

目标跟踪方法和装置、电子设备、存储介质

本申请要求在 2018 年 3 月 6 日提交中国专利局、申请号为 CN201810182797.X、发明名称为“目标跟踪方法和装置、电子设备、程序、存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本公开涉及计算机视觉技术，尤其是一种目标跟踪方法和装置、电子设备、存储介质。

背景技术

目标跟踪问题，通常指在给定视频序列第一帧目标物体位置信息的前提下，对视频中物体后续的位置做出预测；目标跟踪技术在视频监控、无人驾驶、人机交互等应用场景中有广泛的应用。由于单帧的检测技术和分割技术耗时相对较高，目标跟踪技术常用来与检测分割技术一起使用，来提高信息处理的速度，保证更好的实时性。

发明内容

本公开实施例提供一种目标跟踪技术。

根据本公开实施例的一个方面，提供一种目标跟踪方法，包括：

获取目标图像的多个参考图像的特征；

基于所述多个参考图像的特征，确定所述目标图像中跟踪目标的多个初始预测位置；

基于所述多个初始预测位置，确定所述目标图像中跟踪目标的最终位置。

可选地，所述基于所述多个参考图像的特征，确定所述目标图像中跟踪目标的多个初始预测位置，包括：

对所述目标图像进行特征提取，得到所述目标图像的多个特征；

基于所述目标图像的多个特征与所述多个参考图像的特征，得到所述目标图像中跟踪目标的多个初始预测位置。

可选地，基于所述目标图像的多个特征与所述多个参考图像的特征，得到所述目标图像中跟踪目标的多个初始预测位置，包括：

将所述目标图像的多个特征分别与所述多个参考图像的特征中的对应特征进行连接，得到多个连接特征；

基于所述多个连接特征，得到所述目标图像中跟踪目标的多个初始预测位置。

可选地，对所述目标图像进行特征提取，得到所述目标图像的多个特征，包括：

对所述目标图像执行卷积操作，得到中间特征；

对所述中间特征执行多个感兴趣区域对齐操作，得到所述目标图像的多个特征。

可选地，所述对所述中间特征执行多个感兴趣区域对齐操作，得到所述目标图像的多个特征，包括：

基于所述多个参考图像中至少一个参考图像中的跟踪目标的位置，对所述中间特征执行感兴趣区域对齐操作，得到所述目标图像的特征。

可选地，所述基于所述多个参考图像中至少一个参考图像中的跟踪目标的位置，对所述中间特征执行感兴趣区域对齐操作，得到所述目标图像的特征，包括：

将所述多个参考图像中至少一个参考图像中的跟踪目标对应的限位框在所述参考图像中的位置作为感兴趣区域进行感兴趣区域对齐操作，得到所述目标图像的特征。

可选地，将所述多个参考图像中至少一个参考图像中的跟踪目标对应的限位框在所述参考图像中的位置作为感兴趣区域进行感兴趣区域对齐操作，得到所述目标图像的特征之前，还包括：

对所述多个参考图像中至少一个参考图像中的跟踪目标的限位框放大第一预设倍数，得到所述至少一个参考图像中的跟踪目标对应的限位框。

可选地, 在所述对所述目标图像执行卷积操作, 得到中间特征之前, 还包括:

对所述目标图像进行裁剪处理, 获得裁剪图像;

所述对所述目标图像执行卷积操作, 得到中间特征, 包括:

对所述裁剪图像执行卷积操作, 得到中间特征。

可选地, 所述对所述目标图像进行裁剪处理, 获得裁剪图像, 包括:

基于至少一个第一图像中跟踪目标的位置信息, 对所述目标图像进行裁剪处理, 得到裁剪图像, 其中, 所述至少一个第一图像位于所述目标图像之前。

可选地, 所述至少一个第一图像包括所述目标图像的至少一个所述参考图像。

可选地, 所述基于至少一个第一图像中跟踪目标的位置信息, 对所述目标图像进行裁剪处理, 得到裁剪图像, 包括:

基于所述至少一个第一图像中跟踪目标的位置信息, 确定所述裁剪图像的中心点;

基于所述裁剪图像的中心点对所述目标图像进行裁剪处理, 得到所述裁剪图像。

可选地, 所述基于所述裁剪图像的中心点对所述目标图像进行裁剪处理, 得到所述裁剪图像, 包括:

将所述目标图像放大第二预设倍数, 得到放大处理后的所述目标图像;

基于所述裁剪图像的中心点, 对所述放大处理后的所述目标图像进行裁剪处理, 得到所述裁剪图像。

可选地, 所述基于所述多个初始预测位置, 确定所述目标图像中跟踪目标的最终位置, 包括:

分别确定所述多个初始预测位置中至少一个初始预测位置与所述跟踪目标的外观参考位置之间的相似度, 获得相似度大于第一预设阈值的至少一个初始预测位置;

基于所述相似度大于第一预设阈值的至少一个初始预测位置, 确定所述目标图像对应的最终位置。

可选地, 基于所述相似度大于预设阈值的至少一个初始预测位置, 确定所述目标图像对应的最终位置, 包括:

对所述相似度大于第一预设阈值的至少一个初始预测位置进行加权平均, 获得所述目标图像对应的最终位置。

可选地, 所述分别确定所述多个初始预测位置与所述跟踪目标的外观参考位置之间的相似度, 包括:

基于所述初始预测位置对应的限位框与所述跟踪目标的外观参考位置对应的限位框之间的交并比, 确定至少一个所述初始预测位置与所述跟踪目标的外观参考位置之间的相似度。

可选地, 还包括: 将所述确定的目标图像中跟踪目标的最终位置存入第一缓存区, 其中, 所述第一缓存区用于存储所述跟踪目标的外观参考位置。

可选地, 所述基于所述多个初始预测位置, 确定所述目标图像中跟踪目标的最终位置, 包括:

确定所述多个初始预测位置中至少一个初始预测位置的置信度;

基于所述置信度大于第二预设阈值的至少一个所述初始预测位置, 确定所述目标图像中跟踪目标的最终位置。

可选地, 基于所述置信度大于第二预设阈值的至少一个所述初始预测位置, 确定所述目标图像中跟踪目标的最终位置, 包括:

对所述置信度大于第二预设阈值的至少一个所述初始预测位置进行加权平均, 获得所述目标图像中跟踪目标的最终位置。

可选地, 所述确定所述多个初始预测位置中至少一个初始预测位置的置信度, 包括:

利用置信度打分网络分别对所述至少一个初始预测位置进行处理, 得到所述至少一个初始预测位置对应的置信度。

可选地, 所述利用置信度打分网络分别对所述至少一个初始预测位置进行处理, 得到所述至少一个初始预测位置对应的置信度之前, 还包括:

利用所述置信度打分网络对样本图像的至少一个初始预测位置进行处理, 得到对应至少一个所述初始预测位置的预测置信度;

基于所述样本图像的标注位置, 确定所述至少一个初始预测位置的真实置信度;

基于所述至少一个初始预测位置的预测置信度和真实置信度, 训练所述置信度打分网络。

可选地,所述获取目标图像的多个参考图像的特征,包括:

从第二缓存区获取目标图像的多个参考图像的特征。

可选地,所述目标图像与多个所述参考图像之间间隔至少一个第二图像;或者

所述多个参考图像包括相邻的至少两个图像帧;或者

所述多个参考图像中的任意两个参考图像不相邻。

可选地,在基于所述多个初始预测位置,确定所述目标图像中跟踪目标的最终位置之后,还包括:

基于所述目标图像中跟踪目标的最终位置和所述多个参考图像中跟踪目标的位置,确定是否将所述目标图像添加至参考图像列表。

可选地,所述基于所述目标图像中跟踪目标的最终位置和所述多个参考图像中跟踪目标的位置,确定是否将所述目标图像作为新添加的参考图像,包括:

基于所述目标图像中跟踪目标的限位框与所述多个参考图像中跟踪目标的限位框之间的交并比,确定是否将所述目标图像添加至参考图像列表。

可选地,所述基于所述目标图像帧中跟踪目标的限位框与所述多个参考图像中跟踪目标的各个限位框之间的交并比,确定是否将所述目标图像添加至参考图像列表,包括:

分别确定目标图像中跟踪目标的限位框与所述多个参考图像的跟踪目标的限位框之间的交并比;

基于确定的多个交并比,获得平均交并比;

响应于所述平均交并比大于或等于交并比阈值,将所述目标图像添加至参考图像列表。

可选地,在基于所述多个初始预测位置,确定所述目标图像中跟踪目标的最终位置之后,还包括:

若所述目标图像与所述多个参考图像中距离所述目标图像最近的参考图像之间间隔预设个数的图像帧,将所述目标图像添加至参考图像列表。

可选地,还包括:

响应于确定将所述目标图像添加至参考图像列表,在所述第二缓存区缓存所述目标图像的特征。

根据本公开实施例的另一个方面,提供一种目标跟踪装置,包括:

参考获取单元,用于获取目标图像的多个参考图像的特征;

初始预测单元,用于基于所述多个参考图像的特征,确定所述目标图像中跟踪目标的多个初始预测位置;

最终位置单元,用于基于所述多个初始预测位置,确定所述目标图像中跟踪目标的最终位置。

可选地,所述初始预测单元,包括:

特征提取模块,用于对所述目标图像进行特征提取,得到所述目标图像的多个特征;

位置预测模块,用于基于所述目标图像的多个特征与所述多个参考图像的特征,得到所述目标图像中跟踪目标的多个初始预测位置。

可选地,所述位置预测模块,用于将所述目标图像的多个特征分别与所述多个参考图像的特征中的对应特征进行一对一的连接,得到多个连接特征;基于所述多个连接特征,得到所述目标图像中跟踪目标的多个初始预测位置。

可选地,所述特征提取模块,包括:

卷积模块,用于对所述目标图像执行卷积操作,得到中间特征;

对齐模块,用于对所述中间特征执行多个感兴趣区域对齐操作,得到所述目标图像的多个特征。

可选地,所述对齐模块,用于基于所述多个参考图像中至少一个参考图像中的跟踪目标的位置,对所述中间特征执行感兴趣区域对齐操作,得到所述目标图像的特征。

可选地,所述对齐模块,用于将所述多个参考图像中至少一个参考图像中的跟踪目标对应的限位框在所述参考图像中的位置作为感兴趣区域进行感兴趣区域对齐操作,得到所述目标图像的特征。

可选地,所述对齐模块,还用于对所述多个参考图像中至少一个参考图像中的跟踪目标的限位框放大第一预设倍数,得到所述至少一个参考图像中的跟踪目标对应的限位框。

可选地,所述特征提取模块,还包括:

裁剪模块,用于对所述目标图像进行裁剪处理,获得裁剪图像;

所述卷积模块,用于对所述裁剪图像执行卷积操作,得到中间特征。

可选地,所述裁剪模块,用于基于至少一个第一图像中跟踪目标的位置信息,对所述目标图像进行裁剪处理,得到裁剪图像,其中,所述至少一个第一图像位于所述目标图像之前。

可选地,所述至少一个第一图像包括所述目标图像的至少一个所述参考图像。

可选地,所述裁剪模块,包括:

中心确定模块,用于基于所述至少一个第一图像中跟踪目标的位置信息,确定所述裁剪图像的中心点;

裁剪处理模块,用于基于所述裁剪图像的中心点对所述目标图像进行裁剪处理,得到所述裁剪图像。

可选地,所述裁剪处理模块,用于将所述目标图像放大第二预设倍数,得到放大处理后的所述目标图像;基于所述裁剪图像的中心点,对所述放大处理后的所述目标图像进行裁剪处理,得到所述裁剪图像。

可选地,所述最终位置单元,包括:

相似度确定模块,用于分别确定所述多个初始预测位置中至少一个初始预测位置与所述跟踪目标的外观参考位置之间的相似度,获得相似度大于第一预设阈值的至少一个初始预测位置;

第一位置计算模块,用于基于所述相似度大于第一预设阈值的至少一个初始预测位置,确定所述目标图像对应的最终位置。

可选地,所述第一位置计算模块,用于对所述相似度大于第一预设阈值的至少一个初始预测位置进行加权平均,获得所述目标图像对应的最终位置。

可选地,所述相似度确定模块,用于基于所述初始预测位置对应的限位框与所述跟踪目标的外观参考位置对应的限位框之间的交并比,确定至少一个所述初始预测位置与所述跟踪目标的外观参考位置之间的相似度。

可选地,所述最终位置单元,还包括:

位置存储模块,用于将所述确定的目标图像中跟踪目标的最终位置存入第一缓存区,其中,所述第一缓存区用于存储所述跟踪目标的外观参考位置。

可选地,所述最终位置单元,包括:

置信度确定模块,用于确定所述多个初始预测位置中至少一个初始预测位置的置信度;

第二位置计算模块,用于基于所述置信度大于第二预设阈值的至少一个所述初始预测位置,确定所述目标图像中跟踪目标的最终位置。

可选地,所述置信度确定模块,用于对所述置信度大于第二预设阈值的至少一个所述初始预测位置进行加权平均,获得所述目标图像中跟踪目标的最终位置。

可选地,所述置信度确定模块,用于利用置信度打分网络分别对所述至少一个初始预测位置进行处理,得到所述至少一个初始预测位置对应的置信度。

可选地,所述置信度确定模块,还用于:

利用所述置信度打分网络对样本图像的至少一个初始预测位置进行处理,得到对应至少一个所述初始预测位置的预测置信度;

基于所述样本图像的标注位置,确定所述至少一个初始预测位置的真实置信度;

基于所述至少一个初始预测位置的预测置信度和真实置信度,训练所述置信度打分网络。

可选地,所述参考获取单元,用于从第二缓存区获取目标图像的多个参考图像的特征。

可选地,所述目标图像与多个所述参考图像之间间隔至少一个第二图像;或者

所述多个参考图像包括相邻的至少两个图像帧;或者

所述多个参考图像中的任意两个参考图像不相邻。

可选地,还包括:

参考图像确定单元,用于基于所述目标图像中跟踪目标的最终位置和所述多个参考图像中跟踪目标的位置,确定是否将所述目标图像添加至参考图像列表。

可选地,所述参考图像确定单元,用于基于所述目标图像中跟踪目标的限位框与所述多个参考图像

中跟踪目标的限位框之间的交并比，确定是否将所述目标图像添加至参考图像列表。

可选地，所述参考图像确定单元，包括：

交并比确定模块，用于分别确定目标图像中跟踪目标的限位框与所述多个参考图像的跟踪目标的限位框之间的交并比；并基于确定的多个交并比，获得平均交并比；

确定添加模块，用于响应于所述平均交并比大于或等于交并比阈值，将所述目标图像添加至参考图像列表。

可选地，还包括：

参考列表更新单元，用于若所述目标图像与所述多个参考图像中距离所述目标图像最近的参考图像之间间隔预设个数的图像帧，将所述目标图像添加至参考图像列表。

可选地，还包括：

特征缓存单元，用于响应于确定将所述目标图像添加至参考图像列表，在所述第二缓存区缓存所述目标图像的特征。

根据本公开实施例的另一个方面，提供一种电子设备，包括处理器，所述处理器包括如上所述的目标跟踪装置。

根据本公开实施例的另一个方面，提供一种电子设备，包括：存储器，用于存储可执行指令；

以及处理器，用于与所述存储器通信以执行所述可执行指令从而完成如上所述目标跟踪方法。

根据本公开实施例的另一个方面，提供一种计算机存储介质，用于存储计算机可读取的指令，所述指令被处理器执行时，所述处理器执行如上所述目标跟踪方法。

根据本公开实施例的另一个方面，提供一种计算机程序，包括计算机可读代码，当所述计算机可读代码在设备上运行时，所述设备中的处理器执行如上所述目标跟踪方法。

根据本公开实施例的再一个方面，提供一种计算机程序产品，用于存储计算机可读指令，所述指令被执行时使得计算机执行上述任一可能的实现方式中所述的目标跟踪方法。

在一个可选实施方式中，所述计算机程序产品具体为计算机存储介质，在另一个可选实施方式中，所述计算机程序产品具体为软件产品，例如 SDK 等。

根据本公开实施例还提供了另一种目标跟踪方法及其对应的装置和电子设备、计算机存储介质、计算机程序以及计算机程序产品，其中，该方法包括：获取目标图像的多个参考图像的特征；基于所述多个参考图像的特征，确定所述目标图像中跟踪目标的多个初始预测位置；基于所述多个初始预测位置，确定所述目标图像中跟踪目标的最终位置。

基于本公开上述实施例提供了一种目标跟踪方法和装置、电子设备、存储介质，获取目标图像的多个参考图像的特征；基于多个参考图像的特征，确定目标图像中跟踪目标的多个初始预测位置，基于多个初始预测位置，确定目标图像中跟踪目标的最终位置，能够避免目标跟踪对于参考图像的依赖，从而提高目标跟踪的鲁棒性。

下面通过附图和实施例，对本公开的技术方案做进一步的详细描述。

附图说明

构成说明书的一部分的附图描述了本公开的实施例，并且连同描述一起用于解释本公开的原理。

参照附图，根据下面的详细描述，可以更加清楚地理解本公开，其中：

图 1 为本公开实施例提供的目标跟踪方法的流程图。

图 2 为本公开实施例提供的目标跟踪方法一个可选示例的流程图。

图 3 示出了图 2 所示的回归网络中目标帧分支得到目标帧的 k 个 proposal 的一个可选示例的流程图。

图 4 为本公开实施例提供的目标跟踪装置的结构示意图。

图 5 为适于用来实现本公开实施例的终端设备或服务器的电子设备的结构示意图。

具体实施方式

现在将参照附图来详细描述本公开的各种示例性实施例。应注意到：除非另外具体说明，否则在这

些实施例中阐述的部件和步骤的相对布置、数字表达式和数值不限制本公开的范围。

同时，应当明白，为了便于描述，附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

以下对至少一个示例性实施例的描述实际上仅仅是说明性的，决不作为对本公开及其应用或使用的任何限制。

对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论，但在适当情况下，所述技术、方法和设备应当被视为说明书的一部分。

应注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步讨论。

本公开实施例可以应用于计算机系统/服务器，其可与众多其它通用或专用计算系统环境或配置一起操作。适于与计算机系统/服务器一起使用的众所周知的计算系统、环境和/或配置的例子包括但不限于：个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境，等等。

计算机系统/服务器可以在由计算机系统执行的计算机系统可执行指令（诸如程序模块）的一般语境下描述。通常，程序模块可以包括例程、程序、目标程序、组件、逻辑、数据结构等等，它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施，分布式云计算环境中，任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中，程序模块可以位于包括存储设备的本地或远程计算系统存储介质上。

图1为本公开实施例提供的目标跟踪方法的流程图。该方法可以由任意电子设备执行，例如终端设备、服务器、移动设备等等。

步骤110，获取目标图像的多个参考图像的特征。

在一个可选示例中，该步骤110可以由处理器调用存储器存储的相应指令执行，也可以由参考获取单元41执行。

为了对跟踪目标实现跟踪，可选地，通过视频录制等手段采集该跟踪目标的多帧图像，通过确定这些图像中跟踪目标的位置，结合获取图像的时间，即可确定该跟踪目标在时间上的运动轨迹，实现对跟踪目标的跟踪。

在一些实施例中，为了实现对跟踪目标进行跟踪，需要已知的一个或多个图像中跟踪目标的位置，并基于该位置确定当前需要跟踪的跟踪目标；或者，已知跟踪目标为某一特定目标（例如：人脸或人体等），针对已知跟踪目标采集多个图像帧。

可选地，可以基于一个或多个图像中的跟踪目标的已确定的位置，实现对跟踪目标的跟踪，本公开实施例的跟踪目标可以是已知或未知的特定物体或人或其他类型，本公开实施例对此不做限定。

在本公开实施例中，采用包括多个参考图像的参考图像列表。其中，多个参考图像中的任意两个参考图像之间可以相邻或不相邻。例如，该多个参考图像可以为视频流中相邻的K个图像帧；或者，该多个参考图像中的部分相邻参考图像可以在视频流中相邻，另一部分相邻参考图像可以在视频流中不相邻；或者，该多个参考图像中的任意两个相邻参考图像可以在视频流中互不相邻，可选地，多个参考图像之间可以间隔一个或多个图像。例如，多个参考图像中的相邻参考图像之间可以间隔固定数量的图像帧，例如N个图像帧，其中，N为大于或等于2的预设数值。N的数值可以根据应用场景和需求确定，本公开实施例对N的数值不作限定。或者，该多个参考图像中的不同的相邻图像之间也可以间隔不同个数的图像，本公开实施例对此不做限定。

应理解，本文中的“多个参考图像中的相邻图像或相邻参考图像”可以指按照一定排序规则对该多个参考图像进行排序之后相邻的参考图像，其中，该排序规则可以为按照采集时间或检测时间排序，等等，本文中的“视频流中的相邻图像”可以指视频流或视频流片段中的多个图像进行排序之后相邻的图像，该排序可以指视频流中的原本顺序或某种特定排序规则，本公开实施例对排序的具体实现不做限定。

在本公开实施例中，可选地，目标图像与其多个参考图像可以相邻。或者，目标图像与其参考图像

之间可以间隔至少一个图像帧（称为第二图像）。可选地，目标图像可以与其多个参考图像中的任意参考图像之间间隔一个或多个图像；或者，目标图像可以与多个参考图像中的部分参考图像相邻，与多个参考图像中的另一部分参考图像间隔一个或多个图像，本公开实施例对此不做限定。与采用目标图像的前一帧图像做参考图像的方式相比，本公开实施例提供的技术方案可以提升预测精度，这是由于较早的图像的预测更加准确，用较早的图像来做跟踪的参考图像能有效解决误差累积的问题。

在本公开实施例中，可选地，不同图像可以具有不同的参考图像，或者，多个图像可以复用相同的参考图像，例如，目标图像与至少一个第三图像可以具有相同的参考图像，例如，目标图像的多个参考图像与第三图像的多个参考图像完全相同，或者，目标图像的多个参考图像中的部分参考图像也是第三图像的参考图像，本公开实施例对此不做限定。其中，可选地，该至少一个第三图像可以与该目标图像相邻，但本公开实施例不限于此。此时，可选地，可以将参考图像的特征存储至缓存区，在进行图像的跟踪目标的位置预测时，仅需要调用存储的参考图像的特征，使得位置预测的速度得到提升。

在本公开实施例中，可以通过多种方式获取参考图像的特征。在一些实施例中，可以从其他装置处获取参考图像的特征，例如，服务器接收终端设备发送的参考图像的特征，再例如，电子设备从数据库中获取参考图像的特征，其中，该数据库可以设置在电子设备本地或其他设备处。在另一些实施例中，可以在获取到参考图像之后，对参考图像进行特征提取处理，得到参考图像的特征。

在一个或多个可选的实施例中，可以获取电子设备的存储器存储的多个参考图像的特征，例如，从第二缓存区获取目标图像的多个参考图像的特征。可选地，可以将参考图像的特征存储在第二缓存区中，并从第二缓存区获取参考图像的特征，其中，该第二缓存区也可以称为位置模板缓存区（Location Exemplar Buffer），但本公开实施例对第二缓存区的名称不作限定。

在本公开实施例中，第二缓存区可以用于存储参考图像的信息，可选地，可以存储当前作为参考图像的多个图像的信息，即参考图像列表当前包括的多个参考图像的信息，例如，可以存储参考图像的特征，或者还可以进一步存储参考图像中跟踪目标的位置信息，本公开实施例对此不做限定。

在基于深度回归网络实现的目标跟踪方法中，深度回归网络由参考图像分支和目标图像分支组成，参考图像分支利用参考图像特征提取器可以获得有价值的位置模板（location exemplar）信息，指导目标图像分支对目标物体的位置进行预测。可选地，一般采用利用第 $t-1$ 帧预测第 t 帧的策略。然而，事实上，在预测第 t 帧目标物体位置的时候，第 $t-2$ 帧前面的所有图像已经经过了参考图像分支的特征提取器，有效的位置模板已经获得。因此可以进行位置模板的复用，从而提高跟踪系统的速度。

步骤 120，基于多个参考图像的特征，确定目标图像中跟踪目标的多个初始预测位置。

在一个可选示例中，该步骤 120 可以由处理器调用存储器存储的相应指令执行，也可以由初始预测单元 42 执行。

在一些实施例中，可以基于多个参考图像中每个参考图像的特征，得到目标图像中跟踪目标的一个初始预测位置，相应地，多个初始预测位置与多个参考图像一一对应。在另一些实施例中，可以基于多个参考图像中至少两个参考图像的特征，得到目标图像中跟踪目标的一个初始预测位置，相应地，多个初始预测位置的数量与多个参考图像的数量相等或不等，本公开实施例对此不做限定。

在一个或多个可选实施例中，可以包括：

对目标图像进行特征提取，得到目标图像的多个特征；

基于目标图像的多个特征与多个参考图像的特征，得到目标图像中跟踪目标的多个初始预测位置。

该目标图像的多个特征的数量与多个参考图像的数量可以相等或者不等，本公开实施例对此不做限定。

应理解，本公开实施例中的术语“特征”可以指特征向量、特征矩阵、特征张量或特征图，本公开实施例对此不做限定。

在一些实施例中，可以对目标图像的多个特征和多个参考图像的特征进行加权平均或其他类型的一种或多种处理，得到处理结果，并基于处理结果得到目标图像中跟踪目标的多个初始预测位置。

在一些可选实施例中，可以连接或拼接目标图像的多个特征与多个参考图像的特征，得到多个连接特征，并基于多个连接特征或拼接特征，得到目标图像中跟踪目标的多个初始预测位置。

其中,可选地,连接(或拼接)目标图像的多个特征与多个参考图像的特征,可以是目标图像的多个特征分别与多个参考图像的特征中的对应特征进行一对一的连接(或拼接),即将目标图像的多个特征与多个参考图像的特征一一对应地连接(或拼接),以获得多个连接特征或拼接特征。

在一种可选的实现中,这里的连接(或拼接)可以指通道叠加,例如,第一图像帧通过3个通道输出3个特征图,参考图像通过3个通道输出3个特征图,经过特征连接可以得到6个特征图。

在本公开实施例中,基于连接特征得到跟踪目标的位置的方式可以有多种。在一些可选实施例中,可以利用全连接层和/或其他网络层的处理,得到跟踪目标的初始预测位置,本公开实施例对此不做限定。

在本公开实施例中,可以利用多种方式对目标图像进行特征提取,例如,可以利用感兴趣区域(region of interest, RoI)池化(pooling)进行特征提取,或者,也可以利用感兴趣区域(RoI)对齐(Align)进行特征提取,或者,也可以利用其他手段进行特征提取,得到目标图像的多个特征,本公开实施例对此不做限定。

ROI Pooling,顾名思义,是 Pooling 层的一种,而且是针对 RoIs 的 Pooling,他的特点是输入特征图尺寸不固定,但是输出特征图尺寸固定;而如果 ROI 大小为(7,6),而经过 ROI Pooling 池化后获得的大小为(6,6), (7,6)到(6,6)的转换必然带来了边缘某像素的损失。而感兴趣区域对齐(ROI Align)利用双线性插值,将(7,6)的 RoI 插值扩充到(12,12),此时再做(6,6)的 ROI Pooling,能够提高精度。

此外,在本公开实施例中,可选地,可以基于多个参考图像中至少一个参考图像中的跟踪目标的位置,对目标图像进行特征提取,得到目标图像的特征,其中,该至少一个参考图像可以为多个参考图像中的部分或全部图像。例如,可以基于 K 个参考图像中跟踪目标的位置,对目标图像进行特征提取,得到目标图像的 K 个特征, K 为大于或等于 1 的整数,但本公开实施例不限于此。

在一些可选的实施例中,对目标图像进行特征提取,得到目标图像的多个特征,包括:

对目标图像执行卷积操作,得到中间特征;

对中间特征执行多个感兴趣区域对齐操作,得到目标图像的多个特征。

可选地,中间特征可以为特征图或其他表现形式。假设中间特征为特征图,则可选地,可以通过在特征图中选择感兴趣区域(RoI)进行感兴趣区域(RoI)对齐操作,得到目标图像的特征。这样,通过选择多个感兴趣区域(RoI),可以得到目标图像的多个特征。

在本公开实施例中,可以通过多种方式在特征图中选择感兴趣区域(RoI)。可选地,可以基于多个参考图像中至少一个参考图像中的跟踪目标的位置作为感兴趣区域,对中间特征执行感兴趣区域(RoI)对齐操作,得到目标图像的特征。在一些实施例中,可以基于参考图像中跟踪目标的位置,确定感兴趣区域(RoI)。例如,可以将参考图像中跟踪目标的限位框(bounding box)在参考图像中的位置作为感兴趣区域(RoI),或者可以将通过对参考图像中跟踪目标对应的限位框进行平移、放大等一种或多种操作,得到感兴趣区域(RoI),本公开实施例对此不做限定。

在一些实施例中,将多个参考图像中至少一个参考图像中的跟踪目标对应的限位框在参考图像中的位置作为感兴趣区域进行感兴趣区域对齐操作,得到目标图像的特征。其中,跟踪目标对应的限位框可以是跟踪目标的限位框本身,也可以是通过跟踪目标的限位框进行处理得到的,例如,通过对参考图像中跟踪目标的限位框放大第一预设倍数,得到参考图像跟踪目标对应的限位框。

可选地,该第一预设倍数可以根据应用场景或实际需求设定,例如该第一预设倍数等于 2.0,但本公开实施例对该第一预设倍数的数值不作限定。

在确定参考图像中跟踪目标对应的限位框在参考图像中的位置之后,可以将目标图像中处于该位置处的区域作为感兴趣区域。

由于跟踪目标在参考图像到目标图像会存在位移,如果仍以参考图像中的限位框得到目标图像的特征,可能会错过跟踪目标的准确位置;因此,将参考图像中跟踪目标的限位框放大第一预设倍数,以放大后的限位框在参考图像中的位置作为感兴趣区域(RoI)进行感兴趣区域(RoI)对齐操作,能够提高预测准确度。

在本公开实施例中,可以直接对目标图像进行特征提取,例如,对目标图像进行卷积操作。或者,

在对目标图像进行一种或多种预处理后进行特征提取，其中，该预处理包括裁剪、尺寸调整、缩放、旋转、亮度调整等一种或多种，本公开实施例对此不做限定。

在一些可选的实施例中，在对目标图像执行卷积操作，得到中间特征之前，还包括：对目标图像进行裁剪处理，获得裁剪图像。相应地，可以对裁剪图像进行卷积操作，得到中间特征。

可选地，可以基于预设尺寸对目标图像进行裁剪处理，或者，可以基于视频流中的其他图像，对目标图像进行裁剪处理。例如，可以基于至少一个第一图像中跟踪目标的位置信息，对目标图像进行裁剪处理，得到裁剪图像，其中，至少一个第一图像位于目标图像之前。可选地，该至少一个第一图像可以包括与目标图像相邻的至少一个图像；或者，该至少一个第一图像包括目标图像的至少一个参考图像，本公开实施例对此不做限定。

在一些实施例中，跟踪目标在裁剪图像中所占的比例达到特定数值，但本公开实施例对此不做限定。

可选地，可以通过神经网络中的卷积层对裁剪图像执行卷积操作，得到中间特征，本公开实施例对此不做限定。

可选地，在一些实施例中，基于至少一个第一图像中跟踪目标的位置信息，对目标图像进行裁剪处理，得到裁剪图像，包括：

基于至少一个第一图像中跟踪目标的位置信息，确定裁剪图像的中心点；

基于裁剪图像的中心点对目标图像进行裁剪处理，得到裁剪图像。

对目标图像进行裁剪是为了将跟踪目标从目标图像中分离出来，使得到的裁剪图像中跟踪目标所占比例大于一定比例，从而有利于跟踪目标的位置更准确被确定。

可选地，可以基于至少一个第一图像中跟踪目标对应的限位框的中心点的位置，确定裁剪图像的中心点的位置。例如，至少一个第一图像的数量为多个，相应地，将多个第一图像中跟踪目标对应的限位框的中心点的位置的平均值作为裁剪图像的中心点的位置，再例如，至少一个第一图像的个数为1个，相应地，将第一图像中跟踪目标对应的限位框的中心点的位置作为裁剪图像的中心点的位置，或者将第一图像中跟踪目标对应的限位框的中心点的位置进行平移等一种或多种处理，得到裁剪图像的中心点的位置，但本公开实施例不限于此。

可选地，裁剪的过程可以包括两个部分，确定中心点以及确定裁剪图像的长度和宽度；其中，在一个可选例子中，中心点可以基于较为接近目标图像（例如，最靠近目标图像）的参考图像或目标图像的前一帧图像中跟踪目标的位置中心确定，而裁剪图像的长度及宽度可以基于参考图像或前一帧图像中跟踪目标对应的限位框确定。在另一个可选例子中，可以将至少一个第一图像中跟踪目标对应的限位框中心点进行平均处理，得到裁剪图像的中心点，但本公开实施例不限于此。

在一些实施例中，在进行裁剪之前，还可以将目标图像放大第二预设倍数，得到放大处理后的目标图像。相应地，可以基于裁剪图像的中心点，对放大处理后的目标图像进行裁剪处理，得到裁剪图像。

第一预设倍数和第二预设倍数可以相等，也可以不相等。第二预设倍数可以根据实际应用场景或需求确定，例如，第二预设倍数等于3.0，但本公开实施例不限于此。

步骤130，基于多个初始预测位置，确定目标图像中跟踪目标的最终位置。

在一个可选示例中，该步骤130可以由处理器调用存储器存储的相应指令执行，也可以由最终位置单元43执行。

可选地，初始预测位置和/或最终位置可以指跟踪目标对应的限位框的位置，例如人脸框、人体框或特定物体或部位的位置，等等，本公开实施例对跟踪目标的位置的具体形式不做限定。

在一些实施例中，可以对多个初始预测位置进行平均处理，例如，数学平均或加权平均等，得到跟踪目标在目标图像中的最终位置。在另一些实施例中，可以从多个初始预测位置中选取至少一个初始预测位置，并基于至少一个初始预测位置，得到跟踪目标在目标图像中的最终位置。

在一个或多个可选的实施例中，基于多个初始预测位置的置信度，确定目标图像中跟踪目标的最终位置。

可选地，基于多个初始预测位置的置信度，从多个初始预测位置中选取至少一个初始预测位置，并基于至少一个初始预测位置，确定最终位置。例如，从多个初始预测位置中选取置信度较高的至少一个

初始预测位置。其中，该至少一个初始预测位置可以是多个初始预测位置中置信度最高的预设数量个位置，也可以是多个初始预测位置中置信度大于第二预设阈值的位置，本公开实施例对此不做限定。

置信度表示该预测位置对应于该目标图像中跟踪目标的真实位置的概率，在一些实施例中，置信度是通过打分网络获得的，该打分网络为经过训练的神经网络，例如：利用一个二分类网络，当预测位置与真实位置完成重合时，输出真实值为 1，预测值为 0，当预测位置与真实位置不重合时，输出真实值为 0，预测值为 1，等等，本公开实施例对打分网络的具体实现不做限定。

在一些实施例中，将选取的至少一个初始预测位置的平均值作为最终位置，其中，该平均值可以为算数平均值、几何平均值或加权平均值，例如，对选取的至少一个初始预测位置进行加权平均，获得目标图像中跟踪目标的最终位置，其中，初始预测位置的权重值与其置信度相关，置信度越大，权重值越大，置信度越小，权重值越小，但本公开实施例不限于此。

通过置信度可以将得到的初始预测位置中偏离较大的初始预测位置删除，相应地，得到的最终位置更接近该目标图像中的跟踪目标的真实位置，能实现更好的跟踪。

可选地，确定多个初始预测位置中至少一个初始预测位置的置信度，包括：

利用置信度打分网络分别对多个初始预测位置进行处理，得到多个初始预测位置对应的置信度。

其中，置信度打分网络是经过训练获得的，对应输入的初始预测位置训练后的置信度打分网络可输出对应的置信度，可选地，可基于预测置信度和真实置信度获得误差，根据误差利用反向梯度传播方法对打分网络进行训练，训练后的打分网络能更好的对预测位置为真实位置（已知目标位置）的概率进行预测；该打分网络针对需要跟踪的特定物体训练，例如：需要对人脸进行跟踪，就采用人脸图像对该打分网络进行训练。

可选地，在利用置信度打分网络分别对多个初始预测位置进行处理，得到多个初始预测位置对应的置信度之前，还包括：

利用置信度打分网络对样本图像的至少一个初始预测位置进行处理，得到对应至少一个初始预测位置的预测置信度；

基于样本图像的标注位置，确定至少一个初始预测位置的真实置信度；

基于至少一个初始预测位置的预测置信度和真实置信度，训练置信度打分网络。

为了提高置信度网络打分的准确性，需要对置信度打分网络进行训练，训练过程与神经网络的训练过程相同，都是基于已知真实置信度的样本图像对置信度打分网络进行训练，可选地，针对不同的跟踪目标，可采用针对性的训练，例如：对人脸进行跟踪，在训练置信度打分网络过程中，采用已知真实置信度的人脸图像对置信度打分网络进行训练，使得到的置信度打分网络对人脸图像的置信度打分更准确。

基于本公开上述实施例提供的目标跟踪方法，获取目标图像的多个参考图像的特征；基于多个参考图像的特征，确定目标图像中跟踪目标的多个初始预测位置，基于多个初始预测位置，确定目标图像中跟踪目标的最终位置，能够避免目标跟踪对于参考图像的依赖，从而提高目标跟踪的鲁棒性。

在一个或多个可选的实施例中，基于多个初始预测位置与跟踪目标的外观参考位置之间的相似度，确定目标图像中跟踪目标的最终位置。

可选地，基于多个初始预测位置与跟踪目标的外观参考位置之间的相似度，从多个初始预测位置中选取至少一个初始预测位置，例如，从多个初始预测位置中选取与跟踪目标的外观参考位置之间的相似度较高的至少一个初始预测位置，其中，该至少一个初始预测位置可以是多个初始预测位置中与外观参考位置之间的相似度大于第一预设阈值的位置，或者是多个初始预测位置中与外观参考位置之间的相似度最高的预设数量个位置，本公开实施例对此不作限定。

可选地，可以从其他装置获取该跟踪目标的外观参考位置。或者，可以设置用于存储跟踪目标的外观参考位置的第一缓存区，该第一缓存区也可以称为外观模板缓存区（Appearance Exemplar Buffer, AEB）或其他名称，本公开实施例对该第一缓存区的名称不作限定。此时，可以从第一缓存区获取该跟踪目标的外观参考位置。或者也可以通过其他方式得到该跟踪目标的外观参考位置，本公开实施例对得到该跟踪目标的外观参考位置的方式不作限定。

其中，该外观参考位置可以是预设的，或者是跟踪目标在该视频流中的一个或多个图像帧的位置，

或者是依据跟踪目标在该视频流中的一个或多个图像帧的位置进行处理得到的,例如,跟踪目标的外观参考位置是基于位于该目标图像之前的至少一个图像中的跟踪目标的最终位置得到的,在一个例子中,跟踪目标的外观参考位置是位于该目标对象之前的某个参考图像中的跟踪目标的最终位置,例如与该目标图像距离最近的参考图像中的跟踪目标的最终位置,或者,跟踪目标的外观参考位置是通过对于该目标对象之前的至少一个参考图像中的跟踪目标的最终位置进行平均或合并等一种或多种处理得到的。该第一缓存区中缓存的信息可以是固定不变的,或者也可以依据固定间隔的图像帧或者其他触发条件进行更新,本公开实施例对此不做限定。经过相似度筛选获得与该目标图像相关度较高的预测位置,可以节省计算时间,提高跟踪速度的同时,保证了跟踪的准确度。

可选地,基于初始预测位置与外观参考位置之间的距离,确定目标图像中跟踪目标的最终位置。

可选地,可以基于初始预测位置对应的限位框与跟踪目标的外观参考位置对应的限位框之间的交并比,确定至少一个初始预测位置与跟踪目标的外观参考位置之间的相似度。

交并比(intersection-over-union)表示两个区域的交集与并集之比,在本公开实施例中,可以指初始预测位置对应的限位框与外观参考位置对应的限位框的交集和初始预测位置对应的限位框与外观参考位置对应的限位框的并集之比;交并比越大,说明两个限位框越相似,即相似度越大;交并比越小,两个限位框的相似度越小。

可选地,本公开实施例还可以采用其他方式确定初始预测位置与外观参考位置之间的相似度,本公开实施例对此不做限定。

可选地,可以将相似度大于第一预设阈值的至少一个初始预测位置进行加权平均,获得目标图像对应的最终位置。

其中,每个初始预测位置对应一个权重值,相似度越大的初始预测位置对应的权重值越大,反之,相似度越小的初始预测位置对应的权重值越小。

可选地,也可以直接对相似度大于第一预设阈值的至少一个初始预测位置进行平均,得到跟踪目标在目标图像中的位置,此时,该至少一个初始预测位置对应相同的权重。或者,也可以利用其他方式对该至少一个初始预测位置进行处理,得到该跟踪目标在目标图像中的位置,本公开实施例对此不做限定。

在一些可选的实施例,还包括:将确定的目标图像中跟踪目标的最终位置存入第一缓存区,其中,第一缓存区用于存储跟踪目标的外观参考位置。

可选地,在本公开实施例中,可以以覆盖或增加的方式将该跟踪目标在目标图像中的位置存入第一缓存区,本公开实施例对此不做限定。可选地,可以在任何条件下均存储该跟踪目标在目标图像中的位置。或者,在跟踪目标在目标图像中的位置满足一定预设条件时将其存储至第一缓存区,例如,目标图像与第一缓存区存储的外观参考位置对应的图像帧之间的间隔为预设数值,或者跟踪目标在目标图像中的位置满足预设条件,例如目标图像的与第一缓存区存储的外观参考位置之间的差异超过一定阈值,等等,本申请实施例对存储目标图像中跟踪目标的位置的条件不作限定。

在一个或多个可选的实施例中,在确定跟踪目标在目标图像中的最终位置之后,还可以确定是否将目标图像添加至参考图像列表,也就是说,确定是否利用目标图像更新参考图像列表。其中,该参考图像列表可以包括多个参考图像,该参考图像列表可以用于跟踪待处理的图像帧中的跟踪目标。如果一直用同样的参考图像的特征对目标图像进行位置预测,当跟踪目标移动位置较大时,将导致的位置不准确,因此,为了提高预测的准确性,可以对参考图像列表进行定期或不定期的更新,本公开实施例对此不做限定。

在本公开实施例中,可选地,可以基于目标图像中跟踪目标的最终位置和多个参考图像中跟踪目标的位置,确定是否将目标图像添加至参考图像列表,或者,也可以以固定的图像间隔(例如,间隔M个图像帧)更新参考图像列表,此时,可以依据目标图像与多个参考图像之间的间隔是否满足预设条件,确定是否将目标图像添加到参考图像列表,或者,以其他依据确定是否将目标图像添加到参考图像列表。

在实际的目标跟踪中,跟踪目标的运动复杂,其速度往往不是恒定不变的。对于第t+1帧的目标图像,可以通过对跟踪目标在第t帧中的位置和多个参考图像中的位置进行比对,来动态调整存储的参考图像列表,以获得更适合的参考图像的特征。

可选地，如果利用第 t 帧更新了参考图像列表，即将第 t 帧添加到参考图像列表，则第 $t+1$ 帧目标图像与多个参考图像中距离目标图像最近的参考图像之间可以相邻，而如果未将第 t 帧添加到参考图像列表，则参考图像列表并未进行更新，此时，第 $t+1$ 帧目标图像与多个参考图像中距离目标图像最近的参考图像之间可以间隔一个或多个图像。例如，目标图像与多个参考图像中距离目标图像最近的参考图像之间可以间隔 N 个图像帧，其中， N 为大于或等于 2 的预设数值。 N 的数值可以根据应用场景确定，本公开实施例对 N 的数值不作限定。

可选地，基于目标图像中跟踪目标的最终位置和多个参考图像中跟踪目标的位置，确定是否将目标图像添加到参考图像列表，包括：

基于目标图像中跟踪目标的限位框 (bounding box) 与多个参考图像中跟踪目标的限位框 (bounding box) 之间的交并比，确定是否将目标图像添加到参考图像列表。

可选地，分别确定目标图像中跟踪目标的限位框与多个参考图像的跟踪目标的限位框之间的交并比；基于确定的多个交并比，获得平均交并比；响应于平均交并比大于或等于交并比阈值，将目标图像添加到参考图像列表。

可选地，基于多个交并比获得平均交并比，通过对多个交并比求平均获得，或通过对多个交并比加权平均获得，其中，可选地，与目标图像距离越近的参考图像，其交并比对应的权重值越大。

在一些可选实施例中，如果最终位置与多个参考图像中距离目标图像最近的参考图像中的位置的距离较近 (两个限位框之间的交并比大于或等于设定阈值)，则不需要将该目标图像添加到参考图像列表，可以继续用当前的参考图像列表的信息来指导下一帧图像的位置预测。若最终位置与多个参考图像中距离目标图像最近的参考图像中的位置的距离较远 (两个限位框之间的交并比小于设定阈值)，则可以将目标图像添加到参考图像列表。通过这种动态调整机制，可以充分利用先前计算过的特征 (如：中间特征图)，从而减少了很多不必要的特征提取过程。

例如：若利用目标跟踪方法跟踪一个在一段视频片段内绝对位置基本不变的物体，可以不对参考图像进行更新，一直利用同一些参考图像，极大的提高了至少一帧视频图像中跟踪目标的位置速度，进而提高了跟踪效率。

在一个或多个可选的实施例中，若目标图像与多个参考图像中距离目标图像较近 (例如，最近) 的参考图像之间间隔预设个数的图像帧，确定将目标图像添加到参考图像列表。

对于特定的问题 (跟踪特定的目标，如：人脸跟踪)，在应用中可以采用针对性的参考图像更新策略对目标跟踪方法的参考图像进行调整，有着非常强的推广价值。

在一个或多个可选的实施例中，响应于确定将目标图像添加到参考图像列表，在第二缓存区缓存目标图像的特征。

可选地，第二缓存区存储的数据可以不断增加，相应地，参考图像列表包括的参考图像个数可以不断增大，这样，在需要进行位置预测时，可以从第二缓存中调用当前存储的所有参考图像中最新的固定数量的参考图像的特征。可选地，为了节省存储空间，在第二缓存区中可以保存预设数量的参考图像，当参考图像列表中增加一个新的参考图像，可以将存储时间较早 (例如，最早) 的一个或多个参考图像删除，相应地，该参考图像列表中包括的参考图像个数可以保持固定不变，在需要进行位置预测时，只需从缓存中调用当前存储的所有参考图像的特征即可。

图 2 为本公开实施例提供的目标跟踪方法一个可选示例的流程图。如图 2 所示，可以设置有跟踪缓冲区，包括两个部分：外观模板缓冲区 (第一缓存区) 和位置模板缓存区 (第二缓存区)，其中，外观模板缓存区用于缓存跟踪目标的参考位置信息，用于确定跟踪目标的位置之间的相似度；位置模板缓存区用于缓存 k 个参考图像中至少一个参考图像的特征图和 k 个参考图像中至少一个参考图像的跟踪目标的位置，用于进行多帧融合。

另一方面，整个目标跟踪网络可以包括回归网络和相似度判别网络，其中，回归网络用于进行特征提取，相似度判别网络可以用于进行相似度比较。

回归网络可以包括两个分支：参考图像分支和目标帧分支，其中，在参考图像分支，可以将参考图像进行 3 倍放大 (例如 padding)，得到放大之后的图像，并将该放大之后的图像输入到卷积神经网络

(convolutional neural network, CNN), 得到中间特征, 然后, 可以对中间特征进行感兴趣区域对齐 (ROI Align) 或感兴趣区域池化 (ROI Pooling) 处理, 得到参考图像对应的特征图, 基于该特征图, 可以得到跟踪目标在该参考图像中的位置。此外, 还可以将该参考图像对应的特征图和跟踪目标位置存入位置模板缓存区。

在目标帧分支, 可以对目标帧进行裁剪处理, 获得裁剪图像, 并将裁剪图像输入到卷积神经网络 (CNN) 进行处理, 得到中间特征; 然后, 基于位置模板缓存区缓存的 k 个参考图像中跟踪目标的位置, 对中间特征进行感兴趣区域 (RoI) 对齐操作, 得到目标图像的 k 个特征图; 然后, 对目标图像的 k 个特征图和位置模板缓存区缓存的 k 个参考图像的特征图进行拼接处理, 得到 k 个候选 (proposal), 也可以称为 k 个初始预测位置; 最后, 可以利用相似度判断网络对该 k 个候选框分别与外观模板缓存区缓存的外观参考位置分别进行相似度比较, 得到目标帧的最终预测结果。

图 3 示出了图 2 所示的回归网络中目标帧分支得到目标帧的 k 个候选 (proposal) 的一个可选示例的流程图。如图 3 所示, 与参考帧分支关联的位置模板缓存器中存储有 k 个限位框, 可以基于该 k 个限位框进行感兴趣区域对齐 (RoI Align) 处理, 得到多种具有不同尺度的特征图, 并通过多尺度池化进行加权相加, 得到 k 个特征图; 然后, 该 k 个特征图可以与位置模板缓存器中存储的 k 个特征图进行拼接处理, 得到拼接特征图, 然后将该拼接特征图经过一个或多个 (图中示例性地示出了 3 个) 全连接层的处理, 得到 k 个候选 (proposal)。

应理解, 图 2 和图 3 所示的例子仅是为了帮助更好地理解本公开实施例提供的技术方案, 而不应理解为对本公开实施例的限定。本公开实施例在图 2 和图 3 示出的例子的基础上可以进行各种变形和替换, 并且有些动作或流程可以省略, 本公开实施例对此不做限定。

本公开针对各类目标跟踪技术的应用场景, 提出了基于多帧信息的回归网络框架。本公开提供的目标跟踪方法主要提升了目标跟踪的速度与鲁棒性。一方面, 本公开的目标跟踪框架基于深度回归网络的特征复用技术, 利用更有效率的中级特征图管理思路, 增强了整个系统在深度回归网络前馈时的特征复用, 保证了跟踪方法的高效率。另一方面, 本公开提出了一种基于感兴趣区域的特征图及多帧复用和信息融合的策略, 解决了传统回归算法对于参考图像的依赖问题, 使得跟踪系统在测试中表现的更加鲁棒。综合来看, 本公开在保持高速的同时提高了目标跟踪方法的鲁棒性。

用户在使用手机、相机进行人脸及一般物体的检测时, 常见的思路是通过与目标跟踪算法结合起来提高速度, 降低设备功耗。本公开实施例在深度回归网络目标跟踪技术上使用特征缓存和特征复用机制, 可以大幅提升目标跟踪的速度, 可以在特定任务 (人脸跟踪) 上实现 1500fps 的高速率跟踪, 并且, 基于多个参考图像的特征进行跟踪可以使跟踪准确度得到提升。这在实际产品应用中非常关键, 大幅提高了传统算法的前向速度。

本公开的核心是提出高鲁棒性的实用高速目标跟踪算法, 本公开的算法是在深度回归网络特征复用技术的基础之上提出的, 传统的深度回归网络主要由参考图像分支和目标帧分支组成, 通过对两帧间目标物体位移的建模实现对目标帧目标物体位置的预测。本公开提出利用位置模板缓存来存储计算得到的中间特征 (参考图像的特征), 在之后的预测中复用之前缓存的参考图像的特征, 从而能够快速的利用多个参考图像的特征对跟踪目标的位置进行预测。在目标图像分支中, 本公开提出利用感兴趣区域对齐 (RoI Align) 来实现特征共享; 特征缓存和特征共享这两个机制使得快速的多帧融合预测成为可能, 可以实现快速鲁棒的目标跟踪。

本公开实施例的技术方案有较强的扩展性, 系统可以与多种多帧融合的策略相结合; 例如: 利用相似度衡量的方法, 或显示的在特定任务上通过评分的方法来做多帧的结果融合, 均可以达到较好的效果。

其他的目标跟踪算法利用前一帧作为参考图像来预测下一帧的位置的目标跟踪算法, 在行人跟踪中遇到柱子遮挡时, 参考图像遮挡会使得参考图像计算而得的位置模板出现漂移, 无法准确的给出预测。本公开实施例提出的目标跟踪方法通过利用多帧参考图像能够有效解决该问题。

本公开实施例的核心是提出基于特征缓存器的框架, 实现高精度的实用高速目标跟踪方法。本公开通过对回归网络结构的分析, 提出通过特征缓存对回归网络中间特征图的维护和复用机制。传统的深度回归网络主要由参考图像分支和目标帧分支组成, 通过对两帧间目标物体位移的建模实现对目标帧目标

物体位置的预测。本公开提出利用位置模板缓存来存储计算得到的中间特征（参考图像分支计算得到的位置模板），在之后的预测中复用之前缓存的位置模板，从而能够快速利用多个参考图像模板对物体进行预测。本公开提出的方法可以应用于各类目标跟踪场景中，对基于深度回归网络的目标跟踪算法提升运行速度到近原来的两倍，提高了跟踪实时性，降低了设备功耗。

本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

图4为本公开实施例提供的目标跟踪装置的结构示意图。该实施例的装置可用于实现本公开上述各方法实施例。如图4所示，该实施例的装置包括：

参考获取单元41，用于获取目标图像的多个参考图像的特征。

为了对跟踪目标实现跟踪，可选地，通过视频录制等手段采集该跟踪目标的多帧图像，通过确定这些图像中跟踪目标的位置，结合获取图像的时间，即可确定该跟踪目标在时间上的运动轨迹，实现对跟踪目标的跟踪。

在一些实施例中，为了实现对跟踪目标进行跟踪，需要已知的一个或多个图像中跟踪目标的位置，并基于该位置确定当前需要跟踪的跟踪目标；或者，已知跟踪目标为某一特定目标（例如：人脸或人体等），针对已知跟踪目标采集多个图像帧。

可选地，可以基于一个或多个图像中的跟踪目标的已确定的位置，实现对跟踪目标的跟踪，本公开实施例的跟踪目标可以已知或未知的特定物体或人或其他类型，本公开实施例对此不做限定。

在本公开实施例中，采用包括多个参考图像的参考图像列表。其中，多个参考图像中的任意两个参考图像之间可以相邻或不相邻。例如，该多个参考图像可以为视频流中相邻的K个图像帧；或者，该多个参考图像中的部分相邻参考图像可以在视频流中相邻，另一部分相邻参考图像可以在视频流中不相邻；或者，该多个参考图像中的任意两个相邻参考图像可以在视频流中互不相邻，可选地，多个参考图像之间可以间隔一个或多个图像。例如，多个参考图像中的相邻参考图像之间可以间隔固定数量的图像帧，例如N个图像帧，其中，N为大于或等于2的预设数值。N的数值可以根据应用场景和需求确定，本公开实施例对N的数值不作限定。或者，该多个参考图像中的不同的相邻图像之间也可以间隔不同个数的图像，本公开实施例对此不做限定。

可选地，目标图像与多个参考图像之间间隔至少一个图像帧（称为第二图像）；或者，多个参考图像包括相邻的至少两个图像帧；或者，多个参考图像中的任意两个参考图像不相邻。可选地，目标图像可以与其多个参考图像中的任意参考图像之间间隔一个或多个图像；或者，目标图像可以与多个参考图像中的部分参考图像相邻，与多个参考图像中的另一部分参考图像间隔一个或多个图像，本公开实施例对此不做限定。与采用目标图像的前一帧图像做参考图像的方式相比，本公开实施例提供的技术方案可以提升预测精度，这是由于较早的图像的预测更加准确，用较早的图像来做跟踪的参考图像能有效解决误差累积的问题。

在一个或多个可选的实施例中，参考获取单元41可以获取电子设备的存储器存储的多个参考图像的特征，例如，从第二缓存区获取目标图像的多个参考图像的特征。可选地，可以将参考图像的特征存储在第二缓存区中，并从第二缓存区获取参考图像的特征，其中，该第二缓存区也可以称为位置模板缓存区（Location Exemplar Buffer），但本公开实施例对第二缓存区的名称不作限定。

初始预测单元42，用于基于多个参考图像的特征，确定目标图像中跟踪目标的多个初始预测位置。

最终位置单元43，用于基于多个初始预测位置，确定目标图像中跟踪目标的最终位置。

基于本公开上述实施例提供的目标跟踪装置，获取目标图像的多个参考图像的特征；基于多个参考图像的特征，确定目标图像中跟踪目标的多个初始预测位置，基于多个初始预测位置，确定目标图像中跟踪目标的最终位置，能够避免目标跟踪对于参考图像的依赖，从而提高目标跟踪的鲁棒性。

在一个或多个可选实施例中，初始预测单元42，包括：

特征提取模块，用于对目标图像进行特征提取，得到目标图像的多个特征；

位置预测模块，用于基于目标图像的多个特征与多个参考图像的特征，得到目标图像中跟踪目标的

多个初始预测位置。

可选地，在一些可选实施例中，位置预测模块，可以用于连接或拼接目标图像的多个特征与多个参考图像的特征，得到多个连接特征；基于多个连接特征，得到目标图像中跟踪目标的多个初始预测位置。

其中，可选地，连接（或拼接）目标图像的多个特征与多个参考图像的特征，可以是分别将目标图像的多个特征分别与多个参考图像的特征中的对应特征进行一对一的连接（或拼接），即将目标图像的多个特征与多个参考图像的特征一一对应地连接（或拼接），以获得多个连接特征或拼接特征。

在一些可选的实施例中，特征提取模块，包括：

卷积模块，用于对目标图像执行卷积操作，得到中间特征；

对齐模块，用于对中间特征执行多个感兴趣区域对齐操作，得到目标图像的多个特征。

可选地，对齐模块，用于基于参考图像中跟踪目标的位置，对中间特征执行感兴趣区域（RoI）对齐操作，得到目标图像的特征。

在本公开实施例中，可以通过多种方式在特征图中选择感兴趣区域（RoI）。可选地，可以基于参考图像中跟踪目标的位置，对中间特征执行感兴趣区域（RoI）对齐操作，得到目标图像的特征。在一些实施例中，对齐模块，可以用于将参考图像中跟踪目标对应的限位框在参考图像中的位置作为感兴趣区域（RoI）进行感兴趣区域（RoI）对齐操作，得到目标图像的特征。例如，可以将参考图像中跟踪目标对应的限位框（bounding box）在参考图像中的位置作为感兴趣区域（RoI），或者可以将通过对参考图像中跟踪目标对应的限位框进行平移等一种或多种操作，得到感兴趣区域（RoI），本公开实施例对此不做限定。

可选地，对齐模块，还用于对参考图像中跟踪目标的限位框放大第一预设倍数，得到参考图像跟踪目标对应的限位框。

可选地，该第一预设倍数可以根据应用场景或实际需求设定，例如该第一预设倍数等于 2.0，但本公开实施例对该第一预设倍数的数值不作限定。

由于跟踪目标在参考图像到目标图像会存在位移，如果仍以参考图像中的限位框得到目标图像的特征，可能会错过跟踪目标的准确位置；因此，将参考图像中跟踪目标的限位框放大第一预设倍数，以放大后的限位框在参考图像中的位置作为感兴趣区域（RoI）进行感兴趣区域（RoI）对齐操作，能够提高预测准确度。

在一些可选的实施例中，特征提取模块，还包括：

裁剪模块，用于对目标图像进行裁剪处理，获得裁剪图像；

此时，卷积模块，用于对裁剪图像执行卷积操作，得到中间特征。

可选地，可以基于预设尺寸对目标图像进行裁剪处理，或者，可以基于视频流中的其他图像，对目标图像进行裁剪处理。例如，裁剪模块，可以用于基于至少一个第一图像中跟踪目标的位置信息，对目标图像进行裁剪处理，得到裁剪图像，其中，至少一个第一图像位于目标图像之前。可选地，该至少一个第一图像可以包括与目标图像相邻的至少一个图像；或者，该至少一个第一图像包括目标图像的至少一个参考图像，本公开实施例对此不做限定。

可选地，在一些实施例中，裁剪模块，包括：

中心确定模块，用于基于至少一个第一图像中跟踪目标的位置信息，确定裁剪图像的中心点；

裁剪处理模块，用于基于裁剪图像的中心点对目标图像进行裁剪处理，得到裁剪图像。

对目标图像进行裁剪是为了将跟踪目标从目标图像中分离出来，使得到的裁剪图像中跟踪目标所占比例大于一定比例，从而有利于跟踪目标的位置更准确被确定。

可选地，裁剪的过程可以包括两个部分，确定中心点以及确定裁剪图像的长度和宽度；其中，在一个可选例子中，中心点可以基于较为接近目标图像（例如，最靠近目标图像）的参考图像或目标图像的前一帧图像中跟踪目标的位置中心确定，而裁剪图像的长度及宽度可以基于参考图像或前一帧图像中跟踪目标对应的限位框确定。在另一个可选例子中，可以将至少一个第一图像中跟踪目标对应的限位框中心点进行平均处理，得到裁剪图像的中心点，但本公开实施例不限于此。

在一些实施例中，裁剪处理模块，还可以用于将目标图像放大第二预设倍数，得到放大处理后的目

标图像；相应地，可以基于裁剪图像的中心点，对放大处理后的目标图像进行裁剪处理，得到裁剪图像。

在一个或多个可选的实施例中，最终位置单元 43，包括：

相似度确定模块，用于分别确定多个初始预测位置中至少一个初始预测位置与跟踪目标的外观参考位置之间的相似度，获得相似度大于第一预设阈值的至少一个初始预测位置；

第一位置计算模块，用于基于相似度大于第一预设阈值的至少一个初始预测位置，确定目标图像对应的最终位置。

可选地，可以从其他装置获取该跟踪目标的外观参考位置。或者，可以设置用于存储跟踪目标的外观参考位置的第一缓存区，该第一缓存区也可以称为外观模板缓存区（Appearance Exemplar Buffer, AEB）或其他名称，本公开实施例对该第一缓存区的名称不作限定。此时，可以从第一缓存区获取该跟踪目标的外观参考位置。或者也可以通过其他方式得到该跟踪目标的外观参考位置，本公开实施例对得到该跟踪目标的外观参考位置的方式不作限定。

经过相似度筛选获得与该目标图像相关度较高的预测位置，可以节省计算时间，提供跟踪速度的同时，保证了跟踪的准确度。

可选地，第一位置计算模块，用于对相似度大于第一预设阈值的至少一个初始预测位置进行加权平均，获得目标图像对应的最终位置。

其中，每个初始预测位置对应一个权重值，相似度越大的初始预测位置对应的权重值越大，反之，相似度越小的初始预测位置对应的权重值越小。

可选，相似度确定模块，可以用于基于初始预测位置对应的限位框与跟踪目标的外观参考位置对应的限位框之间的交并比，确定至少一个初始预测位置与跟踪目标的外观参考位置之间的相似度。

在一些可选的实施例，最终位置单元 43，还包括：

位置存储模块，用于将确定的目标图像中跟踪目标的最终位置存入第一缓存区，其中，第一缓存区用于存储跟踪目标的外观参考位置。

可选地，在本公开实施例中，可以以覆盖或增加的方式将该跟踪目标在目标图像中的位置存入第一缓存区，本公开实施例对此不做限定。可选地，可以在任何条件下均存储该跟踪目标在目标图像中的位置。或者，在跟踪目标在目标图像中的位置满足一定预设条件时将其存储至第一缓存区，例如，目标图像与第一缓存区存储的外观参考位置对应的图像帧之间的间隔为预设数值，或者跟踪目标在目标图像中的位置满足预设条件，例如目标图像的与第一缓存区存储的外观参考位置之间的差异超过一定阈值，等等，本申请实施例对存储目标图像中跟踪目标的位置的条件不作限定。

在一个或多个可选的实施例中，最终位置单元 43，包括：

置信度确定模块，用于确定多个初始预测位置中至少一个初始预测位置的置信度；

第二位置计算模块，用于基于置信度大于第二预设阈值的至少一个初始预测位置，确定目标图像中跟踪目标的最终位置。

置信度表示该预测位置对应该目标图像中跟踪目标的真实位置的概率，这个置信度是通过打分网络获得的，该打分网络为经过训练的神经网络，例如：利用一个二分类网络，当预测位置与真实位置完成重合时，输出真实值为 1，预测值为 0，当预测位置与真实位置不重合时，输出真实值为 0，预测值为 1。该实施例主要针对一些特定物体的跟踪（如：人脸、人体等）。

可选地，置信度确定模块，用于对置信度大于第二预设阈值的至少一个初始预测位置进行加权平均，获得目标图像中跟踪目标的最终位置。

其中权重值与置信度相关，置信度越大权重值越大，置信度越小，权重值越小。

可选地，置信度确定模块，用于利用置信度打分网络分别对至少一个初始预测位置进行处理，得到至少一个初始预测位置对应的置信度。

其中，置信度打分网络是经过训练获得的，对应输入的初始预测位置训练后的置信度打分网络可输出对应的置信度，可选地，可基于预测置信度和真实置信度获得误差，根据误差利用反向梯度传播方法对打分网络进行训练，训练后的打分网络能更好的对预测位置为真实位置（已知目标位置）的概率进行预测；该打分网络针对需要跟踪的特定物体训练，例如：需要对人脸进行跟踪，就采用人脸图像对该打

分网络进行训练。

可选地，置信度确定模块，还用于：

利用置信度打分网络对样本图像的至少一个初始预测位置进行处理，得到对应至少一个初始预测位置的预测置信度；

基于样本图像的标注位置，确定至少一个初始预测位置的真实置信度；

基于至少一个初始预测位置的预测置信度和真实置信度，训练置信度打分网络。

为了提高置信度网络打分的准确性，需要对置信度打分网络进行训练，训练过程与神经网络的训练过程相同，都是基于已知真实置信度的样本图像对置信度打分网络进行训练，可选地，针对不同的跟踪目标，可采用针对性的训练，例如：对人脸进行跟踪，在训练置信度打分网络过程中，采用已知真实置信度的人脸图像对置信度打分网络进行训练，使得到的置信度打分网络对人脸图像的置信度打分更准确。

在一个或多个可选的实施例中，还包括：

参考图像确定单元，用于基于目标图像中跟踪目标的最终位置和多个参考图像中跟踪目标的位置，确定是否将目标图像添加至参考图像列表。

其中，该参考图像列表可以包括多个参考图像，该参考图像列表可以用于跟踪待处理的图像帧中的跟踪目标。如果一直用同样的参考图像的特征对目标图像进行位置预测，当跟踪目标移动位置较大时，将导致的位置不准确，因此，为了提高预测的准确性，可以对参考图像列表进行定期或不定期的更新，本公开实施例对此不做限定。

可选地，参考图像确定单元，可以用于基于目标图像中跟踪目标的限位框 **bounding box** 与所述多个参考图像中跟踪目标的限位框 **bounding box** 之间的交并比，确定是否将目标图像添加至参考图像列表。

可选地，参考图像确定单元，包括：

交并比确定模块，用于分别确定目标图像中跟踪目标的限位框与多个参考图像的跟踪目标的限位框之间的交并比；并基于确定的多个交并比，获得平均交并比；

确定添加模块，用于响应于平均交并比大于或等于交并比阈值，将目标图像添加至参考图像列表。

在一个或多个可选的实施例中，还包括：

参考列表更新单元，用于若目标图像与多个参考图像中距离目标图像较近（例如，最近）的参考图像之间间隔预设个数的图像帧，将目标图像添加至参考图像列表。

对于特定的问题（跟踪特定的目标，如：人脸跟踪），在应用中可以采用针对性的参考图像更新策略对目标跟踪方法的参考图像进行调整，有着非常强的推广价值。

在一个或多个可选的实施例中，还包括：特征缓存单元，用于响应于确定将目标图像添加至参考图像列表，在第二缓存区缓存目标图像的特征。

可选地，第二缓存区存储的数据可以不断增加，相应地，参考图像列表包括的参考图像个数可以不断增大，这样，在需要进行位置预测时，可以从第二缓存中调用当前存储的所有参考图像中最新的固定数量的参考图像的特征。可选地，为了节省存储空间，在第二缓存区中可以保存预设数量的参考图像，当参考图像列表中增加一个新的参考图像，可以将存储时间较早（例如，最早）的一个或多个参考图像删除，相应地，该参考图像列表中包括的参考图像个数可以保持固定不变，在需要进行位置预测时，只需从缓存中调用当前存储的所有参考图像的特征即可。

根据本公开实施例的另一个方面，提供一种电子设备，包括处理器，处理器包括本公开上述任一实施例的目标跟踪装置。

根据本公开实施例的另一个方面，提供一种电子设备，包括：存储器，用于存储可执行指令；

以及处理器，用于与存储器通信以执行可执行指令从而完成本公开提供的目标跟踪方法上述任一实施例。

根据本公开实施例的另一个方面，提供一种计算机存储介质，用于存储计算机可读取的指令，指令被处理器执行时，该处理器执行本公开提供的目标跟踪方法上述任一实施例。

根据本公开实施例的另一个方面，提供一种计算机程序，包括计算机可读代码，当计算机可读代码在设备上运行时，设备中的处理器执行本公开提供的目标跟踪方法。

根据本公开实施例的再一个方面，提供一种计算机程序产品，用于存储计算机可读指令，所述指令被执行时使得计算机执行上述任一可能的实现方式中所述的目标跟踪方法。

在一个或多个可选实施方式中，本公开实施例还提供了一种计算机程序产品，用于存储计算机可读指令，所述指令被执行时使得计算机执行上述任一实施例所述的目标跟踪方法。

该计算机程序产品可以具体通过硬件、软件或其结合的方式实现。在一个可选例子中，所述计算机程序产品具体体现为计算机存储介质，在另一个可选例子中，所述计算机程序产品具体体现为软件产品，例如软件开发包（Software Development Kit, SDK）等等。

根据本公开实施例还提供了另一种目标跟踪方法及其对应的装置和电子设备、计算机存储介质、计算机程序以及计算机程序产品，其中，该方法包括：获取目标图像的多个参考图像的特征；基于多个参考图像的特征，确定目标图像中跟踪目标的多个初始预测位置；基于多个初始预测位置，确定目标图像中跟踪目标的最终位置。

在一些实施例中，该目标跟踪指示可以具体为调用指令，第一装置可以通过调用的方式指示第二装置执行目标跟踪，相应地，响应于接收到调用指令，第二装置可以执行上述目标跟踪方法中的任意实施例中的步骤和/或流程。

应理解，本公开实施例中的“第一”、“第二”等术语仅仅是为了区分，而不应理解成对本公开实施例的限定。

还应理解，在本公开中，“多个”可以指两个或两个以上，“至少一个”可以指一个、两个或两个以上。

还应理解，对于本公开中提及的任一部件、数据或结构，在没有明确限定或者在前后文给出相反启示的情况下，一般可以理解为一个或多个。

还应理解，本公开对各个实施例的描述着重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，不再一一赘述。

本公开实施例还提供了一种电子设备，例如可以是移动终端、个人计算机（PC）、平板电脑、服务器等。下面参考图 5，其示出了适于用来实现本申请实施例的终端设备或服务器的电子设备 500 的结构示意图：如图 5 所示，计算机系统 500 包括一个或多个处理器、通信部等，所述一个或多个处理器例如：一个或多个中央处理单元（CPU）501，和/或一个或多个图像处理器（GPU）513 等，处理器可以根据存储在只读存储器（ROM）502 中的可执行指令或者从存储部分 508 加载到随机访问存储器（RAM）503 中的可执行指令而执行各种适当的动作和处理。通信部 512 可包括但不限于网卡，所述网卡可包括但不限于 IB（Infiniband）网卡。

处理器可与只读存储器 502 和/或随机访问存储器 530 中通信以执行可执行指令，通过总线 504 与通信部 512 相连、并经通信部 512 与其他目标设备通信，从而完成本申请实施例提供的任一项方法对应的操作，例如，获取目标图像的多个参考图像的特征；基于多个参考图像的特征，确定目标图像中跟踪目标的多个初始预测位置；基于多个初始预测位置，确定目标图像中跟踪目标的最终位置。

此外，在 RAM 503 中，还可存储有装置操作所需的各种程序和数据。CPU501、ROM502 以及 RAM503 通过总线 504 彼此相连。在有 RAM503 的情况下，ROM502 为可选模块。RAM503 存储可执行指令，或在运行时向 ROM502 中写入可执行指令，可执行指令使处理器 501 执行上述通信方法对应的操作。输入/输出（I/O）接口 505 也连接至总线 504。通信部 512 可以集成设置，也可以设置为具有多个子模块（例如多个 IB 网卡），并在总线链接上。

以下部件连接至 I/O 接口 505：包括键盘、鼠标等的输入部分 506；包括诸如阴极射线管（CRT）、液晶显示器（LCD）等以及扬声器等的输出部分 507；包括硬盘等的存储部分 508；以及包括诸如 LAN 卡、调制解调器等的网络接口卡的通信部分 509。通信部分 509 经由诸如因特网的网络执行通信处理。驱动器 510 也根据需要连接至 I/O 接口 505。可拆卸介质 511，诸如磁盘、光盘、磁光盘、半导体存储器等等，根据需要安装在驱动器 510 上，以便于从其上读出的计算机程序根据需要被安装入存储部分 508。

需要说明的，如图 5 所示的架构仅为一种可选实现方式，在实践过程中，可根据实际需要对上述图 5 的部件数量和类型进行选择、删减、增加或替换；在不同功能部件设置上，也可采用分离设置或集成

设置等实现方式，例如 GPU 和 CPU 可分离设置或者可将 GPU 集成在 CPU 上，通信部可分离设置，也可集成设置在 CPU 或 GPU 上，等等。这些可替换的实施方式均落入本公开公开的保护范围。

特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括有形地包含在机器可读介质上的计算机程序，计算机程序包含用于执行流程图所示的方法的程序代码，程序代码可包括对应执行本申请实施例提供的方法步骤对应的指令，例如，获取目标图像的多个参考图像的特征；基于多个参考图像的特征，确定目标图像中跟踪目标的多个初始预测位置；基于多个初始预测位置，确定目标图像中跟踪目标的最终位置。在这样的实施例中，该计算机程序可以通过通信部分 509 从网络上被下载和安装，和/或从可拆卸介质 511 被安装。在该计算机程序被中央处理单元（CPU）501 执行时，执行本申请的方法中限定的上述功能。

本说明书中各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之处，各个实施例之间相同或相似的部分相互参见即可。对于系统实施例而言，由于其与方法实施例基本对应，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

可能以许多方式来实现本公开的方法和装置。例如，可通过软件、硬件、固件或者软件、硬件、固件的任何组合来实现本公开的方法和装置。用于所述方法的步骤的上述顺序仅是为了进行说明，本公开的方法的步骤不限于以上具体描述的顺序，除非以其它方式特别说明。此外，在一些实施例中，还可将本公开实施为记录在记录介质中的程序，这些程序包括用于实现根据本公开的方法的机器可读指令。因而，本公开还覆盖存储用于执行根据本公开的方法的程序的记录介质。

本公开的描述是为了示例和描述起见而给出的，而并不是无遗漏的或者将本公开限于所公开的形式。很多修改和变化对于本领域的普通技术人员而言是显然的。选择和描述实施例是为了更好说明本公开的原理和实际应用，并且使本领域的普通技术人员能够理解本公开从而设计适于特定用途的带有各种修改的各种实施例。