(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 110363140 A
(43)申请公布日 2019.10.22

(21)申请号 201910635697.2

(22)申请日 2019.07.15

(71)申请人 成都理工大学
地址 610059 四川省成都市成华区二仙桥东三路1号

(72)发明人 易诗 谢家海

(74)专利代理机构 成都弘毅天承知识产权代理有限公司 51230
代理人 杨保刚

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

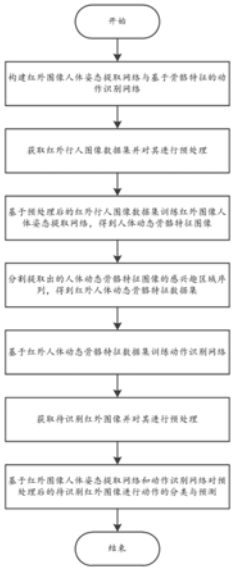
权利要求书2页 说明书9页 附图5页

(54)发明名称

一种基于红外图像的人体动作实时识别方法

(57)摘要

本发明公开了一种基于红外图像的人体动作实时识别方法，涉及人体动作识别技术领域，包括以下步骤：构建红外图像人体姿态提取网络与基于骨骼特征的动作识别网络；获取红外行人图像数据集并对其进行预处理，基于预处理后的红外行人图像数据集训练红外图像人体姿态提取网络，得到人体动态骨骼特征图像；分割人体动态骨骼特征图像的感兴趣区域序列，利用分割结果训练动作识别网络；获取待识别红外图像，基于红外图像人体姿态提取网络和动作识别网络对预处理后的待识别红外图像进行动作的分类与预测。本发明解决了现有的行为识别方法普遍针对可见光环境，在夜间无光或天气恶劣环境下通过红外图像进行人体行为动作识别存在实时性差、识别率低的问题。



CN 110363140 A

1.一种基于红外图像的人体动作实时识别方法,其特征在于,包括以下步骤:

构建红外图像人体姿态提取网络与基于骨骼特征的动作识别网络SaNet;

获取红外行人图像数据集并对其进行预处理,基于预处理后的红外行人图像数据集训练红外图像人体姿态提取网络,得到人体动态骨骼特征图像;

分割提取出的人体动态骨骼特征图像的感兴趣区域序列,得到红外人体动态骨骼特征数据集,基于红外人体动态骨骼特征数据集训练动作识别网络SaNet;

获取待识别红外图像并对其进行预处理,基于红外图像人体姿态提取网络和动作识别网络SaNet对预处理后的待识别红外图像进行动作的分类与预测。

2.根据权利要求1所述的一种基于红外图像的人体动作实时识别方法,其特征在于,红外图像人体姿态提取网络结构由基础网络MS-RsNet与CenterNet构架的检测网络构成。

3.根据权利要求2所述的一种基于红外图像的人体动作实时识别方法,其特征在于,MS-RsNet的获取方式为:在ResNet101网络结构基础上,抽取卷积层3、卷积层4、卷积层5的特征图在三个尺度上的特征输出并融合,形成多尺度金字塔特征提取结构,再将首个卷积层内卷积核换为单通道卷积核,得到多尺度ResNet网络,即基础网络MS-RsNet。

4.根据权利要求1或2所述的一种基于红外图像的人体动作实时识别方法,其特征在于,红外图像人体姿态提取网络训练过程的损失函数定义如下:

$L=L_{det}+L_{off}$

上式中,L_{det}表示中心点的散焦损失,用于训练检测目标边缘与中心点;L_{off}表示中心关键点偏移损失,用于预测偏移值。

5.根据权利要求1所述的一种基于红外图像的人体动作实时识别方法,其特征在于,基于骨骼特征的动作识别网络SaNet由2个卷积层、2个最大池化层、2个全连接层、1个ReLU激活函数、1个平滑层和Softmax分类函数构成,用以识别包括行走、骑车、跑步、跳跃、攀爬、下蹲在内的6种动作。

6.根据权利要求1所述的一种基于红外图像的人体动作实时识别方法,其特征在于,采用背景抑制方法对红外行人图像数据集和待识别红外图像进行预处理,对红外行人图像进行预处理的具体方法如下:

步骤A1:采用多尺度的图像细节提升方法提升红外行人图像中人体细节与背景的对比度,得到细节增强图像;细节增强图像的获取方式如下:

$D=(1-0.5\times sgn(D_1))\times D_1+0.5\times D_2+0.25\times D_3$

上式中,D表示处理后的细节增强图像,sgn()表示符号函数,D_1、D_2、D_3分别表示三个尺度上的细节增强处理,D_1、D_2、D_3的计算方法分别为:

$D_1=I-B_1$、$D_2=I-B_2$、$D_3=I-B_3$

其中,I表示原始图像,中间参数B_1、B_2、B_3的计算方法分别为:

$B_1=G_1*I$、$B_2=G_2*I$、$B_3=G_3*I$

其中,G_1、G_2、G_3分别表示方差为1、2、4的高斯核;

步骤A2:使用双边滤波抑制细节增强图像中对比度低的细节部分,得到滤波图像;滤波图像的获取方式如下:

$$I_p = \frac{1}{W_p}\sum_{q\in S} G_{\delta_s}(\|p-q\|)\, G_{\delta_r}(\|I_p - I_q\|) I_q$$

上式中，p表示图像当前像素点，q表示图像空间邻域像素点，I_p表示处理后得到的滤波图像，“‖ ‖”表示求取两个值之间的欧式距离，I_q表示输入的细节增强图像；G表示高斯核，对于两个参数取值，δ_s表示原始图像斜对角线长度的2％的数值，δ_r表示原始图像梯度值的中值或者平均数；W_p表示权重，其计算方法如下：

$$W_p = \sum_{q\in S} G_{\delta_s}(\|p-q\|)\, G_{\delta_r}(\|I_p - I_q\|)$$

上式中，S表示图像空间域。

7.根据权利要求1所述的一种基于红外图像的人体动作实时识别方法，其特征在于，得到人体动态骨骼特征图像的具体方法如下：

步骤B1：采用CVC-09红外数据集中包含行人的M幅红外图像作为训练集A，以及实际通过红外热成像采集设备采集的红外视频中截取的包含行人的N幅红外图像作为验证集，训练集A与验证集数量比例为5:1；其中，M、N均为常数；

步骤B2：将红外行人图像数据集中的图像样本全部转换为512×512像素的图像并进行图像预处理，利用预处理后的训练集A训练红外图像人体姿态提取网络，选取精度最高的网络模型A，并基于该模型提取出人体动态骨骼特征图像。

8.根据权利要求1所述的一种基于红外图像的人体动作实时识别方法，其特征在于，训练动作识别网络SaNet的具体操作如下：

步骤C1：剪裁提取出的每个人体动态骨骼特征图像的感兴趣区域，形成人体动态骨骼ROI图像序列，依次输入动作识别网络SaNet；

步骤C2：使用Labelimg工具对所有提取的人体动态骨骼ROI图像序列进行动作标注，区分需要识别的6类动作，得到红外人体动态骨骼特征数据集，并将红外人体动态骨骼特征数据集按5:1的比例划分成训练集B和测试集；

步骤C3：将红外人体动态骨骼特征数据集中所有图像样本转换为单通道28×28像素的图像，利用训练集B训练动作识别网络SaNet，选取精度最高的网络模型B，并基于该模型识别分类6种动作。

9.根据权利要求1所述的一种基于红外图像的人体动作实时识别方法，其特征在于，识别待识别红外图像的具体方法如下：

步骤D1：获取待识别红外图像，先将该图像调整为512×512像素，再对调整大小后的待识别红外图像进行预处理，得到预处理图像；

步骤D2：利用红外图像人体姿态提取网络提取预处理图像的动态骨骼特征，得到待识别人体动态骨骼特征图；

步骤D3：截取待识别人体动态骨骼特征图的感兴趣区域，作为动作识别网络SaNet的输入序列，并将待识别人体动态骨骼特征图调整为28×28像素，利用动作识别网络SaNet对大小经调整后的待识别人体动态骨骼特征图进行动作的分类与预测。

一种基于红外图像的人体动作实时识别方法

技术领域

[0001] 本发明涉及人体动作识别技术领域，具体涉及一种基于红外图像的人体动作实时识别方法。

背景技术

[0002] 红外热成像系统成像原理为红外光谱辐射成像，不依赖光源，受天气影响小，探测距离远，在夜间全黑环境下进行目标识别与探测、搜救、军事、行车辅助等领域具有很强应用价值。随着机器视觉与人工智能的快速发展，其运用于红外热成像图像中图像复原、目标跟踪、目标检测与识别、语义分割等方向已取得一定突破。而在夜间无光环境下或/和气候恶劣条件下使用红外热成像对人体行为、动作进行智能化识别与分析的研究还较少，现有的大量行为识别、动作识别技术均基于可见光环境，对于全黑无光及雨雾天气等环境下的动作识别方法缺乏研究与实践。

[0003] 在可见光环境下，具有代表性的行为动作识别方法主要包括ChristophFeichtenhofer等人提出的卷积双光流网络融合视频动作识别方法，Ali Diba等人提出的深时线性编码网络，视频动作识别的时空残差网络等方法，上述方法基本思想均是使用多帧视频信息作为训练输入，使用深度卷积网络提取动作信息，在可见光人体行为公开数据集上取得了良好的识别分类效果。

[0004] 而对于红外热成像视频图像，具有以下特点：(1)图像分辨率较低，目前普遍主流为384×288分辨率；(2)红外图像中目标边缘特征相对可见光图像细节模糊；(3)缺乏色彩特征，红外图像为单通道图像，输出类似于灰度图。因此，提取多帧红外图像信息难度较高，会造成特征难以捕获，连续性差，上述行为动作识别方法对于红外图像的运用效果并不理想。

[0005] Sijie Yan等人提出了一种基于动态骨骼的动作识别方法ST-GCN(时空图卷积网络模型)，该方法提出动态人类骨骼通常能与其他模态相辅相成，传达重要信息，构造一个时空图。ST-GCN的输入是图节点的联合坐标向量，其中，人体关节对应图的节点，身体结构的连通性和时间上的连通性对应图的两类边，ST-GCN可被认为是一个基于图像的CNN模拟，输入由2D图像网格上的像素强度矢量形成。对输入数据应用多层的时空图卷积操作，可以生成更高级别的特征图，再通过标准的SoftMax分类器将其分类到相应的动作类别。整个模型用反向传播进行端对端方式的训练，方法的基本思想为使用姿态提取框架提取视频中人体动态骨骼特征输入后端卷积神经网络进行识别与分类，对于红外图像中人体动作识别具有很强的指导意义，但该方法依然建立于可见光环境下，提取多帧视频特征进行行为预测，且模型构架较为复杂，实时性有待提高。直接使用该方法进行红外图像中人体动作识别效果仍然存在不足。现有基于可见光环境的行为动作识别方法均不满足于实时红外图像人体动作识别需求，因此，设计一种实时的红外热成像人体动作识别方法意义重大。

发明内容

[0006] 本发明的目的在于:为解决现有的通过红外图像进行人体行为动作识别方法识别出动作实时性差、识别率低的问题,提供了一种基于红外图像的人体动作实时识别方法。

[0007] 本发明采用的技术方案如下:

[0008] 一种基于红外图像的人体动作实时识别方法,包括以下步骤:

[0009] 构建红外图像人体姿态提取网络与基于骨骼特征的动作识别网络SaNet;

[0010] 获取红外行人图像数据集并对其进行预处理,基于预处理后的红外行人图像数据集训练红外图像人体姿态提取网络,得到人体动态骨骼特征图像;

[0011] 分割提取出的人体动态骨骼特征图像的感兴趣区域序列,得到红外人体动态骨骼特征数据集,基于红外人体动态骨骼特征数据集训练动作识别网络SaNet;

[0012] 获取待识别红外图像并对其进行预处理,基于红外图像人体姿态提取网络和动作识别网络SaNet对预处理后的待识别红外图像进行动作的分类与预测。

[0013] 进一步地,红外图像人体姿态提取网络结构由基础网络MS-RsNet与CenterNet构架的检测网络构成。

[0014] 进一步地,MS-RsNet的获取方式为:在ResNet101网络结构基础上,抽取卷积层3、卷积层4、卷积层5的特征图在三个尺度上的特征输出并融合,形成多尺度金字塔特征提取结构,再将首个卷积层内卷积核换为单通道卷积核,得到多尺度ResNet网络,即基础网络MS-RsNet。

[0015] 进一步地,红外图像人体姿态提取网络训练过程的损失函数定义如下:

[0016] $L=L_{det}+L_{off}$

[0017] 上式中,L_{det}表示中心点的散焦损失,用于训练检测目标边缘与中心点;L_{off}表示中心关键点偏移损失,用于预测偏移值。

[0018] 进一步地,基于骨骼特征的动作识别网络SaNet由2个卷积层、2个最大池化层、2个全连接层、1个ReLU激活函数、1个平滑层和Softmax分类函数构成,用以识别包括行走、骑车、跑步、跳跃、攀爬、下蹲在内的6种动作。

[0019] 进一步地,采用背景抑制方法对红外行人图像数据集和待识别红外图像进行预处理,对红外行人图像进行预处理的具体方法如下:

[0020] 步骤A1:采用多尺度的图像细节提升方法提升红外行人图像中人体细节与背景的对比度,得到细节增强图像;细节增强图像的获取方式如下:

[0021] $D=(1-0.5\times sgn(D_1))\times D_1+0.5\times D_2+0.25\times D_3$

[0022] 上式中,D表示处理后的细节增强图像,sgn()表示符号函数,D_1、D_2、D_3分别表示三个尺度上的细节增强处理,D_1、D_2、D_3的计算方法分别为:

[0023] $D_1=I-B_1$、$D_2=I-B_2$、$D_3=I-B_3$

[0024] 其中,I表示原始图像,中间参数B_1、B_2、B_3的计算方法分别为:

[0025] $B_1=G_1*I$、$B_2=G_2*I$、$B_3=G_3*I$

[0026] 其中,G_1、G_2、G_3分别表示方差为1、2、4的高斯核;

[0027] 步骤A2:使用双边滤波抑制细节增强图像中对比度低的细节部分,得到滤波图像;滤波图像的获取方式如下:

[0028] $$I_p = \frac{1}{W_p}\sum_{q\in S} G_{\delta_s}(\|p-q\|)\, G_{\delta_r}(\|I_p - I_q\|) I_q$$

[0029] 上式中，p表示图像当前像素点，q表示图像空间邻域像素点，I_p表示处理后得到的滤波图像，“||||”表示求取两个值之间的欧式距离，I_q表示输入的细节增强图像；G表示高斯核，对于两个参数取值，δ_s表示原始图像斜对角线长度的2％的数值，δ_r表示原始图像梯度值的中值或者平均数；W_p表示权重，其计算方法如下：

[0030] $$W_p = \sum_{q\in S} G_{\delta_s}(\|p-q\|)\, G_{\delta_r}(\|I_p - I_q\|)$$

[0031] 上式中，S表示图像空间域(spatial domain)。

[0032] 进一步地，得到人体动态骨骼特征图像的具体方法如下：

[0033] 步骤B1：采用CVC-09红外数据集中包含行人的6500幅红外图像作为训练集A，以及实际通过红外热成像采集设备采集的红外视频中截取的包含行人的1500幅红外图像作为验证集，训练集A与验证集数量比例为5∶1；

[0034] 步骤B2：将红外行人图像数据集中的图像样本全部转换为512×512像素的图像并进行图像预处理，利用预处理后的训练集A训练红外图像人体姿态提取网络，选取精度最高的网络模型A，并基于该模型提取出人体动态骨骼特征图像。

[0035] 进一步地，训练动作识别网络SaNet的具体操作如下：

[0036] 步骤C1：剪裁提取出的每个人体动态骨骼特征图像的感兴趣区域，形成人体动态骨骼ROI图像序列，依次输入动作识别网络SaNet；

[0037] 步骤C2：使用Labelimg工具对所有提取的人体动态骨骼ROI图像序列进行动作标注，区分需要识别的6类动作，得到红外人体动态骨骼特征数据集，并将红外人体动态骨骼特征数据集按5∶1的比例划分成训练集B和测试集；

[0038] 步骤C3：将红外人体动态骨骼特征数据集中所有图像样本转换为单通道28×28像素的图像，利用训练集B训练动作识别网络SaNet，选取精度最高的网络模型B，并基于该模型识别分类6种动作。

[0039] 进一步地，识别待识别红外图像的具体方法如下：

[0040] 步骤D1：获取待识别红外图像，先将该图像调整为512×512像素，再对调整大小后的待识别红外图像进行预处理，得到预处理图像；

[0041] 步骤D2：利用红外图像人体姿态提取网络提取预处理图像的动态骨骼特征，得到待识别人体动态骨骼特征图；

[0042] 步骤D3：截取待识别人体动态骨骼特征图的感兴趣区域，作为动作识别网络SaNet的输入序列，并将待识别人体动态骨骼特征图调整为28×28像素，利用动作识别网络SaNet对大小经调整后的待识别人体动态骨骼特征图进行动作的分类与预测。

[0043] 综上所述，由于采用了上述技术方案，本发明的有益效果是：

[0044] 1、本发明中，提出了一种针对红外热成像图像的人体姿态提取网络，该网络提出了一种有利于提取不同距离人体骨骼特征的多尺度ResNet网络，针对于红外图像单通道数据的特点，在主干网络首个卷积层采用单通道卷积核以降低运算量，提高实时性。并且，检测部分使用了基于CenterNet架构的高效实时性姿态提取检测方法，从而提高了在红外图

像人体姿态提取环节上的提取准确度和提取实时性。

[0045] 2、本发明中，对于提取出的红外图像中人体动态骨骼特征，将感兴趣区域剪裁为图像序列，在动作识别环节，考虑到骨骼特征相对动作的显著性，红外图像中存在的特征提取不连续性，使用了单帧图像动作识别方式并提出了一种基于精简型、轻量化动态骨骼图像对应动作识别的卷积神经网络SaNet，在准确识别骨骼特征对应的动作同时，减少运算量，提高了实时性。

[0046] 3、本发明中，通过红外图像背景抑制的预处理方法突出了红外热成像图像中热源目标的显著性，抑制了背景噪声，提高了后续姿态检测、动作识别的精度。

[0047] 4、本发明中，采集行人图像数据采用红外热成像仪，因此可以应用于夜间无光，存在雨雾等天气影响环境等可见光摄像头和普通数码夜视仪无法应对的探测环境，可在百米左右采集清晰的行人红外光谱成像，进行后期人体行为识别。

[0048] 5、本发明中，运用深度学习技术提取红外图像中人体姿态骨架，通过卷积神经网络对骨架特征进行识别分类，高效实时完成红外热成像中人体动作的识别，对无光、天气较为恶劣环境下的搜救、安防、反恐等领域具备重大应用价值。本方法解决了现有的行为识别方法普遍针对可见光环境，在夜间无光或天气恶劣环境下通过红外图像进行人体行为动作识别存在实时性差、识别率低的问题。

附图说明

[0049] 为了更清楚地说明本发明实施例的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，应当理解，以下附图仅示出了本发明的某些实施例，因此不应被看作是对范围的限定，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他相关的附图。

[0050] 图1为本发明的整体流程图；

[0051] 图2为本发明的多尺度ResNet网络主体结构图；

[0052] 图3为本发明的检测网络CenterNet结构图；

[0053] 图4为本发明的SaNet网络构架图；

[0054] 图5为本发明在全黑环境下使用红外热成像仪采集的红外行人图像；

[0055] 图6为本发明实施例一的部分流程示意图；

[0056] 图7为本发明实施例一实时红外热成像人体动作识别测试结果图。

具体实施方式

[0057] 为了使本发明的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本发明进行进一步详细说明。应当理解，此处所描述的具体实施例仅用以解释本发明，并不用于限定本发明，即所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。通常在此处附图中描述和示出的本发明实施例的组件可以以各种不同的配置来布置和设计。因此，以下对在附图中提供的本发明的实施例的详细描述并非旨在限制要求保护的本发明的范围，而是仅仅表示本发明的选定实施例。基于本发明的实施例，本领域技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0058] 需要说明的是，术语“第一”和“第二”等之类的关系术语仅仅用来将一个实体或者

操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个.....”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0059] 下面结合实施例对本发明的特征和性能作进一步的详细描述。

[0060] 实施例一

[0061] 本发明的较佳实施例，提供了一种基于红外图像的人体动作实时识别方法，如图1所示，包括以下步骤：

[0062] 步骤1：构建红外图像人体姿态提取网络与基于骨骼特征的动作识别网络。

[0063] 红外图像人体姿态提取网络结构由基础网络MS-RsNet与CenterNet构架的检测网络构成。其中：

[0064] (1)多尺度ResNet网络-MS-ResNet

[0065] 多尺度ResNet网络通过检测提取滤波图像(通过对图像预处理可得)中人体的动态骨骼特征，为后续的动作识别提供输入。目前，针对人体姿态提取的框架中，具有代表性的有OpenPose、DensePose、AlphaPose，上述框架均在可见光数据集上取得了良好的检测效果，而针对红外图像中人体动态骨骼提取的准确性与实时性方面，均不能满足需求。

[0066] 为提高多个距离上红外图像中人体的检测与骨骼提取准确率，考虑实时性与检测准确性的均衡，本发明在ResNet101网络结构基础上，抽取卷积层3、卷积层4、卷积层5的特征图在三个尺度上的特征输出并融合，形成多尺度金字塔特征提取结构，以供后续检测部分进行多尺度分类预测，提高各个距离及尺度上检测红外人体和骨骼提取的精度。针对红外图像单通道16bit的图像格式，再将首个卷积层内卷积核换为单通道卷积核，从而减少首个卷积层2/3的运算量，并提高实时性，得到多尺度ResNet网络(MultScale-ResNet，简称MS-ResNet)，如图2所示。

[0067] (2)多尺度ResNet网络提取多尺度特征后，在三个尺度上输入检测部分进行动态骨骼数据的检测提取。检测部分为进一步提高检测准确性与实时性，每一个预测尺度上均使用CenterNet结构。

[0068] 检测网络CenterNet结构由2个卷积归一化残差融合层(Conv-BN-ReLU)、1个左池化层(Left Pooling)、1个右池化层(Right Pooling)、1个顶层池化层(Top Pooling)、1个底层池化层(Bottom Pooling)组成，作用在于预测中心关键点的分支，有利于中心获得更多目标物的中心区域，进而更易感知proposal的中心区域。通过取中心位置横向与纵向响应值的和的最大值实现快速准确地提取红外图像中人体姿态关节点热力图，形成人体动态骨骼图像。检测部分的网络结构如图3所示。

[0069] CenterNet与基于锚点的one-stage方法相近，中心点可看成形状未知的锚点，但其分配的锚点仅仅是放在位置上，没有尺寸框，没有手动设置的阈值做前后景分类。每个目标仅仅有一个正的锚点，因此不会用到NMS，提取关键点特征图上局部峰值点。CenterNet相比较传统目标检测(缩放16倍尺度)而言，使用更大分辨率的输出特征图(缩放了4倍)，因此无需用到多重特征图锚点。使用CenterNet结构提取图像中多个人体的k个2D关节点位置，

令中心点的姿态是kx2维的，然后将每个关键点(关节点对应的点)参数化为相对于中心点的偏移。只用对检测到的目标框中的关节点进行关联。综上，该结构能够在提升姿态检测准确率的同时，极大提高检测实时性。

[0070] 红外图像人体姿态提取网络训练过程的损失函数定义如下：

[0071] $L=L_{det}+L_{off}$

[0072] 上式中，L_{det}表示中心点的散焦损失，用于训练检测目标边缘与中心点；L_{off}表示中心关键点偏移损失，用于预测偏移值。

[0073] (3)基于骨骼特征的动作识别网络SaNet

[0074] 本发明提出了一种动态骨骼到动作的精简型卷积神经网络(Skeleton-action Net，简称SaNet)，用以识别包括行走、骑车、跑步、跳跃、攀爬、下蹲在内的6种动作。SaNet网络构架如图4所示，由2个卷积层、2个最大池化层、2个全连接层、1个ReLU激活函数、1个平滑层和Softmax分类函数构成，网络结构精简，运算量小，可以准确识别骨骼特征对应的动作，满足识别精度与实时性要求。

[0075] 步骤2：获取红外行人图像数据集，基于红外行人图像数据集训练红外图像人体姿态提取网络，得到人体动态骨骼特征图像。

[0076] 步骤B1：本发明的红外行人图像数据集采用CVC-09红外数据集中包含行人的6500幅红外图像作为训练集A，以及通过红外热成像仪采集的红外视频中截取的包含行人的1500幅红外图像作为验证集，红外行人图像数据集总数为8000幅，训练集A与验证集数量比例为5:1。

[0077] 步骤B2：将红外行人图像数据集中的图像样本全部转换为512×512像素的图像并进行图像预处理，利用预处理后的训练集A训练红外图像人体姿态提取网络。红外图像人体姿态提取网络训练过程中，利用验证集验证工作，以评价模型预测性能。训练红外图像人体姿态提取网络时，以100幅图像为一个批次进行小批量训练，每训练一批图像，权值更新一次。权值的衰减速率设为0.0005，动量设置为0.9，初始学习率设为0.001，对红外图像人体姿态提取网络进行20000次迭代，每间隔2000次迭代后保存一次模型，最终根据模型平均精度指标AP(average precision)选取精度最高的模型。经过训练后的模型平均损失(average loss)下降至0.2以下。基于选取出的模型提取出红外图像中的人体动态骨骼特征图像。

[0078] 将红外热成像仪部署于夜间无路灯等光源的全黑环境下，通过红外热成像仪对人体的红外图像进行采集，得到红外行人图像，再对红外行人图像进行预处理，得到滤波图像。现有的通过红外图像进行人体行为动作识别方法需要基于可见光环境下才能实现，导致在无光环境或天气较为恶劣环境下无法对人体行为动作进行识别，本发明通过采用红外热成像仪对图像进行采集，在全黑无光环境甚至气候较为恶劣环境下(例如雨雾天气)情况下，其不依赖光源，受天气影响小，探测距离远，不影响探测效果。

[0079] 本实施例采用目前主流的35镜头，机芯分辨率384×288的户外热像仪在全黑环境下对行人目标的检测距离可达500米，识别距离可达150米。在夜晚无光情况下，对行人目标的识别距离在100米左右，可以采集较为清晰的红外行人图像(即前文所述的红外行人图像)，以便后期图像处理算法的实施。

[0080] 由于红外热成像仪输出为AV格式的单通道信号，因此通过数据采集板卡对其进行

格式转换，转换为单通道数字图像格式，便于后续对图像进行处理，本实施例的采集环境与采集到的红外行人图像如图5所示。

[0081] 对红外行人图像进行预处理，其目的是抑制红外图像背景，突出人体等热源目标，本发明使用背景抑制法降低红外图像中背景对感兴趣人体目标的干扰，提高后续提取处理的准确性，对CVC-09红外数据集中包含的红外图像预处理操作与对红外行人图像预处理操作相同。背景抑制方法考虑抑制效果与实时性，采用两级结构，具体如下：

[0082] 首先，使用多尺度的图像细节提升(multi-scale detail boosting)方法提升红外行人图像中人体细节与背景的对比度，其核心思想为：使用三个尺度的高斯模糊，再和原图做减法，获得不同程度的细节信息，然后通过一定的组合方式把这些细节信息融入到原图中，从而得到加强原图信息的能力，计算公式如下：

[0083] $D=(1-0.5\times \mathrm{sgn}(D_1))\times D_1+0.5\times D_2+0.25\times D_3$

[0084] 上式中，D表示处理后的细节增强图像，sgn()表示符号函数，D_1、D_2、D_3分别表示三个尺度上的细节增强处理，D_1、D_2、D_3的计算方法分别为：

[0085] $D_1=I-B_1$、$D_2=I-B_2$、$D_3=I-B_3$

[0086] 其中，I表示原始图像，中间参数B_1、B_2、B_3的计算方法分别为：

[0087] $B_1=G_1*I$、$B_2=G_2*I$、$B_3=G_3*I$

[0088] 其中，G_1、G_2、G_3分别表示方差为1、2、4的高斯核。

[0089] 然后，使用双边滤波抑制细节增强图像中对比度低的细节部分，即抑制细节增强图像中热源以外的背景。双边滤波(Bilateral filter)是一种非线性的滤波方法，是结合图像的空间邻近度和像素值相似度的一种折中处理，同时考虑空域信息和灰度相似性，达到保边去噪的目的，其计算公式如下：

[0090] $$I_p=\frac{1}{W_p}\sum_{q\in S}G_{\delta_s}(\|p-q\|)\,G_{\delta_r}(\|I_p-I_q\|)I_q$$

[0091] 上式中，p表示图像当前像素点，q表示图像空间邻域像素点，I_p表示处理后得到的滤波图像，“||||”表示求取两个值之间的欧式距离，I_q表示输入的细节增强图像；G表示高斯核，对于两个参数取值，δ_s表示原始图像斜对角线长度的2%的数值，δ_r表示原始图像梯度值的中值或者平均数；W_p表示权重，其计算方法如下：

[0092] $$W_p=\sum_{q\in S}G_{\delta_s}(\|p-q\|)\,G_{\delta_r}(\|I_p-I_q\|)$$

[0093] 上式中，S表示图像空间域(spatial domain)。

[0094] 步骤3：分割提取出的人体动态骨骼特征图像的感兴趣区域序列，得到红外人体动态骨骼特征数据集，通过红外人体动态骨骼特征数据集训练动作识别网络SaNet，具体操作如下：

[0095] 步骤C1：利用步骤2提取出的多个人体动态骨骼特征图像，剪裁感兴趣区域(ROI)，形成一个人体动态骨骼ROI图像序列，依次输入动作识别网络SaNet。

[0096] 动作识别考虑到红外图像特征捕捉连续性较差，因此对红外视频中每一帧图像的骨骼特点进行动作识别分类，而非提取多帧图像进行行为预测。提取红外图像中人体动态骨骼姿态后，由于骨骼特征对应于动作较为明显，因此行为识别模块重点在于提高识别的

实时性,算法的精简性。

[0097] 步骤C2:使用Labelimg工具对红外图像人体姿态提取网络所提取的8000幅人体动态骨骼ROI图像序列进行动作标注,区分需要识别的6类动作,得到红外人体动态骨骼特征数据集,将红外人体动态骨骼特征数据集按5:1的比例划分成训练集B和测试集。

[0098] 步骤C3:作为动作识别网络SaNet的输入,为减少计算量,将红外人体动态骨骼特征数据集中所有图像样本转换为单通道28×28像素的图像,利用训练集B训练动作识别网络SaNet。训练动作识别网络SaNet时,以100幅图像为一个批次进行小批量训练,每训练一批图像,权值更新一次。权值的衰减速率设为0.0005,动量设置为0.9,初始学习率设为0.0001,对动作识别网络SaNet进行20万次迭代,每间隔2万次迭代后保存一次模型,最终根据模型平均精度指标AP(average precision)选取精度最高的模型。经过训练后模型平均损失(average loss)下降至0.05以下。基于选取出的模型识别分类6种动作。

[0099] 步骤4:

[0100] 步骤D1:获取待识别红外图像,先将该图像调整为512×512像素,再对调整大小后的待识别红外图像进行预处理,得到预处理图像。

[0101] 步骤D2:利用红外图像人体姿态提取网络提取预处理图像的动态骨骼特征,得到待识别人体动态骨骼特征图。

[0102] 步骤D3:截取待识别人体动态骨骼特征图的感兴趣区域,作为动作识别网络SaNet的输入序列,将待识别人体动态骨骼特征图调整为28×28像素,利用动作识别网络SaNet对大小经调整后的待识别人体动态骨骼特征图进行动作的分类与预测。

[0103] 采用平均准确率M_P、平均误检率M_F、平均漏检率M_M、平均运算速度M_O几项指标对本发明方法进行评价,各项指标的计算如下式:

[0104] $M_P=T_P/(T_P+F_P)$、$M_F=F_P/(T_N+F_P)$、$M_M=F_N/(T_P+F_N)$

[0105] 上式中,T_P表示红外图像中正确检测出的动作数量,F_N表示红外图像中没有检测出来的动作数量,F_P表示红外图像中误检出来的动作数量,T_N表示红外图像中没有误检的动作数量,M_O由实际测试及训练得到。

[0106] 使用长度为500帧的红外行人视频进行实际测试,对识别的部分动作实际测试识别结果如图7所示。根据实际测试结果,对比目前具有代表性的行为识别框架,对上述指标进行测试,测试结果分析如下表所示:

[0107]

方法名称	M_P (%)	M_F (%)	M_M (%)	M_O (FPS)
卷积双光流网络	36	23	45	35
时空残差网络	42	21	43	37
深时线性编码网络	51	23	35	28
ST-GCN(时空图卷积网络)	65	20	26	12
本发明方法	96	3	5	52

[0108] 由上述实际测试分析可以看出，基于光流法的行为识别方法在对红外热成像中人体动作识别方面各项指标均较低，不满足实时识别要求。ST-GCN网络在识别精度等方面稍好，而实时性较差，不能满足实时识别要求。而本发明通过红外图像背景抑制，改进的红外图像人体姿态提取网络，高效精简的动作识别网络SaNet三个阶段的处理，使红外热成像中人体动作识别的平均准确率达96％，平均错误率仅3％，平均漏检率仅5％，平均处理速度高达52FPS，各项指标对比均最为优良，满足红外热成像中准确实时识别人体动作要求。

[0109] 以上所述仅为本发明的较佳实施例而已，并不用以限制本发明，凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等，均应包含在本发明的保护范围之内。

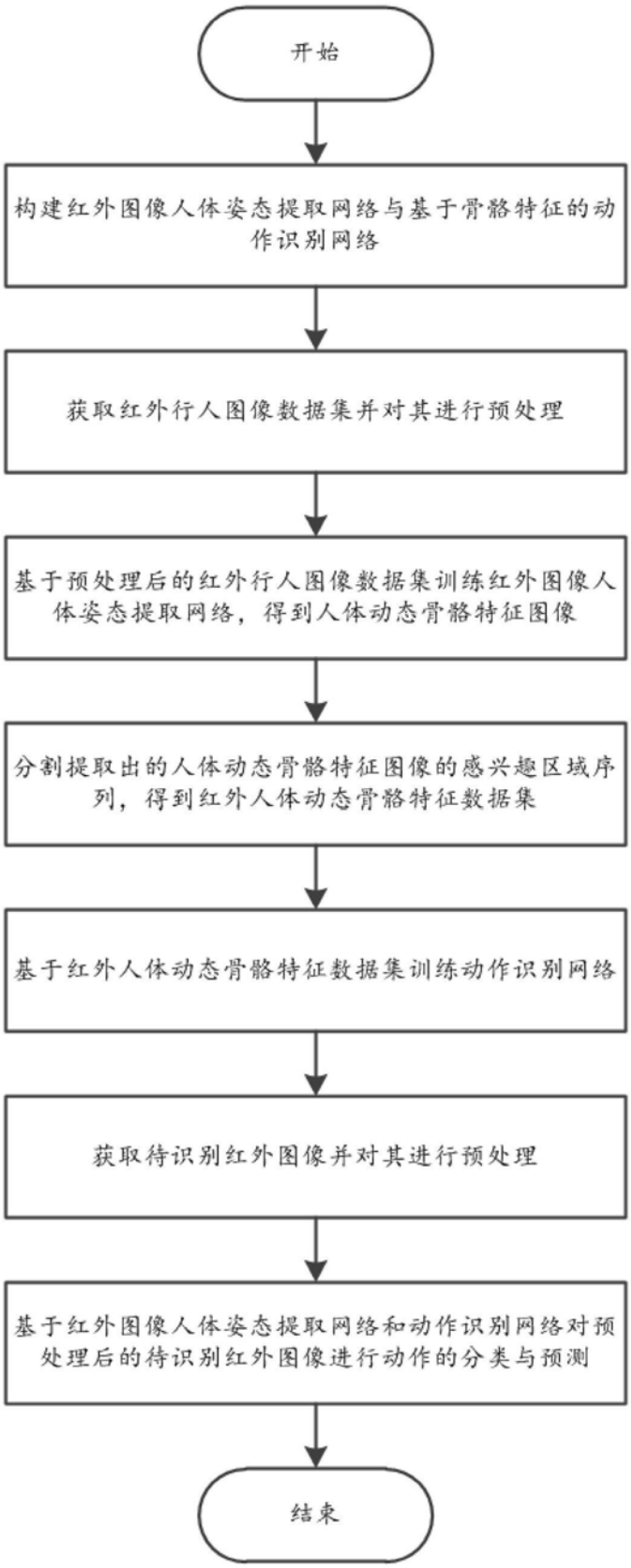

开始
构建红外图像人体姿态提取网络与基于骨骼特征的动作识别网络
获取红外行人图像数据集并对其进行预处理
基于预处理后的红外行人图像数据集训练红外图像人体姿态提取网络，得到人体动态骨骼特征图像
分割提取出的人体动态骨骼特征图像的感兴趣区域序列，得到红外人体动态骨骼特征数据集
基于红外人体动态骨骼特征数据集训练动作识别网络
获取待识别红外图像并对其进行预处理
基于红外图像人体姿态提取网络和动作识别网络对预处理后的待识别红外图像进行动作的分类与预测
结束

图1

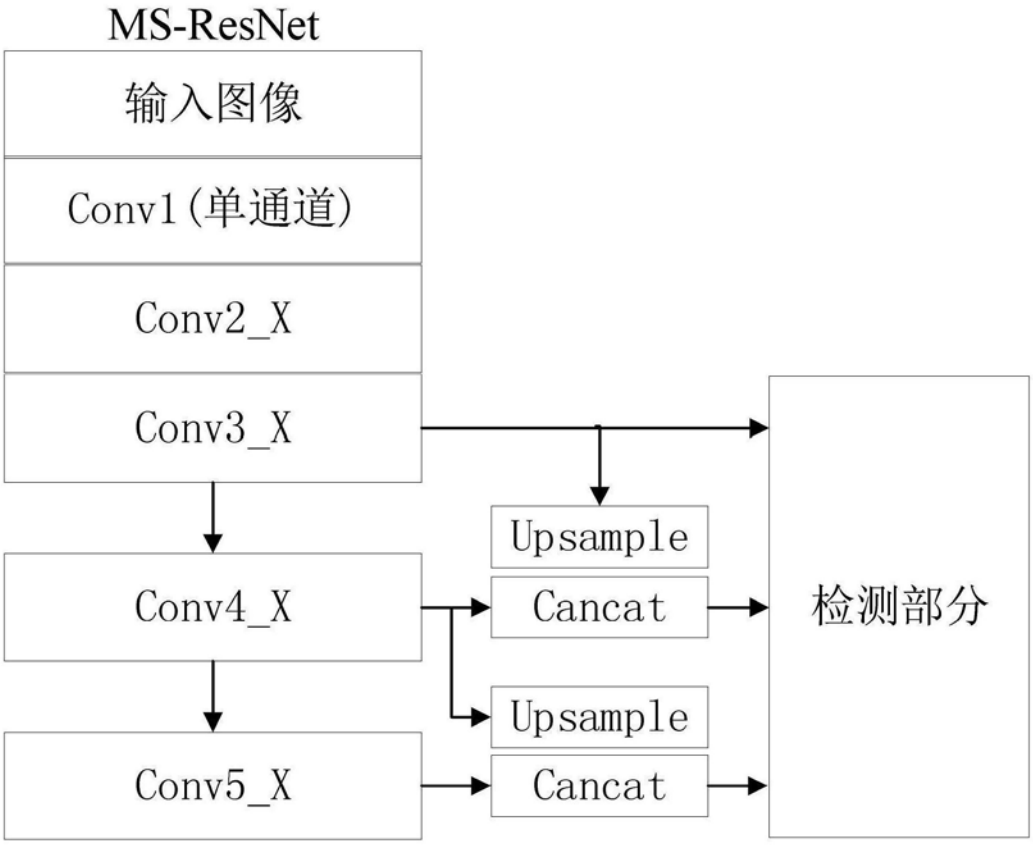

MS-ResNet
输入图像
Conv1(单通道)
Conv2_X
Conv3_X
Conv4_X
Conv5_X
Upsample
Cancat
Upsample
Cancat
检测部分

图2

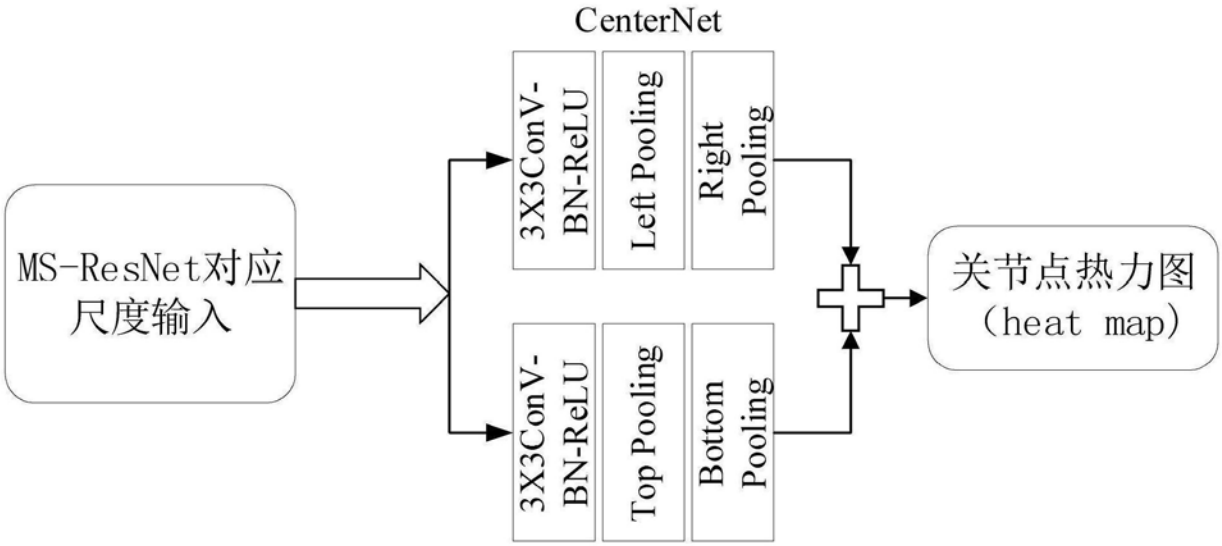

CenterNet
MS-ResNet对应
尺度输入
3X3ConV-
BN-ReLU
Left Pooling
Right
Pooling
3X3ConV-
BN-ReLU
Top Pooling
Bottom
Pooling
关节点热力图
(heat map)

图3

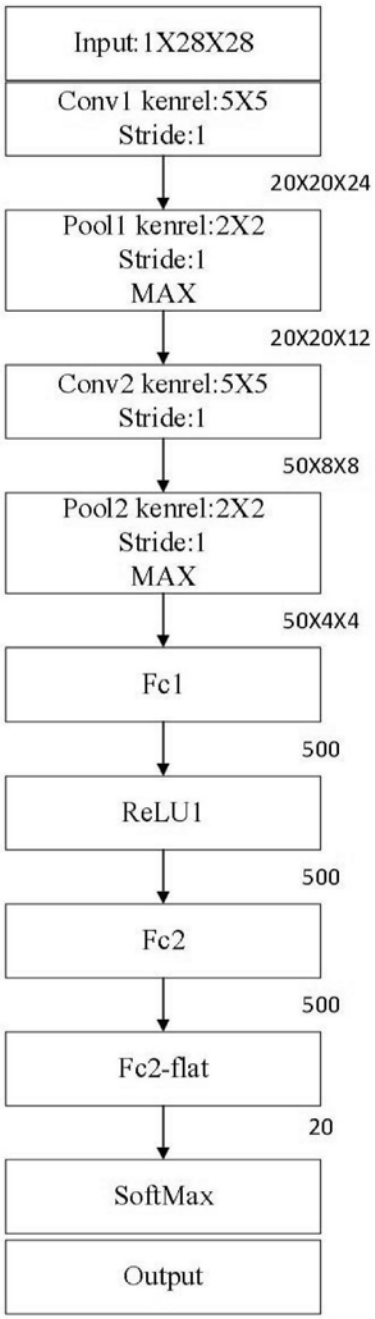

Input:1X28X28
Conv1 kenrel:5X5
Stride:1
20X20X24
Pool1 kenrel:2X2
Stride:1
MAX
20X20X12
Conv2 kenrel:5X5
Stride:1
50X8X8
Pool2 kenrel:2X2
Stride:1
MAX
50X4X4
Fc1
500
ReLU1
500
Fc2
500
Fc2-flat
20
SoftMax
Output

图4

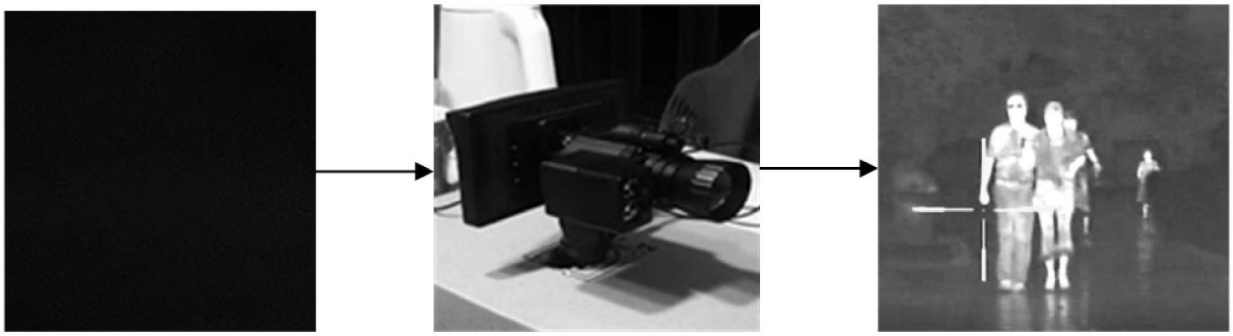


图5

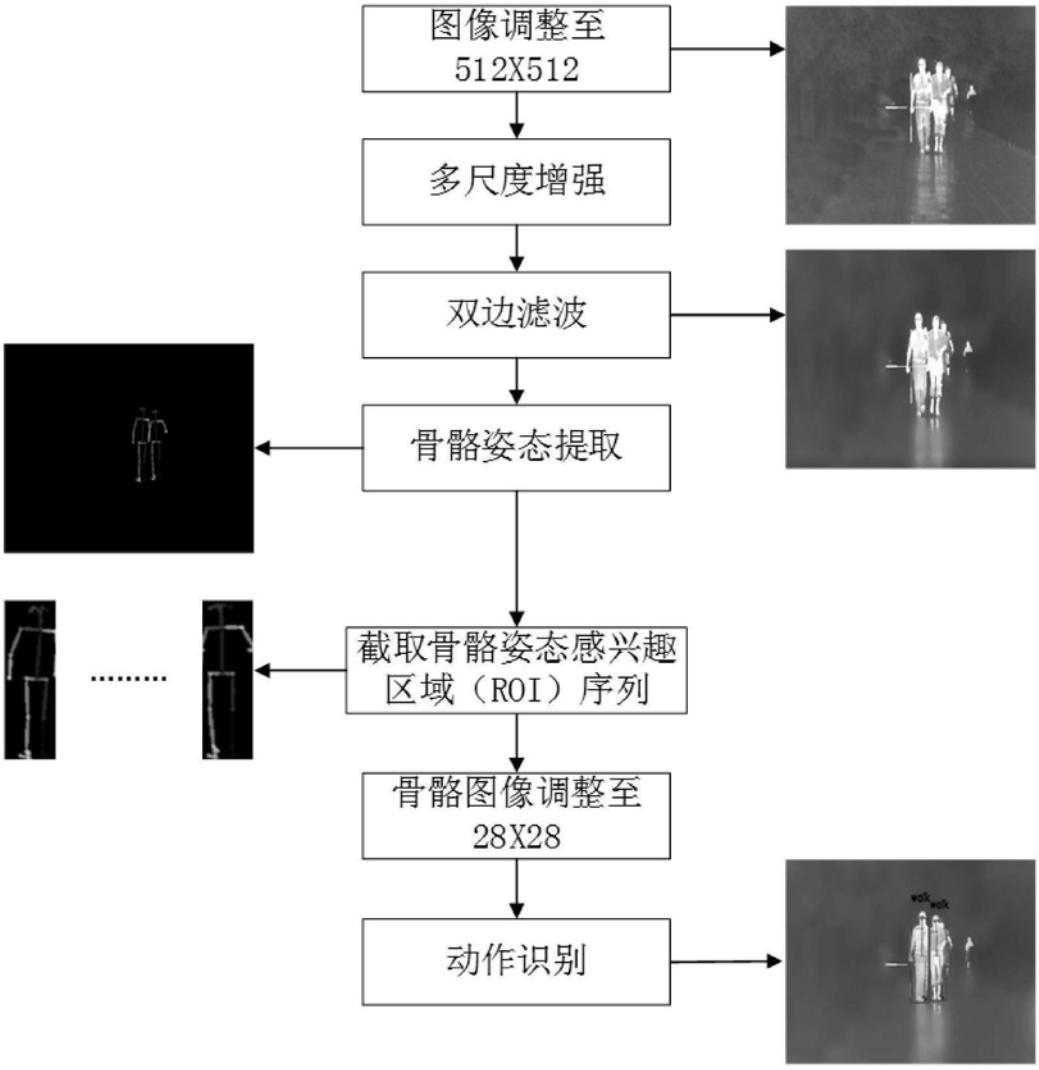

图像调整至
512X512
多尺度增强
双边滤波
骨骼姿态提取
截取骨骼姿态感兴趣
区域（ROI）序列
.....
骨骼图像调整至
28X28
动作识别

图6

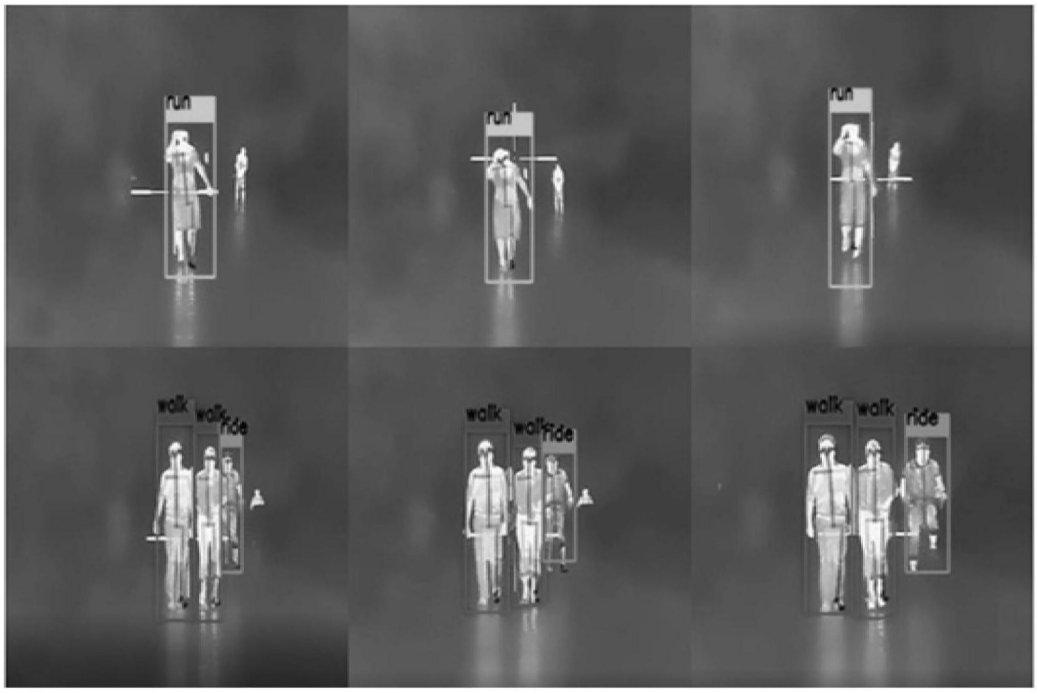


图7