

目标检测跟踪方法、装置、电子设备和存储介质

本申请要求在2020年08月06日提交中国专利局、申请号为202010784786.6的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本申请涉及视频处理技术领域，例如涉及一种目标检测跟踪方法、装置、电子设备和存储介质。

背景技术

随着移动终端硬件的容量的增加和性能的提升，视频和图像在移动终端上被人们广泛的生产、使用和传播，而视频中的目标检测和跟踪作为智能视频内容理解的一部分也受到极大的关注。

目标检测跟踪框架包括检测算法和跟踪算法，检测算法负责定位和识别视频中的目标物体，跟踪算法则负责对视频中的目标物体进行跟踪和平滑。通常在关键帧进行目标检测，在非关键帧（也称跟踪帧）可通过跟踪算法对检测出的目标物体进行跟踪。然而，在处理关键帧时，由于检测算法耗时长，导致关键帧对应的检测和跟踪的耗时峰值陡增。由于移动终端是在对当前关键帧执行的检测算法运行结束后才接收下一视频帧，进而容易造成计算资源有限的移动终端在接收关键帧后发生卡顿等问题，最终降低了用户的使用体验。

发明内容

本申请提供一种目标检测跟踪方法、装置、电子设备和存储介质，以解决相关技术中目标检测跟踪方法应用于移动终端时容易造成移动终端卡顿的问题。

本申请提供了一种目标检测跟踪方法，包括：

将从视频流接收到的每个视频帧输入跟踪网络中得到所述每个视频帧的目标跟踪结果；

在所述每个视频帧为关键帧的情况下，将所述关键帧输入检测网络并在接收指定数量的延迟帧期间控制所述检测网络运行，获得在接收完所述指定数量的延迟帧中的最后一个延迟帧时所述检测网络输出的目标检测结果，其中，所述关键帧为所述视频流中用于目标检测的视频帧，所述指定数量的延迟帧为所述关键帧后连续的视频帧；

根据所述目标检测结果和所述最后一个延迟帧的目标跟踪结果生成最终目

标跟踪结果。

本申请提供了一种目标检测跟踪装置，包括：

跟踪模块，设置为将从视频流接收到的每个视频帧输入跟踪网络中得到所述每个视频帧的目标跟踪结果；

检测模块，设置为在所述每个视频帧为关键帧的情况下，将所述关键帧输入检测网络并在接收指定数量的延迟帧期间控制所述检测网络运行，获得在接收完所述指定数量的延迟帧中的最后一个延迟帧时所述检测网络输出的目标检测结果，其中，所述关键帧为所述视频流中用于目标检测的视频帧，所述指定数量的延迟帧为所述关键帧后连续的视频帧；

检测跟踪结果生成模块，设置为根据所述目标检测结果和所述最后一个延迟帧的目标跟踪结果生成最终目标跟踪结果。

本申请提供了一种电子设备，所述电子设备包括：

一个或多个处理器；

存储装置，设置为存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现上述的目标检测跟踪方法。

本申请提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现上述的目标检测跟踪方法。

附图说明

图1是本申请实施例一提供的一种目标检测跟踪方法的步骤流程图；

图2是本申请实施例二提供的一种目标检测跟踪方法的步骤流程图；

图3A是本申请实施例三提供的一种目标检测跟踪方法的步骤流程图；

图3B是本申请实施例三提供的一种级联检测网络进行目标检测的示意图；

图4A是本申请实施例四提供的一种目标检测跟踪方法的步骤流程图；

图4B是本申请实施例四提供的一种非级联检测网络进行目标检测的示意图；

图5是本申请实施例五提供的一种目标检测跟踪装置的结构框图；

图6是本申请实施例六提供的一种电子设备的结构示意图。

具体实施方式

下面结合附图和实施例对本申请进行说明。

实施例一

图 1 是本申请实施例一提供的一种目标检测跟踪方法的步骤流程图，本申请实施例可适用于对视频中的目标进行检测跟踪的情况，该方法可以由本申请实施例的目标检测跟踪装置来执行，该目标检测跟踪装置可以由硬件或软件来实现，并集成在本申请实施例所提供的电子设备中，如图 1 所示，本申请实施例的目标检测跟踪方法可以包括如下步骤：

S101、将从视频流接收到的每个视频帧输入跟踪网络中得到所述每个视频帧的目标跟踪结果。

本申请实施例中，视频流包含多个视频帧，多个视频帧以一定帧率播放形成视频播放画面，当对视频中的目标进行跟踪时，从解码后的视频流中每接收到一个视频帧，可以将该视频帧输入预先训练好的跟踪网络中得到该视频帧的目标跟踪结果，其中，目标可以是视频中的人脸、车辆、人物、动物等对象，跟踪网络可以是对已检测到的目标、在多个视频帧进行跟踪的网络，目标跟踪结果可以是已跟踪目标列表、已跟踪目标列表中目标的位置、状态、速度、轨迹等，本申请实施例的跟踪网络可以是多种神经网络，本申请实施例对跟踪网络和视频中的目标均不加以限制。

S102、在所述每个视频帧为关键帧时，将所述关键帧输入检测网络并在接收指定数量的延迟帧期间控制所述检测网络运行，获得在接收完所述指定数量的延迟帧中的最后一个延迟帧时所述检测网络输出的目标检测结果。

在视频流的多个视频帧中，关键帧可以是视频流中用于目标检测的视频帧，指定数量的延迟帧为关键帧后连续的视频帧。其中，关键帧可以是用户预先设置的视频帧，例如，用户可以设置一定帧数间隔的视频帧为关键帧，示例性地，可以设置每间隔 10 帧、20 帧等为一个关键帧，还可以直接将视频流中的 I 帧作为关键帧，或者用户根据视频内容设置关键帧，本申请实施例对关键帧的设置不加以限制。

在从视频流中接收视频帧的过程中，当根据用户设置关键帧的方式检测接收到的视频帧为关键帧时，可以确定关键帧之后的连续的、指定数量的视频帧均为延迟帧。将关键帧输入检测网络中对该关键帧中的目标进行检测，在检测网络运行过程中，在接收指定数量的延迟帧期间控制该检测网络运行，从而在接收完最后一个延迟帧时输出目标检测结果。

计算节点可以是控制检测网络运行过程中停止运行的节点，在将关键帧输入检测网络后，在每接收到一个延迟帧时，控制检测网络的多个计算节点依次运行，直到接收完最后一个延迟帧后检测网络运行完最后一个计算节点输出目

标检测结果，示例性地，对于级联检测网络，可以将级联检测网络的每个层级的子网络的输入节点作为一个计算节点，每接收到一个延迟帧时，按照级联检测网络的子网络的顺序依次运行多个层级子网络，以在接收最后一个延迟帧时通过最后层级子网络输出目标检测结果。对于非级联检测网络，可以以运行时间将非级联检测网络分为多个部分网络，每个部分网络的输出作为一个计算节点，即非级联检测网络运行一定时长后暂停并保存中间结果，在接收到一个延迟帧后从暂停位置开始恢复运行非级联检测网络，并在接收最后一个延迟帧后运行完非级联检测网络的剩余部分从而得到目标检测结果。在实际应用中，还可以通过其他方式在接收延迟帧期间控制检测网络运行，本申请实施例对此不加以限制。

S103、根据所述目标检测结果和所述最后一个延迟帧的目标跟踪结果生成最终目标跟踪结果。

在本申请实施例中，检测网络负责识别和定位新目标，跟踪网络基于检测网络的输出进行目标跟踪（Tracking-by-detection）。检测网络输出目标检测结果后，如果检测到新目标则存储到已检测目标列表中，同时，跟踪网络根据当前视频帧和历史已跟踪目标，预测并更新已跟踪目标的当前位置，然后将更新位置后的已跟踪目标列表与已检测目标列表进行关联，并更新已跟踪目标列表的状态信息，该状态信息可以为目标仍然存在、丢失、离开等。在已检测目标为新的目标时将该新的目标增加到已跟踪目标列表中，并记录该新的目标的轨迹、速度等信息，当检测到当前跟踪的目标丢失或者离开，则对目标进行销毁，从已跟踪目标列表中移除该目标，最终输出已跟踪目标列表。

在本申请实施例中，关键帧在输入检测网络的同时也输入跟踪网络中，指定数量的延迟帧在检测网络中被处理期间，每个延迟帧也输入跟踪网络并得到每个延迟帧的目标跟踪结果，当检测网络在最后一个延迟帧的处理结束输出目标检测结果时，可以采用该目标检测结果和最后一个延迟帧的目标跟踪结果生成该延迟帧的最终目标跟踪结果，例如，采用目标检测结果中的已检测目标列表和目标跟踪结果中的已跟踪目标列表进行比对，确定新增目标、丢失目标和离开目标，以此更新已跟踪目标列表，输出该已跟踪目标列表并以该已跟踪目标列表为基础进行后续的目标跟踪。

本申请实施例将从视频流接收到的每个视频帧输入跟踪网络中得到视频帧的目标跟踪结果，而在视频帧为关键帧时，将关键帧输入检测网络并在接收指定数量的延迟帧期间控制检测网络运行，获得在接收完最后一个延迟帧时检测网络输出的目标检测结果，根据目标检测结果和最后一个延迟帧的目标跟踪结果生成最终目标跟踪结果。避免了目标检测耗时长，在关键帧处进行目标检测

后再接收下一视频帧造成耗时峰值陡增、视频卡顿的问题，通过在接收多个延迟帧期间控制检测网络运行，将目标检测的耗时分摊到接收多个延迟帧期间，大大削减了在关键帧处目标检测的耗时峰值，在播放关键帧后能够及时播放下一视频帧，使得视频能够以预设的帧率播放以保证视频流畅地播放。

实施例二

图 2 是本申请实施例二提供的一种目标检测跟踪方法的步骤流程图，本申请实施例在前述实施例一的基础上进行说明，如图 2 所示，本申请实施例的目标检测跟踪方法可以包括如下步骤：

S201、将从视频流接收到的每个视频帧输入跟踪网络中得到所述每个视频帧的目标跟踪结果。

S202、判断所述每个视频帧是否为关键帧。

在本申请实施例中，关键帧可以是用户预先设置用于目标检测的视频帧，在本申请的一个示例中，用户可以设置一定帧数间隔的视频帧为关键帧，在将从视频流接收到的每个视频帧输入跟踪网络的同时，根据一定帧数间隔判断当前接收到的视频帧是否为关键帧，如设置每间隔 15 帧为关键帧，则可以统计接收视频帧的数量，即每接收一个视频帧，视频帧的数量累加 1，当视频帧的数量为 $15N+1$ 时（ N 为自然数），确定该视频帧为关键帧，又或者用户将视频流分割为多个片段，若接收的视频帧为每个片段的第一个视频帧时，该视频帧为关键帧，又或者用户设置视频流的 I 帧为关键帧，则在检测到接收到的视频帧为 I 帧时，该视频帧为关键帧，本申请实施例设置关键帧的方式不同，判断视频帧是否为关键帧的方式也不相同，本申请实施例对此不加以限制。

本申请实施例在确定接收到的视频帧为关键帧时，执行 S203-S207，在确定接收到的视频帧为非关键帧时，执行 S208 直接输出视频帧的跟踪结果。

S203、确定指定数量的延迟帧。

在本申请实施例中，延迟帧为视频流中关键帧之后的视频帧，延迟帧的数量可以根据实际情况预先设定。在本申请的一个示例中，当检测网络为级联检测网络时，关键帧和延迟帧的数量的总和等于级联检测网络中子网络的数量，即延迟帧的数量等于级联检测网络中子网络的数量减去 1（关键帧的数量为 1），当用户确定所使用的级联检测网络后，可以通过级联检测网络中子网络的数量设置延迟帧的数量，例如，在一个示例中，级联检测网络包含 3 个子网络，则可以设置关键帧之后连续的两个视频帧均为延迟帧。在本申请的另一示例中，当检测网络为非级联检测网络时，可以估算该非级联检测网络对关键帧进行目标检测所需要的总时长，通过该总时长预设延迟帧的数量，例如，在一个示例

中，视频流的帧率为 30 传输帧数/每秒（Frames Per Second, FPS），控制非级联检测网络对关键帧进行目标检测所需要的总时长均分到延迟帧上，为避免延迟帧的数量过多，导致目标检测结果距离关键帧中目标的实际位置偏差过大，可以设置延迟帧的数量为 5 帧，即非级联检测网络的目标检测结果延时约 0.17 秒（5/30）。在实际应用中，可以根据实际情况设置延迟帧的数量，本申请实施例对此不加以限制。

S204、确定所述检测网络的多个计算节点。

检测网络的计算节点可以是在检测网络运行过程中控制检测网络暂停运行和恢复运行的节点，该计算节点可以是检测网络中的任意一个节点，计算节点暂停运行后，当检测网络每接收到一个延迟帧时，从暂停的计算节点恢复运行。

在本申请的一个示例中，级联检测网络包含多个子网络，则可以将级联检测网络中子网络的输入节点设置为计算节点，对于非级联检测网络，则通过总时长和延迟帧的数量确定非级联检测网络运行到每个计算节点的时长，该时长对应的结束时间即为计算节点运行结束时间。

S205、在将所述关键帧输入所述检测网络后，每接收到一个延迟帧时控制所述检测网络的多个计算节点依次开始运行，获得在接收完最后一个延迟帧时所述检测网络输出的目标检测结果。

在实际应用中，当视频帧为关键帧时，关键帧先输入检测网络并在接收多个延迟帧期间控制检测网络运行，可以以接收到延迟帧为触发条件触发检测网络的多个计算节点依次运行。

在本申请的一个示例中，检测网络为级联检测网络，计算节点为每个子网络的输入，将关键帧输入级联检测网络的第一子网络后输出候选框，在接收到第一个延迟帧时可以将第一个延迟帧和第一子网络输出的候选框输入到第二子网络中输出候选框，在接收到第二个延迟帧时将第二个延迟帧和第二子网络输出的候选框输入到第三子网络中输出候选框，以此类推，直到接收到最后一个延迟帧输入最后一个子网络后输出的检测框作为目标检测结果。

在本申请的另一个示例中，检测网络为非级联检测网络，根据非级联检测网络对关键帧执行目标检测所需要的总时长和指定数量计算非级联检测网络的每个计算节点的运行结束时间，可以在将关键帧输入非级联检测网络时即计算非级联检测网络的运行时间，当该运行时间为计算节点运行结束时间时，暂停运行该非级联检测网络并保存运行结果，当接收到一个延迟帧时，基于运行结果从该计算节点恢复运行非级联检测网络，以此类推，直到接收最后一个延迟帧后运行完非级联检测网络并输出目标检测结果。

本申请实施例通过设置检测网络的计算节点，并在每接收到一个延迟帧时控制计算节点依次开始运行，以在接收完最后一个延迟帧时输出目标检测结果，从而实现了在接收多个延迟帧期间控制检测网络运行，将目标检测的耗时分摊到接收多个延迟帧期间，大大削减了在关键帧处目标检测的耗时峰值，在播放关键帧后能够及时播放下一视频帧，使得视频能够以预设的帧率播放以保证视频流畅地播放。

S206、采用所述目标检测结果中的已检测目标列表更新所述目标跟踪结果中的已跟踪目标列表。

本申请实施例中，目标检测结果包含已检测目标列表和已检测目标列表中多个目标的位置、轨迹、速度等信息，目标跟踪结果包含已跟踪目标列表和已跟踪目标列表中多个已跟踪目标的状态，状态可以为目标丢失、离开等状态。

当检测网络接输出目标检测结果时，可以将目标检测结果中的已检测目标列表和最后一个延迟帧的目标跟踪结果中的已跟踪目标列表进行比对，从而确定出哪些目标丢失、离开、新增的目标等，以此更新已跟踪目标列表。

S207、更新所述已跟踪目标列表的目标状态。

对于已跟踪目标列表中的已跟踪目标，如果在已检测列表中存在该目标，则可以标记该目标的状态为存在状态，如果在已检测列表中不存在该目标则标记该目标的状态为丢失或者离开。

在本申请的可选实施例中，在已跟踪目标列表中的目标的目标状态为已丢失时，确定关键帧，将关键帧输入检测网络中得到目标检测结果，根据该目标检测结果对目标状态为已丢失的目标的状态进行确认，其中，在已跟踪目标列表中的目标的目标状态为已丢失时，将在最后一个延迟帧之后接收到的第 N 个视频帧确定为关键帧，其中，N 为自然数。即在目标的状态为丢失或者离开时，可以将延迟帧之后的第 N 个视频帧作为关键帧输入检测网络中得到目标检测结果，以通过该目标检测结果对目标状态为已丢失的目标的状态确认，当确认目标丢失或者离开后，从已跟踪目标列表中销毁该目标，避免了检测出错造成漏检或者误检的问题，保证了已跟踪目标列表的准确性。

S208、输出所述每个视频帧的目标跟踪结果。

在本申请实施例中，如果视频帧为非关键帧，则将该视频帧进行目标跟踪得到的目标跟踪结果作为输出。

本申请实施例将从视频流接收到的每个视频帧输入跟踪网络中得到视频帧的目标跟踪结果后，判断该视频帧是否为关键帧，在该视频帧为关键帧时，确定指定数量的延迟帧并确定检测网络的多个计算节点，在将关键帧输入检测网

络后，每接收到一个延迟帧时控制检测网络的多个计算节点依次开始运行，在接收完最后一个延迟帧时检测网络输出目标检测结果，采用目标检测结果中的已检测目标列表更新目标跟踪结果中的已跟踪目标列表，以及更新已跟踪目标列表的目标状态，将已跟踪目标列表作为最终的检测跟踪输出。由于通过设置计算节点在接收多个延迟帧期间控制检测网络运行，将目标检测的耗时分摊到接收多个延迟帧期间，大大削减了在关键帧处目标检测的耗时峰值，在播放关键帧后能够及时播放下一视频帧，使得视频能够以预设的帧率播放以保证视频流畅地播放。

实施例三

图 3A 是本申请实施例三提供的一种目标检测跟踪方法的步骤流程图，本申请实施例在前述实施例一的基础上进行说明，如图 3A 所示，本申请实施例的目标检测跟踪方法可以包括如下步骤：

S301、将从视频流接收到的每个视频帧输入跟踪网络中得到所述每个视频帧的目标跟踪结果。

S302、判断所述每个视频帧是否为关键帧。

S303、获取所述级联检测网络的子网络的第一数量。

在本申请实施例中，检测网络为级联检测网络，该级联检测网络可以是预先训练的检测网络，级联检测网络可以包含多个子网络，则可以获取级联检测网络的子网络的数量，在本申请的一个示例中，级联检测网络可以为多任务卷积神经网络（Multi-Task Convolutional Neural Network, MTCNN），MTCNN 可以包含 3 个子网络，分别为 P-Net、R-Net 和 O-Net，其中，P-Net 为一个全连接网络，通过滑动窗口（窗口大小为 12）从视频帧中提取目标候选框，R-Net（输入视频帧分辨率为 24）和 O-Net（输入视频帧分辨率为 48）作为过滤器，分别过滤前一级子网络的非目标候选框，剩下的候选框即为目标的检测框。在实际应用中级联检测网络还可以是 Faster-基于区域的卷积神经网络（Region based Convolutional Neural Network, RCNN）等其他级联网络。

S304、根据所述第一数量将所述关键帧后连续的多个视频帧均确定为延迟帧。

关键帧后的视频帧可以作为延迟帧，在确定级联检测网络的子网络数量后，可以将关键帧后连续的多个视频帧均作为延迟帧，由于关键帧输入到级联检测网络的第一级子网络，即子网络的数量和延迟帧的数量的差值为 1，示例性地，级联检测网络的子网络数量为 4，则延迟帧的数量为 3，可以将关键帧之后的第 1 个视频帧、第 2 个视频帧、以及第 3 个视频帧均作为延迟帧。

S305、将所述级联检测网络的每个子网络的输入节点设置为计算节点。

在级联检测网络中，每个子网络为独立的网络，可以将每个子网络的输入节点作为计算节点。

S306、在将所述关键帧输入所述级联检测网络的第一层级子网络后，每接收到一个延迟帧时，将上一层级子网络输出的检测结果和所述延迟帧输入当前层级子网络中。

在实际应用中，关键帧输入级联检测网络的第一层级子网络后输出检测结果，该检测结果可以是目标的候选框，对于第二层级子网络，当接收到关键帧之后的第一个延迟帧时，将第一层级子网络输出的候选框和第一个延迟帧输入第二层级子网络中输出目标的候选框，以此类推，每接收到一个延迟帧时，将上一层级子网络输出的检测结果和接收到的延迟帧输入当前层级子网络中。

S307、在接收到最后一个延迟帧后，将当前层级子网络输出的检测结果作为所述级联检测网络的目标检测结果。

在本申请实施例中，每接收到一个延迟帧时触发级联检测网络的多个子网络依次开始计算，当接收到最后一个延迟帧时，当前层级子网络为级联检测网络的最后一个层级子网络，可以将最后一个层级子网络的前一层级子网络输出的候选框和最后一个延迟帧输入到最后一个层级子网络中，最后一个层级子网络输出的检测框即为目标检测结果。

S308、根据所述目标检测结果和所述最后一个延迟帧的目标跟踪结果生成最终目标跟踪结果。

S309、输出所述每个视频帧的目标跟踪结果。

以下结合图 3B 说明级联检测网络在接收延迟帧期间运行的过程，如图 3B 所示，级联检测网络的运行过程如下：

视频流输入后，每个视频帧均输入到跟踪网络中，跟踪网络输出每个视频帧的目标跟踪结果，如果该视频帧为非关键帧，将该视频帧的目标跟踪结果作为最终的目标跟踪结果，如果该视频帧为关键帧，该关键帧输入到级联检测网络的第一级子网络并输出目标的候选框，同时确定关键帧之后的 N 个视频帧均为延迟帧，当接收到第 1 个延迟帧时，将第一级子网络的候选框和第 1 个延迟帧输入到第二级子网络中以使第二级子网络输出候选框，当接收到第 2 个延迟帧时，将第二级子网络的候选框和第 2 个延迟帧输入到第三级子网络中以使第三级子网络输出候选框，以此类推，当接收到第 N 个延迟帧时，将 N 级子网络输出的候选框和第 N 个延迟帧输入第 N+1 级子网络输出检测框作为目标检测结果，将该目标检测结果和第 N 个延迟帧的目标结果关联生成最终的目标跟踪结果。

果。

本申请实施例对于级联检测网络，将从视频流接收到的每个视频帧输入跟踪网络中得到视频帧的目标跟踪结果后，判断视频帧为关键帧时，获取级联检测网络的子网络的第一数量，并根据第一数量将关键帧后连续的多个视频帧均确定为延迟帧，将级联检测网络的每个子网络的输入节点设置为计算节点后，将关键帧输入级联检测网络的第一层级子网络，并在每接收到一个延迟帧时，将上一层级子网络输出的检测结果和延迟帧输入当前层级子网络中，在接收到最后一个延迟帧后，将当前层级子网络输出的检测结果作为检测网络的目标检测结果，根据目标检测结果和最后一个延迟帧的目标跟踪结果生成最终目标跟踪结果。本申请实施例通过级联检测网络的级联数确定延迟帧的数量，并设置每个级联的子网络的输入作为计算节点，每接收一个延迟帧时控制多个级联的子网络依次运行，实现了将目标检测的耗时分摊到接收多个延迟帧期间，大大削减了在关键帧处目标检测的耗时峰值，在播放关键帧后能够及时播放下一视频帧，使得视频能够以预设的帧率播放以保证视频流畅地播放。

实施例四

图 4A 是本申请实施例四提供的一种目标检测跟踪方法的步骤流程图，本申请实施例在前述实施例一的基础上进行说明，如图 4A 所示，本申请实施例的目标检测跟踪方法可以包括如下步骤：

S401、将从视频流接收到的每个视频帧输入跟踪网络中得到所述每个视频帧的目标跟踪结果。

S402、判断所述视频帧是否为关键帧。

S403、确定所述关键帧后连续的指定数量的视频帧均为延迟帧。

在检测网络为非级联检测网络时，由于非级联检测网络的唯一输入为关键帧，如果为非级联检测网络设置的延迟帧的数量过多，可能导致最终得到的目标检测结果距离关键帧中目标的实际位置过大，可以设置较少数量的延迟帧以获得准确的目标检测结果，在一个示例中，延迟帧的数量可以设置为 4 个，也可以根据实际情况设置延迟帧的数量，例如，在视频流中目标稳定时可以设置较多的延迟帧（如一个主播的直播中通常只包含主播一个目标，对主播进行跟踪），在目标不稳定时可以设置较少的延迟帧（如在足球赛事直播中每个视频帧出现的球员不同，对球员进行跟踪），本申请实施例对设置延迟帧的数量不加以限制。

S404、预估所述非级联检测网络对所述关键帧执行目标检测所需要的总时长。

在本申请的可选实施例中，可以根据经验值预先设置非级联检测网络对关键帧进行目标检测的总时长，或者根据关键帧的复杂程度设置非级联检测网络对关键帧进行目标检测的总时长，还可以是将关键帧输入预先训练的时长估计模型中获得非级联检测网络对关键帧进行目标检测的总时长，本申请实施例对预估非级联检测网络对关键帧执行目标检测所需要的总时长的方式不加以限制。

S405、采用所述总时长和所述指定数量计算所述非级联检测网络的每个计算节点的运行结束时间。

计算节点的运行结束时间可以是指非级联检测网络在运行过程中停止运行的时间，即在非级联检测网络的运行过程中，以非级联检测网络开始运行时间为时间起点，当运行时间为计算节点的运行结束时间时，非级联检测网络停止运行并保存中间运行结果，在检测到恢复运行事件时从该计算节点恢复非级联检测网络的运行。在本申请实施例中，可以计算延迟帧的数量和关键帧的数量之和，总时长和该和值的比值即为每个计算节点的运行时长，以关键帧输入非级联检测网络的时间为时间起点，以所述运行时长为时间步长可以计算出每个计算节点的运行结束时间。

在本申请的一个示例中，可以将非级联检测网络的总时长均分到多个延迟帧中，示例性地，延迟帧的数量为 4 个，加上关键帧总共有 5 个视频帧，假设非级联检测网络运行一个关键帧的总时长为 1 秒，则在每个视频帧的预先时长为 200 毫秒（ $1000/5$ ），可以设置非级联检测网络每运行 200 毫秒作为一个计算节点停止运行点，还可以是设置以关键帧输入非级联检测网络后的 0.2 秒、0.4 秒、0.6 秒、0.8 秒为计算节点的运行结束时间。上述数值仅仅是示例，在实际应用中可以根据实际情况设置延迟帧的数量，以保证非级联检测网络针对每个关键帧的运行时长小于或等于相邻两个关键帧的时间间隔。

S406、在将所述关键帧输入所述非级联检测网络后，计算所述非级联网络的运行时间。

示例性地，可以设置一个计时器，当关键帧输入非级联检测网络后，计时器开始计时，该计时器的值记为非级联检测网络的运行时间。

S407、在所述运行时间为所述计算节点运行结束时间时，暂停运行所述非级联检测网络。

当计时器的值为计算节点运行结束时间时，非级联检测网络保存该时刻的检测结果，并暂停运行，其中，检测结果可以是非级联检测网络在该计算节点运行结束时间对关键帧提取到的特征图。在实际应用中，可以通过程序脚本控

制非级联检测网络停止运行，本申请实施例对控制非级联检测网络暂停运行的方式不加以限制。

在本申请实施例中，当计时器的值为计算节点运行结束时间时，计时器暂停计时。

S408、每接收到一个延迟帧时，从暂停的计算节点恢复运行所述非级联检测网络。

非级联检测网络暂停运行后，每接收到一个延迟帧，则读取保存的检测结果，并从上一暂停的计算节点恢复运行非级联检测网络，继续对关键帧提取特征，同时计时器恢复计时，对非级联检测网络的运行时间进行累加，以此类推直到接收到最后一个延迟帧为止。

S409、在接收到最后一个延迟帧时，将所述非级联检测网络恢复运行后输出的检测结果作为所述非级联检测网络的目标检测结果。

在从视频流中接收到最后一个延迟帧后，非级联检测网络恢复最后一部分运行输出最终的目标检测结果。

S410、根据所述目标检测结果和所述最后一个延迟帧的目标跟踪结果生成最终目标跟踪结果。

S411、输出所述每个视频帧的目标跟踪结果。

以下结合图 4B 说明非级联检测网络在接收延迟帧期间运行的过程，如图 4B 所示，非级联检测网络的运行过程如下：

视频流输入后，每个视频帧均输入到跟踪网络中，跟踪网络输出每个视频帧的目标跟踪结果，如果该视频帧为非关键帧，将该视频帧的目标跟踪结果作为最终的目标跟踪结果。

如果该视频帧为关键帧，确定 N 个延迟帧，并估算非级联检测网络对关键帧执行目标检测所需要的总时长，根据总时长和 N 确定每个计算节点的运行结束时间，以运行结束时间将非级联检测网络划分为多个部分的网络。关键帧输入到非级联检测网络的第一部分网络后计算非级联检测网络的运行时间，当运行时间为第一部分网络的运行结束时间时，非级联检测网络暂停运行并输出关键帧的特征图，当接收到第 1 个延迟帧时，以上一暂停运行时的特征图为基础恢复非级联检测网络，即运行第二部分网络，当运行时间为第二部分网络的运行结束时间时，非级联检测网络暂停运行并输出关键帧的特征图，以此类推，当接收到第 N 个延迟帧时，恢复第 $N+1$ 部分网络运行并输出检测框作为目标检测结果，将该目标检测结果和第 N 个延迟帧的目标跟踪结果关联生成最终的目标跟踪结果。

本申请实施例通过非级联检测网络对关键帧执行目标检测所估算的总时长和延迟帧的数量计算每个计算节点的运行结束时间，在非级联检测网络的运行时间为运行结束时间时暂停运行，每接收一个延迟帧时控制多个计算节点依次恢复运行，实现了将目标检测的耗时分摊到接收多个延迟帧期间，大大削减了在关键帧处目标检测的耗时峰值，在播放关键帧后能够及时播放下一视频帧，使得视频能够以预设的帧率播放以保证视频流畅地播放。

实施例五

图 5 是本申请实施例五提供的一种目标检测跟踪装置的结构框图，如图 5 所示，本申请实施例的目标检测跟踪装置可以包括如下模块：

跟踪模块 501，设置为将从视频流接收到的每个视频帧输入跟踪网络中得到所述每个视频帧的目标跟踪结果；检测模块 502，设置为在所述每个视频帧为关键帧时，将所述关键帧输入检测网络并在接收指定数量的延迟帧期间控制所述检测网络运行，获得在接收完所述指定数量的延迟帧中的最后一个延迟帧时所述检测网络输出的目标检测结果，其中，所述关键帧为所述视频流中用于目标检测的视频帧，所述指定数量的延迟帧为所述关键帧后连续的视频帧；检测跟踪结果生成模块 503，设置为根据所述目标检测结果和所述最后一个延迟帧的目标跟踪结果生成最终目标跟踪结果。

本申请实施例所提供的目标检测跟踪装置可执行本申请实施例一到实施例四任一实施例所提供的目标检测跟踪方法，具备执行方法相应的功能模块和效果。

实施例六

图 6 是本申请实施例六提供的一种电子设备的结构示意图。如图 6 所示，该电子设备可以包括：处理器 601、存储器 602、具有触摸功能的显示屏 603、输入装置 604、输出装置 605 以及通信装置 606。该设备中处理器 601 的数量可以是一个或者多个，图 6 中以一个处理器 601 为例。该设备的处理器 601、存储器 602、显示屏 603、输入装置 604、输出装置 605 以及通信装置 606 可以通过总线或者其他方式连接，图 6 中以通过总线连接为例。所述电子设备设置为执行如本申请任一实施例提供的目标检测跟踪方法。

本申请实施例还提供一种计算机可读存储介质，所述存储介质中的指令由设备的处理器执行时，使得设备能够执行如上述方法实施例所述的目标检测跟踪方法。

对于装置、电子设备、存储介质实施例而言，由于其与方法实施例基本相似，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。

在本文的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“示例”、或“一些示例”等的描述意指结合该实施例或示例描述的特征、结构、材料或者特点包含于本申请的至少一个实施例或示例中。在本文中，对上述术语的示意性表述不一定指的是相同的实施例或示例。而且，描述的特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。

权利要求书

1、一种目标检测跟踪方法，包括：

将从视频流接收到的每个视频帧输入跟踪网络中得到所述每个视频帧的目标跟踪结果；

在所述每个视频帧为关键帧的情况下，将所述关键帧输入检测网络并在接收指定数量的延迟帧期间控制所述检测网络运行，获得在接收完所述指定数量的延迟帧中的最后一个延迟帧时所述检测网络输出的目标检测结果，其中，所述关键帧为所述视频流中用于目标检测的视频帧，所述指定数量的延迟帧为所述关键帧后连续的视频帧；

根据所述目标检测结果和所述最后一个延迟帧的目标跟踪结果生成最终目标跟踪结果。

2、根据权利要求1所述的方法，其中，所述在所述每个视频帧为关键帧的情况下，将所述关键帧输入检测网络中并在接收指定数量的延迟帧期间控制所述检测网络运行，获得在接收完所述指定数量的延迟帧中的最后一个延迟帧时所述检测网络输出的目标检测结果，包括：

在所述每个视频帧为关键帧的情况下，确定所述指定数量的延迟帧并根据所述指定数量确定所述检测网络的多个计算节点，或者在所述每个视频帧为关键帧的情况下，确定所述检测网络的多个计算节点并根据所述检测网络的多个计算节点确定所述指定数量的延迟帧；

在将所述关键帧输入所述检测网络后，每接收到一个延迟帧时控制所述检测网络的多个计算节点依次开始运行，获得在接收完所述最后一个延迟帧时所述检测网络输出的所述目标检测结果。

3、根据权利要求2所述的方法，其中，所述检测网络为包括多个层级子网络的级联检测网络；

所述在所述每个视频帧为关键帧的情况下，确定所述检测网络的多个计算节点，包括：

获取所述级联检测网络的子网络的第一数量；

将所述第一数量的子网络中的每个子网络的输入节点设置为计算节点。

4、根据权利要求3所述的方法，其中，所述根据所述检测网络的多个计算节点确定所述指定数量的延迟帧，包括：

根据所述第一数量，将所述关键帧后连续的多个视频帧确定为所述指定数量的延迟帧。

5、根据权利要求4所述的方法，其中，所述每接收到一个延迟帧时控制所

述检测网络的多个计算节点依次开始运行，获得在接收完所述最后一个延迟帧时所述检测网络输出的所述目标检测结果，包括：

在将所述关键帧输入所述级联检测网络的第一层级子网络后，每接收到一个延迟帧时，将上一层级子网络输出的检测结果和所述延迟帧输入当前层级子网络中；

在接收最后一个延迟帧后，将当前层级子网络输出的检测结果作为所述级联检测网络的目标检测结果。

6、根据权利要求2所述的方法，其中，所述检测网络为非级联检测网络；

所述在所述每个视频帧为关键帧的情况下，确定所述指定数量的延迟帧，包括：

确定所述关键帧后连续的指定数量的视频帧为所述指定数量的延迟帧。

7、根据权利要求6所述的方法，其中，所述根据所述指定数量确定所述检测网络的多个计算节点，包括：

预估所述非级联检测网络对所述关键帧执行目标检测所需要的总时长；

采用所述总时长和所述指定数量计算所述非级联检测网络的每个计算节点的运行结束时间。

8、根据权利要求7所述的方法，其中，所述在将所述关键帧输入所述检测网络后，每接收到一个延迟帧时控制所述检测网络的多个计算节点依次开始运行，获得在接收完所述最后一个延迟帧时所述检测网络输出的目标检测结果，包括：

在将所述关键帧输入所述非级联检测网络后，计算所述非级联网络的运行时间；

在所述运行时间为所述计算节点运行结束时间的情况下，暂停运行所述非级联检测网络；

每接收到一个延迟帧时，从暂停的计算节点恢复运行所述非级联检测网络；

在接收到所述最后一个延迟帧时，将所述非级联检测网络恢复运行后所述非级联检测网络输出的检测结果作为所述非级联检测网络的目标检测结果。

9、根据权利要求1-8任一项所述的方法，其中，在所述将所述关键帧输入检测网络中并在接收指定数量的延迟帧期间控制所述检测网络运行，获得在接收完最后一个延迟帧时所述检测网络输出的目标检测结果之前，还包括：

判断所述每个视频帧是否为关键帧；

响应于所述每个视频帧为关键帧，执行将所述关键帧输入检测网络并在接收指定数量的延迟帧期间控制所述检测网络运行，获得在接收完所述指定数量的延迟帧中的最后一个延迟帧时所述检测网络输出的目标检测结果的步骤；所述方法还包括：

响应于所述每个视频帧不为关键帧，输出所述每个视频帧的目标跟踪结果。

10、根据权利要求 1-8 任一项所述的方法，其中，所述根据所述目标检测结果和所述最后一个延迟帧的目标跟踪结果生成最终目标跟踪结果，包括：

采用所述目标检测结果中的已检测目标列表更新所述目标跟踪结果中的已跟踪目标列表；

更新所述已跟踪目标列表的目标状态。

11、根据权利要求 10 所述的方法，还包括：

在所述已跟踪目标列表中的目标的目标状态为已丢失的情况下，确定关键帧；

将确定的所述关键帧输入所述检测网络中得到目标检测结果。

12、根据权利要求 11 所述的方法，其中，所述在所述已跟踪目标列表中的目标的目标状态为已丢失的情况下，确定关键帧，包括：

在所述已跟踪目标列表中的目标的目标状态为已丢失的情况下，将在所述最后一个延迟帧之后接收到的第 N 个视频帧确定为所述关键帧，其中，N 为自然数。

13、一种目标检测跟踪装置，包括：

跟踪模块，设置为将从视频流接收到的每个视频帧输入跟踪网络中得到所述每个视频帧的目标跟踪结果；

检测模块，设置为在所述每个视频帧为关键帧的情况下，将所述关键帧输入检测网络并在接收指定数量的延迟帧期间控制所述检测网络运行，获得在接收完所述指定数量的延迟帧中的最后一个延迟帧时所述检测网络输出的目标检测结果，其中，所述关键帧为所述视频流中用于目标检测的视频帧，所述指定数量的延迟帧为所述关键帧后连续的视频帧；

检测跟踪结果生成模块，设置为根据所述目标检测结果和所述最后一个延迟帧的目标跟踪结果生成最终目标跟踪结果。

14、一种电子设备，包括：

至少一个处理器；

存储装置，设置为存储至少一个程序；

当所述至少一个程序被所述至少一个处理器执行，使得所述至少一个处理器实现如权利要求 1-12 中任一项所述目标检测跟踪方法。

15、一种计算机可读存储介质，存储有计算机程序，其中，所述程序被处理器执行时实现如权利要求 1-12 中任一项所述目标检测跟踪方法。