

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 5 月 22 日 (22.05.2020)



(10) 国际公布号
WO 2020/098076 A1

(51) 国际专利分类号:
G06K 9/00 (2006.01)

(21) 国际申请号: PCT/CN2018/123080

(22) 国际申请日: 2018 年 12 月 24 日 (24.12.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811339121.3 2018 年 11 月 12 日 (12.11.2018) CN

(71) 申请人: 平安科技(深圳)有限公司 (PING AN TECHNOLOGY(SHENZHEN)CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 杨国青 (YANG, Guoqing); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳众鼎专利商标代理事务所(普通合伙) (SHENZHEN ZHONGDING INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市龙岗区龙城街道中心城清林路 546 号城投商务中心 4 层/B, Guangdong 518172 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,

(54) Title: METHOD AND APPARATUS FOR POSITIONING TRACKING TARGET, DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 跟踪目标的定位方法、装置、设备及存储介质

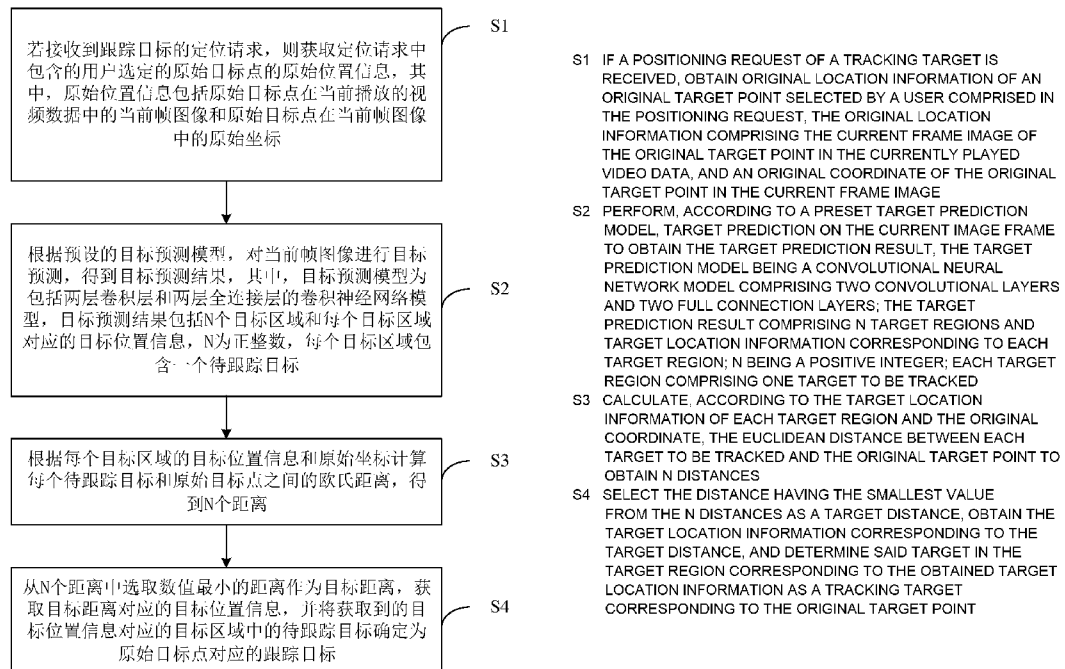


图 2

(57) Abstract: The present application relates to the technical field of image processing, and provides a method and apparatus for positioning a tracking target, a device, and a storage medium. The method comprises: if a positioning request of a tracking target is received, obtaining original location information of an original target point selected by a user comprised in the positioning request; performing, according to a preset target prediction model, target prediction on the current image frame to obtain the target prediction result; calculating, according to target location information of each target region and an original coordinate, the Euclidean distance

MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

between each target to be tracked and the original target point to obtain N distances; and selecting the distance having the smallest value from the N distances as a target distance, obtaining the target location information corresponding to the target distance, and determining said target in the target region corresponding to the obtained target location information as a tracking target corresponding to the original target point. The present application can ensure the accuracy of positioning the tracking target, and improve the positioning efficiency thereof.

(57) 摘要: 本申请涉及图像处理技术领域, 提供了一种跟踪目标的定位方法、装置、设备及存储介质, 该方法包括: 若接收到跟踪目标的定位请求, 则获取定位请求中包含的用户选定的原始目标点的原始位置信息; 根据预设的目标预测模型, 对当前帧图像进行目标预测, 得到目标预测结果; 根据每个目标区域的目标位置信息和原始坐标计算每个待跟踪目标和原始目标点之间的欧氏距离, 得到N个距离; 从N个距离中选取数值最小的距离作为目标距离, 获取目标距离对应的目标位置信息, 并将获取到的目标位置信息对应的目标区域中的待跟踪目标确定为原始目标点对应的跟踪目标。本申请能够保证对跟踪目标的定位的准确率和提高对跟踪目标的定位效率。

跟踪目标的定位方法、装置、设备及存储介质

本申请以 2018 年 11 月 12 日提交的申请号为 201811339121.3，名称为“跟踪目标的定位方法、装置、设备及存储介质”的中国发明专利申请为基础，并要求其优先权。

技术领域

本申请涉及图像处理技术领域，尤其涉及一种跟踪目标的定位方法、装置、设备及存储介质。

背景技术

进行长距离的跟踪目标的定位一直是视频监控中最具挑战性的任务之一，传统的跟踪目标的定位方法中，有的是通过将图像中需要跟踪的目标的坐标信息写入文件，在跟踪启动时通过读取文件获取初始化的目标信息，这种方法的周期长，获取初始化的目标信息的效率不高，或者是通过暂停视频画面，采用人工画图的方式在图像中画出需跟踪的目标的位置，并在跟踪启动后，根据人工画出的位置信息，进行跟踪目标的初始化，由于人工画出的位置信息的准确率不高，且容易包含大量背景信息，以及造成跟踪的目标被背景截断或遮挡，导致跟踪目标的难度增大，获取初始化的跟踪目标的定位信息的效率不高。

发明内容

基于此，有必要针对上述技术问题，提供一种跟踪目标的定位方法、装置、设备及存储介质，以解决跟踪目标的定位效率和准确率均不高的问题。

一种跟踪目标的定位方法，包括：

若接收到跟踪目标的定位请求，则获取所述定位请求中包含的用户选定的原始目标点的原始位置信息，其中，所述原始位置信息包括所述原始目标点在当前播放的视频数据中的当前帧图像和所述原始目标点在所述当前帧图像中的原始坐标；

根据预设的目标预测模型，对所述当前帧图像进行目标预测，得到目标预测结果，其中，所述目标预测模型为包括两层卷积层和两层全连接层的卷积神经网络模型，所述目标预测结果包括 N 个目标区域和每个所述目标区域对应的目标位置信息，N 为正整数，每个所述目标区域包含一个待跟踪目标；

根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离，得到 N 个距离；

从所述 N 个距离中选取数值最小的距离作为目标距离，获取所述目标距离对应的目标位置信息，并将获取到的目标位置信息对应的目标区域中的待跟踪目标确定为所述原始目标点对应的跟踪目标。

一种跟踪目标的定位装置，包括：

请求接收模块，用于若接收到跟踪目标的定位请求，则获取所述定位请求中包含的用户选定的原始目标点的原始位置信息，其中，所述原始位置信息包括所述原始目标点在当前播放的视频数据中的当前帧图像和所述原始目标点在所述当前帧图像中的原始坐标；

目标预测模块，用于根据预设的目标预测模型，对所述当前帧图像进行目标预测，得到目标预测结果，其中，所述目标预测模型为包括两层卷积层和两层全连接层的卷积神经网络模型，所述目标预测结果包括 N 个目标区域和每个所述目标区域对应的目标位置信息，N 为正整数，每个所述目标区域包含一个待跟踪目标；

距离计算模块，用于根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离，得到 N 个距离；

目标确认模块，用于从所述 N 个距离中选取数值最小的距离作为目标距离，获取所述目标距离对应的目标位置信息，并将获取到的目标位置信息对应的目标区域中的待跟踪目标确定为所述原始目标点对应的跟踪目标。

一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现上述跟踪目标的定位方法。

一个或多个存储有计算机可读指令的非易失性可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行时实现上述跟踪目标的定位方法。

本申请的一个或多个实施例的细节在下面的附图和描述中提出，本申请的其他特征和优点将从说明书、附图以及权利要求变得明显。

附图说明

为了更清楚地说明本申请实施例的技术方案，下面将对本申请实施例的描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 是本申请一实施例中跟踪目标的定位方法的一应用环境示意图；

图 2 是本申请一实施例中跟踪目标的定位方法的一流程图；

图 3 是本申请一实施例中跟踪目标的定位方法中步骤 S3 的实现流程图；

图 4 是本申请一实施例中跟踪目标的定位方法中构建目标预测模型的一流程图；

图 5 是本申请一实施例中跟踪目标的定位方法中构建卷积神经网络目标检测算法的结构的一实现流程图；

图 6 是本申请一实施例中跟踪目标的定位方法中确认跟踪目标的一实现流程图；

图 7 是本申请一实施例中跟踪目标的定位装置的一示意图；

图 8 是本申请一实施例中计算机设备的一示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

图 1 示出了本申请实施例提供的应用环境，该应用环境包括服务端和客户端，其中，服务端和客户端之间通过网络进行连接，客户端用于检测用户的触摸操作，根据该触摸操作生成定位请求，并将定位请求发送到服务端，客户端具体可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备；服务端用于处理视频数据，服务端具体可以用独立的服务器或者多个服务器组成的服务器集群实现。本申请实施例提供的跟踪目标的定位方法应用于服务端。

请参阅图 2，图 2 示出本实施例提供的跟踪目标的定位方法的实现流程。详述如下：

S1：若接收到跟踪目标的定位请求，则获取定位请求中包含的用户选定的原始目标点的原始位置信息，其中，原始位置信息包括原始目标点在当前播放的视频数据中的当前帧图像和原始目标点在当前帧图像中的原始坐标。

在本实施例中，原始目标点是用户在正在播放的视频画面中，选定的一个想要跟踪的目标所在的位置区域上的一个点。客户端通过检测到用户的触摸操作、点击操作或者其他操作，生成跟踪目标的定位请求，并将定位请求发送到服务端进行处理。

其中，触摸操作是指在显示设备上通过人手触摸屏幕的方式来实现的操作，显示设备能够根据人手触摸的图标或菜单位置来定位选择信息输入，或者根据人手触摸的位置，检测触摸的位置的坐标。点击操作是指在显示设备通过光标点击的方式来实现的操作，显示设备能够根据光标点击的图标或菜单位置来定位选择信息输入，或者根据光标点击的位置，检测光标点击的坐标。

需要说明的是，用于获取原始目标点在视频数据中的原始位置信息的显示设备具体可以是触摸屏，也可以是其他设备，此处不做具体限制。当前帧图像是指原始目标点所在的当前正在播放的画面，可以通过显示设备中的部件进行锁定。

优选地，本实施例中采用的是触摸屏的显示设备用于获取原始目标点在视频数据中的原始位置信息，其中，触摸屏由触摸检测部件和触摸屏控制器组成，即触摸检测部件和触摸屏控制器安装在显示设备中，触摸检测部件用于检测用户触摸的位置和锁定当前正在播放的视频数据的当前帧图像，生成触摸信息，然后，将触摸信息发送至触摸屏控制器，而触摸屏控制器的主要作用是从触摸点检测装置上接收触摸信息，并将触摸信息中的用户触摸的位置转换成触点坐标，该触点坐标即为原始目标点的原始坐标。

具体地，在实时播放的视频中，由于用户想要选定视频画面中的某一目标，则可以对播放视频的显示设备进行触摸操作，当客户端检测到用户在显示设备上的触摸操作时，则可以根据该触摸操作确定当前正在播放的视频数据的当前帧图像，以及用户选定的原始目标点在当前帧图像中的原始坐标，并生成包含当前帧图像和原始坐标的定位请求，客户端将该定位请求发送给服务端。

服务端接收到该定位请求，读取该定位请求中携带的原始目标点的原始位置信息。

S2: 根据预设的目标预测模型，对当前帧图像进行目标预测，得到目标预测结果，其中，目标预测模型为包括两层卷积层和两层全连接层的卷积神经网络模型，目标预测结果包括 N 个目标区域和每个目标区域对应的目标位置信息， N 为正整数，每个目标区域包含一个待跟踪目标。

在本实施例中，预设的目标预测模型是根据实际业务需求进行训练得到的，用于预测原始目标点所在的当前帧图像中候选的目标区域，其中，候选的目标区域是在当前帧图像中包含了可移动对象的区域，每个可移动对象对应一个目标区域，该目标区域以矩形框的形式显示，目标区域中的可移动对象即为待跟踪目标，目标区域的目标位置信息包括区域的左上角坐标、区域的宽和高。目标预测模型的卷积层用于对当前帧图像进行图像特征提取，全连接层根据提取到的图像特征，输出图像特征对应的目标区域，以及目标区域对应的目标位置信息。

具体地，将当前帧图像输入预设的目标预测模型中，通过模型中的卷积层对当前帧图像进行图像特征提取，然后，将图像特征通过模型的全连接层，输出图像特征对应的目标区域，以及目标区域对应的目标位置信息，作为目标预测结果。

例如，将一帧图像输入预设的目标预测模型中，通过模型中的卷积层对这帧图像进行图像特征提取，然后，将提取到的图像特征通过模型的全连接层，将不属于图像背景的可移动对象在以矩形框的形式显示出来，假设得到预测得到 6 个可移动对象，则目标预测结果为 6 个目标区域，以及该 6 个目标区域对应的目标位置信息。

需要说明的是，通过训练好的目标预测模型对当前帧图像进行目标预测，通过目标预测模型中的卷积层对当前帧图像进行图像特征提取，能够准确区分出当前帧图像中的可移动对象和背景，避免人工画出的预测目标的位置信息不准确，预测目标易被背景遮挡的情况，从而提高对跟踪目标预测的准确率。

S3: 根据每个目标区域的目标位置信息和原始坐标计算每个待跟踪目标和原始目标点之间的欧氏距离，得到 N 个距离。

在本实施例中，由于欧氏距离是一个常用的距离定义，可用于表示在 m 维空间中多个

点之间的真实距离，或者向量的自然长度，以及在二维空间中的两点之间的欧氏距离就是两点之间的实际距离，因此，通过计算每个待跟踪目标和原始目标点之间的欧氏距离，能够最快速、直观地反映出每个待跟踪目标和原始目标点之间的位置关系。

具体地，将步骤 S2 中得到的每个目标区域的目标位置信息和步骤 S1 中得到的原始目标点的原始坐标，代入欧氏距离公式中，经过计算可以得到目标区域中每个待跟踪目标和原始目标点之间的 N 个欧氏距离。

例如，继续使用步骤 S2 中的例子，若通过目标预测模型预测得到 6 个目标区域和该 6 个目标区域对应的目标位置信息，则将这 6 个目标区域的目标位置信息和原始坐标，代入欧氏距离公式中，能够计算出 6 个待跟踪目标和原始目标点之间的欧氏距离，得到 6 个距离。

S4：从 N 个距离中选取数值最小的距离作为目标距离，获取目标距离对应的目标位置信息，并将获取到的目标位置信息对应的目标区域中的待跟踪目标确定为原始目标点对应的跟踪目标。

在本实施例中，由于在步骤 S3 中计算得到 N 个欧氏距离，能够直观地反映出每个待跟踪目标和原始目标点之间的位置关系，所以在通常情况下，可以理解为待跟踪目标的位置最接近原始目标点，则该待跟踪目标就是原始目标点对应的跟踪目标。目标距离即原始目标点位置与最接近的待跟踪目标之间的距离。

具体地，将步骤 S3 中得到的 N 个距离中任意两个距离进行比较，取数值小的距离，再将该数值小的距离与其他距离进行比较，以此类推，直至比较出 N 个距离中数值最小的距离，将该数值最小的距离作为目标距离，并将该数值最小的距离对应的目标区域中的待跟踪目标作为原始目标点对应的跟踪目标。

例如，继续使用步骤 S3 中的例子，假设计算得到的 6 个距离，分别为 45.63、56.03、78.24、34.33、40.55 和 63.06，将 6 个距离中任意两个距离进行比较，如取 45.63 与 56.03 比较，得到 45.63 小于 56.03，则选出 45.63，然后再将 45.63 与其他距离进行比较，如将 45.63 与 40.55 比较，得到 45.63 大于 40.55，则选出 40.55，以此类推，直到比较得出数值最小的距离为 34.33，则将距离 34.33 对应的目标区域中的待跟踪目标作为原始目标点对应的跟踪目标。

在本实施例中，通过从接收到的跟踪目标的定位请求中获取用户选定的原始目标点的原始位置信息，并根据预设的目标预测模型，对原始位置信息中的视频数据的当前帧图像进行目标预测，得到目标预测结果，能够准确区分出当前帧图像中的目标和背景，避免人工画出的预测目标的位置信息不准确，预测目标易被背景遮挡的情况，从而提高对跟踪目标预测的准确率，同时，根据每个目标区域的目标位置信息和原始坐标计算每个待跟踪目标和原始目标点之间的欧氏距离，然后，选取数值最小的距离作为目标距离，获取该目标距离对应的目标位置信息，并将获取到的目标位置信息对应的目标区域中的待跟踪目标确定为原始目标点对应的跟踪目标，这种在保证当前帧图像中的预测出的目标区域较为准确的基础上，计算待跟踪目标和原始目标点之间最小距离，来确定跟踪目标的方法，能够保证对跟踪目标的定位的准确率，该方法运算量小，操作简便，并且用户不需要暂停当前正在播放的视频，缩短了操作周期，从而提高对跟踪目标的定位效率。

在一实施例中，如图 3 所示，步骤 S3 中，目标区域对应的目标位置信息包括目标区域的左上角坐标、目标区域的宽和高，即根据每个目标区域的目标位置信息和原始坐标计算每个待跟踪目标和原始目标点之间的欧氏距离，得到 N 个距离具体包括如下步骤：

S301：针对每个目标位置信息，按照如下中心点坐标公式，计算第 i 个目标位置信息对应的目标区域的中心点坐标 (x_{i_0}, y_{i_0}) ：

$$(x_{i_0}, y_{i_0}) = \left(x_i + \frac{w_i}{2}, y_i + \frac{h_i}{2} \right)$$

其中, i 为正整数, 且 $0 < i \leq N$, x_i 为第 i 个目标位置信息对应的目标区域的左上角坐标的横坐标, y_i 为左上角坐标的纵坐标, w_i 为第 i 个目标位置信息对应的目标区域的宽和 h_i 为第 i 个目标位置信息对应的目标区域的高。

具体地, 由于目标区域是以矩形框的形式显示的, 且目标区域的位置信息包括矩形框的左上角的坐标、矩形框的宽和高, 为了能准确反映出每个目标区域中的待跟踪目标与原始目标点之间的位置关系, 需要将目标区域的位置信息代入中心点坐标公式, 计算出每个目标区域的中心点坐标, 因此, 当前帧图像上预测出 N 个目标区域, 按照中心点坐标公式, 分别计算第一个至第 N 个目标区域的中心点坐标, 得到 N 个中心点坐标, 方便后续能够根据得到的 N 个目标区域的中心点坐标, 快速计算出每个待跟踪目标和原始目标点之间的欧氏距离。

例如, 假设当前帧图像上预测出的一个目标区域, 该目标区域的左上角坐标为 (50, 50), 目标区域的宽为 80, 高为 60, 那么将目标区域的位置信息代入中心坐标公式进行计算, 得到该目标区域的中心点坐标为 (70, 80)。

S302: 针对每个中心点坐标, 计算第 i 个中心点坐标 (x_{i_0}, y_{i_0}) 与原始坐标 $(\Delta x, \Delta y)$ 之间的欧氏距离 d_i :

$$d_i = \sqrt{(\Delta x - x_{i_0})^2 + (\Delta y - y_{i_0})^2}。$$

具体地, 将步骤 S301 中得到的每个中心点坐标和原始坐标, 分别代入欧氏距离公式中进行计算, 能够快速计算出每个目标区域中的待跟踪目标和原始目标点之间的欧氏距离。

例如, 继续使用步骤 S301 中的例子, 得到的一个目标区域的中心点坐标为 (90, 80), 假设原点坐标为 (110, 130), 将中心点坐标和原点坐标, 代入欧氏距离公式中进行计算, 得到欧氏距离 d_1 约等于 53.85, 假如计算得到另外两个中心点坐标分别为 (60, 90) 和 (100, 70), 将这两个中心点坐标和原点坐标, 分别代入欧氏距离公式中进行计算, 中心点坐标 (60, 90) 对应的欧氏距离 d_2 约等于 64.03, 中心点坐标 (100, 70) 对应的欧氏距离 d_3 约等于 60.83。

在本实施例, 对步骤 S2 中得到的目标区域的目标位置信息, 采用中心点坐标公式来计算目标区域的中心点坐标, 然后, 将得到的每个中心点坐标和原点坐标, 分别代入欧氏距离公式中进行计算, 能够快速计算出每个目标区域中的待跟踪目标和原始目标点之间的欧氏距离。

在一实施例中, 如图 4 所示, 在步骤 S2 之前, 该跟踪目标的定位方法还包括如下步骤:

S5: 获取样本训练图像数据。

在本实施例中, 样本训练图像数据包括 MsCoCo 数据集和 CalTech 数据集。其中, MsCoCo 数据集为网络共享的公有数据集, 在该数据集中包括了 328000 个影像、2500000

个标签和包含了 91 类目标的图像，这个数据集以场景理解（scene understanding）为目标，图像中的目标通过精确的分割进行位置的标定，从复杂的日常场景中截取。Caltech 数据集为网络共享的公有数据集主要包括 2 类，一类是 Caltech-101，包含了 101 类的图像，每类大约有 40~800 张图像，大部分是 50 张/类，每张图像的大小大约是 300x200；另一类是 Caltech-256，和 Caltech-101 相似，包含了 30607 张图像。

具体地，获取样本训练图像数据的方式具体可以包括但不限于接收用户上传的样本训练图像数据，或者是自动定时从第三方视频图像数据库获取样本训练图像数据等，还可以是其他获取方式，此处不做限制，其中，第三方视频图像数据库具体可以是网络视频平台的数据库。

S6: 基于深度学习框架构建卷积神经网络目标检测算法的结构。

在本实施例中，深度学习框架包括 caffe 或 caffe2 等，其中，caffe2 是轻量级、模块化和可扩展的一个框架，包括存储接口（blob）、层结构（layer）和结构链接（net）。其中，blob 是整个框架的标准的数组结构和统一存储接口。layer 作为建模和计算的层结构基础，net 作为 layer 的集合和链接。blob 的细节描述了信息是怎样在 layers 和 nets 间存储和交换的。caffe2 具体是通过使用自己的建模方法将网络一层一层定义出来的，即将网络由输入数据到损失层把整个模型自底向上的定义出来。数据和偏导数等信息在网络中前向、后向流动，而 caffe2 使用 blob 存储、交换和操纵这些信息。

优选地，本实施例采用基于 caffe2 框架构建卷积神经网络目标检测算法(Faster-RCNN)的结构，具体可以通过 caffe2 框架中的 layer 对卷积神经网络目标检测算法的结构进行每一个层结构的定义，然后，通过结构链接（net）将定义好的每一个层结构进行链接，其次，通过 caffe2 框架中的 blob 将链接好的层结构进行存储，得到搭建好的 Faster-RCNN 的结构，在 caffe2 框架构建的 Faster-RCNN 的结构，逻辑结构层次分明，参数定义清晰，能够根据参数定义快速锁定结构中的参数位置，便于 Faster-RCNN 的结构的修改和完善。

在本实施例中，构建的卷积神经网络目标检测算法的结构是由两层卷积层和两层全连接层构成，按照从前至后的顺序，分别是第一层为第一卷积层，第二层为第二卷积层，第三层为第一全连接层和第四层为第二全连接层。

其中，第一卷积层能够用于对输入的图像数据进行卷积，对图像数据进行初步特征提取，具体可以通过将该卷积层的卷积核窗口大小设为 11*11，相邻局部接受域的中心距离设为 4，能够提取出输入的图像数据对应的 96 个特征图，然后，第一卷积层连接有一个下采样层，能够将每个特征图经过一个下采样层进行降维，以压缩特征图，保证提取的特征的完整性，下采样层的核窗口大小通常设为 3*3，相邻局部接受域的中心距离为 2，最后，能够将经下采样层得到降维后的特征图输入到第二层卷积层。

第二卷积层，能够用于对第一卷积层中得到特征图进行卷积，是对第一卷积层得到特征图进行进一步地特征提取，提高特征的提取精度，将第二卷积层的卷积核窗口大小设为 5*5，相邻局部接受域的中心距离设为 2，能够输出 256 个特征图，第二卷积层连接有一个下采样层，将每个特征图经过一个下采样层进行降维，以压缩特征图，保证提取的特征的完整性和精度，下采样层的核窗口大小设为 3*3，相邻局部接受域的中心距离为 2，能够将经下采样层得到降维后的特征图输入到第一全连接层。

第三层是第一全连接层，该层有 1000 个神经元，能够用于将第二卷积层中得到的特征图变换成一个 n 维列向量，能够将 n 维列向量与第一全连接层的权值矩阵和偏置进行非线性变换得到一个 1000 维列向量输入到第二全连接层；

第四层是第二全连接层，该层有 m 个神经元，能够用于将第一全连接层中得到的 1000 维列向量与第二全连接层的权值矩阵和偏置进行非线性变换得到一个 m 维特征向量，其中，m 的数值是根据图像上的目标个数，具体可以通过更新第二全连接层的权值矩阵和偏置进行确定，即 m 个目标对应设置有 m 个神经元。

S7: 在卷积神经网络目标检测算法的结构中, 使用样本训练图像数据对卷积神经网络目标检测算法的结构进行训练, 得到目标预测模型。

在本实施例中, 使用样本训练图像数据对卷积神经网络目标检测算法的结构进行训练, 是在 MsCoCo 数据集和 CalTech 数据集上通过使用共享特征交替训练的方式, 即对两个数据集中相同的图片进行特征提取和对卷积神经网络目标检测算法的结构中的权值矩阵和偏置进行更新, 直到得到相同的目标数, 则可以确定目标预测模型中应该使用的权值和偏置, 得到一个目标预测模型。其中, 目标预测模型包括两个文件构成, 一个文件是卷积神经网络目标检测算法的结构, 另外一个文件是卷积神经网络目标检测算法的结构中各个每个层的权值 (weight) 和偏置 (bias), 权值即权值矩阵中的权重值, 偏置即权值矩阵所对应的偏移量。

具体地, 训练的过程是将样本训练图像数据在卷积神经网络目标检测算法的结构中提取到的特征, 通过传统的随机梯度下降算法, 将卷积神经网络目标检测算法的结构的每个层中包含的代价函数, 按照层结构从前至后的方向, 计算每个层对应的代价函数值, 该代价函数值可以用于后续每个层中的误差灵敏度的计算, 然后, 通过传统的反向传播算法, 将卷积神经网络目标检测算法的结构的每个层, 按照层结构从后至前的方向, 计算每个层中的误差灵敏度, 最后, 将计算得到的每个层的误差灵敏度中包含的权重值和偏移量, 用于更新层结构中的原始的权重值和偏移量, 得到目标预测模型。

在本实施例中, 由于目标检测算法的结构对样本训练图像数据具有很好的拟合性, 因此, 基于 caffe2 框架构建卷积神经网络目标检测算法的结构, 在构建好的卷积神经网络目标检测算法的结构中利用设置好的可滑动的卷积核窗口对输入的样本训练图像数据进行卷积的操作来实现对样本训练图像数据的特征提取, 能够保证对该样本训练图像数据的特征提取完整性, 且在提取图像特征的过程中, 对图像形变以及平移等几何变换具有高度不变性, 从而能够保证训练得到的目标预测模型预测到的图像的的目标的特征的完整性, 能够保证了后续对图像中的目标信息识别的准确率。

在一实施例中, 如图 5 所示, 步骤 S6 中, 即基于深度学习框架构建卷积神经网络目标检测算法的结构具体包括如下步骤:

S601: 获取卷积神经网络目标检测算法的结构的定义信息。

在本实施例中, 不同的网络结构所需的配置参数的可能不同, 网络结构的结构定义也可能不同, 针对不同网络结构的构建需求, 预先设置每个网络结构对应的层结构包含的参数名及参数名对应的具体参数数据, 从而根据待构建网络结构的参数名能够快速从定义信息中获取该待构建网络结构所需配置的参数数据, 以使得后续在构建网络结构的过程中能够快速填入层结构中。

例如, 层结构中参数名为层的名字、层的类型或输出数据的类型定义等, 参数名对应的具体参数数据, 如“层的名字”对应的具体参数为“conv2 (第二卷积层)”、“层的类型”对应的具体参数为“convolution (卷积)”或“输出数据的类型定义”对应的具体参数为“data (数据)”等。

S602: 在深度学习框架中按照预设的网络定义的结构, 将定义信息填入对应的卷积层和全连接层的层结构中, 生成卷积神经网络目标检测算法的结构。

在本实施例中, 预设的网络定义结构包括层的名字、层的类型、与层的类型相对应的层参数、输出数据的类型定义、滤波参数中滤波的学习率和滤波参数中滤波的衰减率等。

例如, 若预设的网络定义结构的一个层结构包括层的名字、层的类型、滤波参数中滤波的学习率和滤波参数中滤波的衰减率等, 则模型定义结构的层结构可以设置为:

```
layer{
  name:" "
  type:" "
```

```

param{
  lr_mult:" "
  decay_mult:" "
}

```

当层的名字为“conv1（第一卷积层）”、层的类型为“convolution（卷积）”、滤波参数中滤波的学习率为“1”和滤波参数中滤波的衰减率为“1”时，生成的结构描述为：

```

layer{
  name:"conv1"
  type:"convolution"
  param{
    lr_mult:"1"
    decay_mult:"1"
  }
}

```

在本实施例中，通过获取到的卷积神经网络目标检测算法的结构的定义信息，将定义信息中与预设的网络定义的层结构中有相同参数名对应的具体参数数据，添加到预设的网络定义的层结构中，即能够在深度学习框架中快速生成与定义信息对应的卷积层和全连接层的层结构中，生成卷积神经网络目标检测算法的结构，操作简便快捷，不易出错，还可以通过参数名实现对具体参数数据的快速修改，便于完善卷积神经网络目标检测算法的结构。

在一实施例中，如图 6 所示，在步骤 S3 之后，该跟踪目标的定位方法还包括如下步骤：

S8：将 N 个距离中小于预设阈值的距离的目标位置信息对应的目标区域中的待跟踪目标反馈给用户进行确认。

具体地，为了进一步地提高对跟踪目标的定位的准确性，能够通过设定一个距离阈值，将小于该距离阈值的距离对应的目标区域中的待跟踪目标发给用户进行确认，以增加目标跟踪定位的准确率。

S9：接收用户返回的确认结果，并根据确认结果，将用户确认的跟踪目标作为原始目标点对应的跟踪目标。

具体地，当用户接收到步骤 S8 中得到的小于预设阈值的距离对应的目标区域中的待跟踪目标，用户可以根据当前帧图像的原始目标点，从这些待跟踪目标中选定一个跟踪目标，作为确认结果并返回，以实现准确地定位出跟踪目标。

在本实施例中，通过将小于该阈值的距离对应的目标区域中的待跟踪目标发给用户进行确认，以使用户根据当前帧图像的原始目标点，从这些待跟踪目标中选定一个跟踪目标，作为确认结果并返回，能够实现准确地确定跟踪目标的定位，从而提高了对跟踪目标的定位的准确率。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本申请实施例的实施过程构成任何限定。

在一实施例中，提供一种跟踪目标的定位装置，该跟踪目标的定位装置与上述实施例中跟踪目标的定位方法一一对应。如图 7 所示，该跟踪目标的定位装置包括请求接收模块 701、目标预测模块 702、距离计算模块 703 和目标确认模块 704。各功能模块详细说明如下：

请求接收模块 701，用于若接收到跟踪目标的定位请求，则获取定位请求中包含的用户选定的原始目标点的原始位置信息，其中，原始位置信息包括原始目标点在当前播放的

视频数据中的当前帧图像和原始目标点在当前帧图像中的原始坐标；

目标预测模块 702，用于根据预设的目标预测模型，对当前帧图像进行目标预测，得到目标预测结果，其中，目标预测模型为包括两层卷积层和两层全连接层的卷积神经网络模型，目标预测结果包括 N 个目标区域和每个目标区域对应的目标位置信息，N 为正整数，每个目标区域包含一个待跟踪目标；

距离计算模块 703，用于根据每个目标区域的目标位置信息和原始坐标计算每个待跟踪目标和原始目标点之间的欧氏距离，得到 N 个距离；

目标确认模块 704，用于从 N 个距离中选取数值最小的距离作为目标距离，获取目标距离对应的目标位置信息，并将获取到的目标位置信息对应的目标区域中的待跟踪目标确定为原始目标点对应的跟踪目标。

进一步地，距离计算模块 703 包括：

坐标计算单元 7031，用于针对每个目标位置信息，按照如下中心点坐标公式，计算第 i 个目标位置信息对应的目标区域的中心点坐标 (x_{i_0}, y_{i_0}) ：

$$(x_{i_0}, y_{i_0}) = \left(x_i + \frac{w_i}{2}, y_i + \frac{h_i}{2} \right)$$

其中，i 为正整数，且 $0 < i \leq N$ ， x_i 为第 i 个目标位置信息对应的目标区域的左上角坐标的横坐标， y_i 为左上角坐标的纵坐标， w_i 为第 i 个目标位置信息对应的目标区域的宽和 h_i 为第 i 个目标位置信息对应的目标区域的高；

距离计算单元 7032，用于针对每个中心点坐标，计算第 i 个中心点坐标 (x_{i_0}, y_{i_0}) 与原始坐标 $(\Delta x, \Delta y)$ 之间的欧氏距离 d_i ：

$$d_i = \sqrt{(\Delta x - x_{i_0})^2 + (\Delta y - y_{i_0})^2}。$$

进一步地，该跟踪目标的定位装置还包括：

数据获取模块 705，用于获取样本训练图像数据；

结构构建模块 706，用于基于深度学习框架构建卷积神经网络目标检测算法的结构；

模型训练模块 707，用于在卷积神经网络目标检测算法的结构中，使用样本训练图像数据对卷积神经网络目标检测算法的结构进行训练，得到目标预测模型。

进一步地，结构构建模块 706 包括：

定义获取单元 7061，用于获取卷积神经网络目标检测算法的结构的定义信息；

结构生成单元 7062，用于在深度学习框架中按照预设的网络定义的层结构，将定义信息填入对应的卷积层和全连接层的层结构中，生成卷积神经网络目标检测算法的结构。

进一步地，该跟踪目标的定位装置还包括：

目标反馈模块 708，用于将 N 个距离中小于预设阈值的距离的目标位置信息对应的目标区域中的待跟踪目标反馈给用户进行确认；

结果接收模块 709，用于接收用户返回的确认结果，并根据确认结果，将用户确认的跟踪目标作为原始目标点对应的跟踪目标。

关于跟踪目标的定位装置的具体限定可以参见上文中对于跟踪目标的定位方法的限定，在此不再赘述。上述跟踪目标的定位装置中的各个模块可全部或部分通过软件、硬件

及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 8 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于保存视频数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种跟踪目标的定位方法。

在一个实施例中，提供了一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机可读指令，处理器执行计算机可读指令时实现上述实施例跟踪目标的定位方法的步骤，例如图 2 所示的步骤 S1 至步骤 S4。或者，处理器执行计算机可读指令时实现上述实施例中跟踪目标的定位装置的各模块/单元的功能，例如图 7 所示模块 701 至模块 704 的功能。为避免重复，这里不再赘述。

在一个实施例中，提供了一种非易失性存储介质，其上存储有计算机可读指令，计算机可读指令被处理器执行时实现上述方法实施例中跟踪目标的定位方法，或者，该计算机可读指令被处理器执行时实现上述装置实施例中跟踪目标的定位装置中各模块/单元的功能。为避免重复，这里不再赘述。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程 ROM（PROM）、电可编程 ROM（EPROM）、电可擦除可编程 ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM 以多种形式可得，诸如静态 RAM（SRAM）、动态 RAM（DRAM）、同步 DRAM（SDRAM）、双数据率 SDRAM（DDRSDRAM）、增强型 SDRAM（ESDRAM）、同步链路（Synchlink）、DRAM（SLDRAM）、存储器总线（Rambus）直接 RAM（RDRAM）、直接存储器总线动态 RAM（DRAM）、以及存储器总线动态 RAM（RDRAM）等。

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，仅以上述各功能单元、模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能单元、模块完成，即将所述装置的内部结构划分成不同的功能单元或模块，以完成以上描述的全部或者部分功能。

以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解，其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

权 利 要 求 书

1. 一种跟踪目标的定位方法，其特征在于，所述跟踪目标的定位方法包括：

若接收到跟踪目标的定位请求，则获取所述定位请求中包含的用户选定的原始目标点的原始位置信息，其中，所述原始位置信息包括所述原始目标点在当前播放的视频数据中的当前帧图像和所述原始目标点在所述当前帧图像中的原始坐标；

根据预设的目标预测模型，对所述当前帧图像进行目标预测，得到目标预测结果，其中，所述目标预测模型为包括两层卷积层和两层全连接层的卷积神经网络模型，所述目标预测结果包括 N 个目标区域和每个所述目标区域对应的目标位置信息， N 为正整数，每个所述目标区域包含一个待跟踪目标；

根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离，得到 N 个距离；

从所述 N 个距离中选取数值最小的距离作为目标距离，获取所述目标距离对应的目标位置信息，并将获取到的目标位置信息对应的目标区域中的待跟踪目标确定为所述原始目标点对应的跟踪目标。

2. 如权利要求 1 所述的跟踪目标的定位方法，其特征在于，所述目标区域对应的目标位置信息包括目标区域的左上角坐标、所述目标区域的宽和高，所述根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离，得到 N 个距离包括：

针对每个所述目标位置信息，按照如下中心点坐标公式，计算第 i 个所述目标位置信息对应的目标区域的中心点坐标 (x_{i_0}, y_{i_0}) ：

$$(x_{i_0}, y_{i_0}) = \left(x_i + \frac{w_i}{2}, y_i + \frac{h_i}{2} \right)$$

其中， i 为正整数，且 $0 < i \leq N$ ， x_i 为第 i 个所述目标位置信息对应的所述目标区域的左上角坐标的横坐标， y_i 为所述左上角坐标的纵坐标， w_i 为所述第 i 个所述目标位置信息对应的目标区域的宽和 h_i 为所述第 i 个所述目标位置信息对应的目标区域的高；

针对每个所述中心点坐标，计算第 i 个所述中心点坐标 (x_{i_0}, y_{i_0}) 与所述原始坐标 $(\Delta x, \Delta y)$ 之间的欧氏距离 d_i ：

$$d_i = \sqrt{(\Delta x - x_{i_0})^2 + (\Delta y - y_{i_0})^2}。$$

3. 如权利要求 1 所述的跟踪目标的定位方法，其特征在于，所述根据预设的目标预测模型，对所述当前帧图像进行目标预测，得到目标预测结果之前，所述跟踪目标的定位方法还包括：

获取样本训练图像数据；

基于深度学习框架构建卷积神经网络目标检测算法的结构；

在所述卷积神经网络目标检测算法的结构中，使用所述样本训练图像数据对所述卷积神经网络目标检测算法的结构进行训练，得到所述目标预测模型。

4. 如权利要求 3 所述的跟踪目标的定位方法，其特征在于，所述基于深度学习框架构

建卷积神经网络目标检测算法的结构包括：

获取所述卷积神经网络目标检测算法的结构的定义信息；

在所述深度学习框架中按照预设的网络定义的层结构，将所述定义信息填入对应的所述卷积层和所述全连接层的层结构中，生成所述卷积神经网络目标检测算法的结构。

5. 如权利要求 1 所述的跟踪目标的定位方法，其特征在于，在所述根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离，得到 N 个距离之后，所述跟踪目标的定位方法还包括：

将所述 N 个距离中小于预设阈值的所述距离的目标位置信息对应的目标区域中的待跟踪目标反馈给用户进行确认；

接收所述用户返回的确认结果，并根据所述确认结果，将所述用户确认的跟踪目标作为所述原始目标点对应的跟踪目标。

6. 一种跟踪目标的定位装置，其特征在于，所述跟踪目标的定位装置包括：

请求接收模块，用于若接收到跟踪目标的定位请求，则获取所述定位请求中包含的用户选定的原始目标点的原始位置信息，其中，所述原始位置信息包括所述原始目标点在当前播放的视频数据中的当前帧图像和所述原始目标点在所述当前帧图像中的原始坐标；

目标预测模块，用于根据预设的目标预测模型，对所述当前帧图像进行目标预测，得到目标预测结果，其中，所述目标预测模型为包括两层卷积层和两层全连接层的卷积神经网络模型，所述目标预测结果包括 N 个目标区域和每个所述目标区域对应的目标位置信息， N 为正整数，每个所述目标区域包含一个待跟踪目标；

距离计算模块，用于根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离，得到 N 个距离；

目标确认模块，用于从所述 N 个距离中选取数值最小的距离作为目标距离，获取所述目标距离对应的目标位置信息，并将获取到的目标位置信息对应的目标区域中的待跟踪目标确定为所述原始目标点对应的跟踪目标。

7. 如权利要求 6 所述的跟踪目标的定位装置，其特征在于，所述距离计算模块包括：

坐标计算单元，用于针对每个所述目标位置信息，按照如下中心点坐标公式，计算第 i 个所述目标位置信息对应的目标区域的中心点坐标 (x_{i_0}, y_{i_0}) ：

$$(x_{i_0}, y_{i_0}) = \left(x_i + \frac{w_i}{2}, y_i + \frac{h_i}{2} \right)$$

其中， i 为正整数，且 $0 < i \leq N$ ， x_i 为第 i 个所述目标位置信息对应的所述目标区域的左上角坐标的横坐标， y_i 为所述左上角坐标的纵坐标， w_i 为所述第 i 个所述目标位置信息对应的目标区域的宽和 h_i 为所述第 i 个所述目标位置信息对应的目标区域的高；

距离计算单元，用于针对每个所述中心点坐标，计算第 i 个所述中心点坐标 (x_{i_0}, y_{i_0}) 与所述原始坐标 $(\Delta x, \Delta y)$ 之间的欧氏距离 d_i ：

$$d_i = \sqrt{(\Delta x - x_{i_0})^2 + (\Delta y - y_{i_0})^2}。$$

8. 如权利要求 6 所述的跟踪目标的定位装置，其特征在于，所述跟踪目标的定位装置还包括：

数据获取模块, 用于获取样本训练图像数据;

结构构建模块, 用于基于深度学习框架构建卷积神经网络目标检测算法的结构;

模型训练模块, 用于在所述卷积神经网络目标检测算法的结构中, 使用所述样本训练图像数据对所述卷积神经网络目标检测算法的结构进行训练, 得到所述目标预测模型。

9. 如权利要求 7 所述的跟踪目标的定位装置, 其特征在于, 所述结构构建模块包括:

定义获取单元, 用于获取卷积神经网络目标检测算法的结构的定义信息;

结构生成单元, 用于在深度学习框架中按照预设的网络定义的层结构, 将定义信息填入对应的卷积层和全连接层的层结构中, 生成卷积神经网络目标检测算法的结构。

10. 如权利要求 6 所述的跟踪目标的定位装置, 其特征在于, 所述跟踪目标的定位装置还包括:

目标反馈模块, 用于将 N 个距离中小于预设阈值的距离的目标位置信息对应的目标区域中的待跟踪目标反馈给用户进行确认;

结果接收模块, 用于接收用户返回的确认结果, 并根据确认结果, 将用户确认的跟踪目标作为原始目标点对应的跟踪目标。

11. 一种计算机设备, 包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令, 其特征在于, 所述处理器执行所述计算机可读指令时实现如下步骤:

若接收到跟踪目标的定位请求, 则获取所述定位请求中包含的用户选定的原始目标点的原始位置信息, 其中, 所述原始位置信息包括所述原始目标点在当前播放的视频数据中的当前帧图像和所述原始目标点在所述当前帧图像中的原始坐标;

根据预设的目标预测模型, 对所述当前帧图像进行目标预测, 得到目标预测结果, 其中, 所述目标预测模型为包括两层卷积层和两层全连接层的卷积神经网络模型, 所述目标预测结果包括 N 个目标区域和每个所述目标区域对应的目标位置信息, N 为正整数, 每个所述目标区域包含一个待跟踪目标;

根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离, 得到 N 个距离;

从所述 N 个距离中选取数值最小的距离作为目标距离, 获取所述目标距离对应的目标位置信息, 并将获取到的目标位置信息对应的目标区域中的待跟踪目标确定为所述原始目标点对应的跟踪目标。

12. 如权利要求 11 所述的计算机设备, 其特征在于, 所述目标区域对应的目标位置信息包括目标区域的左上角坐标、所述目标区域的宽和高, 所述根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离, 得到 N 个距离包括:

针对每个所述目标位置信息, 按照如下中心点坐标公式, 计算第 i 个所述目标位置信息对应的目标区域的中心点坐标 (x_{i_0}, y_{i_0}) :

$$(x_{i_0}, y_{i_0}) = \left(x_i + \frac{w_i}{2}, y_i + \frac{h_i}{2} \right)$$

其中, i 为正整数, 且 $0 < i \leq N$, x_i 为第 i 个所述目标位置信息对应的所述目标区域的左上角坐标的横坐标, y_i 为所述左上角坐标的纵坐标, w_i 为所述第 i 个所述目标位置信息对应的目标区域的宽和 h_i 为所述第 i 个所述目标位置信息对应的目标区域的高;

针对每个所述中心点坐标,计算第 i 个所述中心点坐标 (x_{i_0}, y_{i_0}) 与所述原始坐标 $(\Delta x, \Delta y)$ 之间的欧氏距离 d_i :

$$d_i = \sqrt{(\Delta x - x_{i_0})^2 + (\Delta y - y_{i_0})^2}。$$

13. 如权利要求 11 所述的计算机设备,其特征在于,所述根据预设的目标预测模型,对所述当前帧图像进行目标预测,得到目标预测结果之前,所述跟踪目标的定位方法还包括:

获取样本训练图像数据;

基于深度学习框架构建卷积神经网络目标检测算法的结构;

在所述卷积神经网络目标检测算法的结构中,使用所述样本训练图像数据对所述卷积神经网络目标检测算法的结构进行训练,得到所述目标预测模型。

14. 如权利要求 13 所述的计算机设备,其特征在于,所述基于深度学习框架构建卷积神经网络目标检测算法的结构包括:

获取所述卷积神经网络目标检测算法的结构的定义信息;

在所述深度学习框架中按照预设的网络定义的层结构,将所述定义信息填入对应的所述卷积层和所述全连接层的层结构中,生成所述卷积神经网络目标检测算法的结构。

15. 如权利要求 11 所述的计算机设备,其特征在于,在所述根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离,得到 N 个距离之后,所述跟踪目标的定位方法还包括:

将所述 N 个距离中小于预设阈值的所述距离的目标位置信息对应的目标区域中的待跟踪目标反馈给用户进行确认;

接收所述用户返回的确认结果,并根据所述确认结果,将所述用户确认的跟踪目标作为所述原始目标点对应的跟踪目标。

16. 一个或多个存储有计算机可读指令的非易失性可读存储介质,其特征在于,所述计算机可读指令被一个或多个处理器执行时,使得所述一个或多个处理器执行如下步骤:

若接收到跟踪目标的定位请求,则获取所述定位请求中包含的用户选定的原始目标点的原始位置信息,其中,所述原始位置信息包括所述原始目标点在当前播放的视频数据中的当前帧图像和所述原始目标点在所述当前帧图像中的原始坐标;

根据预设的目标预测模型,对所述当前帧图像进行目标预测,得到目标预测结果,其中,所述目标预测模型为包括两层卷积层和两层全连接层的卷积神经网络模型,所述目标预测结果包括 N 个目标区域和每个所述目标区域对应的目标位置信息, N 为正整数,每个所述目标区域包含一个待跟踪目标;

根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离,得到 N 个距离;

从所述 N 个距离中选取数值最小的距离作为目标距离,获取所述目标距离对应的目标位置信息,并将获取到的目标位置信息对应的目标区域中的待跟踪目标确定为所述原始目标点对应的跟踪目标。

17. 如权利要求 16 所述的非易失性可读存储介质,其特征在于,所述目标区域对应的目标位置信息包括目标区域的左上角坐标、所述目标区域的宽和高,所述根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离,得到 N 个距离包括:

针对每个所述目标位置信息,按照如下中心点坐标公式,计算第 i 个所述目标位置信息对应的目标区域的中心点坐标 (x_{i_0}, y_{i_0}) :

$$(x_{i_0}, y_{i_0}) = \left(x_i + \frac{w_i}{2}, y_i + \frac{h_i}{2} \right)$$

其中, i 为正整数, 且 $0 < i \leq N$, x_i 为第 i 个所述目标位置信息对应的所述目标区域的左上角坐标的横坐标, y_i 为所述左上角坐标的纵坐标, w_i 为所述第 i 个所述目标位置信息对应的目标区域的宽和 h_i 为所述第 i 个所述目标位置信息对应的目标区域的高;

针对每个所述中心点坐标, 计算第 i 个所述中心点坐标 (x_{i_0}, y_{i_0}) 与所述原始坐标 $(\Delta x, \Delta y)$ 之间的欧氏距离 d_i :

$$d_i = \sqrt{(\Delta x - x_{i_0})^2 + (\Delta y - y_{i_0})^2}。$$

18. 如权利要求 16 所述的非易失性可读存储介质, 其特征在于, 所述根据预设的目标预测模型, 对所述当前帧图像进行目标预测, 得到目标预测结果之前, 所述跟踪目标的定位方法还包括:

获取样本训练图像数据;

基于深度学习框架构建卷积神经网络目标检测算法的结构;

在所述卷积神经网络目标检测算法的结构中, 使用所述样本训练图像数据对所述卷积神经网络目标检测算法的结构进行训练, 得到所述目标预测模型。

19. 如权利要求 18 所述的非易失性可读存储介质, 其特征在于, 所述基于深度学习框架构建卷积神经网络目标检测算法的结构包括:

获取所述卷积神经网络目标检测算法的结构的定义信息;

在所述深度学习框架中按照预设的网络定义的层结构, 将所述定义信息填入对应的所述卷积层和所述全连接层的层结构中, 生成所述卷积神经网络目标检测算法的结构。

20. 如权利要求 16 所述的非易失性可读存储介质, 其特征在于, 在所述根据每个所述目标区域的所述目标位置信息和所述原始坐标计算每个待跟踪目标和所述原始目标点之间的欧氏距离, 得到 N 个距离之后, 所述跟踪目标的定位方法还包括:

将所述 N 个距离中小于预设阈值的所述距离的目标位置信息对应的目标区域中的待跟踪目标反馈给用户进行确认;

接收所述用户返回的确认结果, 并根据所述确认结果, 将所述用户确认的跟踪目标作为所述原始目标点对应的跟踪目标。

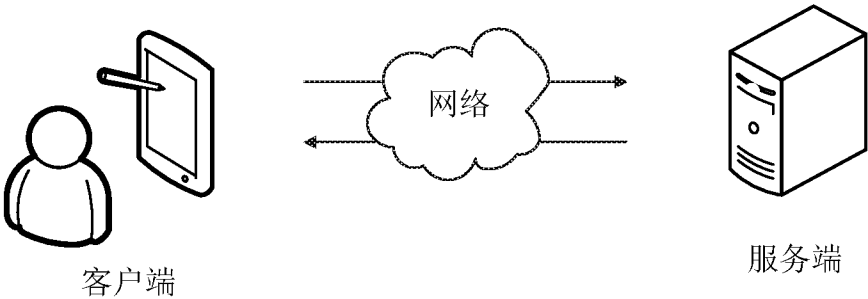


图 1

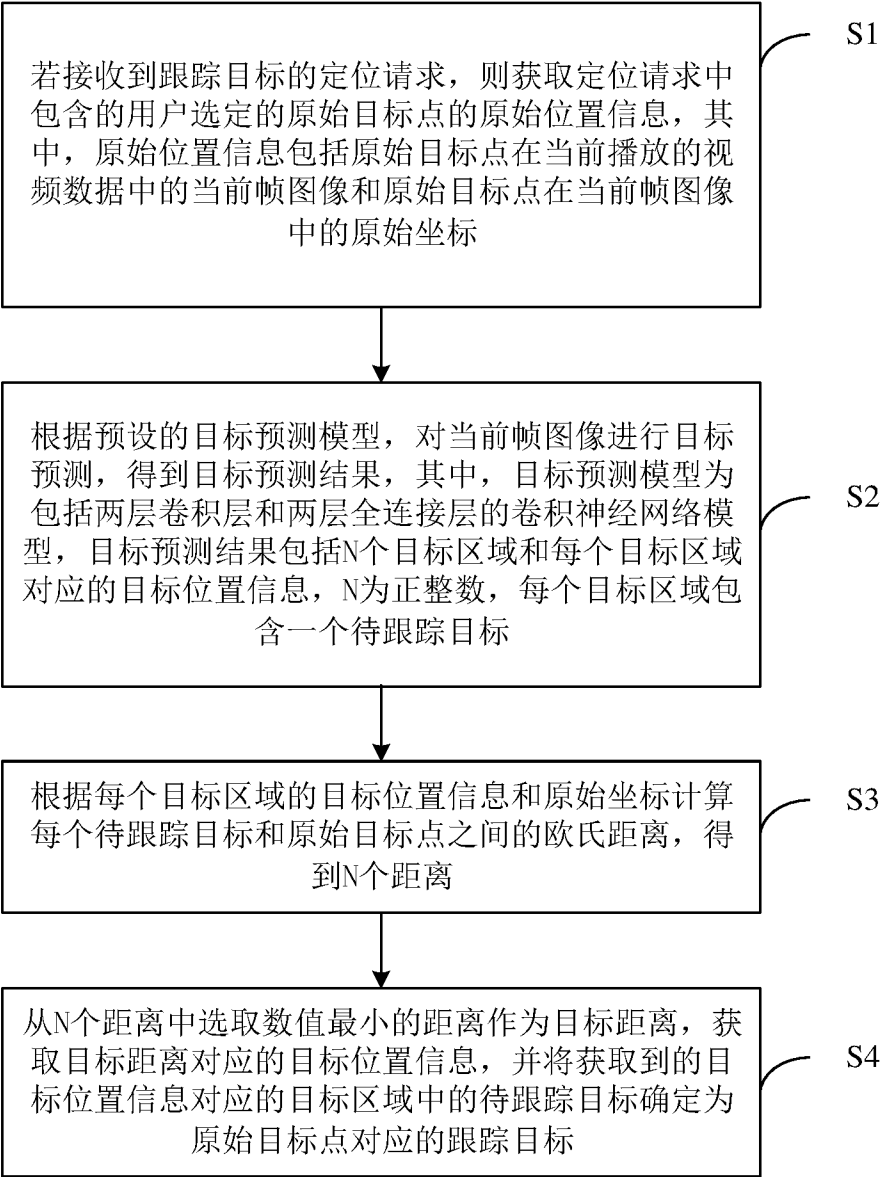


图 2

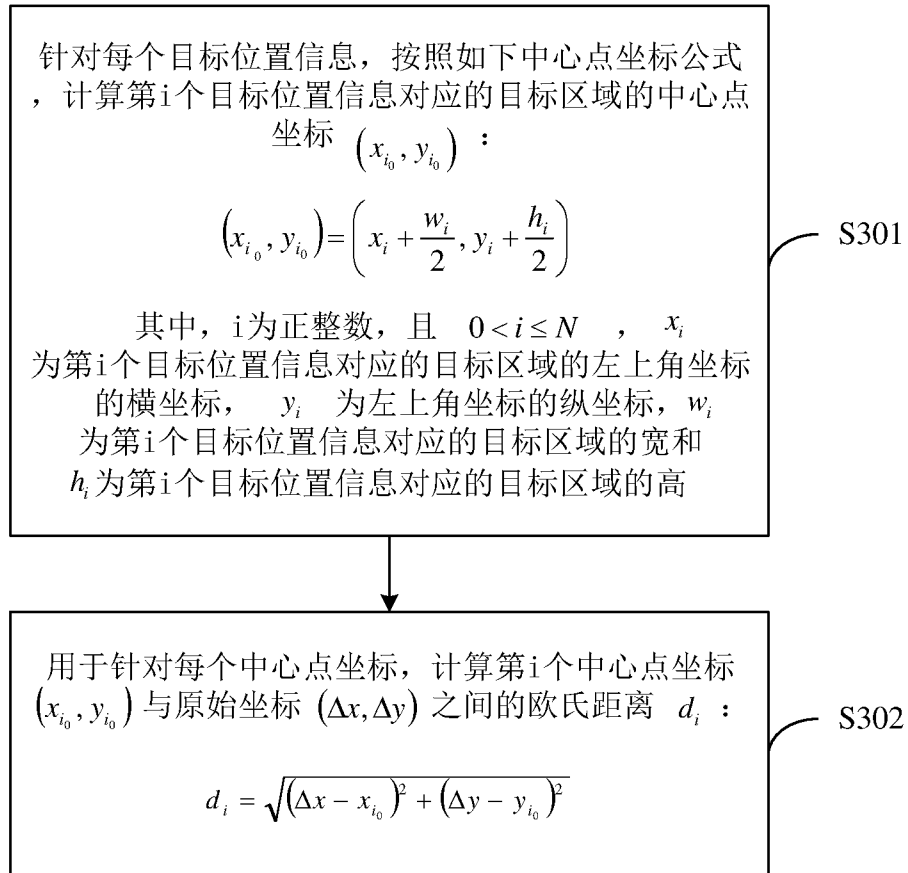


图 3

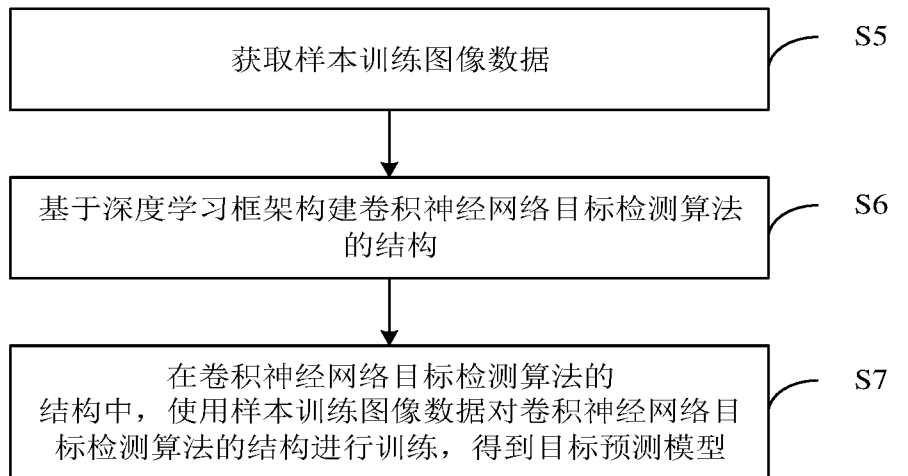


图 4

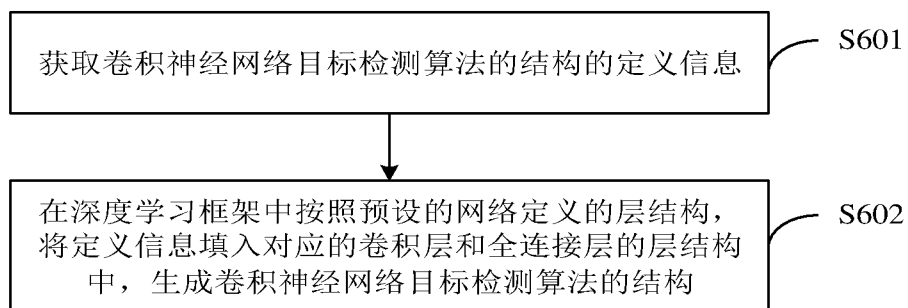


图 5

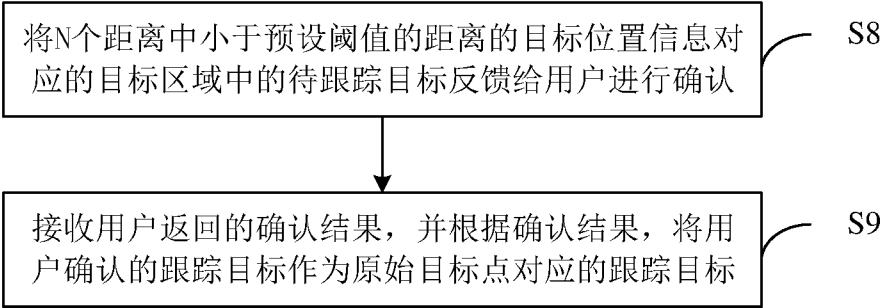


图 6

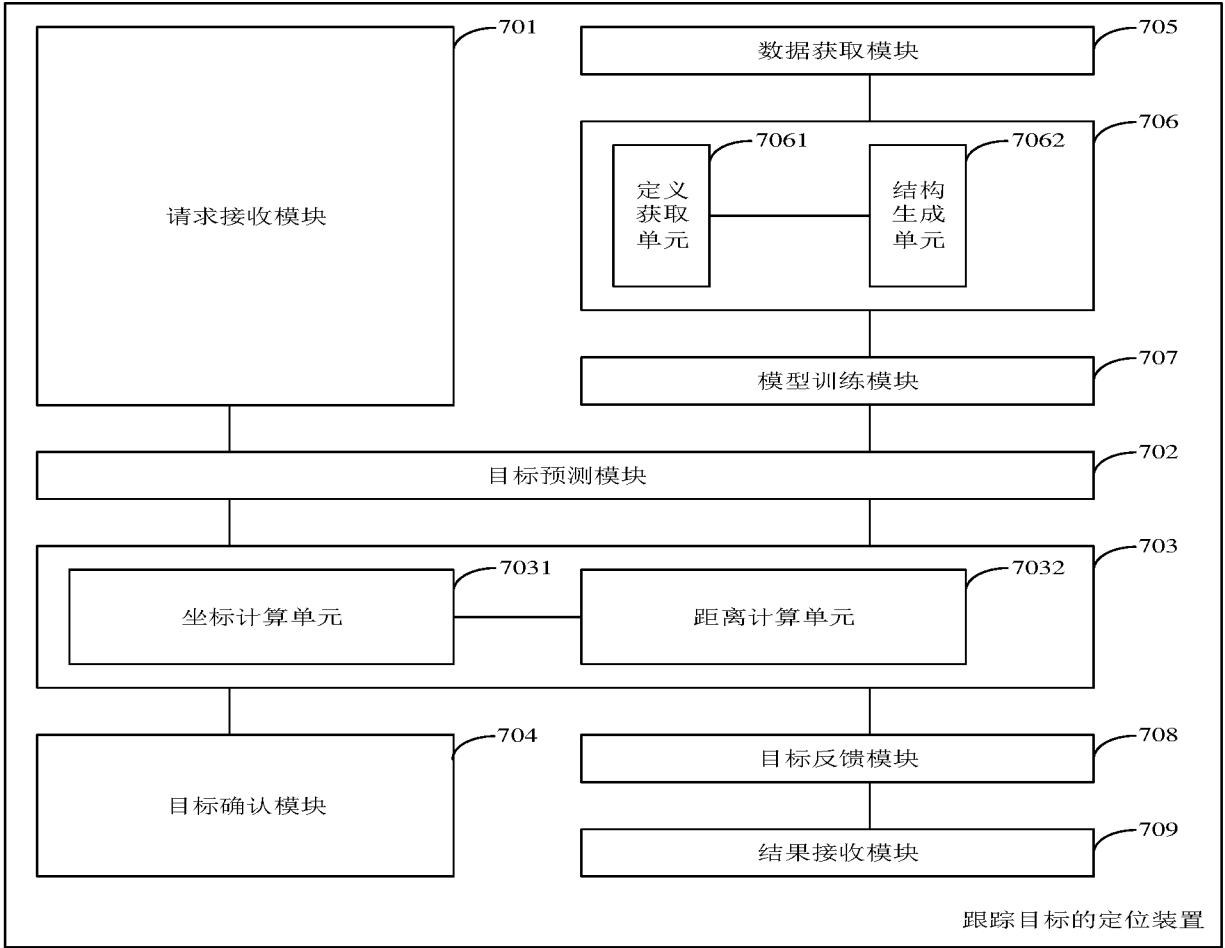


图 7

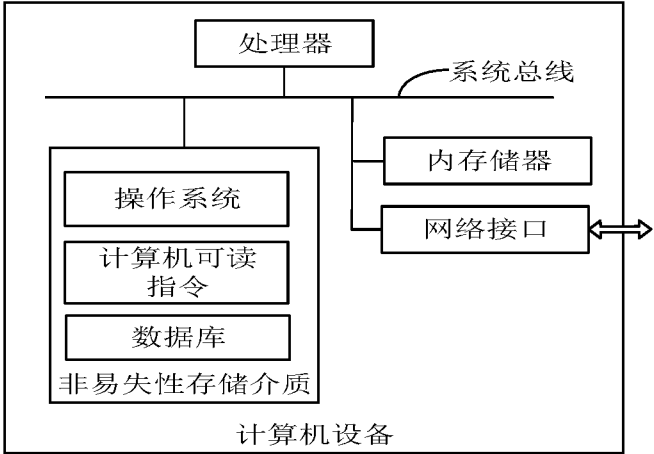


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/123080

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; G06K; G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, SIPOABS, CNTXT, CNKI, DWPI: 目标, 跟踪, 原始, 帧, 视频, 预测, 卷积, 欧式距离, target, tracking, original, frame, video, prediction, convolution, Euclidean distance

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 107679455 A (PING'AN TECHNOLOGY (SHENZHEN) CO., LTD.) 09 February 2018 (2018-02-09) entire document	1-20
A	CN 108230353 A (BEIJING SENSETIME SCIENCE TECHNOLOGY LTD.) 29 June 2018 (2018-06-29) entire document	1-20
A	CN 108268823 A (NINEBOT (BEIJING) TECH CO., LTD.) 10 July 2018 (2018-07-10) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

09 August 2019

Date of mailing of the international search report

16 August 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/123080

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107679455	A	09 February 2018	WO	2019041519	A1	07 March 2019
CN	108230353	A	29 June 2018	WO	2018157735	A1	07 September 2018
CN	108268823	A	10 July 2018	WO	2018121287	A1	05 July 2018

国际检索报告

国际申请号

PCT/CN2018/123080

A. 主题的分类

G06K 9/00 (2006.01) i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F; G06K; G06T

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS, SIPOABS, CNTXT, CNKI, DWPI: 目标, 跟踪, 原始, 帧, 视频, 预测, 卷积, 欧式距离, target, tracking, original, frame, video, prediction, convolution, Euclidean distance

C. 相关文件

类 型*

引用文件, 必要时, 指明相关段落

相关的权利要求

A

CN 107679455 A (平安科技深圳有限公司) 2018年 2月 9日 (2018 - 02 - 09)
全文

1-20

A

CN 108230353 A (北京市商汤科技开发有限公司) 2018年 6月 29日 (2018 - 06 - 29)
全文

1-20

A

CN 108268823 A (纳恩博北京科技有限公司) 2018年 7月 10日 (2018 - 07 - 10)
全文

1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 8月 9日

国际检索报告邮寄日期

2019年 8月 16日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

徐薇

传真号 (86-10)62019451

电话号码 86-(010)-62411668

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/123080

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	107679455	A	2018年 2月 9日	W0	2019041519	A1	2019年 3月 7日
CN	108230353	A	2018年 6月 29日	W0	2018157735	A1	2018年 9月 7日
CN	108268823	A	2018年 7月 10日	W0	2018121287	A1	2018年 7月 5日