

视频物体加速检测方法、装置、服务器及存储介质

[0001] 本申请以2019年6月17日提交的申请号为201910523099.6，名称为“视频物体加速检测方法、装置、服务器及存储介质”的中国发明申请为基础，并要求其优先权。

技术领域

[0002] 本申请涉及基于人工智能的图像识别技术领域，具体涉及一种视频物体加速检测方法、装置、服务器及存储介质。

[0003] 背景技术

[0004] 科技的发展使得摄像设备得到了普及，应运而生海量的图像数据、视频数据，其中视频数据也受到了广泛的关注，许多领域都需要用到目标的检测和追踪，例如监控视频，无人机追踪等。

[0005] 传统的目标检测方法需要对视频中逐帧做如下处理：1) 根据每一帧图像的骨干网络生成特征图；2) 在所述特征图的金字塔上划分锚点，做前景背景分类和坐标框回归；3) 将前一步骤中获得的高得分前景特征图送入分类网络，做二次分类和坐标回归并输出结果。

[0006] 发明人意识到上述方法在处理720p以上的图片时由于检测锚点过多，会出现检测速度慢的问题，往往很难达到实时检测的效果，因此只能通过跳帧来解决，从而降低了检测精度。

[0007]

[0008] 发明内容

[0009] 鉴于以上内容，有必要提出一种视频物体加速检测方法、装置、服务器及存储介质，能够加快侦测视频图像中的目标的速度。

[0010] 一种视频物体加速检测方法，包括：

[0011] 将视频图像中的所有帧图像分为预设组帧图像，其中，每一组帧图像包括关键帧图像和非关键帧图像；

[0012] 识别所述关键帧图像中的检测目标，并在所述检测目标上标记检测框；

- [0013] 识别所述预设组帧图像中的非关键帧图像中的检测目标，将所述关键帧图中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述非关键帧图像中的检测目标；
- [0014] 根据所述预选框裁剪所述非关键帧图像得到多个预选框图像；
- [0015] 将所述多个预选框图像输入至非关键帧检测网络，以得到所述检测目标在所述预选框中的位置；及
- [0016] 根据所述检测目标在所述预选框中的位置将所述检测目标映射至所述非关键帧对应的原图。
- [0017] 一种视频物体加速检测装置，包括：
- [0018] 初始化模块，用于将视频图像中的所有帧图像分为预设组帧图像，其中，每一组帧图像包括关键帧图像和非关键帧图像；
- [0019] 识别模块，用于识别所述关键帧图像中的检测目标，并在所述检测目标上标记检测框；
- [0020] 所述识别模块，还用于识别预设组帧图像中的非关键帧图像中的检测目标，将所述关键帧图中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述非关键帧图像中的检测目标；
- [0021] 裁剪模块，用于根据所述预选框裁剪所述非关键帧图像得到多个预选框图像；
- [0022] 处理模块，用于将所述多个预选框图像输入至非关键帧检测网络，以得到所述检测目标在所述预选框中的位置；及
- [0023] 所述处理模块，还用于根据所述检测目标在所述预选框中的位置将所述检测目标映射至所述非关键帧对应的原图。
- [0024] 一种服务器，所述服务器包括处理器和存储器，所述处理器用于执行所述存储器中存储的计算机可读指令时实现如下步骤：
- [0025] 将视频图像中的所有帧图像分为预设组帧图像，其中，每一组帧图像包括关键帧图像和非关键帧图像；
- [0026] 识别所述关键帧图像中的检测目标，并在所述检测目标上标记检测框；
- [0027] 识别所述预设组帧图像中的非关键帧图像中的检测目标，将所述关键帧图中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述非关键帧图

像中的检测目标；

- [0028] 根据所述预选框裁剪所述非关键帧图像得到多个预选框图像；
- [0029] 将所述多个预选框图像输入至非关键帧检测网络，以得到所述检测目标在所述预选框中的位置；及
- [0030] 根据所述检测目标在所述预选框中的位置将所述检测目标映射至所述非关键帧对应的原图。
- [0031] 一个或多个存储有计算机可读指令的可读存储介质，所述计算机可读存储介质存储有计算机可读指令，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：
- [0032] 将视频图像中的所有帧图像分为预设组帧图像，其中，每一组帧图像包括关键帧图像和非关键帧图像；
- [0033] 识别所述关键帧图像中的检测目标，并在所述检测目标上标记检测框；
- [0034] 识别所述预设组帧图像中的非关键帧图像中的检测目标，将所述关键帧图中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述非关键帧图像中的检测目标；
- [0035] 根据所述预选框裁剪所述非关键帧图像得到多个预选框图像；
- [0036] 将所述多个预选框图像输入至非关键帧检测网络，以得到所述检测目标在所述预选框中的位置；及
- [0037] 根据所述检测目标在所述预选框中的位置将所述检测目标映射至所述非关键帧对应的原图。
- [0038] 上述视频物体加速检测方法、装置、系统及存储介质，本发明通过将视频图像中的所有帧图像分为预设组帧图像，其中，每一组帧图像包括关键帧图像和非关键帧图像；利用关键帧图像中的目标的检测框在非关键帧图像中生成预选框，再通过侦测预选框中目标的位置来实现对视频中目标物体的侦测。充分利用了视频连续帧的空间相关性，在不损失精度的情况下，极大减少了锚点分类的时间消耗，且在基本不降低检测性能的情况下，大幅优化了检测速度。另外可以通过调节关键帧与非关键帧的比例，可以在不改变算法的基础上，实现不同硬件场景的需求。

[0039]

[0040] 附图说明

[0041] 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术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的前提下，还可以根据提供的附图获得其他的附图。

[0042] 图1是本申请实施例一提供的视频物体加速检测方法的流程图。

[0043] 图2是本申请实施例二提供的视频物体加速检测装置较佳实施例中的功能模块图。

[0044] 图3是本申请实施例三提供的服务器的示意图。

[0045] 如下具体实施方式将结合上述附图进一步说明本申请。

[0046] 具体实施方式

[0047] 为了能够更清楚地理解本申请的上述目的、特征和优点，下面结合附图和具体实施例对本申请进行详细描述。需要说明的是，在不冲突的情况下，本申请的实施例及实施例中的特征可以相互组合。

[0048] 在下面的描述中阐述了很多具体细节以便于充分理解本申请，所描述的实施例仅仅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做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。

[0049] 除非另有定义，本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同。本文中在本申请的说明书中所使用的术语只是为了描述具体的实施例的目的，不是旨在于限制本申请。

[0050] 本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”和“第三”等是用于区别不同对象，而非用于描述特定顺序。此外，术语“包括”以及它们任何变形，意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元，而是可选地还包括没有列出的步骤或单元，或可选地还包括对于这些过程、方法、产品或设备固有的其它步骤或单元。

- [0051] 本申请实施例的视频物体加速检测方法应用在由至少一个服务器和通过网络与所述服务器进行连接的移动终端所构成的硬件环境中。网络包括但不限于：广域网、城域网或局域网。本申请实施例的视频物体加速检测方法可以由服务器来执行，也可以由移动终端来执行；还可以是由服务器和移动终端共同执行。
- [0052] 所述对于需要进行视频物体加速检测方法的服务器，可以直接在服务器上集成本申请的方法所提供的视频物体加速检测功能，或者安装用于实现本申请的方法的客户端。再如，本申请所提供的方法还可以以软件开发工具包（Software Development Kit, SDK）的形式运行在服务器等设备上，以SDK的形式提供视频物体加速检测功能的接口，服务器或其他设备通过提供的接口即可实现视频物体加速检测功能。
- [0053] 实施例一
- [0054] 图1是本申请实施例一提供的视频物体加速检测方法的流程图。根据不同的需求，该流程图中的执行顺序可以改变，某些步骤可以省略。
- [0055] 步骤S1，将视频图像中的所有帧图像分为预设组帧图像，其中，每一组帧图像包括关键帧图像和非关键帧图像。
- [0056] 具体地，可以将所述视频图像中的所有帧图像分为预设组帧图像，其中每一组帧图像包括连续预设帧图像，所述连续预设帧图像中的第一帧图像为关键帧图像，其余帧图像为非关键帧图像。
- [0057] 例如，将具有五十帧图像的视频图像分为十组帧图像，将连续五帧图像分为一组，所述五帧图像中第一帧图像为关键帧图像，第二至第五帧图像为非关键帧图像。每一张所述关键帧图像对应了所述视频图像中的一张原图，且每一张所述非关键帧图像也对应所述视频图像中的一张原图，所述原图为720p以上的图片。
- [0058] 可以理解的是，所述每一组帧图像中的图像帧数并不限定，且所述每一组帧图像中的关键帧图像与非关键帧图像的数量也不限定。
- [0059] 当需要精确检测视频图像中的目标时，所述预设组帧图像的数量可以适当增加，对应每一组帧图像中的图像帧数适当减少。例如，将具有六十帧图像的视频图像分为组二十帧图像，将所述视频图像中的所有帧图像中连续三帧图像分为

一组。

[0060] 当对视频图像中的目标检测精度要求不高时，所述预设组帧图像的数量可以适当减少，对应每一组帧图像中的图像帧数适当增加。例如，将具有六十帧图像的视频图像分为组十帧图像，将所述视频图像中的所有帧图像中连续六帧图像分为一组。

[0061] 在本实施方式中，在将视频图像中的所有帧图像分为预设组帧图像之前，所述视频物体加速检测方法还包括：

[0062] 接收摄像头采集的视频图像，并对所述视频图像进行解码的步骤。

[0063] 在本实施方式中，通过摄像头采集视频图像。所述摄像头与所述服务器之间通过有线或无线网络通信连接。所述摄像头将采集的视频图像通过有线或无线网络发送至所述服务器。

[0064] 步骤S2，识别所述关键帧图像中的检测目标，并在所述检测目标上标记检测框。

[0065] 在本实施方式中，所述检测目标可以是人、或动物、或交通工具等。所述关键帧图像中的所述检测目标可以包括多个，每一个所述检测目标对应一个检测框。可以理解的是，根据所述检测目标的不同，对应的检测框大小也不相同。

[0066] 在本实施方式中，所述识别所述关键帧图像中的检测目标包括：

[0067] 通过不同窗口大小的滑窗对所述关键帧图像按预设顺序进行滑动；

[0068] 每次滑动时将当前窗口对应的图像输入至预先训练好的分类器；及

[0069] 当所述分类器处理所述当前窗口对应的图像后输出分类概率大于预设值，则确认识别所述关键帧图像中的检测目标，输出检测目标的标记。

[0070] 例如，首先对当前帧图像进行不同窗口大小的滑窗进行从左往右、从上到下的滑动。每次滑动时候对当前窗口执行分类器(分类器是事先训练好的)。如果当前窗口得到较高的分类概率，则认为检测到了目标。对每个不同窗口大小的滑窗都进行检测后，会得到不同窗口检测到的目标标记，这些窗口大小会存在重复较高的部分，最后采用非极大值抑制(Non-Maximum Suppression, NMS)的方法进行筛选以得到所述关键帧图像中的检测目标。最终，经过NMS筛选后获得检测到的目标。

- [0071] 筛选所述检测目标的标记以得到所述关键帧图像中的检测目标包括：
- [0072] 将所有滑窗输入至分类器后得到的概率按照降序排列，选取最大分类概率对应的滑窗；
- [0073] 分别计算其他滑窗与所述最大分类概率对应的滑窗之间的重叠面积，并判断所述重叠面积是否大于预设面积；
- [0074] 若所述重叠面积小于或等于所述预设面积，保留对应的滑窗并输出所述滑窗对应的检测目标；
- [0075] 若所述重叠面积大于所述预设面积，删除对应的滑窗；及
- [0076] 重复上述步骤，直到所有其他滑窗都比对完成。
- [0077] 步骤S3，所述识别预设组帧图像中的非关键帧图像中的检测目标，将所述关键帧图像中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述非关键帧图像中的检测目标。
- [0078] 可以理解的是，所述识别当前非关键帧图像中的检测目标的方法可以与所述识别所述关键帧图像中的检测目标的方法一致，在此不再赘述。
- [0079] 在本实施方式中，识别每一组帧图像中每一非关键帧图像中的检测目标。所述前一帧图像可以是关键帧图像，也可以是非关键帧图像。
- [0080] 所述识别预设组帧图像中的非关键帧图像中的检测目标，将所述关键帧图像中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述非关键帧图像中的检测目标包括如下步骤：
- [0081] 识别所述预设组帧图像中当前非关键帧图像中的检测目标；
- [0082] 将所述当前非关键帧图像之前的关键帧图像中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述当前非关键帧图像中的检测目标；
- [0083] 识别所述当前非关键帧图像之后的非关键帧图像中的检测目标；
- [0084] 将所述当前非关键帧图像中的预选框进行预设比例放大后生成第一预选框，通过所述第一预选框标记所述当前非关键帧图像之后的非关键帧图像中的检测目标，依此类推，直到标记完所述预设组帧图像中的所有非关键帧图像中的检测目标。
- [0085] 例如，若连续五帧图像分为一组，所述五帧图像中第一帧图像为关键帧图像，

第二至第五帧图像为非关键帧图像。识别第一帧为关键帧图像中的检测目标，并在所述检测目标上标记检测框；识别第二帧为非关键帧图像中的检测目标，将所述关键帧图像中的检测框进行预设比例放大后生成第一预选框，通过所述第一预选框标记所述第二帧图像中的检测目标；接着识别第三帧为非关键帧图像中的检测目标，将所述第二帧图像（非关键帧图像）中的检测框（即第一预选框）进行预设比例放大后生成第二预选框，通过所述第二预选框标记所述第三帧图像中的检测目标，依此类推，将每一组帧图像中的非关键帧图像中的检测目标都通过预选框进行标记。

[0086] 例如，若连续五帧图像分为一组，所述五帧图像中第二帧图像为关键帧图像，第一帧和第三至第五帧图像为非关键帧图像。识别第二帧为关键帧图像中的检测目标，并在所述检测目标上标记检测框；识别第三帧为非关键帧图像中的检测目标，将所述关键帧图像中的检测框进行预设比例放大后生成第一预选框，通过所述第一预选框标记所述第三帧图像中的检测目标；接着识别第四帧为非关键帧图像中的检测目标，将所述第三帧图像（非关键帧图像）中的检测框（即第一预选框）进行预设比例放大后生成第二预选框，通过所述第二预选框标记所述第四帧图像中的检测目标，依此类推，将每一组帧图像中的非关键帧图像中的检测目标都通过预选框进行标记。

[0087] 也就是说，并非限定每一组帧图像中的第一帧为关键帧图像，其他帧图像也可以作为关键帧图像。

[0088] 步骤S4，根据所述预选框裁剪所述非关键帧图像得到多个预选框图像。

[0089] 例如，所述非关键帧图像中的检测目标包括五个，通过五个预选框分别标记所述五个检测目标，根据所述五个预选框裁剪所述非关键帧图像可以得到五个预选框图像。

[0090] 在处理720p以上的图片时，由于图片太大，处理起来计算量大且费时。因此，在本实施方式中，可以将所述非关键帧图像根据所述预选框进行裁剪得到预选框图像，再将所述预选框图像输入至非关键帧网络中处理。不仅可以大图片转换成小图片处理，还可以去除大图片中的非检测目标区域，从而可以提高处理速度。

- [0091] 步骤S5, 将所述多个预选框图像输入至非关键帧检测网络, 以得到所述检测目标在所述预选框中的位置。
- [0092] 在本实施方式中, 所述非关键帧检测网络可以是预先训练的目标检测模型。其中, 所述目标检测模型用于表征预选框图像与所述预选框图像包括的检测目标的位置信息的对应关系。
- [0093] 当所述预选框图像输入至所述目标检测模型时, 可以确定所述预选框中的检测目标的位置信息。即, 所述非关键帧网络是用于获取检测目标在图像中的位置的深度学习神经网络。相比传统神经网络, 深度神经网络具有更多神经元, 能够对大规模不同位置、形态的区域目标图像样本进行学习; 同时在数据准备阶段, 通过增强变换方法可以对同一图像进行翻转变换、色调变换、尺度变换、噪声扰动、颜色变换等操作, 从而生成多种不同环境特性的训练图像, 使得神经网络检测器对于不同场景, 不同环境, 不同形状的区域目标都能够进行性能稳定的检测, 保证充足的样本数据量, 避免过拟合出现。另外, 神经网络的Dropout机制能够在训练过程的前馈运算时随机剔除部分神经元, 从而减少神经网络之间的相互依赖关系, 让神经网络具备更强的泛化能力, 避免过拟合。所述深度学习神经网络可以是卷积神经网络、深度信念网络和堆叠自动编码器。
- [0094] 作为实例, 所述目标检测模型可以是技术人员基于对大量的预选框图像和预选框图像包括的检测目标的对应关系表, 该对应关系表可以反映预选框图像和预选框图像包括的检测目标的对应关系。所述目标检测模型也可以是技术人员基于现有的人工神经网络进行有监督地训练而得到的图像识别模型。
- [0095] 优选地, 在将所述多个预选框图像输入至非关键帧检测网络之前, 所述视频物体加速检测方法还包括: 将所述多个预选框图像进行缩放处理。
- [0096] 可以理解的是, 由于检测目标大小不一, 其对应的预选框大小也各不相同。为了快速处理图像, 需将所述多个预选框图像进行缩放处理成大小近似的图像, 然后将同一大小后图像输入至非关键帧检测网络。
- [0097] 在本实施方式中, 将所述非关键帧图像中裁剪出的预选框图像批量输入至非关键帧检测网络之前, 可以将所述预选框图像缩放处理成大小相似的图像。如此, 实现将统一大小后图像输入至非关键帧检测网络, 便于所述非关键帧检测网

络快速检测所述检测目标在所述预选框中的位置。

[0098] 例如，所述当前非关键帧图像中的检测目标包括五个，通过五个预选框分别标记所述五个检测目标，根据所述五个预选框裁剪所述非关键帧图像可以得到五个预选框图像。所述五个预选框图像大小不一，需要进行缩放处理成相似大小的五个预选框图像。

[0099] 步骤S6，根据所述预选框中的位置将所述检测目标映射至所述非关键帧对应的原图。

[0100] 可以理解的是，所述检测目标在所述预选框中的位置是视频中的原图经过缩放处理后检测到的位置，因此，需要其进行映射至所述非关键帧对应的原图中以完成对视频图像中的目标的侦测。

[0101] 综上所述，本申请提供的视频物体加速检测方法，包括将视频图像中的所有帧图像分为预设组帧图像，其中，每一组帧图像包括关键帧图像和非关键帧图像；识别所述关键帧图像中的检测目标，并在所述检测目标上标记检测框；识别当前非关键帧图像中的检测目标，将所述当前非关键帧的前一帧图像中的检测目标对应的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述当前非关键帧图像中的检测目标；根据所述预选框裁剪所述当前非关键帧图像得到多个预选框图像；将所述多个预选框图像输入至非关键帧检测网络，以得到所述检测目标在所述预选框中的位置；及根据所述预选框中的位置将所述检测目标映射至所述非关键帧对应的原图。

[0102] 主要针对检测网络在720p以上场景中由于检测锚点过多导致的检测速度慢问题，通过将视频图像分成关键帧图像和非关键帧图像，并对所述非关键帧图像进行简化运算，实现了无需对视频中的每一帧图像都进行精确的目标检测处理，充分利用了视频连续帧的空间相关性，在不损失精度的情况下，极大减少了锚点分类的时间消耗，且在基本不降低检测性能的情况下，大幅优化了检测速度。另外可以通过调节关键帧与非关键帧的比例，可以在不改变算法的基础上，应对不同硬件场景的需求。

[0103] 实施例二

[0104] 图2为本申请视频物体加速检测装置较佳实施例中的功能模块图。

- [0105] 在一些实施例中，所述视频物体加速检测装置20（下文简称为“检测装置20”）运行于服务器中。所述检测装置20可以包括多个由程序代码段所组成的功能模块。所述检测装置20中的各个程序段的程序代码可以存储于存储器中，并由至少一个处理器所执行，以执行（详见图1及其相关描述）视频物体加速检测功能。
- [0106] 本实施例中，所述检测装置20根据其所执行的功能，可以被划分为多个功能模块。所述功能模块可以包括：初始化模块201、识别模块202、裁剪模块203及处理模块204。本申请所称的模块是指一种能够被至少一个处理器所执行并且能够完成固定功能的一系列计算机可读指令，其存储在存储器中。在一些实施例中，关于各模块的功能将在后续的实施例中详述。
- [0107] 所述初始化模块201用于将视频图像中的所有帧图像分为预设组帧图像，其中，每一组帧图像包括关键帧图像和非关键帧图像。
- [0108] 具体地，可以将所述视频图像中的所有帧图像分为预设组帧图像，其中每一组帧图像包括连续预设帧图像，所述连续预设帧图像中的第一帧图像为关键帧图像，其余帧图像为非关键帧图像。
- [0109] 例如，将具有五十帧图像的视频图像分为十组帧图像，将连续五帧图像分为一组，所述五帧图像中第一帧图像为关键帧图像，第二至第五帧图像为非关键帧图像。每一张所述关键帧图像对应了所述视频图像中的一张原图，且每一张所述非关键帧图像也对应所述视频图像中的一张原图，所述原图为720p以上的图片。
- [0110] 可以理解的是，所述每一组帧图像中的图像帧数并不限定，且所述每一组帧图像中的关键帧图像与非关键帧图像的数量也不限定。
- [0111] 当需要精确检测视频图像中的目标时，所述预设组帧图像的数量可以适当增加，对应每一组帧图像中的图像帧数适当减少。例如，将具有六十帧图像的视频图像分为组二十帧图像，将所述视频图像中的所有帧图像中连续三帧图像分为一组。
- [0112] 当对视频图像中的目标检测精度要求不高时，所述预设组帧图像的数量可以适当减少，对应每一组帧图像中的图像帧数适当增加。例如，将具有六十帧图像

的视频图像分为组十帧图像，将所述视频图像中的所有帧图像中连续六帧图像分为一组。

- [0113] 在本实施方式中，在将视频图像中的所有帧图像分为预设组帧图像之前，所述视频物体加速检测装置20还可以：接收摄像头采集的视频图像，并对所述视频图像进行解码。
- [0114] 在本实施方式中，通过摄像头采集视频图像。所述摄像头与所述服务器之间通过有线或无线网络通信连接。所述摄像头将采集的视频图像通过有线或无线网络发送至所述服务器。
- [0115] 所述识别模块202用于识别所述关键帧图像中的检测目标，并在所述检测目标上标记检测框。
- [0116] 在本实施方式中，所述检测目标可以是人、或动物、或交通工具等。所述关键帧图像中的所述检测目标可以包括多个，每一个所述检测目标对应一个检测框。可以理解的是，根据所述检测目标的不同，对应的检测框大小也不相同。
- [0117] 在本实施方式中，所述识别所述关键帧图像中的检测目标包括：
- [0118] 通过不同窗口大小的滑窗对所述关键帧图像按预设顺序进行滑动；
- [0119] 每次滑动时将当前窗口对应的图像输入至预先训练好的分类器；及
- [0120] 当所述分类器处理所述当前窗口对应的图像后输出分类概率大于预设值，则确认识别所述关键帧图像中的检测目标，输出检测目标的标记。
- [0121] 例如，首先对当前帧图像进行不同窗口大小的滑窗进行从左往右、从上到下的滑动。每次滑动时候对当前窗口执行分类器(分类器是事先训练好的)。如果当前窗口得到较高的分类概率，则认为检测到了目标。对每个不同窗口大小的滑窗都进行检测后，会得到不同窗口检测到的目标标记，这些窗口大小会存在重复较高的部分，最后采用非极大值抑制(Non-Maximum Suppression, NMS)的方法进行筛选以得到所述关键帧图像中的检测目标。最终，经过NMS筛选后获得检测到的目标。
- [0122] 筛选所述检测目标的标记以得到所述关键帧图像中的检测目标包括：
- [0123] 将所有滑窗输入至分类器后得到的概率按照降序排列，选取最大分类概率对应的滑窗；

- [0124] 分别计算其他滑窗与所述最大分类概率对应的滑窗之间的重叠面积，并判断所述重叠面积是否大于预设面积；
- [0125] 若所述重叠面积小于或等于所述预设面积，保留对应的滑窗并输出所述滑窗对应的检测目标；
- [0126] 若所述重叠面积大于所述预设面积，删除对应的滑窗；及
- [0127] 重复上述步骤，直到所有其他滑窗都比对完成。
- [0128] 所述识别模块202还用于识别所述预设组帧图像中的非关键帧图像中的检测目标，将所述关键帧图中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述非关键帧图像中的检测目标。
- [0129] 可以理解的是，所述识别当前非关键帧图像中的检测目标的方法可以与所述识别所述关键帧图像中的检测目标的方法一致，在此不再赘述。
- [0130] 在本实施方式中，识别每一组帧图像中每一非关键帧图像中的检测目标。所述前一帧图像可以是关键帧图像，也可以是非关键帧图像。
- [0131] 所述识别预设组帧图像中的非关键帧图像中的检测目标，将所述关键帧图中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述非关键帧图像中的检测目标包括如下步骤：
- [0132] 识别所述预设组帧图像中当前非关键帧图像中的检测目标；
- [0133] 将所述当前非关键帧图像之前的关键帧图像中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述当前非关键帧图像中的检测目标；
- [0134] 识别所述当前非关键帧图像之后的非关键帧图像中的检测目标；
- [0135] 将所述当前非关键帧图像中的预选框进行预设比例放大后生成第一预选框，通过所述第一预选框标记所述当前非关键帧图像之后的非关键帧图像中的检测目标，依此类推，直到标记完所述预设组帧图像中的所有非关键帧图像中的检测目标。
- [0136] 例如，若连续五帧图像分为一组，所述五帧图像中第一帧图像为关键帧图像，第二至第五帧图像为非关键帧图像。识别第一帧为关键帧图像中的检测目标，并在所述检测目标上标记检测框；识别第二帧为非关键帧图像中的检测目标，将所述关键帧图像中的检测框进行预设比例放大后生成第一预选框，通过所述

第一预选框标记所述第二帧图像中的检测目标；接着识别第三帧为非关键帧图像中的检测目标，将所述第二帧图像（非关键帧图像）中的检测框（即第一预选框）进行预设比例放大后生成第二预选框，通过所述第二预选框标记所述第三帧图像中的检测目标，依此类推，将每一组帧图像中的非关键帧图像中的检测目标都通过预选框进行标记。

[0137] 例如，若连续五帧图像分为一组，所述五帧图像中第二帧图像为关键帧图像，第一帧和第三至第五帧图像为非关键帧图像。识别第二帧为关键帧图像中的检测目标，并在所述检测目标上标记检测框；识别第三帧为非关键帧图像中的检测目标，将所述关键帧图像中的检测框进行预设比例放大后生成第一预选框，通过所述第一预选框标记所述第三帧图像中的检测目标；接着识别第四帧为非关键帧图像中的检测目标，将所述第三帧图像（非关键帧图像）中的检测框（即第一预选框）进行预设比例放大后生成第二预选框，通过所述第二预选框标记所述第四帧图像中的检测目标，依此类推，将每一组帧图像中的非关键帧图像中的检测目标都通过预选框进行标记。

[0138] 也就是说，并非限定每一组帧图像中的第一帧为关键帧图像，其他帧图像也可以作为关键帧图像。

[0139] 所述裁剪模块203用于根据所述预选框裁剪所述非关键帧图像得到多个预选框图像。

[0140] 例如，所述非关键帧图像中的检测目标包括五个，通过五个预选框分别标记所述五个检测目标，根据所述五个预选框裁剪所述非关键帧图像可以得到五个预选框图像。

[0141] 在处理720p以上的图片时，由于图片太大，处理起来计算量大且费时。因此，在本实施方式中，可以将所述非关键帧图像根据所述预选框进行裁剪得到预选框图像，再将所述预选框图像输入至非关键帧网络中处理。不仅可以大图片转换成小图片处理，还可以去除大图片中的非检测目标区域，从而可以提高处理速度。

[0142] 所述处理模块204用于将所述多个预选框图像输入至非关键帧检测网络，以得到所述检测目标在所述预选框中的位置。

- [0143] 在本实施方式中，所述非关键帧检测网络可以是预先训练的目标检测模型。其中，所述目标检测模型用于表征预选框图像与所述预选框图像包括的检测目标的位置信息的对应关系。
- [0144] 当所述预选框图像输入至所述目标检测模型时，可以确定所述预选框中的检测目标的位置信息。即，所述非关键帧网络是用于获取检测目标在图像中的位置的深度学习神经网络。
- [0145] 相比传统神经网络，深度神经网络具有更多神经元，能够对大规模不同位置、形态的区域目标图像样本进行学习；同时在数据准备阶段，通过增强变换方法可以对同一图像进行翻转变换、色调变换、尺度变换、噪声扰动、颜色变换等操作，从而生成多种不同环境特性的训练图像，使得神经网络检测器对于不同场景，不同环境，不同形状的区域目标都能够进行性能稳定的检测，保证充足的样本数据量，避免过拟合出现。另外，神经网络的Dropout机制能够在训练过程的前馈运算时随机剔除部分神经元，从而减少神经网络之间的相互依赖关系，让神经网络具备更强的泛化能力，避免过拟合。所述深度学习神经网络可以是卷积神经网络、深度信念网络和堆叠自动编码器。
- [0146] 作为实例，所述目标检测模型可以是技术人员基于对大量的预选框图像和预选框图像包括的检测目标的对应关系表，该对应关系表可以反映预选框图像和预选框图像包括的检测目标的对应关系。所述目标检测模型也可以是技术人员基于现有的人工神经网络进行有监督地训练而得到的图像识别模型。
- [0147] 优选地，在将所述多个预选框图像输入至非关键帧检测网络之前，所述视频物体加速检测装置20还可以：将所述多个预选框图像进行缩放处理。
- [0148] 可以理解的是，由于检测目标大小不一，其对应的预选框大小也各不相同。为了快速处理图像，需将所述多个预选框图像进行缩放处理成大小近似的图像，然后将同一大小后图像输入至非关键帧检测网络。
- [0149] 在本实施方式中，将所述非关键帧图像中裁剪出的预选框图像批量输入至非关键帧检测网络之前，可以将所述预选框图像缩放处理成大小相似的图像。如此，实现将统一大小后图像输入至非关键帧检测网络，便于所述非关键帧检测网络快速检测所述检测目标在所述预选框中的位置。

- [0150] 例如，所述当前非关键帧图像中的检测目标包括五个，通过五个预选框分别标记所述五个检测目标，根据所述五个预选框裁剪所述当前非关键帧图像可以得到五个预选框图像。所述五个预选框图像大小不一，需要进行缩放处理成相似大小的五个预选框图像。
- [0151] 所述处理模块204还用于根据所述预选框中的位置将所述检测目标映射至所述非关键帧对应的原图。
- [0152] 可以理解的是，所述检测目标在所述预选框中的位置是视频中的原图经过缩放处理后检测到的位置，因此，需要其进行映射至所述非关键帧对应的原图中以完成对视频图像中的目标的侦测。
- [0153] 综上所述，本申请提供的视频物体加速检测装置20，包括初始化模块201、识别模块202、裁剪模块203及处理模块204。所述初始化模块201用于将视频图像中的所有帧图像分为预设组帧图像，其中，每一组帧图像包括关键帧图像和非关键帧图像；所述识别模块202用于识别所述关键帧图像中的检测目标，并在所述检测目标上标记检测框；所述识别模块202还用于识别所述预设组帧图像中的非关键帧图像中的检测目标，将所述关键帧图中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述非关键帧图像中的检测目标；所述裁剪模块203用于根据所述预选框裁剪所述非关键帧图像得到多个预选框图像；所述处理模块204用于将所述多个预选框图像输入至非关键帧检测网络，以得到所述检测目标在所述预选框中的位置；及所述处理模块204还用于根据所述预选框中的位置将所述检测目标映射至所述非关键帧对应的原图。
- [0154] 主要针对检测网络在720p以上场景中由于检测锚点过多导致的检测速度慢问题，通过将视频图像分成关键帧图像和非关键帧图像，并对所述非关键帧图像进行简化运算，实现了无需对视频中的每一帧图像都进行精确的目标检测处理，充分利用了视频连续帧的空间相关性，在不损失精度的情况下，极大减少了锚点分类的时间消耗，且在基本不降低检测性能的情况下，大幅优化了检测速度。另外可以通过调节关键帧与非关键帧的比例，可以在不改变算法的基础上，应对不同硬件场景的需求。
- [0155] 上述以软件功能模块的形式实现的集成的单元，可以存储在一个计算机可读取

存储介质中。上述软件功能模块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，双屏设备，或者网络设备等）或处理器（processor）执行本申请各个实施例所述方法的部分。

[0156] 实施例三

[0157] 图3为本申请实施例三提供的服务器的示意图。

[0158] 所述服务器3包括：数据库31、存储器32、至少一个处理器33、存储在所述存储器32中并可在所述至少一个处理器33上运行的计算机可读指令34及至少一条通讯总线35。

[0159] 所述至少一个处理器33执行所述计算机可读指令34时实现上述视频物体加速检测方法实施例中的步骤。

[0160] 示例性的，所述计算机可读指令34可以被分割成一个或多个模块/单元，所述一个或者多个模块/单元被存储在所述存储器32中，并由所述至少一个处理器33执行，以完成本申请。所述一个或多个模块/单元可以是能够完成特定功能的一系列计算机可读指令指令段，该指令段用于描述所述计算机可读指令34在所述服务器3中的执行过程。

[0161] 所述服务器3是一种能够按照事先设定或存储的指令，自动进行数值计算和/或信息处理的设备，其硬件包括但不限于微处理器、专用集成电路（应用程序lication Specific Integrated Circuit, ASIC）、可编程门阵列（Field-Programmable Gate Array, FPGA）、数字处理器（Digital Signal Processor, DSP）、嵌入式设备等。本领域技术人员可以理解，所述示意图3仅仅是服务器3的示例，并不构成对服务器3的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如所述服务器3还可以包括输入输出设备、网络接入设备、总线等。

[0162] 所述数据库（Database）31是按照数据结构来组织、存储和管理数据的建立在所述服务器3上的仓库。数据库通常分为层次式数据库、网络式数据库和关系式数据库三种。在本实施方式中，所述数据库31用于存储所述视频图像。

[0163] 所述至少一个处理器33可以是中央处理单元（Central Processing Unit, CPU），还可以是其他通用处理器、数字信号处理器（Digital Signal Processor

，DSP）、专用集成电路（Application Specific Integrated Circuit, ASIC）、现成可编程门阵列（Field-Programmable Gate Array, FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。该处理器33可以是微处理器或者该处理器33也可以是任何常规的处理器等，所述处理器33是所述服务器3的控制中心，利用各种接口和线路连接整个服务器3的各个部分。

[0164] 所述存储器32可用于存储所述计算机可读指令34和/或模块/单元，所述处理器33通过运行或执行存储在所述存储器32内的计算机可读指令和/或模块/单元，以及调用存储在存储器32内的数据，实现所述服务器3的各种功能。所述存储器32可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序（比如声音播放功能、图像播放功能等）等；存储数据区可存储根据服务器3的使用所创建的数据（比如音频数据、电话本等）等。此外，存储器32可以包括高速随机存取存储器，还可以包括非易失性存储器，例如硬盘、内存、插接式硬盘，智能存储卡（Smart Media Card, SMC），安全数字（Secure Digital, SD）卡，闪存卡（Flash Card）、至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

[0165] 所述存储器32中存储有程序代码，且所述至少一个处理器33可调用所述存储器32中存储的程序代码以执行相关的功能。例如，图2中所述的各个模块（初始化模块201、识别模块202、裁剪模块203及处理模块204）是存储在所述存储器32中的程序代码，并由所述至少一个处理器33所执行，从而实现所述各个模块的功能以达到视频物体加速检测目的。

[0166] 所述初始化模块201用于将视频图像中的所有帧图像分为预设组帧图像，其中，每一组帧图像包括关键帧图像和非关键帧图像；

[0167] 所述识别模块202用于识别所述关键帧图像中的检测目标，并在所述检测目标上标记检测框；

[0168] 所述识别模块202还用于识别所述预设组帧图像中的非关键帧图像中的检测目标，将所述关键帧图中的检测框进行预设比例放大后生成预选框，通过所述预选框标记所述非关键帧图像中的检测目标；

[0169] 所述裁剪模块203用于根据所述预选框裁剪所述非关键帧图像得到多个预选框

图像；所述处理模块204用于将所述多个预选框图像输入至非关键帧检测网络，以得到所述检测目标在所述预选框中的位置；及

[0170] 所述处理模块204还用于根据所述预选框中的位置将所述检测目标映射至所述非关键帧对应的原图。

[0171] 所述服务器3集成的模块/单元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，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请实现上述实施例方法中的全部或部分流程，也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计算机可读存储介质中，该计算机可读指令在被处理器执行时，可实现上述各个方法实施例的步骤。其中，所述计算机可读指令包括计算机可读指令代码，所述计算机可读指令代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括：能够携带所述计算机可读指令代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、电载波信号、电信信号以及软件分发介质等。需要说明的是，所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减，例如在某些司法管辖区，根据立法和专利实践，计算机可读介质不包括电载波信号和电信信号。

[0172] 尽管未示出，所述服务器3还可以包括给各个部件供电的电源（比如电池），优选的，电源可以通过电源管理系统与所述至少一个处理器33逻辑相连，从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。电源还可以包括一个或一个以上的直流或交流电源、再充电系统、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。所述服务器3还可以包括蓝牙模块、Wi-Fi模块等，在此不再赘述。

[0173] 在一实施例中，提供一个或多个存储有计算机可读指令的可读存储介质，所述计算机可读存储介质存储有计算机可读指令，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实施例中视频物体加速检测方法，为避免重复，这里不再赘述。本实施例中的可读存储介质包

括非易失性可读存储介质和易失性可读存储介质。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，该计算机可读指令可存储于一非易失性可读存储介质也可以存储在易失性可读存储介质中，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。

[0174] 应该了解，所述实施例仅为说明之用，在专利申请范围上并不受此结构的限制。

[0175] 在本申请所提供的几个实施例中，应该理解到，所揭露的电子设备和方法，可以通过其它的方式实现。例如，以上所描述的电子设备实施例仅仅是示意性的，例如，所述单元的划分，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。

[0176] 另外，在本申请各个实施例中的各功能单元可以集成在相同处理单元中，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相同单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用硬件加软件功能模块的形式实现。

[0177] 对于本领域技术人员而言，显然本申请不限于上述示范性实施例的细节，而且在不背离本申请的精神或基本特征的情况下，能够以其他的具体形式实现本申请。因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本申请的范围由所附权利要求而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本申请内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。此外，显然“包括”一词不排除其他单元或，单数不排除复数。系统权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一，第二等词语用来表示名称，而并不表示任何特定的顺序。

[0178] 最后应说明的是，以上实施例仅用以说明本申请的技术方案而非限制，尽管参照较佳实施例对本申请进行了详细说明，本领域的普通技术人员应当理解，可以对本申请的技术方案进行修改或等同替换，而不脱离本申请技术方案的精神范围。

发明概述

技术问题

问题的解决方案

发明的有益效果