(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109753906 A
(43)申请公布日 2019.05.14

(21)申请号 201811594841.4

(22)申请日 2018.12.25

(71)申请人 西北工业大学
地址 710072 陕西省西安市友谊西路127号

(72)发明人 王琦 李学龙 林维

(74)专利代理机构 西北工业大学专利中心 61204
代理人 刘新琼

(51)Int.Cl.
G06K 9/00(2006.01)
G06K 9/62(2006.01)
G06N 3/04(2006.01)

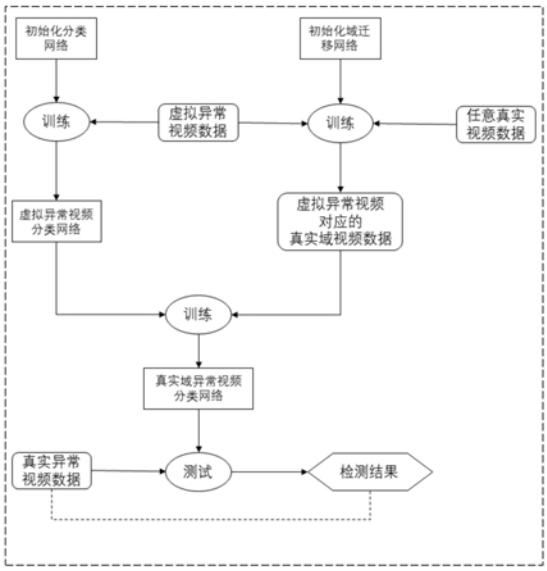
权利要求书1页 说明书4页 附图2页

(54)发明名称

基于域迁移的公共场所异常行为检测方法

(57)摘要

本发明涉及一种基于域迁移的公共场所异常行为检测方法,利用虚拟世界的模拟创建出大量虚拟异常时间视频,解决了异常事件的多样性但是数据不足的问题,又用域迁移的方法将虚拟数据迁移到真实情况下,提高分类检测网络在正式监控视频中的适应性,有效提升训练网络的可用性。



CN 109753906 A

1.一种基于域迁移的公共场所异常行为检测方法,其特征在于步骤如下:

步骤1:利用已有的虚拟图像产品生成虚拟异常数据,虚拟异常数据包括不同的异常类别和正常类别数据,各个类别的数据数量相同;

步骤2:使用步骤1生成的虚拟异常数据训练视频分类网络,得到一个虚拟异常数据分类网络;

步骤3:利用生成的虚拟异常数据和采集的真实数据,训练域迁移网络,获得虚拟异常视频数据对应的真实域视频数据;所述的域迁移网络为改进的cycle-GAN,改进方法:将cycle-GAN网络中,所有的2D卷积结构,都改为面向视频数据的3D卷积结构,3D卷积结构的计算方法为:

$$V_{ij}^{xyz} = b_{ij} + \sum_{m} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} W_{ijm}^{pqr} V_{(i-1)m}^{(x+p)(y+q)(z+r)}$$

其中P、Q、R分别表示上一层网络输出的特征图的长宽高,m表示上已成网络输出的特征图数量。最终计算得到在该卷积模块W下,下一层网络中对应的特征图V,b为偏移量,i,j第i层第j个3d卷积结构,x,y,z长宽高上的坐标值;

步骤4:利用步骤3得到的真实域异常数据对步骤2得到的虚拟异常数据分类网络进行进一步分类训练,训练过程和步骤2相同,从而得到真实域的异常视频分类网络;

步骤5:将待测试的真实的异常数据输入到步骤4训练得到的网络模型,利用softmax函数获得该输入视频在隶属于各个异常类别中的概率,取最大值的类别作为该段视频的异常类型即可。

2.根据权利要求1所述的一种基于域迁移的公共场所异常行为检测方法,其特征在于所述的步骤2中的视频分类网络为3DresNet或时空双流视频分类网络。

基于域迁移的公共场所异常行为检测方法

技术领域

[0001] 本发明属于计算机视觉，视频监控领域。针对视频监控的公共场所，检测出视频当中发生的如打架、逃散等异常行为。

背景技术

[0002] 如今遍及城市公共区域的摄像头每时每刻都在产生无数的监控视频，如果可以通过自动化的方法对采集到的视频进行异常行为的检测，那么这对于公共安全事件的发生具有极强的预防作用。但是由于异常行为的发生频率远小于正常行为发生的频率，以及异常行为的多样性，使得异常事件的检测变得非常困难。

[0003] 目前公共场所中异常行为的检测方法有两种：第一种是R.Mehran等人在文献“R.Mehran,A.Oyama,and M.Shah,Abnormal crowd behavior detection using social force model,Computer Vision and Pattern Recognition,2009.CVPR 2009.IEEE Conference on,pp.935-942,2009.”中提出的基于社会力模型的方法，它将行人看作一个个的移动点，将人与人之间的交互看作点与点之间的作用力，通过发现异常的粒子移动来检测视频中的异常行为。

[0004] 第二种方法是基于光流法的方法，例如“Y.Yu,W.Shen,H.Huang,and Z.Zhang,Abnormal event detection in crowded scenes using two sparse dictionaries with saliency,Journal of Electronic Imaging,vol.26,no.3,pp.033013,2017.”中提出的方法，通过结合多尺度光流直方图和多尺度梯度直方图来获取一种行人的表面与动作特征，在传统的只包含正常特征的稀疏模型中加入异常特征构建字典。此外，将测试样本的显著性与正常字典和异常字典上的稀疏重构代价相结合，测量测试样本的正常程度。

[0005] 这些方法都有其局限性，粒子点模型并不能捕捉人物的动作特征，基于光流的特征字典并不能保证所有异常行为均能存在于字典当中。

发明内容

[0006] 要解决的技术问题

[0007] 为了避免现有技术的不足之处，本发明提出一种基于域迁移的公共场所异常行为检测方法。

[0008] 技术方案

[0009] 一种基于域迁移的公共场所异常行为检测方法，其特征在于步骤如下：

[0010] 步骤1：利用已有的虚拟图像产品生成虚拟异常数据，虚拟异常数据包括不同的异常类别和正常类别数据，各个类别的数据数量相同；

[0011] 步骤2：使用步骤1生成的虚拟异常数据训练视频分类网络，得到一个虚拟异常数据分类网络；

[0012] 步骤3：利用生成的虚拟异常数据和采集的真实数据，训练域迁移网络，获得虚拟异常视频数据对应的真实域视频数据；所述的域迁移网络为改进的cycle-GAN，改进方法：

将cycle-GAN网络中，所有的2D卷积结构，都改为面向视频数据的3D卷积结构，3D卷积结构的计算方法为：

[0013] $$V_{ij}^{xyz} = b_{ij} + \sum_{m}\sum_{p=0}^{P-1}\sum_{q=0}^{Q-1}\sum_{r=0}^{R-1} W_{ijm}^{pqr} V_{(i-1)m}^{(x+p)(y+q)(z+r)}$$

[0014] 其中P、Q、R分别表示上一层网络输出的特征图的长宽高，m表示上已成网络输出的特征图数量。最终计算得到在该卷积模块W下，下一层网络中对应的特征图V，b为偏移量，i，j第i层第j个3d卷积结构，x，y，z长宽高上的坐标值；

[0015] 步骤4：利用步骤3得到的真实域异常数据对步骤2得到的虚拟异常数据分类网络进行进一步分类训练，训练过程和步骤2相同，从而得到真实域的异常视频分类网络；

[0016] 步骤5：将待测试的真实的异常数据输入到步骤4训练得到的网络模型，利用softmax函数获得该输入视频在隶属于各个异常类别中的概率，取最大值的类别作为该段视频的异常类型即可。

[0017] 所述的步骤2中的视频分类网络为3DresNet或时空双流视频分类网络。

[0018] 有益效果

[0019] 本发明提出的一种基于域迁移的公共场所异常行为检测方法，利用虚拟世界的模拟创建出大量虚拟异常时间视频，解决了异常事件的多样性但是数据不足的问题，又用域迁移的方法将虚拟数据迁移到真实情况下，提高分类检测网络在正式监控视频中的适应性，有效提升训练网络的可用性。

附图说明

[0020] 图1为本发明的模型、数据流动图；

[0021] 图2为域迁移网络的数据流动图。

具体实施方式

[0022] 现结合实施例、附图对本发明作进一步描述：

[0023] 本发明提出一种基于域迁移的公共场景异常行为检测方法，以解决异常行为多样性、频率低等现象导致异常行为检测的困难性。整个技术方案包括以下步骤：

[0024] 1.利用已有的虚拟图像产品，如游戏、CG等创建出虚拟的场景、任务、模型与异常相关的动作，录制虚拟世界中的异常行为。

[0025] 2.在捕获到大量录制好的虚拟视频数据后，利用这些数据训练一个视频分类的深度神经网络，该网络可以有效区分虚拟数据集中的异常行为类别(如打架、逃散等)以及正常情况。

[0026] 3.利用一些现实中的监控视频，这些视频不必一定有异常事件的发生。利用这些视频和已有的虚拟视频之间的相互转化关系，学习一个域迁移网络，进行无监督的视频域迁移，将虚拟的视频迁移到和现实场景非常相似且逼真的真实视频域，获得大量的包含异常行为的监控视频。

[0027] 4.利用迁移后的视频作为数据集，再次训练(2)中获得的分类神经网络，来提高该神经网络在跨域后，即真实数据域中的适应能力，提高该网络应用到真实视频监控中的检

测能力。

[0028] 5.在实际运用过程中,每次可将固定时间长度的监控视频实时传入到训练好的神经网络中,获得捕获到的短视频在各个异常类别与正常情况下的分类概率,取概率最高的类别作为该段视频的类别。利用检测结果属于哪一种级别的异常或是正常行为,来确定监控下是否有异常行为的发生。

[0029] 本发明的具体实现步骤如下:

[0030] 步骤1,首先需要准备好一个无监督的域迁移网络,该网络的类型为"J.Zhu,T.Park,P.Isola,and A.A.Efros,Unpaired image-to-image translation using cycle-consistent adversarial networks,arXiv preprint,2017."中提到的cycle-GAN。不同的是,应当将其进行一些修改,使得它可以处理视频域的数据(cycle-GAN只能处理图像)。修改的方法是,将cycle-GAN网络中,所有的2D卷积结构,都改为面向视频数据的3D卷积结构。3D卷积结构的计算方法为:

[0031] $$V_{ij}^{xyz} = b_{ij} + \sum_{m}\sum_{p=0}^{P-1}\sum_{q=0}^{Q-1}\sum_{r=0}^{R-1} W_{ijm}^{pqr} V_{(i-1)m}^{(x+p)(y+q)(z+r)}$$

[0032] 其中P、Q、R分别表示上一层网络输出的特征图的长宽高,m表示上已成网络输出的特征图数量。最终计算得到在该卷积模块W下,下一层网络中对应的特征图V。同时在虚拟世界中模拟并录制相关的异常事件视频数据,在图中表示为圆角方框,即虚拟异常视频数据。这些数据包含打架,追逐,逃散,枪击,奔跑,逮捕等不同的异常类别和正常类别数据。各个类别的时评数据量大致相同。最后,还需要一部分真实的视频监控数据,用来表达现实场景中的监控视频是怎么样的,这些时评数据不需要标注,并且对视频内容也没有限制。

[0033] 步骤2,初始化一个视频分类网络,该网络可以是3DResNet,也可以是时空双流视频分类网络或者是其他已有的视频分类网络。这里我们采用的是已有的3DResNet,它来自于"K.Hara,H.Kataoka,and Y.Satoh,"Learning spatio-temporal features with 3D residual networks for action recognition,"Proceedings of the ICCV Workshop on Action,Gesture,and Emotion Recognition,vol.2,no.3,pp.4,2017"。这个网络是2015年提出的网络结构----ResNet的改进版本,它的改进方法和步骤一中阐述的相同,即将2D的卷积结构改为3D的卷积结构。

[0034] 步骤3,利用采集到的虚拟异常数据和任意真实数据,对一个域迁移网络进行训练,并获得虚拟异常视频数据对应的真实域视频数据。如图2所示,假定S_{real}、R_{real}分别为我们采集到的虚拟异常数据和任意真实数据,将其传送到生成网络G_{StoR}和G_{RtoS}中得到R_{fake}和S_{fake},再分别传入G_{RtoS}和G_{StoR}中,获得与S_{real}、R_{real}对应的视频,经过一致性对比以及鉴别器D_R和D_S的鉴别来提高域迁移后视频的逼真度。

[0035] 整个过程可以用下式表示:

[0036] $$\min_{G_{StoR},G_{RtoS}} \max_{D_S,D_R} L(G_{StoR}, G_{RtoS}, D_R, D_S)$$

[0037] 即在训练生成器的过程中,致力于最小化鉴别器的值与最大化一致性对比;在鉴别器训练过程中则最大化鉴别器的值。最终得到的R_{fake}便可以看作是图1中虚拟异常视频对应的真实域视频数据。

[0038] 步骤4,利用步骤3得到的真实域异常数据对步骤2得到的网络进行进一步分类训练,过程和2相同,从而得到真实域的异常视频分类网络。

[0039] 步骤5,实际测试过程中,将真实的异常数据输入到步骤4训练得到的网络模型,利用softmax函数获得该输入视频在隶属于各个异常类别中的概率,取最大值的类别作为该段视频的异常类型即可。

[0040] 本发明的效果可以通过以下仿真实验做进一步的说明。

[0041] 1.仿真条件

[0042] 本发明以四块GeForce GTX 1080 Ti GPU作为硬件基础,以64位Ubuntu 16.04 LTS系统上利用3.5.4版本的python编程语言,0.4.1版本的Pytorch和9.2版本的CUDA作为软件环境进行整个发明的实际演练。

[0043] 2.仿真内容

[0044] 首先,使用模拟得到的虚拟视频数据集和一些视频数据集中拿到的视频数据按照图1训练,最后得到真实域异常视频分类网络。并且使用了“K.Hara,H.Kataoka,and Y.Satoh,Learning spatio-temporal features with 3D residual networks for action recognition,Proceedings of the ICCV Workshop on Action,Gesture,and Emotion Recognition,vol.2,no.3,pp.4,2017.”和我们自己设计的网络进行对比,以及未经域迁移数据训练的模型和经过了域迁移数据训练的模型结果对比。评断标准有二,一是视频的分类准确率,二是错分类严重性(MISE,misclassification severity)。后者将异常类别按其严重性进行分级,然后计算错分类以后的严重程度。结果如下:

[0045] 表1:四个模型在真实数据集上的测试结果

[0046]

Accuracy(%)	3D ResNet	本发明
域迁移前	19.51	17.07
域迁移后	21.14	26.02

[0047] 从表1可以看出,本发明的网络在经过域迁移后在真实数据集上的分类准确率有了显著提升。并且本发明提出的域迁移技术对3DResNet的性能也有一定的效果提升,使其对公共场所异常行为检测具有更高的预报准确性。

[0048] 表2:四个模型在真实数据集上的错分类严重性

[0049]

MISE	3D ResNet	本发明
域迁移前	3.48	3.45
域迁移后	3.45	2.74

[0050] 从表2看出,我们的方法在错分类严重性上也具有最低的值,也印证了本发明对公共场所异常行为检测具有较低的错误分类严重性。

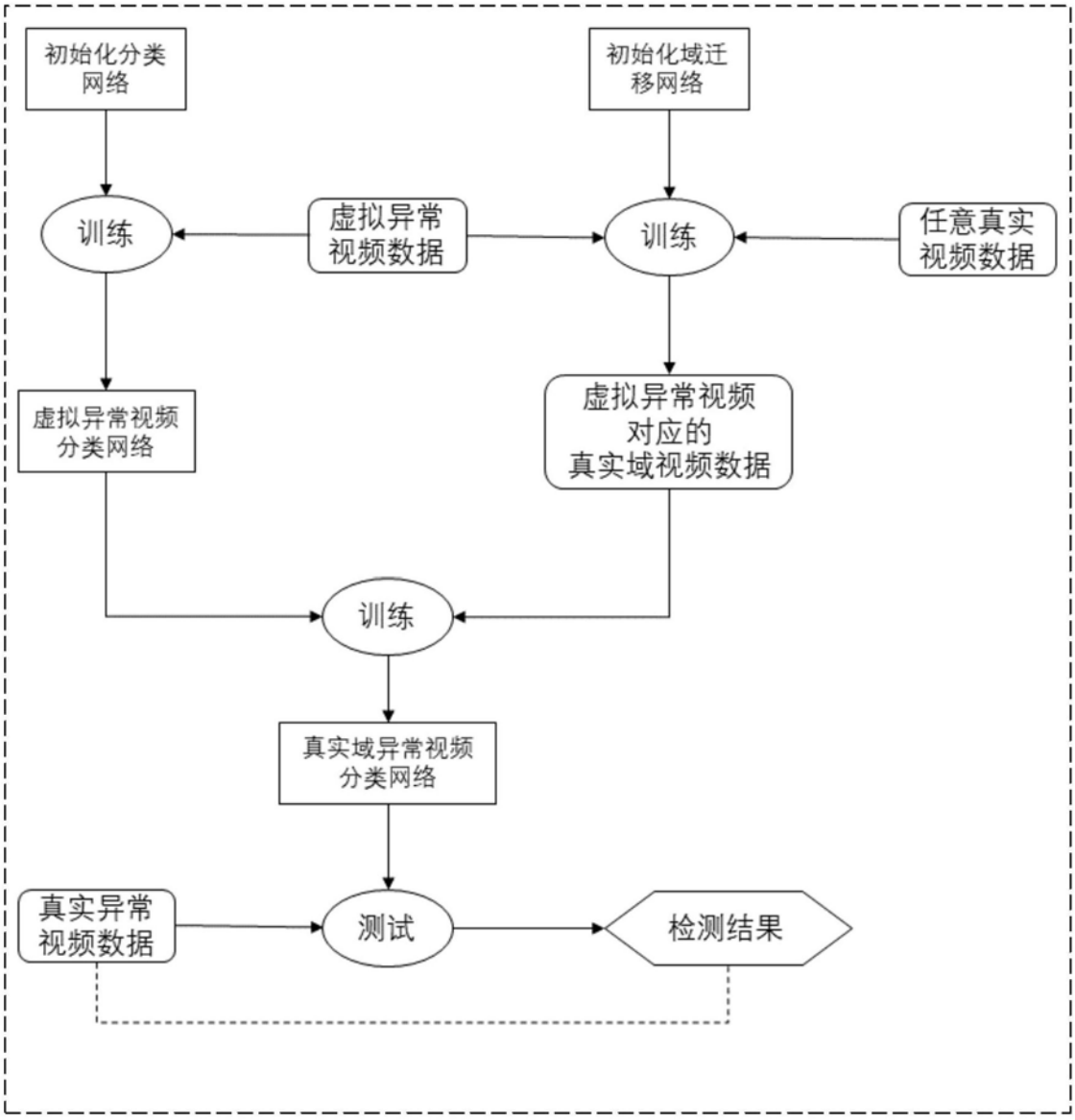

初始化分类网络
初始化域迁移网络
训练
虚拟异常视频数据
训练
任意真实视频数据
虚拟异常视频分类网络
虚拟异常视频对应的真实域视频数据
训练
真实域异常视频分类网络
真实异常视频数据
测试
检测结果

图1

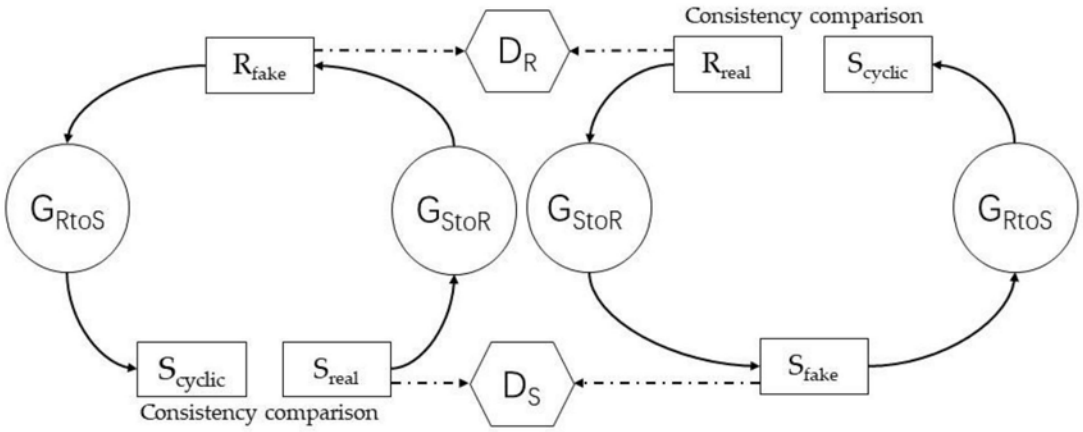

Consistency comparison
R_{fake}
D_R
R_{real}
S_{cyclic}
G_{RtoS}
G_{StoR}
G_{StoR}
G_{RtoS}
S_{cyclic}
S_{real}
D_S
S_{fake}
Consistency comparison

图2