

目标跟踪方法、装置、电子设备及存储介质

相关申请的交叉引用

本公开基于申请号为 202010011243.0、申请日为 2020 年 01 月 06 日的中国专利申请提出，并要求该中国专利申请的优先权，该中国专利申请的全部内容在此以全文引入的方式引入本公开。

技术领域

本公开涉及计算机技术、图像处理领域，尤其涉及一种目标跟踪方法、装置、电子设备及计算机可读存储介质。

背景技术

视觉目标跟踪是计算机视觉中的一个重要研究方向，可以广泛的应用于各种场景，例如，机器自动跟踪、视频监控、人机交互、无人驾驶等。视觉目标跟踪任务就是在给定某视频序列中初始帧中的目标对象大小与位置的情况下，预测后续帧中该目标对象的大小与位置，从而得到整个视频序列内的目标的运动轨迹。

在实际跟踪预测的工程中，由于视角、光照、尺寸、遮挡等等不确定干扰因素的影响，跟踪过程极易产生漂移和丢失的情况。不仅如此，跟踪技术往往需要较高的简易性和实时性，以满足实际移动端部署和应用的需求。

发明内容

有鉴于此，本公开实施例至少提供一种目标跟踪方法、装置、电子设备及计算机可读存储介质。

第一方面，本公开实施例提供了一种目标跟踪方法，包括：

获取视频图像；

针对除所述视频图像中的参考帧图像之后的待跟踪图像，生成所述待跟踪图像中的搜索区域与所述参考帧图像中的目标图像区域之间的图像相似性特征图；其中，所述目标图像区域内包含待跟踪对象；

根据所述图像相似性特征图，确定所述搜索区域中的待定位区域的定位位置信息；

响应于在所述搜索区域中确定出所述待定位区域的定位位置信息，根据确定的待定位区域的定位位置信息确定所述待跟踪对象在包含所述搜索区域的待跟踪图像中的检测框。

在一种可能的实施方式中，根据所述图像相似性特征图，确定所述搜索区域中的待定位区域的定位位置信息，包括：根据所述图像相似性特征图，预测所述待定位区域的尺寸信息；根据所述图像相似性特征图，预测所述搜索区域的特征图中的每个特征像素点的概率值，一个特征像素点的概率值表征所述搜索区域中与该特征像素点对应的像素点位于所述待定位区域内的几率；根据所述图像相似性特征图，预测所述搜索区域中与每个所述特征像素点对应的像素点与所述待定位区域的位置关系信息；从预测的概率值中选取所述概率值最大的特征像素点所对应的所述搜索区域中的像素点作为目标像素点；基于所述目标像素点、所述目标像素点与所述待定位区域的位置关系信息、以及所述待定位区域的尺寸信息，确定所述待定位区域的定位位置信息。

在一种可能的实施方式中，根据以下步骤从所述参考帧图像中提取所述目标图像区域：确定所述待跟踪对象在所述参考帧图像中的检测框；基于所述参考帧图像中的所述检测框的尺寸信息，确定所述参考帧图像中的所述检测框对应的第一延伸尺寸信息；基于所述第一延伸尺寸信息，以所述参考帧图像中的所述检测框为起始位置向周围延伸，

得到所述目标图像区域。

在一种可能的实施方式中,根据以下步骤从待跟踪图像中提取搜索区域:获取在所述视频图像中当前帧待跟踪图像的前一帧待跟踪图像中,所述待跟踪对象的检测框;基于所述前一帧待跟踪图像中的所述待跟踪对象的检测框的尺寸信息,确定所述前一帧待跟踪图像中的所述待跟踪对象的检测框对应的第二延伸尺寸信息;基于所述第二延伸尺寸信息和所述前一帧待跟踪图像中的所述待跟踪对象的检测框的尺寸信息,确定当前帧待跟踪图像中的搜索区域的尺寸信息;以所述前一帧待跟踪图像中的所述待跟踪对象的检测框的中心点为当前帧待跟踪图像中的搜索区域的中心,根据当前帧待跟踪图像中的搜索区域的尺寸信息确定所述搜索区域。

在一种可能的实施方式中,所述生成所述待跟踪图像中的搜索区域与所述参考帧图像中的目标图像区域之间的图像相似性特征图,包括:将所述搜索区域缩放至第一预设尺寸,以及,将所述目标图像区域缩放至第二预设尺寸;生成所述搜索区域中的第一图像特征图,以及所述目标图像区域中的第二图像特征图;所述第二图像特征图的尺寸小于所述第一图像特征图的尺寸;确定所述第二图像特征图与所述第一图像特征图中的每个子图像特征图之间的相关性特征;所述子图像特征图与所述第二图像特征图的尺寸相同;基于确定的多个相关性特征,生成所述图像相似性特征图。

在一种可能的实施方式中,所述目标跟踪方法由跟踪定位神经网络执行;其中所述跟踪定位神经网络由标注有目标对象的检测框的样本图像训练得到。

在一种可能的实施方式中,上述目标跟踪方法还包括训练所述跟踪定位神经网络的步骤:获取样本图像,所述样本图像包括参考帧样本图像和待跟踪的样本图像;将所述样本图像输入待训练的跟踪定位神经网络,经过所述待训练的跟踪定位神经网络对输入的样本图像进行处理,预测所述目标对象在所述待跟踪的样本图像中的检测框;基于所述待跟踪的样本图像中标注的检测框和所述待跟踪的样本图像中预测的检测框,调整所述待训练的跟踪定位神经网络的网络参数。

在一种可能的实施方式中,将所述待跟踪的样本图像中的待定位区域的定位位置信息作为所述待跟踪的样本图像中预测的检测框的位置信息,所述基于所述待跟踪的样本图像中标注的检测框和所述待跟踪的样本图像中预测的检测框,调整所述待训练的跟踪定位神经网络的网络参数,包括:基于所述待跟踪的样本图像中预测的检测框的尺寸信息、所述待跟踪的样本图像中搜索区域中每个像素点位于所述待跟踪的样本图像中预测的检测框内的预测概率值、所述待跟踪的样本图像中搜索区域中每个像素点与所述待跟踪的样本图像中预测的检测框的预测位置关系信息、所述待跟踪的样本图像中标注的检测框的标准尺寸信息、所述待跟踪的样本图像中标准搜索区域中每个像素点是否位于标注的检测框中的信息、所述待跟踪的样本图像中标准搜索区域中每个像素点与所述待跟踪的样本图像中标注的检测框的标准位置关系信息,调整所述待训练的跟踪定位神经网络的网络参数。

第二方面,本公开实施例提供了一种目标跟踪装置,包括:

图像获取模块,配置为获取视频图像;

相似性特征提取模块,配置为针对除所述视频图像中的参考帧图像之后的待跟踪图像,生成所述待跟踪图像中的搜索区域与所述参考帧图像中的目标图像区域之间的图像相似性特征图;其中,所述目标图像区域内包含待跟踪对象;

定位模块,配置为根据所述图像相似性特征图,确定所述搜索区域中的待定位区域的定位位置信息;

跟踪模块,配置为响应于在所述搜索区域中确定出所述待定位区域的定位位置信息,根据确定的待定位区域的定位位置信息确定所述待跟踪对象在包含所述搜索区域的待

跟踪图像中的检测框。

第三方面，本公开实施例提供了一种电子设备，包括：处理器、存储器和总线，所述存储器存储有所述处理器可执行的机器可读指令，当电子设备运行时，所述处理器与
5 所述存储器之间通过总线通信，所述机器可读指令被所述处理器执行时执行如上述目标跟踪方法的步骤。

第四方面，本公开实施例还提供一种计算机可读存储介质，该计算机可读存储介质上存储有计算机程序，该计算机程序被处理器运行时执行如上述目标跟踪方法的步骤。

本公开实施例上述装置、电子设备、和计算机可读存储介质，至少包含与本公开实施例上述方法的任一方面或任一方面的任一实施方式的技术特征实质相同或相似的技术特征，因此关于上述装置、电子设备、和计算机可读存储介质的效果描述，可以参见
10 上述方法内容的效果描述，这里不再赘述。

附图说明

为了更清楚地说明本公开实施例的技术方案，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公开实施例的某些实施例，因此不应被
15 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。

图1示出了本公开实施例提供了一种目标跟踪方法的流程图；

图2示出了本公开实施例中的确定待定位区域的中心点的示意图；

图3示出了本公开实施例提供的另一种目标跟踪方法中提取目标图像区域的流程图；

20 图4示出了本公开实施例提供的再一种目标跟踪方法中提取搜索区域的流程图；

图5示出了本公开实施例提供的再一种目标跟踪方法中生成图像相似性特征图的流程图；

图6示出了本公开实施例的再一种目标跟踪方法中生成图像相似性特征图的示意图；

图7示出了本公开实施例的再一种目标跟踪方法中训练跟踪定位神经网络的流程图；

25 图8A示出了本公开实施例提供了一种目标跟踪方法的流程示意图；

图8B示出了本公开实施例提供了一种定位目标的流程示意图；

图9示出了本公开实施例提供了一种目标跟踪装置的结构示意图；

图10示出了本公开实施例提供了一种电子设备的结构示意图。

具体实施方式

30 为使本公开实施例的目的、技术方案和优点更加清楚，下面将结合本公开实施例中的附图，对本公开实施例中的技术方案进行清楚、完整地描述，应当理解，本公开实施例中附图仅起到说明和描述的目的，并不用于限定本公开实施例的保护范围。另外，应当理解，示意性的附图并未按实物比例绘制。本公开实施例中使用的流程图示出了根据本公开实施例的一些实施例实现的操作。应该理解，流程图的操作可以不按顺序实现，
35 没有逻辑的上下文关系的步骤可以反转顺序或者同时实施。此外，本领域技术人员在本公开实施例内容的指引下，可以向流程图添加一个或多个其他操作，也可以从流程图中移除一个或多个操作。

另外，所描述的实施例仅仅是本公开实施例的一部分实施例，而不是全部的实施例。通常在此处附图中描述和示出的本公开实施例的组件可以以各种不同的配置来布置和
40 设计。因此，以下对在附图中提供的本公开实施例的实施例的详细描述并非旨在限制要求保护的本公开实施例的范围，而是仅仅表示本公开实施例的选定实施例。基于本公开实施例，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，都属于本公开实施例保护的范围。

需要说明的是，本公开实施例中将会用到术语“包括”，用于指出其后所声明的特征

的存在,但并不排除增加其它的特征。

本公开实施例针对视觉目标跟踪,提供了一种可以有效降低跟踪过程中进行预测计算的复杂度的方案,可以基于待跟踪图像中的搜索区域与参考帧图像中的目标图像区域(包含待跟踪对象)之间的图像相似性特征图,来预测待跟踪对象在上述待跟踪图像中的位置信息(实际实施中预测待跟踪对象所在待定位区域的位置信息),即预测所述待跟踪对象在所述待跟踪图像中的检测框。详细实施过程将在下述实施例中详述。

如图1所示,本公开实施例提供了一种目标跟踪方法,该方法应用于对待跟踪对象进行跟踪定位的终端设备上,该终端设备可以为用户设备(User Equipment, UE)、移动设备、用户终端、终端、蜂窝电话、无绳电话、个人数字处理(Personal Digital Assistant, PDA)、手持设备、计算设备、车载设备、可穿戴设备等。在一些可能的实现方式中,该目标跟踪方法可以通过处理器调用存储器中存储的计算机可读指令的方式来实现。该方法可以包括如下步骤:

S110、获取视频图像;

这里,视频图像是需要对待跟踪对象进行定位和跟踪的图像序列。

视频图像包括参考帧图像和至少一帧待跟踪图像。参考帧图像是包括待跟踪对象的图像,可以是视频图像中的第一帧图像,当然也可以是视频图像中的其他帧图像。待跟踪图像为需要在其中搜索和定位待跟踪对象的图像。参考帧图像中待跟踪对象的位置和大小,即检测框是已经确定了的,而待跟踪图像中的定位区域或检测框并没有确定,是需要计算和预测的区域,也称为待定位区域,或待跟踪图像中的检测框。

S120、针对除所述视频图像中的参考帧图像之后的待跟踪图像,生成所述待跟踪图像中的搜索区域与所述参考帧图像中的目标图像区域之间的图像相似性特征图;其中,所述目标图像区域内包含待跟踪对象;

在执行此步骤之前,需要从待跟踪图像中提取搜索区域,从参考帧图像中提取目标图像区域。目标图像区域中包括待跟踪对象的检测框;搜索区域中包括未完成定位的待定位区域。定位区域的位置即为待跟踪对象的位置。

在提取得到搜索区域和目标图像区域之后,可以从搜索区域中和目标图像区域中分别提取图像特征,之后基于搜索区域对应的图像特征和目标图像区域的图像特征,确定搜索区域与目标图像区域之间的图像相似性特征,即确定搜索区域与目标图像区域之间的图像相似性特征图。

S130、根据所述图像相似性特征图,确定所述搜索区域中的待定位区域的定位位置信息;

这里,基于上一步骤中生成的图像相似性特征图,可以预测搜索区域的特征图中的每个特征像素点的概率值,以及搜索区域中与每个所述特征像素点对应的像素点与所述待定位区域的位置关系信息。

上述一个特征像素点的概率值表征所述搜索区域中与该特征像素点对应的像素点位于所述待定位区域内的几率。

上述位置关系信息可以是待跟踪图像中的搜索区域中的像素点与所述待跟踪图像中的待定位区域的中心点的偏差信息。例如,以待定位区域的中心点为坐标中心建立坐标系,则该位置关系信息包括对应的像素点在该建立的坐标系中的坐标信息。

这里,基于上述概率值能够确定出搜索区域中概率最大的位于待定位区域内的像素点。之后基于该像素点的位置关系信息,就能够较为准确的确定搜索区域中的待定位区域的定位位置信息。

上述定位位置信息可以包括待定位区域的中心点的坐标等信息,在实际实施时,可以基于搜索区域中概率最大的位于待定位区域内的像素点的坐标信息,和该像素点与待

定位区域的中心点的偏差信息，来确定待定位区域的中心点的坐标信息。

应当说明的是，此步骤确定了搜索区域中的待定位区域的定位位置信息，但在实际应用中，搜索区域中可能存在待定位区域，也可能不存在待定位区域。如果搜索区域中不存在待定位区域，则无法确定待定位区域的定位位置信息，即无法确定待定位区域的中心点的坐标等信息。

S140、响应于在所述搜索区域中确定出所述待定位区域的定位位置信息，根据确定的待定位区域的定位位置信息，确定所述待跟踪对象在包含所述搜索区域的待跟踪图像中的检测框。

在搜索区域中存在待定位区域的情况下，此步骤，根据确定的待定位区域的定位位置信息，确定所述待跟踪对象在包含所述搜索区域的待跟踪图像中的检测框。这里，可以将待跟踪图像中的待定位区域的定位位置信息，作为所述待跟踪图像中预测的检测框的位置信息。

上述实施例从待跟踪图像中提取搜索区域，从参考帧图像中提取目标图像区域，之后基于提取的两部分图像区域之间的图像相似性特征图，来预测或确定待跟踪图像中的待定位区域的定位位置信息，即确定待跟踪对象在包含所述搜索区域的待跟踪图像中的检测框，使得参与预测检测框的像素点的数量有效减少。本公开实施例不仅能够提高预测的效率和实时性，并且能够减低预测计算的复杂度，从而使用于预测待跟踪对象的检测框的神经网络的网络架构得到简化，更加适用于对实时性和网络结构简易性要求均较高的移动端。

在一些实施例中，在确定所述待定位区域在所述搜索区域中的定位位置信息之前，上述目标跟踪方法还包括：预测所述待定位区域的尺寸信息。这里，可以基于上面生成的图像相似性特征图，预测搜索区域中每个像素点对应的待定位区域的尺寸信息。在实际实施时，该尺寸信息可以包括待定位区域的高度值和宽度值。

在确定了搜索区域中每个像素点对应的待定位区域的尺寸信息之后，上述根据所述图像相似性特征图，确定所述搜索区域中的待定位区域的定位位置信息的过程可以通过如下步骤实现：

步骤一、根据所述图像相似性特征图，预测所述搜索区域的特征图中的每个特征像素点的概率值，一个特征像素点的概率值表征所述搜索区域中与该特征像素点对应的像素点位于所述待定位区域内的几率。

步骤二、根据所述图像相似性特征图，预测所述搜索区域中与每个所述特征像素点对应的像素点与所述待定位区域的位置关系信息。

步骤三、从预测的概率值中选取所述概率值最大的特征像素点所对应的所述搜索区域中的像素点作为目标像素点。

步骤四、基于所述目标像素点、所述目标像素点与所述待定位区域的位置关系信息、以及所述待定位区域的尺寸信息，确定所述待定位区域的定位位置信息。

上述步骤利用搜索区域中最有可能位于待定位区域中的像素点即目标像素点与所述待定位区域的位置关系信息，和该目标像素点在搜索区域中的坐标信息，能够确定待定位区域的中心点坐标。之后，再结合该目标像素点对应的待定位区域的尺寸信息，能够提高确定的搜索区域中待定位区域的准确度，即能够提高对待跟踪对象进行跟踪和定位的准确度。

如图2所示，图2中的极大值点即为最有可能位于待定位区域中的像素点，即概率值最大的目标像素点。基于极大值点的坐标 (x^m, y^m) 、极大值点与所述待定位区域的位置关系信息即偏差信息 $O^m = (O_x^m, O_y^m)$ ，就能确定待定位区域的中心点 (x_c^t, y_c^t) 的坐标。其中， O_x^m 为极大值点与待定位区域的中心点在横轴方向上的距离， O_y^m 为极大值点

与待定位区域的中心点在纵轴方向上的距离。在定位待定位区域的过程中，可以利用如下公式（1）至（5）实现：

$$x_c^t = x^m + O_x^m \quad (1);$$

$$y_c^t = y^m + O_y^m \quad (2);$$

$$w_t = w^m \quad (3);$$

$$h_t = h^m \quad (4);$$

$$R_t = (x_c^t, y_c^t, w_t, h_t) \quad (5);$$

其中， x_c^t 表示待定位区域的中心点的横坐标， y_c^t 表示待定位区域的中心点的纵坐标， x^m 表示极大值点的横坐标， y^m 表示极大值点的纵坐标， O_x^m 表示极大值点与待定位区域的中心点在横轴方向上的距离， O_y^m 表示极大值点与待定位区域的中心点在纵轴方向上的距离， w_t 表示待定位区域定位完成后的宽度值， h_t 表示待定位区域定位完成后的高度值， w^m 表示预测得到待定位区域的宽度值， h^m 表示预测得到待定位区域的高度值， R_t 表示待定位区域定位完成后的位置信息。

上述实施例，在得到搜索区域与目标图像区域之间的图像相似性特征图之后，基于该图像相似性特征图能够从搜索区域中筛选出位于待定位区域内的概率值最大的目标像素点，基于对应的概率值最大的目标像素点在搜索区域中的坐标信息、该像素点对与待定位区域的位置关系信息和该像素点对应的待定位区域的尺寸信息，来确定待定位区域的定位位置信息，能够提高确定的定位位置信息的准确度。

在一些实施例中，如图3所示，可以根据以下步骤从所述参考帧图像中提取所述目标图像区域：

S310、确定所述待跟踪对象在所述参考帧图像中的检测框；

上述检测框是已经定位完成的、包括待跟踪对象的图像区域。在实施时，上述检测框可以是一个矩形的图像框 $R_{t_0} = (x_c^{t_0}, y_c^{t_0}, w_{t_0}, h_{t_0})$ ，其中， R_{t_0} 表示检测框的位置信息， $x_c^{t_0}$ 表示检测框的中心点的横坐标， $y_c^{t_0}$ 表示检测框的中心点的纵坐标， w_{t_0} 表示检测框的宽度值， h_{t_0} 表示检测框的高度值。

S320、基于所述参考帧图像中的所述检测框的尺寸信息，确定所述参考帧图像中的所述检测框对应的第一延伸尺寸信息。

这里可以基于第一延伸尺寸信息对检测框进行延伸处理，可以利用如下公式（6）计算，即将检测框的高度和检测框的宽度之间的平均值作为第一延伸尺寸信息：

$$pad_h = pad_w = (w_{t_0} + h_{t_0}) / 2 \quad (6);$$

其中， pad_h 表示检测框在检测框的高度上需要延伸的长度， pad_w 表示检测框在检测框的宽度上需要延伸的长度； w_{t_0} 表示检测框的宽度值， h_{t_0} 表示检测框的高度值。

在对检测框进行延伸的时候，可以在检测框的高度方向的两边分别延伸上面计算得到的数值的一半，在检测框的宽度方向的两边分别延伸上面计算得到的数值的一半。

S330、基于所述第一延伸尺寸信息，以所述参考帧图像中的所述检测框为起始位置向周围延伸，得到所述目标图像区域。

这里，基于第一延伸尺寸信息对检测框进行延伸，可以直接得到目标图像区域。当然，对检测框进行延伸后，还可以对延伸后的图像进行进一步地处理，以得到目标图像区域，或者并不基于第一延伸尺寸信息对检测框进行延伸，只是基于第一延伸尺寸信息确定目标图像区域的尺寸信息，之后基于确定的目标图像区域的尺寸信息对检测框进行延伸来直接得到目标图像区域。

基于待跟踪对象在参考帧图像中的大小和位置，即待跟踪对象在参考帧图像中的检测框的尺寸信息，对检测框进行延伸，得到的目标图像区域不仅包括待跟踪对象，还包括待跟踪对象周边的区域，从而能够确定包括较多图像内容的目标图像区域。

在一些实施例中,上述基于所述第一延伸尺寸信息,以所述参考帧图像中的所述检测框为起始位置向周围延伸,得到所述目标图像区域,可以通过如下步骤实现:

基于所述检测框的尺寸信息和所述第一延伸尺寸信息,确定目标图像区域的尺寸信息;基于所述检测框的中心点和目标图像区域的尺寸信息,确定将所述检测框延伸后的所述目标图像区域。

在实施时,可以利用如下公式(7)确定目标图像区域的尺寸信息,即分别对检测框的宽度 w_{t_0} 延伸固定尺寸 pad_w ,对检测框的高度 h_{t_0} 延伸固定尺寸 pad_h ,然后对延伸后的宽度和高度求算术平方根,得到的结果作为目标图像区域的宽度(或高度),也就是说,目标图像区域为高度和宽度相等的正方形区域:

$$Rect_{t_0}^w = Rect_{t_0}^h = \sqrt{(w_{t_0} + pad_w) * (h_{t_0} + pad_h)} \quad (7);$$

其中, $Rect_{t_0}^w$ 表示目标图像区域的宽度值, $Rect_{t_0}^h$ 表示目标图像区域的高度值; pad_h 表示检测框在检测框的高度上需要延伸的长度, pad_w 表示检测框在检测框的宽度上需要延伸的长度; w_{t_0} 表示检测框的宽度值, h_{t_0} 表示检测框的高度值。

在确定了目标图像区域的尺寸信息之后,就可以以检测框的中心点为中心点,按照确定的尺寸信息,直接对检测框进行延伸,得到目标图像区域;或以检测框的中心点为中心点,按照确定的尺寸信息,在检测框按照第一延伸尺寸信息延伸后的图像中截取目标图像区域。

上述实施例基于检测框的尺寸信息和所述第一延伸尺寸信息,在对检测框进行延伸的基础上,可以在延伸的图像上截取一个正方形的目标图像区域,从而使得到的目标图像区域不包括过多的除待跟踪对象以外的其他图像区域。

在一些实施例中,如图4所示,可以根据以下步骤从待跟踪图像中提取搜索区域:

S410、获取在所述视频图像中当前帧待跟踪图像的前一帧待跟踪图像中,所述待跟踪对象的检测框。

这里,当前帧待跟踪图像的前一帧待跟踪图像中的检测框,是已经定位完成的待跟踪对象所在的图像区域。

S420、基于所述待跟踪对象的检测框的尺寸信息,确定所述待跟踪对象的检测框对应的第二延伸尺寸信息。

这里,基于检测框的尺寸信息确定第二延伸尺寸信息的算法与上述实施例中确定第一延伸尺寸信息的步骤相同。这里不再赘述。

S430、基于所述第二延伸尺寸信息和所述待跟踪对象的检测框的尺寸信息,确定当前帧待跟踪图像中的搜索区域的尺寸信息。

这里,可以通过如下步骤确定搜索区域的尺寸信息:

基于所述第二延伸尺寸信息和所述前一帧待跟踪图像中的检测框的尺寸信息,确定待延伸搜索区域的尺寸信息;基于所述待延伸搜索区域的尺寸信息、所述搜索区域对应的第一预设尺寸、以及所述目标图像区域对应的第二预设尺寸,确定所述搜索区域的尺寸信息;其中,所述搜索区域为将所述待延伸搜索区域延伸后得到的。

上述确定待延伸搜索区域的尺寸信息的计算方法与上述实施例中的基于所述检测框的尺寸信息和所述第一延伸尺寸信息,确定目标图像区域的尺寸信息的计算方法相同,这里不再赘述。

上述基于所述待延伸搜索区域的尺寸信息、所述搜索区域对应的第一预设尺寸、以及所述目标图像区域对应的第二预设尺寸,确定将所述待延伸搜索区域延伸后的所述搜索区域的尺寸信息,可以利用如下公式(8)和(9)计算:

$$SeachRect_{t_i} = Rect_{t_i} + 2 * Pad_{margin} \quad (8);$$

$$Pad_{margin} = (Size_s - Size_t) / (2 * \frac{Size_t}{Rect_{t_i}}) \quad (9);$$

其中, $SeachRect_t$ 表示搜索区域的尺寸信息, $Rect_t$ 表示待延伸搜索区域的尺寸信息, Pad_{margin} 表示所述待延伸搜索区域需要延伸的尺寸, $Size_s$ 表示搜索区域对应的第一预设尺寸, $Size_t$ 表示目标图像区域对应的第二预设尺寸。这里基于公式 (7) 可知, 搜索区域和目标图像区域均为高度和宽度相等的正方形区域, 因此这里的尺寸为对应的图像区域的高度和宽度对应的像素数量。

本步骤中, 基于待延伸搜索区域的尺寸信息、所述搜索区域对应的第一预设尺寸、以及所述目标图像区域对应的第二预设尺寸, 对搜索区域进行进一步的延伸, 从而能够进一步增大搜索区域。较大的搜索区域能够提高对待跟踪对象进行跟踪定位的成功率。

S440、以所述待跟踪对象的检测框的中心点为当前帧待跟踪图像中的搜索区域的中心, 根据当前帧待跟踪图像中的搜索区域的尺寸信息确定所述搜索区域。

在实施时, 可以以所述前一帧待跟踪图像中的检测框的中心点的坐标, 作为当前帧待跟踪图像中的初始定位区域的中心点, 以前一帧待跟踪图像中的检测框的尺寸信息, 作为当前帧待跟踪图像中的初始定位区域的尺寸信息, 确定当前帧待跟踪图像中的初始定位区域。之后, 可以基于第二延伸尺寸信息对初始定位区域进行延伸处理, 再按照上述待延伸搜索区域的尺寸信息, 从延伸后的图像中截取待延伸搜索区域。之后, 基于上述待延伸搜索区域延伸后的尺寸信息, 对待延伸搜索区域进行延伸后得到搜索区域。

当然, 也可以以所述前一帧待跟踪图像中的检测框的中心点, 作为当前帧待跟踪图像中搜索区域的中心点, 按照计算得到的上述搜索区域的尺寸信息, 直接在当前帧待跟踪图像上截图搜索区域。

基于前一帧待跟踪图像中确定的检测框的尺寸信息, 确定第二延伸尺寸信息, 基于第二延伸尺寸信息能够为当前帧待跟踪图像确定一个较大的搜索区域, 较大的搜索区域能够提高确定的待定位区域的定位位置信息的准确度, 即能够提高对待跟踪对象进行跟踪定位的成功率。

在一些实施例中, 生成所述图像相似性特征图之前, 上述目标跟踪方法还可以包括如下步骤:

将所述搜索区域缩放至第一预设尺寸, 以及, 将所述目标图像区域缩放至第二预设尺寸。

这里, 将搜索区域和目标图像区域设置为对应的预设尺寸, 能够控制生成图像相似性特征图中的中像素点的数量, 从而能够控制计算的复杂度。

在一些实施例中, 如图5所示, 上述生成所述待跟踪图像中的搜索区域与所述参考帧图像中的目标图像区域之间的图像相似性特征图, 可以通过如下步骤实现:

S510、生成所述搜索区域中的第一图像特征图, 以及所述目标图像区域中的第二图像特征图; 所述第二图像特征图的尺寸小于所述第一图像特征图的尺寸。

这里, 可以利用深度卷积神经网络提取搜索区域中的图像特征和目标图像区域中的图像特征, 分别得到上述第一图像特征图和第二图像特征图。

如图6中, 第一图像特征图61的宽度值和高度值均为8个像素点, 第二图像特征图62的宽度值和高度值均为4个像素点。

S520、确定所述第二图像特征图与所述第一图像特征图中的每个子图像特征图之间的相关性特征; 所述子图像特征图与所述第二图像特征图的尺寸相同。

如图6所示, 可以将第二图像特征图62按照从左到右、从上到下的顺序在第一图像特征图61上移动, 将第二图像特征图62在第一图像特征图61中的各个正投影区域作为各个子图像特征图。

在实施时, 可以利用相关 (correlation) 计算, 确定第二图像特征图与子图像特征图之间的相关性特征。

S530、基于确定的多个相关性特征，生成所述图像相似性特征图。

如图6所示，基于第二图像特征图与各个子图像特征图之间的相关性特征，生成的图像相似性特征图63的宽度值和高度值均为5个像素点。

上述图像相似性特征图中，每个像素点对应的相关性特征，即可表征第一图像特征图中一个子区域（即子图像特征图）与第二图像特征图之间的图像相似性的程度。基于该图像相似性的程度能够准确的筛选出搜索区域中的位于待定位区域内的概率最大的像素点，继而基于该概率值最大的像素点的信息能够有效提高确定的待定位区域的定位位置信息的准确度。

上述实施例的目标跟踪方法中，对获取的视频图像进行处理，得到每帧待跟踪图像中的待定位区域的定位位置信息，以及确定所述待跟踪对象在包含所述搜索区域的待跟踪图像中的检测框的过程，可以利用跟踪定位神经网络来完成，所述跟踪定位神经网络由标注有目标对象的检测框的样本图像训练得到。

上述目标跟踪方法中利用了跟踪定位神经网络，确定待定位区域的定位位置信息，即确定所述待跟踪对象在包含所述搜索区域的待跟踪图像中的检测框。由于简化了计算方法，因此使得跟踪定位神经网络的结构得到了简化，从而更易于部署在移动终端上。

本公开实施例还提供了训练上述跟踪定位神经网络的方法，如图7所示，包括如下步骤：

S710、获取样本图像，所述样本图像包括参考帧样本图像和待跟踪的样本图像。

样本图像中包括参考帧样本图像和至少一帧待跟踪的样本图像。参考帧样本图像中包括待跟踪对象的、已经确定了定位位置信息的检测框。待跟踪的样本图像中的待定位区域的定位位置信息没有确定，需要跟踪定位神经网络来预测或确定。

S720、将所述样本图像输入待训练的跟踪定位神经网络，经过所述待训练的跟踪定位神经网络对输入的样本图像进行处理，预测所述目标对象在所述待跟踪的样本图像中的检测框。

S730、基于所述待跟踪的样本图像中标注的检测框，和所述待跟踪的样本图像中预测的检测框，调整所述待训练的跟踪定位神经网络的网络参数。

在实施时，将所述待跟踪的样本图像中的待定位区域的定位位置信息，作为所述待跟踪的样本图像中预测的检测框的位置信息。

上述基于所述待跟踪的样本图像中标注的检测框，和所述待跟踪的样本图像中预测的检测框，调整所述待训练的跟踪定位神经网络的网络参数，可以通过如下步骤实现：

基于所述预测的检测框的尺寸信息、所述待跟踪的样本图像中搜索区域中每个像素点位于所述预测的检测框内的预测概率值、所述待跟踪的样本图像中搜索区域中每个像素点与所述预测的检测框的预测位置关系信息、所述标注的检测框的标准尺寸信息、所述待跟踪的样本图像中标准搜索区域中每个像素点是否位于标注的检测框中的信息、所述标准搜索区域中每个像素点与所述标注的检测框的标准位置关系信息，调整所述待训练的跟踪定位神经网络的网络参数。

其中，所述标准尺寸信息、所述标准搜索区域中每个像素点是否位于标注的检测框中的信息、所述标准搜索区域中每个像素点与所述标注的检测框的标准位置关系信息，均可以根据所述标注的检测框确定。

上述预测位置关系信息为对应的像素点与预测的检测框的中心点的偏差信息，可以包括对应的像素点与该中心点的距离在横轴方向上的分量，和对应的像素点与该中心点的距离在纵轴方向上的分量。

上述像素点是否位于标注的检测框中的信息，可以利用对象的像素点位于标注的检测框内的标准值 L_p 确定：

$$L_p = \begin{cases} 1, L_p^i \in R_t \\ 0, L_p^i \notin R_t \end{cases} \quad (10);$$

其中, R_t 表示待跟踪的样本图像中的检测框, L_p^i 表示搜索区域中的从左到右、从上到下第*i*个位置处的像素点位于检测框 R_t 内的标准值。标准值 L_p 为0表示像素点位于检测框 R_t 外, 标准值 L_p 为1表示像素点位于检测框 R_t 内。

5 在实施时, 可以采用交叉熵损失函数对 L_p 和预测概率值进行约束, 构建一个子损失函数 $Loss_{cls}$, 如公式(11)所示:

$$Loss_{cls} = \sum_{i \in k_p} L_p^i \log y_i^1 + \sum_{i \in k_n} L_p^i \log y_i^0 \quad (11);$$

其中, k_p 表示属于标注的检测框内的像素点的集合, k_n 表示属于标注的检测框外的像素点的集合, y_i^1 表示像素点*i*属于预测的检测框内的预测概率值, y_i^0 表示像素点*i*属于预测的检测框外的预测概率值。

10 在实施时, 可以采用光滑后的L1范数损失函数(smoothL1)来确定标准位置关系信息和预测位置关系信息之间的子损失函数 $Loss_{offset}$:

$$Loss_{offset} = \text{smoothL1}(L_o - Y_o) \quad (12);$$

其中, Y_o 表示预测位置关系信息, L_o 表示标准位置关系信息。

15 标准位置关系信息 L_o 为像素点与标注的检测框的中心点的真实偏差信息, 可以包括像素点与标注的检测框的中心点的距离在横轴方向上的分量 L_{ox} 和像素点与标注的检测框的中心点的距离在纵轴方向上的分量 L_{oy} 。

基于上述公式(11)生成的子损失函数和上述公式(12)生成的子损失函数, 可以构建一个综合的损失函数, 如下公式(13)所示:

$$20 \quad Loss_{all} = Loss_{cls} + \lambda_1 * Loss_{offset} \quad (13);$$

其中, λ_1 为一个预设的权重系数。

进一步, 可以结合上述预设的检测框尺寸信息, 调整所述待训练的跟踪定位神经网络中的网络参数, 可以利用上面的公式(11)、(12)建立子损失函数 $Loss_{cls}$ 和子损失函数 $Loss_{offset}$ 。

25 可以利用如下公式(14)建立关于预测的检测框尺寸信息的子损失函数 $Loss_{w,h}$:

$$Loss_{w,h} = \text{smoothL1}(L_w - Y_w) + \text{smoothL1}(L_h - Y_h) \quad (14);$$

其中, L_w 表示标准尺寸信息中的宽度值, L_h 表示标准尺寸信息中的高度值, Y_w 表示检测框的预测尺寸信息中的宽度值, Y_h 表示检测框的预测尺寸信息中的高度值。

30 基于上述 $Loss_{cls}$ 、 $Loss_{offset}$ 和 $Loss_{w,h}$ 三个子损失函数可以构建一个综合的损失函数 $Loss_{all}$, 可以如下公式(15)所示:

$$Loss_{all} = Loss_{cls} + \lambda_1 * Loss_{offset} + \lambda_2 * Loss_{w,h} \quad (15);$$

其中, λ_1 为一个预设的权重系数, λ_2 为另一个预设的权重系数。

上述实施例在训练跟踪定位神经网络的过程中, 进一步结合预测得到的检测框的尺寸信息和待跟踪的样本图像中检测框的标准尺寸信息, 来构造损失函数, 利用该损失函数能够进一步提高训练得到跟踪定位神经网络的计算准确度。利用预测得到的概率值、位置关系信息、预测的检测框的尺寸信息与样本图像的对应的标准值构建损失函数来训练跟踪定位神经网络, 训练的目标是使构建的损失函数取值最小, 从而有利于提高训练得到的跟踪定位神经网络计算的准确度。

40 目标跟踪方法根据观测模型的类别可以分为生成式方法和判别式方法。近年来主要以深度学习和相关滤波为主的判别式跟踪方法占据了主流的地位, 并使目标跟踪技术获得了突破性的进展。特别是以深度学习得到的图像特征为基础各类判别式方法, 在跟踪性能上达到了领先的水平。深度学习方法利用其在大规模图像数据上端到端学习训练得

到的高效特征表达能力,使得目标跟踪算法更加精准和快速。

基于深度学习方法跨域跟踪方法(MDNet),通过大量的离线学习和在线更新策略,学习得到针对目标和非目标的高精度分类器,并对后续帧内的对象进行分类判别和框调整,最终得到跟踪结果。这类完全基于深度学习的跟踪方法,在跟踪精度上提升巨大但是实时性能较差,例如每秒传输帧数(Frames Per Second, FPS)为1。同年提出的GOTURN方法,通过深度卷积神经网络提取相邻帧图像的特征,并学习目标特征相对于前一帧的位置变化从而完成后续帧的目标定位操作。该方法在保持一定精度的同时获得了较高的实时性能,例如100FPS。尽管基于深度学习的跟踪方法在速度和精度上都有较好的表现,但是更深的网络结构如VGG(Visual Geometry Group, 计算机视觉组)、ResNet等网络带来的计算复杂度,使得精度更高的跟踪算法难以应用到实际生产中。

针对任意指定目标对象的跟踪,目前存在的方法,主要包括逐帧检测、相关滤波以及基于深度学习的实时跟踪算法等等。这些方法在实时性、精度和结构复杂性上均有一定不足,不能很好地适应复杂的跟踪场景和实际移动端应用。基于检测分类方式的跟踪方法例如MDNet等等方法,需要在线学习,很难达到实时要求;基于相关滤波和基于检测的跟踪算法在预测位置之后,微调上一帧目标框形状,产生的框不够精准;基于区域候选框如RPN(Region Proposal Network, 区域生成网络)的方法产生的框冗余较多,计算复杂。

本公开实施例期望提供一种目标跟踪方法,在具备较高精度的同时从算法的实时性方面进行优化。

图8A为本公开实施例提供的一种目标跟踪方法的流程示意图,如图8所示,所述方法包括以下步骤:

步骤S810,对目标图像区域和搜索区域进行特征提取。

其中,本公开实施例跟踪的目标图像区域在初始帧(第一帧)以目标框的形式给出。搜索区域则根据上一帧目标的跟踪位置和大小,拓展一定空间区域得到。将截取的目标区域和搜索区域经过放缩固定不同尺寸后,经过同一个预训练过的深度卷积神经网络,提取得到两者各自的图像特征。也就是以目标所在的图像和待跟踪的图像作为输入,经过卷积神经网络,输出目标图像区域的特征和搜索区域的特征。下面对这些操作进行说明。

首先,获取目标图像区域:本公开实施例跟踪的对象是视频数据,一般在跟踪的第一帧(初始帧)中给以矩形框的方式出目标区域中心的位置信息,如 $R_{t_0} = (x_c^{t_0}, y_c^{t_0}, w_{t_0}, h_{t_0})$,以该目标区域中心所在的位置为中心位置,按照目标长宽填充(pad_w, pad_h)后截取一个面积不变的正方形区域 $Rect_{t_0}$,得到目标图像区域。

其次,获取搜索区域:根据前一帧跟踪结果 $R_{t_{i-1}}$ (初始帧为给定的目标框 R_{t_0}),在当前帧的 t_i 中以 $R_{t_{i-1}}$ 的位置为中心,经过和目标图像区域相同的处理得到正方形区域 $Rect_{t_i}$ 。为了尽可能的包含目标对象,在该正方形区域的基础上添加一个更大的内容信息区域,得到搜索区域。

然后,缩放获取的图像得到输入图像:本公开实施例中采用边长为 $Size_s = 255$ 像素的图像作为搜索区域的输入,采用 $Size_t = 127$ 的图像作为目标图像区域的输入。对搜索区域 $SeachRect_{t_i}$ 放缩至固定大小 $Size_s$ 和目标图像区域 $Rect_{t_0}$ 放缩至固定大小 $Size_t$ 。

最后,特征提取:采用深度卷积神经网络分别对放缩后的输入图像提取特征,得到目标特征 F_t 和搜索区域的特征 F_s 。

步骤S820,计算搜索区域的相似度特征。

输入目标特征 F_t 和搜索区域特征 F_s ,如图6所示,将 F_t 以滑动窗的方式在 F_s 上移动,并对搜索子区域(与目标特征大小相同的子区域)和目标特征做相关计算。最终得到搜

索区域的相似度特征 F_c 。

步骤S830, 定位目标。

该过程将相似度度量特征 F_c 作为输入, 最后输出目标点分类结果 Y 、偏差回归结果 $Y_o = (Y_{ox}, Y_{oy})$ 、以及目标框长宽结果 Y_w, Y_h 。

- 5 定位目标的流程如图8B所示, 将相似度度量特征81送入目标点分类分支82得到目标点分类结果83, 目标点分类结果83预测每个点对应的搜索区域是否为待搜索的目标区域。将相似度度量特征81送入回归分支84得到目标点的偏差回归结果85和目标框的长宽回归结果86。偏差回归结果85预测目标点到目标中心点的偏差。长宽回归结果86对目标框的长宽进行预测。最终结合相似度最高的目标点位置信息和偏差信息得到目标中心点位置, 再根据目标框的长宽预测结果给出该处位置最终的目标框结果。下面对算法训练和定位两个过程分别进行描述。

- 10 算法训练过程: 算法采用反向传播的方式, 端到端的训练特征提取网路, 以及后续的分类和回归分支。特征图上的目标点对应的类别标签 L_p 由上述公式(10)确定。目标点分类结果 Y 上的每个位置均输出一个二分类结果, 判断该位置是否属于目标框内。算法采用交叉熵损失函数对 L_p 和 Y 进行约束, 针对距离中心点的偏差和长宽回归输出的损失函数采取smoothL1计算。根据以上定义好的损失函数, 通过梯度反向传播的计算方式训练网络参数。模型训练完成后, 固定网络参数, 将预处理好的动作区域图像, 输入到网络中前馈, 预测当前帧目标点分类结果 Y 、偏差回归结果 Y_o 以及目标框长宽结果 Y_w, Y_h 。

- 20 算法定位过程: 从分类结果 Y 上取极大值点 y^m 所在的位置 x^m 和 y^m , 以及该点预测得到的偏差 $O^m = (O_x^m, O_y^m)$ 和预测得到的长宽信息 w^m, h^m , 然后利用公式(1)至(5)计算新一帧的目标区域 R_t 。

- 25 本公开实施例首先确定待跟踪图像中的搜索区域与参考帧中的目标图像区域之间的图像相似性特征图, 之后基于图像相似性特征来预测或确定待跟踪图像中的待定位区域的定位位置信息, 即确定待跟踪对象在包含搜索区域的待跟踪图像中的检测框, 使得参与预测待跟踪对象的检测框的像素点的数量有效减少, 不仅能够提高预测的效率和实时性, 并且能够减低预测计算的复杂度, 从而可以简化预测待跟踪对象的检测框的神经网络的网络架构, 更加适用于对实时性和网络结构简易性要求较高的移动端。

- 30 本公开实施例利用端到端训练的方式对预测目标进行充分训练, 不需要在线更新, 实时性更高。同时通过网络直接预测目标框的点位置、偏差和长宽, 通过计算可以直接获得最终目标框信息, 结构更为简单有效, 不存在候选框的预测过程, 更适应移动端的算法需求, 并且在提升了精度的同时保持了跟踪算法的实时性。可以利用本公开实施例提供的算法进行移动端以及嵌入式设备的跟踪算法应用, 例如终端设备中的人脸跟踪, 无人机下的目标跟踪等场景。利用该算法配合移动或者嵌入式设备完成难以人为跟拍的高速运动, 以及对指定对象的实时智能跟随和方向矫正跟踪任务。

- 35 对应于上述目标跟踪方法, 本公开实施例还提供了一种目标跟踪装置, 该装置应用于需要进行目标跟踪的终端设备上, 并且该装置及其各个模块能够执行与上述目标跟踪方法的相同的方法步骤, 并且能够达到相同或相似的有益效果, 因此对于重复的部分不再赘述。

- 40 如图9所示, 本公开实施例提供的目标跟踪装置包括:

图像获取模块910, 配置为获取视频图像;

相似性特征提取模块920, 配置为针对除所述视频图像中的参考帧图像之后的待跟踪图像, 生成所述待跟踪图像中的搜索区域与所述参考帧图像中的目标图像区域之间的图像相似性特征图; 其中, 所述目标图像区域内包含待跟踪对象;

定位模块930,配置为根据所述图像相似性特征图,确定所述搜索区域中的待定位区域的定位位置信息;

跟踪模块940,配置为响应于在所述搜索区域中确定出所述待定位区域的定位位置信息,根据确定的待定位区域的定位位置信息,确定所述待跟踪对象在包含所述搜索区域的待跟踪图像中的检测框。

在一些实施例中,所述定位模块930,配置为:根据所述图像相似性特征图,预测所述待定位区域的尺寸信息;根据所述图像相似性特征图,预测所述搜索区域的特征图中的每个特征像素点的概率值,一个特征像素点的概率值表征所述搜索区域中与该特征像素点对应的像素点位于所述待定位区域内的几率;根据所述图像相似性特征图,预测所述搜索区域中与每个所述特征像素点对应的像素点与所述待定位区域的位置关系信息;从预测的概率值中选取所述概率值最大的特征像素点所对应的所述搜索区域中的像素点作为目标像素点;基于所述目标像素点、所述目标像素点与所述待定位区域的位置关系信息、以及所述待定位区域的尺寸信息,确定所述待定位区域的定位位置信息。

在一些实施例中,所述相似性特征提取模块920,配置为利用以下步骤从所述参考帧图像中提取所述目标图像区域:确定所述待跟踪对象在所述参考帧图像中的检测框;基于所述参考帧图像中的所述检测框的尺寸信息,确定所述参考帧图像中的所述检测框对应的第一延伸尺寸信息;基于所述第一延伸尺寸信息,以所述参考帧图像中的所述检测框为起始位置向周围延伸,得到所述目标图像区域。

在一些实施例中,所述相似性特征提取模块920,配置为利用以下步骤从待跟踪图像中提取搜索区域:获取在所述视频图像中当前帧待跟踪图像的前一帧待跟踪图像中,所述待跟踪对象的检测框;基于所述待跟踪对象的检测框的尺寸信息,确定所述待跟踪对象的检测框对应的第二延伸尺寸信息;基于所述第二延伸尺寸信息和所述待跟踪对象的检测框的尺寸信息,确定当前帧待跟踪图像中的搜索区域的尺寸信息;以所述待跟踪对象的检测框的中心点为当前帧待跟踪图像中的搜索区域的中心,根据当前帧待跟踪图像中的搜索区域的尺寸信息确定所述搜索区域。

在一些实施例中,所述相似性特征提取模块920,配置为:将所述搜索区域缩放至第一预设尺寸,以及,将所述目标图像区域缩放至第二预设尺寸;生成所述搜索区域中的第一图像特征图,以及所述目标图像区域中的第二图像特征图;所述第二图像特征图的尺寸小于所述第一图像特征图的尺寸;确定所述第二图像特征图与所述第一图像特征图中的每个子图像特征图之间的相关性特征;所述子图像特征图与所述第二图像特征图的尺寸相同;基于确定的多个相关性特征,生成所述图像相似性特征图。

在一些实施例中,所述目标跟踪装置利用跟踪定位神经网络确定所述待跟踪对象在包含所述搜索区域的待跟踪图像中的检测框;其中所述跟踪定位神经网络由标注有目标对象的检测框的样本图像训练得到。

在一些实施例中,所述目标跟踪装置还包括模型训练模块950,配置为:获取样本图像,所述样本图像包括参考帧样本图像和待跟踪的样本图像;将所述样本图像输入待训练的跟踪定位神经网络,经过所述待训练的跟踪定位神经网络对输入的样本图像进行处理,预测所述目标对象在所述待跟踪的样本图像中的检测框;基于所述待跟踪的样本图像中标注的检测框和所述待跟踪的样本图像中预测的检测框,调整所述待训练的跟踪定位神经网络的网络参数。

在一些实施例中,将所述待跟踪的样本图像中的待定位区域的定位位置信息作为所述待跟踪的样本图像中预测的检测框的位置信息,所述模型训练模块950在基于所述待跟踪的样本图像中标注的检测框和所述待跟踪的样本图像中预测的检测框,调整所述待训练的跟踪定位神经网络的网络参数的情况下,配置为:基于所述预测的检测框的尺寸

信息、所述待跟踪的样本图像中搜索区域中每个像素点位于所述预测的检测框内的预测概率值、所述待跟踪的样本图像中搜索区域中每个像素点与所述预测的检测框的预测位置关系信息、所述标注的检测框的标准尺寸信息、所述待跟踪的样本图像中标准搜索区域中每个像素点是否位于标注的检测框中的信息、所述标准搜索区域中每个像素点与所述标注的检测框的标准位置关系信息，调整所述待训练的跟踪定位神经网络的网络参数。

本公开实施例上述目标跟踪装置在预测检测框过程中执行的实施方式可以参见上述目标跟踪方法的描述，实施过程与上述相似，这里不再赘述。

本公开实施例公开了一种电子设备，如图10所示，包括：处理器1001、存储器1002和总线1003，所述存储器1002存储有所述处理器1001可执行的机器可读指令，当电子设备运行时，所述处理器1001与所述存储器1002之间通过总线1003通信。

所述机器可读指令被所述处理器1001执行时执行以下目标跟踪方法的步骤：获取视频图像；针对除所述视频图像中的参考帧图像之后的待跟踪图像，生成所述待跟踪图像中的搜索区域与所述参考帧图像中的目标图像区域之间的图像相似性特征图；其中，所述目标图像区域内包含待跟踪对象；根据所述图像相似性特征图，确定所述搜索区域中的待定位区域的定位位置信息；响应于在所述搜索区域中确定出所述待定位区域的定位位置信息，根据确定的待定位区域的定位位置信息，确定所述待跟踪对象在包含所述搜索区域的待跟踪图像中的检测框。

除此之外，机器可读指令被处理器1001执行时，还可以执行上述方法部分描述的任一实施方其中的方法内容，这里不再赘述。

本公开实施例还提供的一种对应于上述方法及装置的计算机程序产品，包括存储了程序代码的计算机可读存储介质，程序代码包括的指令可用于执行前面方法实施例中的方法，实现过程可参见方法实施例，在此不再赘述。

上文对各个实施例的描述倾向于强调各个实施例之间的不同之处，其相同或相似之处可以相互参考，为了简洁，本文不再赘述。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统 and 装置的工作过程，可以参考方法实施例中的对应过程，本公开实施例中不再赘述。在本公开实施例所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。以上所描述的装置实施例仅仅是示意性的，例如，所述模块的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，又例如，多个模块或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些通信接口，装置或模块的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的模块可以是或者也可以不是物理上分开的，作为模块显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本公开实施例各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。

所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个处理器可执行的非易失的计算机可读存储介质中。基于这样的理解，本公开实施例的技术方案本质上或者说对相关技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本公开实施例各个实施例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、ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

5 以上仅为本公开实施例的实施方式，但本公开实施例的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本公开实施例揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本公开实施例的保护范围之内。因此，本公开实施例的保护范围应以权利要求要求的保护范围为准。

工业实用性

10 本公开实施例中，利用端到端训练的方式对预测目标框进行充分训练，不需要在线更新，实时性更高。同时通过跟踪网络直接预测目标框的点位置、偏差和长宽结果，从而直接获得最终目标框信息。网络结构更为简单有效，不存在候选框的预测过程，更适应移动端的算法需求，并且在提升了精度的同时保持了跟踪算法的实时性。