



(12) 发明专利申请

(10) 申请公布号 CN 113781475 A

(43) 申请公布日 2021.12.10

(21) 申请号 202111204050.8

(22) 申请日 2021.10.15

(71) 申请人 中航华东光电有限公司

地址 241000 安徽省芜湖市高新技术产业
开发区天井山路46号

(72) 发明人 张骏 张鹏 张政 白云飞
张恺翔 孙瑞

(74) 专利代理机构 北京润平知识产权代理有限
公司 11283

代理人 董杰

(51) Int.Cl.

G06T 7/00 (2017.01)

G06T 7/13 (2017.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

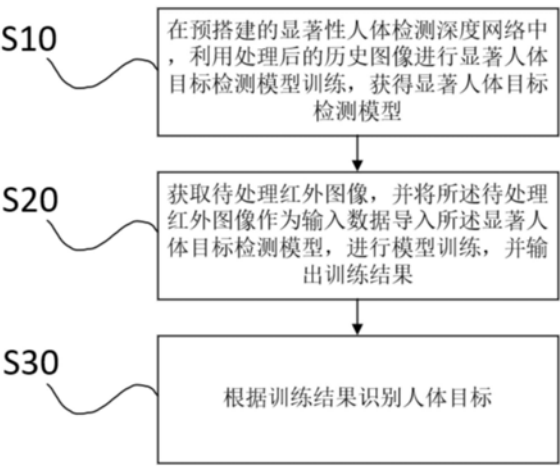
权利要求书2页 说明书7页 附图2页

(54) 发明名称

热红外图像显著人体目标检测方法及系统

(57) 摘要

本发明实施例提供一种热红外图像显著人体目标检测方法及系统,属于图像处理及识别技术领域。所述方法包括:在预搭建的显著性人体检测深度网络中,利用处理后的历史图像进行显著人体目标检测模型训练,获得显著人体目标检测模型;其中,所述处理后的历史图像为基于预设数据增强算法对包含人体目标的热红外图像历史信息进行处理生成;获取待处理红外图像;将所述待处理红外图像作为输入数据导入所述显著人体目标检测模型,进行模型训练,并输出训练结果;根据训练结果识别人体目标。本发明方案实现了在红外图像中进行准确人体目标识别,提高红外图像人体显著目标识别准确性和智能性。



CN 113781475 A

1. 一种热红外图像显著人体目标检测方法,其特征在于,所述方法包括:

在预搭建的显著性人体检测深度网络中,利用处理后的历史图像进行显著人体目标检测模型训练,获得显著人体目标检测模型;其中,所述处理后的历史图像为基于预设数据增强算法对包含人体目标的热红外图像历史信息进行处理生成;

获取待处理红外图像;

将所述待处理红外图像作为输入数据导入所述显著人体目标检测模型,进行模型训练,并输出训练结果;

根据训练结果识别人体目标。

2. 根据权利要求1所述的方法,其特征在于,所述预设数据增强算法至少包括:颜色变换、几何变换和裁剪操作。

3. 根据权利要求1所述的方法,其特征在于,所述预搭建的显著性人体检测深度网络基于HED边缘检测模型搭建。

4. 根据权利要求1所述的方法,其特征在于,所述预搭建的显著性人体检测深度网络包括:主干网络、编解码层和融合层。

5. 根据权利要求4所述的方法,其特征在于,所述方法还包括:

构建所述主干网络,包括:

删除VGG网络的全连接层;

在删除全连接层的VGG网络中,增加多个卷积,获得主干网络。

6. 根据权利要求4所述的方法,其特征在于,所述编解码层为单一尺度的U型编解码或不同尺度之间的FPN连接编解码。

7. 根据权利要求4所述的方法,其特征在于,所述融合层为基于HED模型融合预设损失函数进行构建的;所述融合层包括上采样运算和Concat运算;所述上采样运算的计算公式为:

$$\begin{cases} W_i = W_{input} / 2^i \\ H_i = H_{input} / 2^i \end{cases}$$

其中, W_{input} 为输入图像的宽度;

H_{input} 为输入图像的高度;

W_i, H_i 分别为编解码层输出的图像的宽度和高度。

8. 根据权利要求7所述的方法,其特征在于,所述预设损失函数表达式为:

$$L = \sum_{k=1}^K a_k l^k = \sum_{k=1}^K a_k (l_{bce}^k) = \sum_{k=1}^K \frac{(l_{bce}^k)}{K}$$

其中, l^k 为第k个输出结果的损失;

K表示输出结果的个数;

a_k 为每个损失函数的权重。

9. 一种热红外图像显著人体目标检测系统,其特征在于,所述系统包括:

采集单元,用于获取包含人体目标的热红外图像历史信息,以及用于获取待处理红外图像;

处理单元,用于在预搭建的显著性人体检测深度网络中,利用处理后的历史图像进行显著人体目标检测模型训练,获得显著人体目标检测模型;

训练单元,用于将所述待处理红外图像作为输入数据导入所述显著人体目标检测模型,进行模型训练,并输出训练结果;

识别单元,用于根据训练结果识别人体目标。

10.一种计算机可读储存介质,该计算机可读存储介质上储存有指令,其在计算机上运行时使得计算机执行权利要求1-8中任一项权利要求所述的热红外图像显著人体目标检测方法。

热红外图像显著人体目标检测方法及系统

技术领域

[0001] 本发明涉及图像处理及识别技术领域,具体地涉及一种热红外图像显著人体目标检测方法及一种热红外图像显著人体目标检测系统。

背景技术

[0002] 显著人体目标检测(SOD)是分割出图像中最具吸引力的视觉目标,在视觉追踪、图像分割等领域有着广泛的应用。传统的显著性目标检测算法多采用超像素相似度、直方图、像素梯度比等手工特征的方法进行检测。但是在小物体、物体遮蔽的情况下,检测效果较差。而热红外图像,因为其特殊的成像原理,使其应用广泛,尤其在军事侦察、资源勘探等领域。但是热红外图像缺点明显,如分辨率差、对比度低、信噪比低、区域边界模糊、灰度分布与目标反射特征无线性关系等。因此,热红外图像的显著目标检测难度更高。现有的显著性目标检测算法无法满足热红外图像的人体目标检测准确性需求,所以,需要创造一种新的热红外图像显著人体目标检测方法。

发明内容

[0003] 本发明实施方式的目的是提供一种热红外图像显著人体目标检测方法及系统,以至少解决现有的显著性目标检测算法无法满足热红外图像的人体目标检测准确性需求的问题。

[0004] 为了实现上述目的,本发明第一方面提供一种热红外图像显著人体目标检测方法,所述方法包括:在预搭建的显著性人体检测深度网络中,利用处理后的历史图像进行显著人体目标检测模型训练,获得显著人体目标检测模型;其中,所述处理后的历史图像为基于预设数据增强算法对包含人体目标的热红外图像历史信息进行处理生成;获取待处理红外图像;将所述待处理红外图像作为输入数据导入所述显著人体目标检测模型,进行模型训练,并输出训练结果;根据训练结果识别人体目标。

[0005] 可选的,所述预设数据增强算法至少包括:颜色变换、几何变换和裁剪操作。

[0006] 可选的,所述预搭建的显著性人体检测深度网络基于HED边缘检测模型搭建。

[0007] 可选的,所述预搭建的显著性人体检测深度网络包括:主干网络、编解码层和融合层。

[0008] 可选的,所述方法还包括:构建所述主干网络;删除VGG网络的全连接层;在删除全连接层的VGG网络中,增加多个卷积,获得主干网络。

[0009] 可选的,所述编解码层为单一尺度的U型编解码或不同尺度之间的FPN连接编解码。

[0010] 可选的,所述融合层为基于HED模型融合预设损失函数进行构建的;所述融合层包括上采样运算和Concat运算;所述上采样运算的计算公式为:

$$[0011] \quad \begin{cases} W_i = W_{input} / 2^i \\ H_i = H_{input} / 2^i \end{cases}$$

[0012] 其中, W_{input} 为输入图像的宽度; H_{input} 为输入图像的高度; W_i, H_i 分别为编解码层输出的图像的宽度和高度。

[0013] 可选的, 所述预设损失函数表达式为:

$$[0014] \quad L = \sum_{k=1}^K a_k l^k = \sum_{k=1}^K a_k (l_{bce}^k) = \sum_{k=1}^K \frac{(l_{bce}^k)}{K}$$

[0015] 其中, l_k 为第 k 个输出结果的损失; K 表示输出结果的个数; a_k 为每个损失函数的权重。

[0016] 本发明第二方面提供一种热红外图像显著人体目标检测系统, 所述系统包括: 采集单元, 用于获取包含人体目标的热红外图像历史信息, 以及用于获取待处理红外图像; 处理单元, 用于在预搭建的显著性人体检测深度网络中, 利用处理后的历史图像进行显著人体目标检测模型训练, 获得显著人体目标检测模型; 其中, 所述处理后的历史图像为基于预设数据增强算法对包含人体目标的热红外图像历史信息进行处理生成; 训练单元, 用于将所述待处理红外图像作为输入数据导入所述显著人体目标检测模型, 进行模型训练, 并输出训练结果; 识别单元, 用于根据训练结果识别人体目标。

[0017] 另一方面, 本发明提供一种计算机可读存储介质, 该计算机可读存储介质上储存有指令, 其在计算机上运行时使得计算机执行上述的热红外图像显著人体目标检测方法。

[0018] 通过上述技术方案, 对大量包含人体目标的热红外图像历史信息进行数据增强, 然后在预搭建的显著性人体检测深度网络中进行显著人体目标检测模型训练。基于训练获得的显著人体目标检测模型, 在后续的红外图像中, 进行人体识别。实现了在红外图像中进行人体目标识别, 提高红外图像人体显著目标识别准确性和智能性。

[0019] 本发明实施方式的其它特征和优点将在随后的具体实施方式部分予以详细说明。

附图说明

[0020] 附图是用来提供对本发明实施方式的进一步理解, 并且构成说明书的一部分, 与下面的具体实施方式一起用于解释本发明实施方式, 但并不构成对本发明实施方式的限制。在附图中:

[0021] 图1是本发明一种实施方式提供的热红外图像显著人体目标检测方法的步骤流程图;

[0022] 图2是本发明一种实施方式提供的显著人体目标检测模型训练的步骤流程图;

[0023] 图3是本发明一种实施方式提供的热红外图像显著人体目标检测系统的系统结构图。

[0024] 附图标记说明

[0025] 10-采集单元; 20-处理单元; 30-训练单元; 40-识别单元。

具体实施方式

[0026] 以下结合附图对本发明的具体实施方式进行详细说明。应当理解的是,此处所描述的具体实施方式仅用于说明和解释本发明,并不用于限制本发明。

[0027] 图3是本发明一种实施方式提供的热红外图像显著人体目标检测系统的系统结构图。如图3所示,本发明实施方式提供一种热红外图像显著人体目标检测系统,所述系统包括:采集单元10,用于获取包含人体目标的热红外图像历史信息,以及用于获取待处理红外图像;处理单元20,用于在预搭建的显著性人体检测深度网络中,利用处理后的历史图像进行显著人体目标检测模型训练,获得显著人体目标检测模型;训练单元30,用于将所述待处理红外图像作为输入数据导入所述显著人体目标检测模型,进行模型训练,并输出训练结果;识别单元40,用于根据训练结果识别人体目标。

[0028] 图1是本发明一种实施方式提供的热红外图像显著人体目标检测方法的方法流程图。如图1所示,本发明实施方式提供一种热红外图像显著人体目标检测方法,所述方法包括:

[0029] 步骤S10:在预搭建的显著性人体检测深度网络中,利用处理后的历史头像进行显著人体目标检测模型训练,获得显著人体目标检测模型。具体的,如图2,包括以下步骤:

[0030] 步骤S101:获取包含人体目标的热红外图像历史信息,并基于预设数据增强算法进行历史图像处理。

[0031] 具体的,显著人体目标检测(SOD)是分割出图像中最具吸引力的视觉目标,其在视觉追踪,图像分割等领域有着广泛的应用。受限于传统显著性目标检测算法检测效果差、分辨率低和对比度低等诸多问题,本发明方案采用以VGG、MobileNet网络作为主干网络(backbone)。以U型的空洞卷积作为编解码层,再使用多层解码特征融合的方式获取最终的检测结果。众所周知,想要获得贴近实际的训练模型,需要大量包含目标规律的历史数据作为训练样本。所以,本发明方案在进行模型训练前,需要获取大量的历史数据作为训练样本。这些历史数据便是资料库中,包含人体目标的热红外图像。这些图像被广泛存储在公网库、消防库和监控设备库中,通过对应设置的采集单元10,与这些资源库进行对接,并通过这些资源库进行关联图像下载。获得大量的历史图像信息后,因为不同采集设备和存储设备中存储的包含人体目标的热红外图像存在分辨率、对比度、信噪比、区域边界、灰度分布和反特征均存在较大区别。若将这些图像信息直接用于模型训练,势必会造成互相干扰,且训练获得的训练模型与实际需求必定存在较大区别。为了避免这种情况发生,优选的,获得历史图像信息后,需要通过预设的数据增强算法进行图像处理。这不仅能够使得图像中的特征得到增强,使得后续训练效果更高,还能将所有历史图像信息处理为同一标准下的图像,避免训练干扰。

[0032] 优选的,预设数据增强算法至少包括:颜色变换、几何变换和裁剪操作。其中,核心增强算法为几何变换和裁剪操作,主要针对红外图像的不同形态、角度的物体情况,通过随机翻转和几何变化使得图像特征增强,然后采用裁剪操作模拟出有障碍物遮蔽的显著性人体检测目标。

[0033] 步骤S102:基于历史图像处理进行模型训练。

[0034] 具体的,进行模型训练,需要对应的深度网络支撑。本发明方案基于显著人体目标检测的需求特性,设计了适配显著人体目标检测模型训练的深度网络,称为显著性人体检

测深度网络,具体的,包含以下几个部分:

[0035] 1) 主干网络:

[0036] 具体的,本发明方案的主干网络是基于传统VGG网络进行搭建的。VGG网络采用连续的几个3x3的卷积核代替AlexNet中的较大卷积核,分别为11x11、7x7、5x5。对于给定的感受野(与输出有关的输入图片的局部大小),采用堆积的小卷积核是优于采用大的卷积核,因为多层非线性层可以增加网络深度来保证学习更复杂的模式,还因为参数更少所以代价还比较小。简单来说,在VGG中,使用了3个3x3卷积核来代替7x7卷积核,使用了2个3x3卷积核来代替5x5卷积核,这样做的主要目的是在保证具有相同感知野的条件下,提升了网络的深度,在一定程度上提升了神经网络的效果。其中,VGG网络最后存在至少一个全连接层。全连接层在图片切割方面存在一定的弊端,优选的,通过conv取代原来的全连接层,可以避免这种问题。因为conv可以通过调整stride完成运算,所以conv在图片切割方面表现优秀。例如,训练时输入的图片大小是 256×256 ,在输入fc或者conv之前得到 $7 \times 7 \times 512$ 的feature map,而若输入的图片大小变为 512×512 ,在输入fc或者conv之前会得到 $14 \times 14 \times 512$ 的feature map,那么fc就会因为权重矩阵已经固定而无法处理这样的feature map。基于此,本发明方案丢弃原始VGG网络最后的FC全连接层,改为增加2个卷积操作分布为Conv2d(input=512,output=1024,kernel_size=3),Conv2d(input=1024,output=512,kernel_size=1),改进后的网络可以兼顾不同尺寸人体目标的检测。本发明设计的主干网络分层输入/输出情况如表1:

[0037]

运算名称	输入尺寸	输出尺寸	是否输出
Input Data	224*224*3		否
Conv_1 (3*3)	224*224*3	224*224*64	否
Conv_2 (3*3)	224*224*64	224*224*64	否
MaxPool (2*2)	224*224*64	112*112*64	否
Conv_3 (3*3)	112*112*64	112*112*128	否
Conv_4 (3*3)	112*112*128	112*112*128	是
MaxPool (2*2)	112*112*128	56*56*128	否
Conv_5 (3*3)	56*56*128	56*56*256	否
Conv_6 (3*3)	56*56*256	56*56*256	否
Conv_7 (3*3)	56*56*256	56*56*256	是
MaxPool (2*2)	56*56*256	28*28*256	否
Conv_8 (3*3)	28*28*256	28*28*512	否
Conv_9 (3*3)	28*28*512	28*28*512	否
Conv_10 (3*3)	28*28*512	28*28*512	是
MaxPool (2*2)	28*28*512	14*14*512	否
Conv_11 (3*3)	14*14*512	14*14*512	否
Conv_12 (3*3)	14*14*512	14*14*512	否
Conv_13 (3*3)	14*14*512	14*14*512	是
MaxPool (2*2)	14*14*512	7*7*512	否
Conv_add1 (3*3)	7*7*512	7*7*1024	否

Conv_add2 (3*3)	7*7*1024	7*7*512	是
-----------------	----------	---------	---

[0038] 表1主干网络VGG网络分层输入/输出

[0039] 2) 编解码层:

[0040] 具体的,本发明进行编解码层设计时,存在两种设计思路,第一种是单一尺度的U型编解码方式;第二种是不同尺度之间的FPN连接方式。首先,进行单一尺度的U型编解码方式设计时,设计的U型编解码网络对比传统的U-Net网络结构,U型编解码网络丢弃了Skip-Connections,使用空洞卷积(Dilation Conv)代替了Conv,减少了网络的深度。传统的U-Net网络使用Skip-Connections用于增加全局上下文的信息,兼顾浅层特征和深层特征,而代价就是网络的数据量在不断的增大。而这里丢弃了Skip-Connections,但为了保证全局上下文信息不会丢失,本发明采用Dilation Conv替代Conv,通过设置逐步增大的空洞系数来增加感受野的范围,以达到增加全局上下文信息的效果。

[0041] 然后是不同尺度之间的FPN连接方式,FPN结构主要用于多尺度变化的目标检测,人体目标的尺寸变化较大,随着网络深度增加和多次下采样运算,小目标信息损失严重,无法在像素级别进行准确区分,易导致出现误检测和漏检测。FPN结构自顶向下,深层特征流向浅层网络,可以兼顾深度特征和浅层特征,弥补了小目标信息丢失的缺陷,极大程度避免了误检测和漏检测的情况出现。

[0042] 3) 融合层:

[0043] 具体的,本发明方案的融合是基于HED网络的设计理念,结合特定的损失函数预测出精确的边缘信息进行构建的。U型编解码网络中,单一尺度的编解码层的输入和输出尺寸一致,而每一尺度的编解码层的输出尺寸与输入图像的尺寸相比都不相同,其原因是特征提取层中5个MaxPool (2*2) 的运算。因此输出融合是,必须先进行上采样(UpSample)运算将所有的输出尺寸统一到输入图像的尺寸大小。假设输入图像为I,宽度为 W_{input} ,高度为 H_{input} 。通过特征提取,编解码层运算后,输出的结果分别为 $O_i, i \in [1, 5]$,对应的尺寸如下:

$$[0044] \quad \begin{cases} W_i = W_{input} / 2^i \\ H_i = H_{input} / 2^i \end{cases}$$

[0045] 其中, W_i, H_i 分别为编解码层输出的宽度和高度。因此每一层输出的上采样率分别为 $2^i, i \in [1, 5]$,上采样运算后得到输出张量尺寸为 $[1, W_{input}, H_{input}]$ 。

[0046] 然后再使用Concat运算,所有上采样后的特征合并成一个 $[5, W_{input}, H_{input}]$ 的张量。再通过一个核为 $1*1$ Conv进行降维操作,最后使用Sigmoid作为激活函数输出最终的检测结果。

[0047] 其中,本发明方案的损失函数采用多层混合损失函数的加权线性形式,公式为:

$$[0048] \quad L = \sum_{k=1}^K \alpha_k l^k$$

[0049] 其中, l^k 为第k输出结果的损失; K 表示输出的个数; α_k 为每个损失函数的权重。本网络中,设定 $\alpha_k = 1/K, K=5$,包括编解码层输出结果和融合后结果。本发明方案中,使用二值交叉熵作为单层的损失函数,公式为:

[0050] $l^k = l_{bce}^k$

[0051] 其中, l_{bce}^k 对应Pixel-level, 其中 l_{bce}^k 是常用的二值交叉熵, 公式为:

[0052] $l_{bce} = - \sum_{(x,y)} [G(x,y) \log(S(x,y)) + (1-G(x,y)) \log(1-S(x,y))]$

[0053] 其中, $G(x,y) \in \{0,1\}$ 表示 (x,y) 位置像素是否为Ground Truth Label (GT Label), $S(x,y)$ 表示预测出 (x,y) 像素点为检测物的概率。 l_{bce} 通过计算每个像素的二值分类熵, 区分前景和背景的概率。基于此, 本发明方案的损失函数采用多层混合损失函数的加权线性形式还可以表示为:

[0054]
$$L = \sum_{k=1}^K a_k l^k = \sum_{k=1}^K a_k (l_{bce}^k) = \sum_{k=1}^K \frac{(l_{bce}^k)}{K}$$

[0055] 其中, 损失函数以线性组合的形式, 对多输出结果进行调整, 最大程度的保证了融合后的结果在边界上的精准性。

[0056] 通过上述设计的显著性人体检测深度网络, 将处理后的包含人体目标的热红外图像作为输入数据, 进行显著人体目标检测模型训练, 获得显著人体目标检测模型。

[0057] 步骤S20: 获取待处理红外图像, 并将所述待处理红外图像作为输入数据导入所述显著人体目标检测模型, 进行模型训练, 并输出训练结果。

[0058] 具体的, 获得显著人体目标检测模型, 便可以基于此模型进行后续红外图像检测, 识别图像中是否存在人体, 并在存在人体是, 定位人体的位置已经凸显人体轮廓。采集单元10通过红外图像获取模块进行待处理红外图像采集, 然后将采集到的红外图像传输到训练单元30。训练单元30将采集的红外图像作为显著人体目标检测模型的输入数据, 基于模型流程进行红外图像分析, 直到完成模型训练, 输出训练结果。

[0059] 步骤S30: 根据训练结果识别人体目标。

[0060] 具体的, 通过显著人体目标检测模型输出训练结果, 输出单元根据训练结果判断结果中是否存在人体, 以及人体的实际位置的轮廓, 然后输出勾画人体轮廓的识别结果到监测端, 供相关人员进行查看。

[0061] 在本发明实施例中, 通过设计的类HED的热红外显著性人体目标网络对热红外图像中的人体目标进行检测。通过该网络可以清晰的检测出红外图像中的人体目标, 而且在有遮挡的情况和不同尺寸目标检测上都可以清晰检测。该发明可以应用到消防救援, 军事目标检测等多领域。

[0062] 在一种可能的实施方式中, 本发明方案提出的热红外图像显著人体目标检测方法及系统被用于消防救援。采集单元10在消防救援采集现场进行无差别红外图像采集, 图像采集覆盖救援现场的各个位置和角度, 然后这些数据自动导入处理单元20, 处理单元20进行图像的预设预处理, 然后将预处理后的红外图像发送到训练单元30, 训练单元30基于预设的显著人体目标检测模型进行人体识别。若在某张红外图像中识别到人体, 输出单元定位存在人体的红外图像的序号, 并行采集该图像的采集位置的采集角度, 在输出包含人体图像信息的同时, 触发报警信息。然后将触发报警信息的红外图像标注处理, 独立推送到各

救援人员的监测点,提醒救援人员前往定位的图像采集位置进行人员救援,提高消防救援现场的救援效率。

[0063] 本发明实施方式还提供一种计算机可读储存介质,该计算机可读存储介质上储存有指令,其在计算机上运行时使得计算机执行上述的热红外图像显著人体目标检测方法。

[0064] 本领域技术人员可以理解实现上述实施方式的方法中的全部或部分步骤是通过程序来指令相关的硬件来完成,该程序存储在一个存储介质中,包括若干指令用以使得单片机、芯片或处理器(processor)执行本发明各个实施方式所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0065] 以上结合附图详细描述了本发明的可选实施方式,但是,本发明实施方式并不限于上述实施方式中的具体细节,在本发明实施方式的技术构思范围内,可以对本发明实施方式的技术方案进行多种简单变型,这些简单变型均属于本发明实施方式的保护范围。另外需要说明的是,在上述具体实施方式中所描述的各个具体技术特征,在不矛盾的情况下,可以通过任何合适的方式进行组合。为了避免不必要的重复,本发明实施方式对各种可能的组合方式不再另行说明。

[0066] 此外,本发明的各种不同的实施方式之间也可以进行任意组合,只要其不违背本发明实施方式的思想,其同样应当视为本发明实施方式所公开的内容。

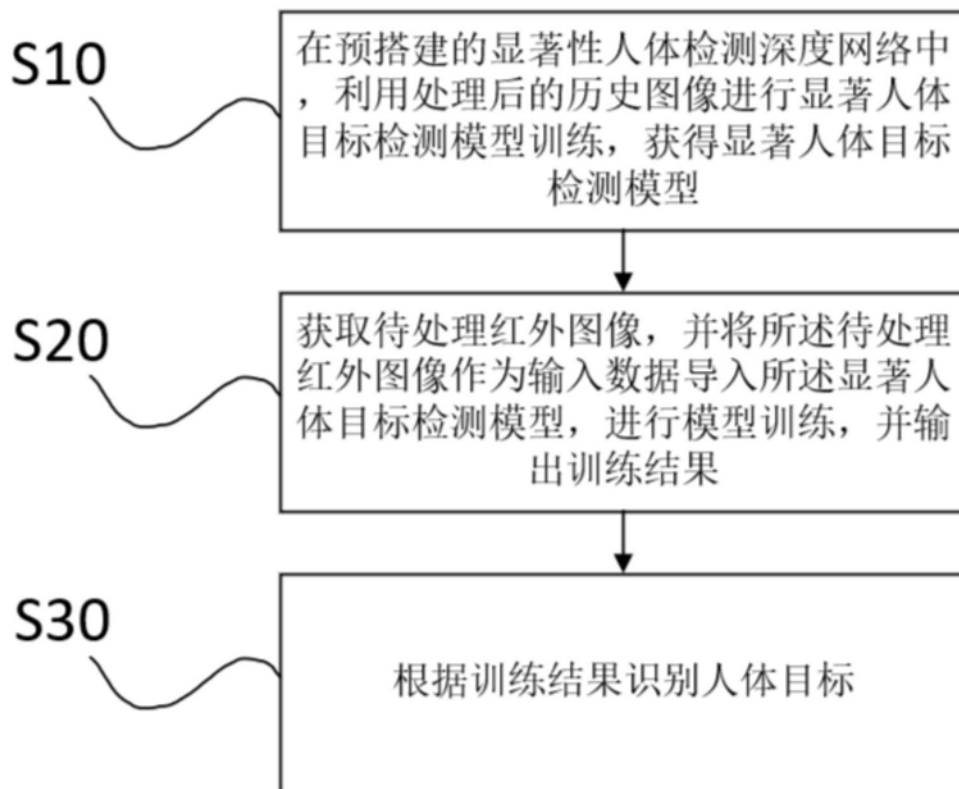


图1

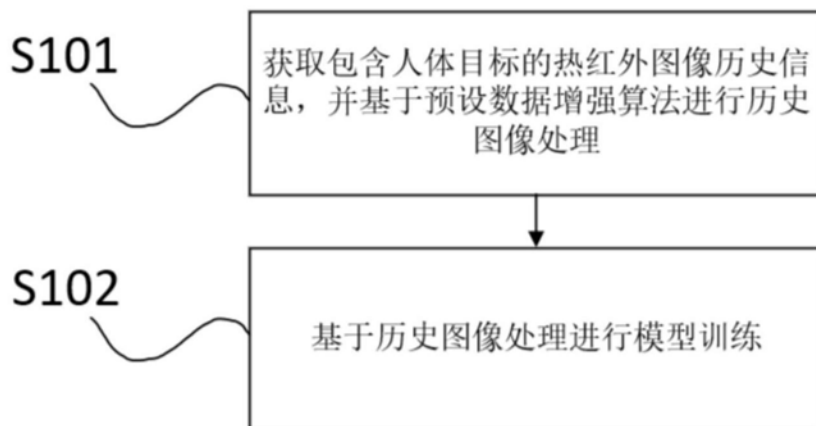


图2



图3