



(12) 发明专利

(10) 授权公告号 CN 112183313 B

(45) 授权公告日 2022.03.11

(21) 申请号 202011030237.6

(22) 申请日 2020.09.27

(65) 同一申请的已公布的文献号  
申请公布号 CN 112183313 A

(43) 申请公布日 2021.01.05

(73) 专利权人 武汉大学  
地址 430067 湖北省武汉市武昌区八一路  
299号

(72) 发明人 王波 张迎晨 马富齐 罗鹏  
周胤宇 张天 王红霞 马恒瑞  
李怡凡 张嘉鑫

(74) 专利代理机构 苏州导思知识产权代理事务  
所(普通合伙) 32425  
代理人 龚建良

(51) Int.Cl.

- G06V 40/20 (2022.01)
- G06V 10/94 (2022.01)
- G06V 10/774 (2022.01)
- G06V 10/82 (2022.01)
- G06K 9/62 (2022.01)
- G06F 16/75 (2019.01)
- G06N 3/04 (2006.01)
- G06N 3/08 (2006.01)

审查员 苏晓燕

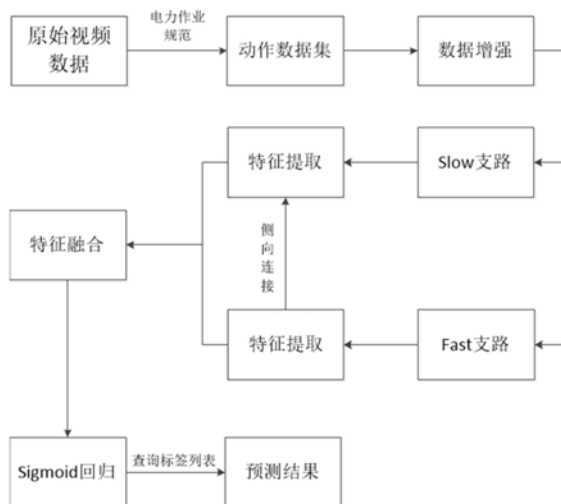
权利要求书2页 说明书4页 附图1页

(54) 发明名称

一种基于SlowFast的电力作业现场动作识别方法

(57) 摘要

本发明涉及一种电力作业现场动作识别方法,该方法基于SlowFast算法,属于深度学习视频识别技术,利用计算机视觉技术和计算机强大的计算能力对作业现场的视频图像进行分析,识别作业人员的动作,并与规范作业流程中包含的动作进行对比,起到实时监控的作用,能有效提高现场作业监控质量,降低现场作业安全风险。



1. 一种基于SlowFast的电力作业现场动作识别方法,包括以下步骤:

第一步,采集作业现场视频图像样本,针对每种作业类型进行视频采集;

第二步,人工视频预处理及标注,通过人工将采集得到的视频数据以人为画面中心进行裁剪,将视频尺寸进行归一化处理,按作业类型进行分类,通过视频剪辑,提取每种作业过程中的动作,按动作先后顺序将剪辑后的视频片段进行标注;

第三步,构建模型,利用Pytorch框架将视频中包含的图像信息、纹理特征、边缘信息、光流信息进行提取,将提取出的特征信息输入到神经网络,由神经网络计算模型参数,并输出基于视频的动作识别结果;

第四步,模型训练,基于已处理好的带标注的视频数据进行神经网络训练,通过验证集和测试集识别结果优化模型参数,对特殊的识别困难样本,可进行数据集补充以及修正标注,将新数据输入模型继续训练,通过专业的图形计算加速卡辅助计算,多次迭代后获得高精度动作识别模型;

第五步,模型评估,通过采集现场作业新视频数据,或现场实地测试动作模型,验证模型识别效果;

其中,实现步骤一的具体操作为:

准备5个可见光摄像头,布置在作业现场,围绕作业人员,从5个视角拍摄作业人员的作业画面,每种作业重复多次,并拍摄倒闸、接地、验电作业三类作业视频;

其中,实现步骤二的具体操作为:

(1) 将步骤一中采集的视频数据,按作业类型进行分类;

(2) 从电力现场作业规范中获取每种作业的规范动作,并制作成标签列表,例如倒闸作业的标签列表中:行标题为作业步骤序号,列标题为作业步骤具体动作名称;

(3) 依据标签列表对每段完整的作业视频进行动作切分,每个视频片段包含的动作对应于标签列表中的具体动作名称;

(4) 将剪辑后的视频片段,按照作业类型已经动作标签重新整理;

其中,实现步骤三的具体操作为:

(1) 在超算平台上安装算法实现需要的开发环境;

(2) SlowFast网络在时空中通过定向滤波捕获特征信息,时间与空间通过视频帧提取间隔来区分,将视频帧按16帧间隔提取,即获取图像跨度较大的少量帧作为Slow支路的输入,将视频帧按2帧间隔提取,即获取图像跨度小的大量帧作为Fast支路的输入,综合两条路径提取的特征信息能有效分析动作包含的语义信息,实现对动作时空的分析;

(3) SlowFast网络在空间维度上保持视频原始帧作为输入,即保留了空间分辨率1080P和可见光RGB颜色信息;

(4) SlowFast网络在空间维度上,输入的视频帧较少,但其需要分析提取的关键信息多且复杂,细粒程度更高,从而会产生大量的计算,大约占据80%算力,在时间维度上,输入的视频帧较多,但其需要分析提取的关键信息少且简单,细粒程度较低,从而不会产生大量的计算,只消耗约20%的算力;

(5) SlowFast网络中包含单向的由Fast支路指向Slow支路的侧向连接通道,将两种特征信息进行融合,两条支路输入的视频帧数不同,其特征图的维度也不同,在进行侧向连接时,对Fast支路的特征图利用 $5*1*1$ 的3D卷积核进行尺度变换,Slow支路的特征图求和,实

现特征融合；

(6) 将两条支路提取的特征信息进行串联后,输入全连接层进一步提取特征；

(7) 将(6)中提取后的特征输入sigmoid回归层进行回归计算,得到预测值；

(8) 根据标签列表查询预测值对应的动作标签,即为预测动作；

其中,实现步骤四的具体操作为：

(1) 将整理好的动作数据集输入模型中,模型自动提取视频帧；

(2) 将(1)中提取的视频帧进行数据增强,增强方法包含高斯模糊、随机光照、水平翻转；

(3) 将(2)中增强后的数据,依预设分别输入Slow支路与Fast支路；

(4) 两条支路进行一系列特征提取与特征融合后,将特征向量输入sigmoid回归层进行回归计算,得到预测值；

(5) 根据标签列表查询预测值对应的动作标签,即为预测动作；

(6) 在验证集将(5)中得到的预测动作标签与真实动作标签对比,计算验证集预测精度；

(7) 依据(6)中获得的验证集精度,对模型参数进行调整,迭代训练；

(8) 多轮训练后获得的理想模型参数在测试集上的预测精度即为最终模型的预测精度；

其中,实现步骤五的具体操作为：

(1) 在现场布置摄像头拍摄作业人员的作业画面,将视频流通过数据传输接口直接输入模型中；

(2) 由模型实时预测作业人员的动作,并记录预测结果；

(3) 由(2)中的预测结果分析模型实际应用效果,并以此为依据,增加数据集以及调整模型参数。

## 一种基于SlowFast的电力作业现场动作识别方法

### 技术领域

[0001] 本发明涉及电力作业安全管控技术领域,具体是一种基于SlowFast的电力作业现场动作识别方法。

### 背景技术

[0002] 在电力生产过程中,作业现场安全监督对于保证工作人员的安全具有重要意义。电力生产作业是相对比较复杂的工序,其中涉及很多的危险性工序。电力安全规程已经对电力作业提出了安全规范,要求在作业期间预防触电事故,做好安全措施。但是目前很多电力作业人员在工作中未树立正确的安全观念,缺乏足够的防范意识,严重影响作业的安全性与可靠性。若相关作业人员安全意识淡薄,未按照目前的电力作业特点、规律和现场实际情况合理参与到工作中,会严重影响电力作业的安全性和设备的安全运行。

[0003] 目前电力现场作业一般采用人工安全监护和视频监控的方式,但监护人员和作业人员一样容易受到外界因素影响,注意力可能会不集中,继而可能引发安全事故。综上所述,目前的电力现场作业人员安全监管所存在的问题可总结为如下几点:

[0004] (1) 现有现场作业的安全监督办法虽然较为完善,但仍需相关安全管理人员人为执行,因此并不能保证百分之百的落到实处,监督人员也不能保证能够全方位的实现安全监督管理办法的所有条例,也不能保证监督人员对目前开展的作业过程进行实时监控与风险预警信息实时反馈。

[0005] (2) 视频监控系统为安全监督提供了有效辅助,但实际的监控任务仍需要较多的人工工作来完成,监控图像的内容分析仍需依靠人工实时观看、手动调阅。并且现有的视频监控系统所提供的信息是原始的视频数据,监控系统通常只是录制视频图像,用作事后取证,不能充分发挥监控系统应有的实时主动的监督作用。

[0006] 本发明基于深度学习视频识别技术,利用计算机视觉技术和计算机强大的计算能力对作业现场的视频图像进行分析,识别作业人员的动作,并与规范作业流程中包含的动作进行对比,起到实时监控的作用。SlowFast算法是一种双数据流结构的视频识别算法,其快、慢两个数据流分别处理视频中的时间、空间信息,并在数据流之间构建数据传播通道,提升模型对时空信息的交叉感知能力,最后将双数据流提取的特征信息进行融合并识别。

### 发明内容

[0007] 为了改进现有技术存在的不足,本发明提出了一种基于SlowFast的电力作业现场动作识别方法。基于SlowFast算法与电力现场作业规范,训练算法模型,实现对常见的倒闸、接地、验电作业的动作识别。

[0008] 本发明提出一种基于SlowFast的电力作业现场动作识别方法,包括以下步骤:

[0009] 第一步,采集作业现场视频图像样本,针对每种作业类型进行视频采集;

[0010] 第二步,人工视频预处理及标注,通过人工将采集得到的视频数据以人为画面中心进行裁剪,将视频尺寸进行归一化处理,按作业类型进行分类,通过视频剪辑,提取每种

作业过程中的动作,按动作先后顺序将剪辑后的视频片段进行标注;

[0011] 第三步,构建模型,利用Pytorch框架将视频中包含的图像信息、纹理特征、边缘信息、光流信息进行提取,将提取出的特征信息输入到神经网络,由神经网络计算模型参数,并输出基于视频的动作识别结果;

[0012] 第四步,模型训练,基于已处理好的带标注的视频数据进行神经网络训练,通过验证集和测试集识别结果优化模型参数,对特殊的识别困难样本,可进行数据集补充以及修正标注,将新数据输入模型继续训练,通过专业的图形计算加速卡辅助计算,多次迭代后获得高精度动作识别模型;

[0013] 第五步,模型评估,通过采集现场作业新视频数据,或现场实地测试动作模型,验证模型识别效果。

[0014] 所述的第二步人工视频预处理及标注中涉及的视频剪辑与标注信息,是以电力现场作业规范为依据,将每种完整的作业过程包含的动作,按照作业规范中的动作描述进行划分,通过视频剪辑获得大量不同动作描述的动作视频片段,并将作业规范中的动作描述作为标注信息对视频片段进行分类标注。

[0015] 所述第三步构建模型将动作识别任务看作对动作时空的分析,并在时空中通过定向滤波捕获特征信息。时间与空间通过视频帧提取间隔来区分,借鉴生物视觉的生理特性,动作空间分析是指分析对象的颜色、纹理、光线等变化缓慢的信息,动作时间分析是指分析对象快速运动部分的位置、姿态、方向等高速变化的信息。将视频帧按较大间隔提取,即获取图像跨度较大的少量帧作为动作空间分析的输入,将视频帧按较小间隔提取,即获取图像跨度小的大量帧作为动作时间分析的输入。综合两条路径提取的特征信息能有效分析动作包含的语义信息,实现对动作时空的分析。

[0016] 所述第三步构建模型采用的SlowFast网络在空间维度上保持视频原始帧作为输入,即保留了空间分辨率和可见光颜色信息,这有助于Slow支路提取动作空间信息;在时间维度上输入的视频帧会进行预处理,包括降低空间分辨率和去除颜色信息,这种处理方式不仅可以减少Fast支路对空间信息的捕获能力,同时也能增强其对时间信息的捕获能力。

[0017] 所述的SlowFast网络在空间维度上,输入的视频帧较少,但其需要分析提取的关键信息多且复杂,细粒程度更高,从而会产生大量的计算,需要占据更多的算力,在时间维度上,输入的视频帧较多,但其需要分析提取的关键信息少且简单,细粒程度较低,从而不会产生大量的计算,只消耗较少的算力。

[0018] 所述的SlowFast网络中包含单向的由Fast支路指向Slow支路的侧向连接通道,目的是为了将两种特征信息进行融合,由于两条支路输入的视频帧数不同,其特征图的维度也不同,在进行侧向连接时,需要对Fast支路的特征图利用3D卷积运算进行尺度变换,匹配Slow支路的特征图维度,才能实现特征融合。

[0019] 所述的SlowFast网络在模型训练过程中,为了获得更好的鲁棒模型,会对输入数据进行增强,增强方式法包含高斯模糊、随机光照、水平翻转。

[0020] 所述的SlowFast网络在模型训练过程中,包含3D卷积运算,采用多GPU并行计算的方式,能有效提高模型训练效率,使得模型参数快速收敛到理想的最小值,在验证集与测试集上获得更高的识别精度。

## 附图说明

[0021] 图1为本发明模型结构示意图

## 具体实施方式

[0022] 下面将结合附图对本发明的具体实施方式作进一步的描述：

[0023] 本发明提出一种基于SlowFast的电力作业现场动作识别方法，包括以下步骤：

[0024] 第一步，采集作业现场视频图像样本，针对每种作业类型进行视频采集；

[0025] 第二步，人工视频预处理及标注，通过人工将采集得到的视频数据以人为画面中心进行裁剪，将视频尺寸进行归一化处理，按作业类型进行分类，通过视频剪辑，提取每种作业过程中的动作，按动作先后顺序将剪辑后的视频片段进行标注；

[0026] 第三步，构建模型，利用Pytorch框架将视频中包含的图像信息、纹理特征、边缘信息、光流信息进行提取，将提取出的特征信息输入到神经网络，由神经网络计算模型参数，并输出基于视频的动作识别结果；

[0027] 第四步，模型训练，基于已处理好的带标注的视频数据进行神经网络训练，通过验证集和测试集识别结果优化模型参数，对特殊的识别困难样本，可进行数据集补充以及修正标注，将新数据输入模型继续训练，通过专业的图形计算加速卡辅助计算，多次迭代后获得高精度动作识别模型；

[0028] 第五步，模型评估，通过采集现场作业新视频数据，或现场实地测试动作模型，验证模型识别效果。

[0029] 实现步骤一的具体操作为：

[0030] 准备5个可见光摄像头，布置在作业现场，围绕作业人员，从5个视角拍摄作业人员的作业画面，每种作业重复多次，并拍摄倒闸、接地、验电作业三类作业视频。

[0031] 实现步骤二的具体操作为：

[0032] (1) 将步骤一中采集的视频数据，按作业类型进行分类；

[0033] (2) 从电力现场作业规范中获取每种作业的规范动作，并制作成标签列表，例如倒闸作业的标签列表中：行标题为作业步骤序号，列标题为作业步骤具体动作名称；

[0034] (3) 依据标签列表对每段完整的作业视频进行动作切分，每个视频片段包含的动作对应于标签列表中的具体动作名称；

[0035] (4) 将剪辑后的视频片段，按照作业类型已经动作标签重新整理。

[0036] 实现步骤三的具体操作为：

[0037] (1) 在超算平台上安装算法实现需要的开发环境；

[0038] (2) SlowFast网络在时空中通过定向滤波捕获特征信息。时间与空间通过视频帧提取间隔来区分。将视频帧按16帧间隔提取，即获取图像跨度较大的少量帧作为Slow支路的输入，将视频帧按2帧间隔提取，即获取图像跨度小的大量帧作为Fast支路的输入。综合两条路径提取的特征信息能有效分析动作包含的语义信息，实现对动作时空的分析；

[0039] (3) SlowFast网络在空间维度上保持视频原始帧作为输入，即保留了空间分辨率1080P和可见光RGB颜色信息；

[0040] (4) SlowFast网络在空间维度上，输入的视频帧较少，但其需要分析提取的关键信息多且复杂，细粒程度更高，从而会产生大量的计算，大约占据80%算力，在时间维度上，

输入的视频帧较多,但其需要分析提取的关键信息少且简单,细粒程度较低,从而不会产生大量的计算,只消耗约20%的算力;

[0041] (5) SlowFast网络中包含单向的由Fast支路指向Slow支路的侧向连接通道,目的是为了将两种特征信息进行融合,由于两条支路输入的视频帧数不同,其特征图的维度也不同,在进行侧向连接时,需要对Fast支路的特征图利用 $5*1*1$ 的3D卷积核进行尺度变换,Slow支路的特征图求和,实现特征融合;

[0042] (6) 将两条支路提取的特征信息进行串联后,输入全连接层进一步提取特征;

[0043] (7) 将(6)中提取后的特征输入sigmoid回归层进行回归计算,得到预测值;

[0044] (8) 根据标签列表查询预测值对应的动作标签,即为预测动作。

[0045] 实现步骤四的具体操作为:

[0046] (1) 将整理好的动作数据集输入模型中,模型自动提取视频帧;

[0047] (2) 将(1)中提取的视频帧进行数据增强,增强方法包含高斯模糊、随机光照、水平翻转。

[0048] (3) 将(2)中增强后的数据,依预设分别输入Slow支路与Fast支路;

[0049] (4) 两条支路进行一系列特征提取与特征融合后,将特征向量输入sigmoid回归层进行回归计算,得到预测值;

[0050] (5) 根据标签列表查询预测值对应的动作标签,即为预测动作;

[0051] (6) 在验证集将(5)中得到的预测动作标签与真实动作标签对比,计算验证集预测精度;

[0052] (7) 依据(6)中获得的验证集精度,对模型参数进行调整,迭代训练;

[0053] (8) 多轮训练后获得的理想模型参数在测试集上的预测精度即为最终模型的预测精度。

[0054] 实现步骤五的具体操作为:

[0055] (1) 在现场布置摄像头拍摄作业人员的作业画面,将视频流通过数据传输接口直接输入模型中;

[0056] (2) 由模型实时预测作业人员的动作,并记录预测结果;

[0057] (3) 由(2)中的预测结果分析模型实际应用效果,并以此为依据,增加数据集以及调整模型参数。

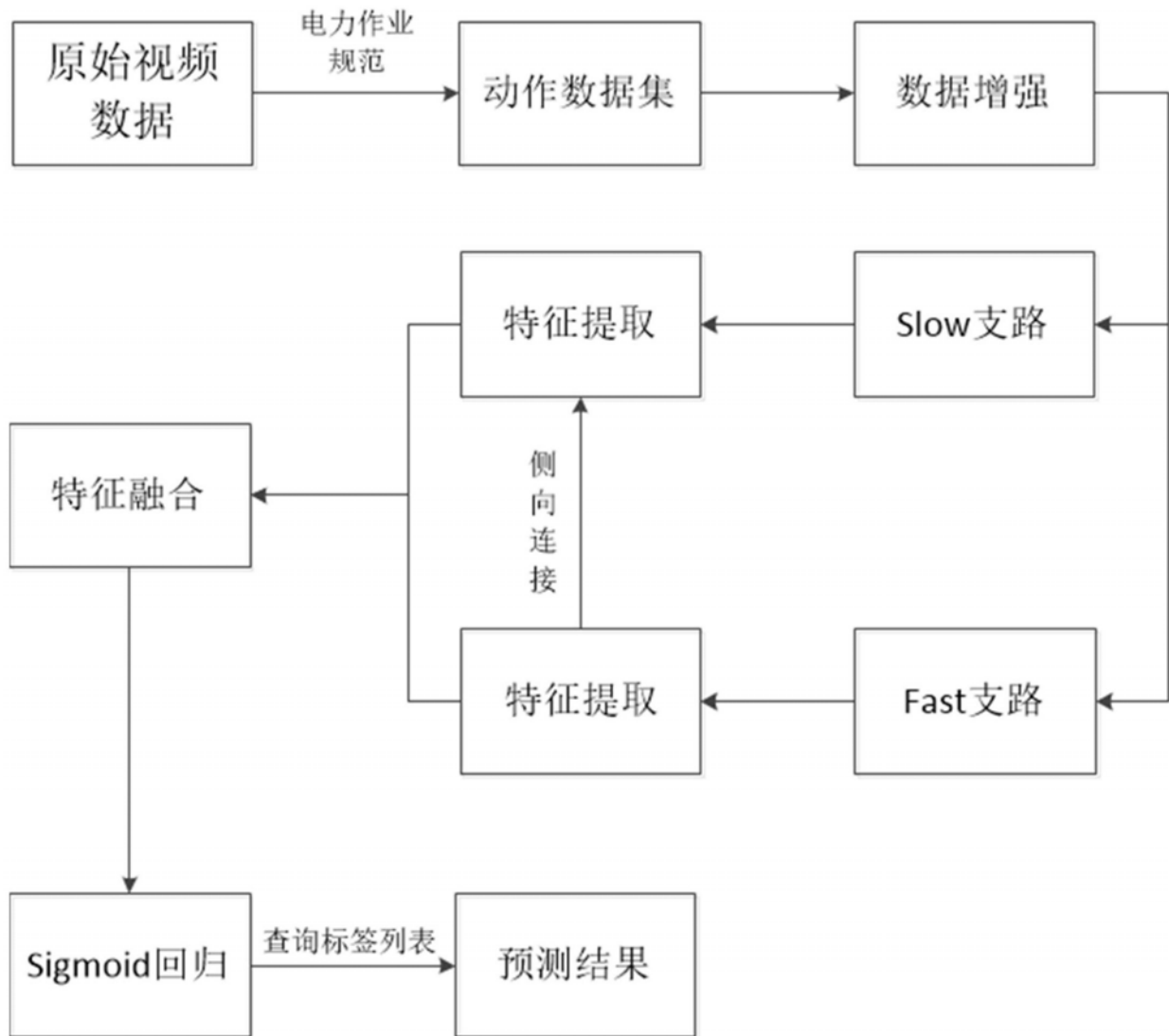


图1