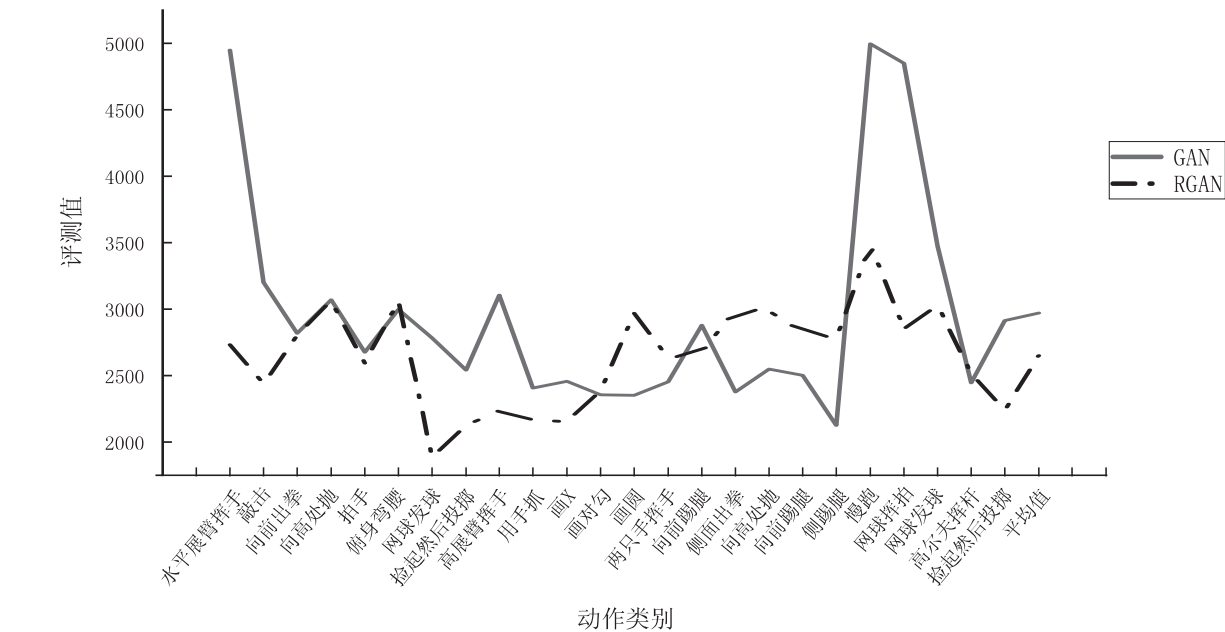


图5 MSR Action 3D 实验结果

从图 5 可以看出,在使用 MSR Action 3D 的实验中,可以根据平均值进行排序后的结果对比, RGAN 表现良好并没有和以上两种算法相差太大,评价值为 2 651.757 254。虽然面临着和 UT-Kinect 相同的问题,但是表现依旧强于传统 GAN 算法,所以, RGAN 算法是一种相对较好的数据集生成算法。

4 结束语

从整体上来说, RGAN 在数据集扩展的实验中表现良好,符合最初的期望。虽然面临着对比实验较少的问题,但是生成的数据集基本上符合预期用于行为识别训练的目标,在实验结果中表现优于传统生成对抗网络。同时,该算法在复杂行为的生成过程中容易产生对行为识别正确结果偏离的问题还需要进一步改进。

参考文献:

- [1] NGUYEN L T, NGUYEN N T, VO B, et al. Efficient method for updating class association rules in dynamic datasets with record deletion[J]. *Applied Intelligence*, 2018, 48(6): 1491–1505.
- [2] WEI Wei. Information retrieval in biomedical research: from articles to datasets[D]. San Diego: University of California, 2017.
- [3] HINTON G E, OSINDER S, TEH Y W. A fast learning algorithm for deep belief nets[J]. *Neural Computation*, 2006, 18(7): 1527–1554.
- [4] 郑胤, 陈权崎, 章毓晋. 深度学习及其在目标和行为识别中的新进展[J]. *中国图象图形学报*, 2014, 19(2): 175–184.
- [5] 裴晓敏, 范慧杰, 唐延东. 时空特征融合深度学习网络人体行为识别方法[J]. *红外与激光工程*, 2018, 47(2): 0203007–1–0203007–6.
- [6] 单言虎, 张彰, 黄凯奇. 人的视觉行为识别研究回顾、现状及展望[J]. *计算机研究与发展*, 2016, 53(1): 93–112.
- [7] SHAIKHINA T, KHOVANOV A. Handling limited datasets with neural networks in medical applications: a small-data approach[J]. *Artificial Intelligence in Medicine*, 2017, 75: 51–63.
- [8] SCHULDT C, LAPTEV I, CAPUTO B. Recognizing human actions: a local SVM approach[C]//International conference on pattern recognition. Cambridge: IEEE, 2004: 32–36.
- [9] LAPTEV I. Local spatio-temporal image features for motion interpretation[D]. Stockholm: Royal Institute of Technology, 2004.
- [10] LAPTEV I, LINDBERG T. Velocity adaptation of space-time interest points[C]//International conference on pattern recognition. Cambridge: IEEE, 2004: 52–56.
- [11] WEINLAND D, BOYER E, RONFARD R. Action recognition from arbitrary views using 3D exemplars[C]//International conference on computer vision. Rio de Janeiro: IEEE, 2007: 1–7.
- [12] ZHANG Zhang, TAO Dacheng. Slow feature analysis for human action recognition[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2012, 34(3): 436–450.
- [13] MARSZALEK M, LAPTEV I, SCHMID C. Actions in context[C]//IEEE conference on computer vision and pattern recognition. Miami, FL, USA: IEEE, 2009: 2929–2936.
- [14] LIU Jingen, LUO Jiebo, SHAH M. Recognizing realistic actions from videos[C]//IEEE conference on computer vision and pattern recognition. Miami, FL, USA: IEEE, 2009: 1996–2003.
- [15] SOOMRO K, ZAMIR A R, SHAH M. UCF101: a dataset of 101 human actions classes from videos in the wild[J]. *Computer Science*, 2012, 4(2): 1212–1219.
- [16] 张营营. 生成对抗网络模型综述[J]. *电子设计工程*, 2018, 26(5): 34–37.
- [17] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]//Empirical methods on natural language processing. [s. l.]: [s. n.], 2014.
- [18] 赵树阳, 李建武. 基于生成对抗网络的低秩图像生成方法[J]. *自动化学报*, 2018, 44(5): 829–839.
- [19] 曹志义, 牛少彰, 张继威. 基于半监督学习生成对抗网络的人脸还原算法研究[J]. *电子与信息学报*, 2018, 40(2): 323–330.
- [20] SUNDERMEYER M, SCHLÜTER R, NEY H. LSTM neural networks for language modeling[EB/OL]. [2014-02-10]. <http://www-i6.informatik.rwth-aachen.de/publications/download/820/Sundermeyer-2012.pdf>.
- [21] GIRSHICK R. Fast R-CNN[J]. *Computer Science*, 2015, 7(3): 1356–1361.
- [22] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN features off-the-shelf: an astounding baseline for recognition[C]//IEEE conference on computer vision and pattern recognition workshops. [s. l.]: IEEE, 2014: 806–813.