



(12) 发明专利

(10) 授权公告号 CN 112215171 B

(45) 授权公告日 2024. 01. 26

(21) 申请号 202011104694.5

G06V 20/40 (2022.01)

(22) 申请日 2020.10.15

G06V 10/764 (2022.01)

(65) 同一申请的已公布的文献号

G06V 10/766 (2022.01)

申请公布号 CN 112215171 A

G06N 3/0464 (2023.01)

G06N 3/082 (2023.01)

(43) 申请公布日 2021.01.12

(73) 专利权人 腾讯科技(深圳)有限公司

(56) 对比文件

CN 109559300 A, 2019.04.02

地址 518057 广东省深圳市南山区高新区

CN 111738133 A, 2020.10.02

科技中一路腾讯大厦35层

葛明进; 孙作雷; 孔薇. 基于anchor-free的交通场景目标检测技术. 计算机工程与科学. 2020, (第04期), 全文.

(72) 发明人 诸加丹

审查员 聂一琴

(74) 专利代理机构 北京派特恩知识产权代理有

限公司 11270

专利代理师 赵翠萍 张颖玲

(51) Int. Cl.

G06V 10/25 (2022.01)

G06V 10/82 (2022.01)

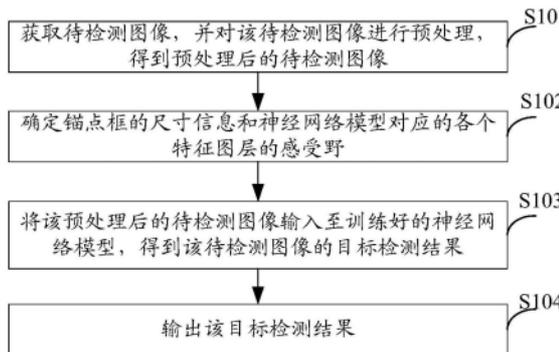
权利要求书2页 说明书18页 附图8页

(54) 发明名称

目标检测方法、装置、设备及计算机可读存储介质

(57) 摘要

本申请提供了一种目标检测方法、装置、设备及计算机可读存储介质,该方法包括:获取待检测图像,并对该待检测图像进行预处理,得到预处理后的待检测图像;将该预处理后的待检测图像输入至训练好的神经网络模型,得到该待检测图像的目标检测结果,该训练好的神经网络模型对应的各个特征图层的感受野与锚点框的尺寸信息满足匹配条件;输出该目标检测结果。通过本申请,能够提高目标检测的检测精度,从而实现精准推荐或者自动智能审核。



1. 一种目标检测方法,其特征在于,包括:

获取待检测图像,并对所述待检测图像进行预处理,得到预处理后的待检测图像;

获取多个样本图像和所述多个样本图像的标注信息,所述标注信息至少包括标注对象对应的标注框的尺寸信息;对多个标注框的尺寸信息进行聚类处理,得到N个锚点框的尺寸信息,N为大于1的整数;获取预设的神经网络模型;基于所述神经网络模型参数信息,按照

$$r_0 = \frac{\left(\sum_{l=1}^L ((k_l - 1) \prod_{i=1}^{l-1} s_i) + 1 \right)}{\sqrt{L}}$$

确定所述神经网络模型中各个特征图层对应的感受野大小, r_0

表示当前特征图层每个点的感受野,L表示当前特征图层之前的网络的层数, k_1 表示第1层的卷积核大小, s_i 表示i层卷积核的步长;

将所述预处理后的待检测图像输入至训练好的神经网络模型,得到所述待检测图像的目标检测结果,所述训练好的神经网络模型对应的各个特征图层的所述感受野与所述锚点框的尺寸信息满足匹配条件;

输出所述目标检测结果。

2. 根据权利要求1中所述的方法,其特征在于,所述方法还包括:

基于所述各个特征图层对应的感受野大小和N个锚点框的尺寸信息,确定各个锚点框对应的特征图层;

将各个锚点框分配至对应的特征图层。

3. 根据权利要求2中所述的方法,其特征在于,所述基于所述各个特征图层对应的感受野大小和N个锚点框的尺寸信息,确定各个锚点框对应的特征图层,包括:

基于各个特征图层对应的感受野大小和第i个锚点框的尺寸信息,确定第i个锚点框与各个感受野大小的各个差值,i为正整数;

将各个差值中的最小差值对应的特征图层确定为第i目标特征图层;

确定第i目标层为第i个锚点框对应的特征图层,并确定所述第i目标特征图层的感受野与所述第i个锚点框的尺寸信息满足匹配条件。

4. 根据权利要求2或3所述的方法,其特征在于,所述方法还包括:

将各个样本图像分别输入所述神经网络模型,得到各个样本图像对应的预测结果;

利用各个样本图像对应的预测结果和各个样本图像的标注信息对所述神经网络模型进行反向传播训练,以对所述神经网络模型的参数进行调整,以得到训练好的神经网络模型。

5. 根据权利要求1中所述的方法,其特征在于,所述获取待检测图像,包括:

获取待检测的视频文件;

对所述待检测的视频文件进行解析,得到各个视频帧图像;

将所述各个视频帧图像确定为各个待检测图像。

6. 根据权利要求5中所述的方法,特征在于,所述方法还包括:

获取所述各个视频帧图像的目标检测结果;

基于各个视频帧图像的目标检测结果,确定所述待检测的视频文件中目标检测结果为包括目标对象的视频帧图像的图像个数;

基于所述图像个数确定所述待检测的视频文件的检测结果,其中,当所述图像个数大

于预设的个数阈值时,确定所述待检测的视频文件的检测结果为包括目标对象。

7.根据权利要求6中所述的方法,其特征在於,所述方法还包括:

当所述视频文件的检测结果为包括目标对象时,确定所述视频文件为目标视频文件;
在确定达到预设的删除时机时,删除所述目标视频文件。

8.根据权利要求6中所述的方法,其特征在於,所述方法还包括:

当所述视频文件的检测结果为包括目标对象时,确定所述视频文件为目标视频文件;
在确定达到预设的推荐时机时,确定目标推荐对象;
向所述目标推荐对象推荐该目标视频文件。

9.一种目标检测装置,其特征在於,包括:

第一获取模块,用于获取待检测图像,并对所述待检测图像进行预处理,得到预处理后的待检测图像;

第一确定模块,用于获取多个样本图像和所述多个样本图像的标注信息,所述标注信息至少包括标注对象对应的标注框的尺寸信息;对多个标注框的尺寸信息进行聚类处理,得到N个锚点框的尺寸信息,N为大于1的整数;获取预设的神经网络模型;基于所述神经网络模型的参数信息,按照

$$r_0 = \frac{\left(\sum_{l=1}^L ((k_l - 1) \prod_{i=1}^{l-1} s_i) + 1 \right)}{\sqrt{L}}$$

确定所述神经网络模型中各个特征图层

对应的感受野大小, r_0 表示当前特征图层每个点的感受野,L表示当前特征图层之前的网络的层数, k_1 表示第1层的卷积核大小, s_i 表示i层卷积核的步长;

目标检测模块,用于将所述预处理后的待检测图像输入至训练好的神经网络模型,得到所述待检测图像的目标检测结果,

所述训练好的神经网络模型对应的各个特征图层的所述感受野与所述锚点框的尺寸信息满足匹配条件;

输出模块,用于输出所述目标检测结果。

10.一种目标检测设备,其特征在於,包括:

存储器,用于存储可执行指令;

处理器,用于执行所述存储器中存储的可执行指令时,实现权利要求1至8任一项所述的方法。

11.一种计算机可读存储介质,其特征在於,存储有可执行指令,用于被处理器执行时,实现权利要求1至8任一项所述的方法。

目标检测方法、装置、设备及计算机可读存储介质

技术领域

[0001] 本申请涉及图像处理技术,尤其涉及一种目标检测方法、装置、设备及计算机可读存储介质。

背景技术

[0002] 图像分类、目标检测及分割是计算机视觉领域的三大任务,其中目标检测广泛应用于机器人导航、智能视频监控、航空航天等诸多领域,通过计算机视觉减少对人力资本的消耗,具有重要的现实意义。因此,目标检测也就成为了近年来理论和应用的研究热点。

[0003] 由于深度学习的广泛运用,目标检测算法得到了较为快速的发展。目前基于深度神经网络模型的目标检测方法又可以分为anchor-free和基于锚点(anchor-base),其中anchor-based算法在anchor的分配上具有局限性,从而降低了检测结果的精度。

发明内容

[0004] 本申请实施例提供一种目标检测方法、装置及计算机可读存储介质,能够提高目标检测的检测精度。

[0005] 本申请实施例的技术方案是这样实现的:

[0006] 本申请实施例提供一种目标检测方法,包括:

[0007] 获取待检测图像,并对该待检测图像进行预处理,得到预处理后的待检测图像;

[0008] 确定锚点框的尺寸信息和神经网络模型对应的各个特征图层的感受野;

[0009] 将该预处理后的待检测图像输入至训练好的神经网络模型,得到该待检测图像的目标检测结果,

[0010] 该训练好的神经网络模型对应的各个特征图层的感受野与锚点框的尺寸信息满足匹配条件;

[0011] 输出该目标检测结果。

[0012] 本申请实施例提供一种目标检测装置,包括:。

[0013] 第一获取模块,用于获取待检测图像,并对该待检测图像进行预处理,得到预处理后的待检测图像;

[0014] 第一确定模块,用于确定锚点框的尺寸信息和神经网络模型对应的各个特征图层的感受野;

[0015] 目标检测模块,用于将该预处理后的待检测图像输入至训练好的神经网络模型,得到该待检测图像的目标检测结果,

[0016] 该训练好的神经网络模型对应的各个特征图层的感受野与锚点框的尺寸信息满足匹配条件;

[0017] 输出模块,用于输出该目标检测结果。

[0018] 在一些实施例中,该第一确定模块,还用于:

[0019] 获取多个样本图像和该多个样本图像的标注信息,该标注信息至少包括标注对象

对应的标注框的尺寸信息；

[0020] 对多个标注框的尺寸信息进行聚类处理,得到N个锚点框的尺寸信息,N 为大于1的整数。

[0021] 在一些实施例中,该第一确定模块,还用于:

[0022] 获取预设的神经网络模型;

[0023] 基于该神经网络模型的参数信息确定该神经网络模型中各个特征图层对应的感受野大小;

[0024] 在一些实施例中,该装置还包括:

[0025] 第二确定模块,用于基于该各个特征图层对应的感受野大小和N个锚点框的尺寸信息,确定各个锚点框对应的特征图层;

[0026] 锚点分配模块,用于将各个锚点框分配至对应的特征图层。

[0027] 在一些实施例中,该第二确定模块,还用于:

[0028] 基于各个特征图层对应的感受野大小和第i个锚点框的尺寸信息,确定第 i个锚点框与各个感受野大小的各个差值,i为正整数;

[0029] 将各个差值中的最小差值对应的特征图层确定为第i目标特征图层;

[0030] 确定第i目标层为第i个锚点框对应的特征图层,并且确定该第i目标特征图层的感受野与该第i个锚点框的尺寸信息满足匹配条件。

[0031] 在一些实施例中,该装置还包括:

[0032] 输入模块,用于将各个样本图像分别输入该神经网络模型,得到各个样本图像对应的预测结果;

[0033] 训练模块,用于利用各个样本图像对应的预测结果和各个样本图像的标注信息对该神经网络模型进行反向传播训练,以对该神经网络模型的参数进行调整,以得到训练好的神经网络模型。

[0034] 在一些实施例中,该第一获取模块还用于:

[0035] 获取待检测的视频文件;

[0036] 对该待检测的视频文件进行解析,得到各个视频帧图像;

[0037] 将该各个视频帧图像确定为各个待检测图像。

[0038] 在一些实施例中,该装置还包括:

[0039] 第二获取模块,用于获取该各个视频帧图像的目标检测结果;

[0040] 第三确定模块,用于基于各个视频帧图像的目标检测结果,确定该待检测的视频文件中目标检测结果为包括目标对象的视频帧图像的图像个数;

[0041] 第四确定模块,用于基于该图像个数确定该待检测的视频文件的检测结果,其中,当该图像个数大于预设的个数阈值时,确定该待检测的视频文件的检测结果为包括目标对象。

[0042] 在一些实施例中,该装置还包括:

[0043] 第五确定模块,用于当该视频文件的检测结果为包括目标对象时,确定该视频文件为目标视频文件;

[0044] 删除模块,用于在确定达到预设的删除时机时,删除该目标视频文件。

[0045] 在一些实施例中,该装置还包括:

- [0046] 第六确定模块,用于当该视频文件的检测结果为包括目标对象时,确定该视频文件为目标视频文件;
- [0047] 第七确定模块,用于在确定达到预设的推荐时机时,确定目标推荐对象;
- [0048] 推荐模块,用于向该目标推荐对象推荐该目标视频文件。
- [0049] 本申请实施例提供一种目标检测设备,包括:
- [0050] 存储器,用于存储可执行指令;
- [0051] 处理器,用于执行该存储器中存储的可执行指令时,实现本申请实施例提供的方法。
- [0052] 本申请实施例提供一种计算机可读存储介质,存储有可执行指令,用于引起处理器执行时,实现本申请实施例提供的方法。
- [0053] 本申请实施例具有以下有益效果:
- [0054] 在本申请实施例提供的目标检测方法中,在获取到待检测图像之后,首先对该待检测图像进行预处理,从而将待检测图像处理为训练好的神经网络模型所要求的尺寸和格式,得到预处理后的待检测图像,然后将预处理后的待检测图像输入至训练好的神经网络模型,得到待检测图像的目标检测结果,并输出该目标检测结果,在本申请实施例中,该训练好的神经网络模型对应的各个特征图层的感受野与锚点框的尺寸信息满足匹配条件,使得在目标检测时所进行的分类和回归都能在跟目标对象大小一致的感受野上进行,从而极大的提高检测性能和检测准确度。

附图说明

- [0055] 图1A为相关技术中Yolo的检测框架示意图;
- [0056] 图1B为相关技术中retinaNet的检测框架示意图;
- [0057] 图1C为feature map的点和输入图像的关系示意图;
- [0058] 图1D为不同层feature map中感受野示意图;
- [0059] 图2为本申请实施例提供的目标检测系统的网络架构示意图;
- [0060] 图3为本申请实施例提供的服务器300的结构示意图;
- [0061] 图4为本申请实施例提供的目标检测方法的一种实现流程示意图;
- [0062] 图5为本申请实施例提供的为特征图层分配锚点框的实现流程示意图;
- [0063] 图6为本申请实施例提供的目标检测方法的再一种实现流程示意图;
- [0064] 图7为本申请实施例提供的目标检测方法的检测框架示意图;
- [0065] 图8为本申请实施例提供的目标检测网络的结构示意图。

具体实施方式

[0066] 为了使本申请的目的、技术方案和优点更加清楚,下面将结合附图对本申请作进一步地详细描述,所描述的实施例不应视为对本申请的限制,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它实施例,都属于本申请保护的范围。

[0067] 在以下的描述中,涉及到“一些实施例”,其描述了所有可能实施例的子集,但是可以理解,“一些实施例”可以是所有可能实施例的相同子集或不同子集,并且可以在不冲突的情况下相互结合。

[0068] 如果申请文件中出现“第一/第二”的类似描述则增加以下的说明,在以下的描述中,所涉及的术语“第一\第二\第三”仅仅是是区别类似的对象,不代表针对对象的特定排序,可以理解地,“第一\第二\第三”在允许的情况下可以互换特定的顺序或先后次序,以使这里描述的本申请实施例能够以除了在这里图示或描述的以外的顺序实施。

[0069] 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。本文中所使用的术语只是为了描述本申请实施例的目的,不是旨在限制本申请。

[0070] 对本申请实施例进行进一步详细说明之前,对本申请实施例中涉及的名词和术语进行说明,本申请实施例中涉及的名词和术语适用于如下的解释。

[0071] 1) 锚点框(anchor),为目标检测中的参考框,能够带来先验知识,一般使用一个矩形的边框来表示。

[0072] 2) 特征图(feature map),一般指的是卷积网络里面的卷积层的输出。

[0073] 3) 特征金字塔,是用于检测不同尺度的对象的识别系统中的基本组件,在特征金字塔中低分辨率、高语义信息的高层特征和高分辨率、低语义信息的低层特征可以是自上而下或者自下而上的侧边连接,使得所有尺度下的特征都有丰富的语义信息。

[0074] 4) 特征金字塔网络,是根据特征金字塔概念设计的特征提取器,目的是提高精度和速度,它替代了类如Faster R-CNN中的特征提取器,并且生成更高质量的特征图金字塔。

[0075] 5) 感受野,卷积神经网络每一层输出的特征图(feature map)上的像素点在输入图片上映射的区域大小,也可以认为是神经网络中神经元“看到的”输入区域,在卷积神经网络中,feature map上某个元素的计算受输入图像上某个区域的影响,这个区域即该元素的感受野。

[0076] 6) 损失函数(Loss Function),又称为代价函数(cost function)是将随机事件或其有关随机变量的取值映射为非负实数以表示该随机事件的“风险”或“损失”的函数。在应用中,损失函数通常作为学习准则与优化问题相联系,即通过最小化损失函数求解和评估模型。例如在统计学和机器学习中被用于模型的参数估计,是机器学习模型的优化目标。

[0077] 云技术(Cloud technology)是指在广域网或局域网内将硬件、软件、网络等系列资源统一起来,实现数据的计算、储存、处理和共享的一种托管技术。

[0078] 云技术(Cloud technology)基于云计算商业模式应用的网络技术、信息技术、整合技术、管理平台技术、应用技术等的总称,可以组成资源池,按需所用,灵活便利。云计算技术将变成重要支撑。技术网络系统的后台服务需要大量的计算、存储资源,如视频网站、图片类网站和更多的门户网站。伴随着互联网行业的高度发展和应用,将来每个物品都有可能存在自己的识别标志,都需要传输到后台系统进行逻辑处理,不同程度级别的数据将会分开处理,各类行业数据皆需要强大的系统后盾支撑,只能通过云计算来实现。此外,还涉及到云测试、云监管等方面。

[0079] 人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说,人工智能是计算机科学的一个综合技术,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法,使机器具有感知、推理与决策的功能。

[0080] 人工智能技术是一门综合学科,涉及领域广泛,既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0081] 机器学习(Machine Learning,ML)是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、式教学习等技术。

[0082] 随着人工智能技术研究和进步,人工智能技术在多个领域展开研究和应用,例如常见的智能家居、智能穿戴设备、虚拟助理、智能音箱、智能营销、无人驾驶、自动驾驶、无人机、机器人、智能医疗、智能客服等,相信随着技术的发展,人工智能技术将在更多的领域得到应用,并发挥越来越重要的价值。

[0083] 为了更好地理解本申请实施例中提供的目标检测方法,首先对相关技术中目标检测方法及存在的缺点进行说明。

[0084] 经典的anchor based目标检测方法,如Yolo(You Only Look Once)系列和retianNet,是单阶段的主流检测算法,具有快速高效的优点。之所以统称为anchor based方法,是因为在分类和回归计算中,利用的是anchor(先验的包围盒(bbox, bounding box)的尺寸信息)来作为回归坐标回归的参考系统。其中,图1A为相关技术中Yolo的检测框架示意图,如图1A所示,在Yolo系列中,特征提取网络100在得到特征向量后,利用特征金字塔网络101(FPN,Feature Pyramid Networks)对特征向量进行卷积,得到3个feature map,从而得到特征金字塔后,anchor按尺寸排序后(一般地,anchor总数为9),不重复的均匀分配到特征图层进行对应的坐标回归和分类。例如,可以为最底层的特征图层分配 anchor [1:3],为中间层的特征图层分配anchor [4:6],为最高层的特征图层分配 anchor [7:9],图1A中示例性示出了中间一层的特征图层的检测头102,负责 anchor [4:6]的检测,其中,中间层的检测结果103中w*h表示特征图层的尺寸, num_anchors表示anchor的个数(此例中num_anchors=3), (4+1)中4表示的是预测框的4个坐标,1表示前景的概率,num_classes表示类别数目。

[0085] 图1B为相关技术中retinaNet的检测框架示意图,如图1B所示,该检测框架包括特征提取网络110、一个骨干网络和两个特定任务子网络组成的单一网络,其中,特征提取网络110可以是ResNet网络、骨干网络也即FPN网络111,负责得到特征金字塔,两个特定任务子网络也即分类子网络112(class subnets)和盒子子网络113(box subnets),其中,分类子网络在特征金字塔网络的输出上执行卷积对象分类,盒子子网络执行卷积边界框回归,并且为了提高召回率,对每一层的特征图层都负责全部的预定义anchor个数k(一般固定的k=9)的分类和坐标回归。其中图2中的K为num_classes,即类别总数。可以看到目前的基于anchor的检测方法都是在每一层特征图层上分配固定的anchor个数。

[0086] 在上述相关技术中anchorbased的目标检测方法中,通过全卷积网络得到的feature map,然后通过feature map直接在预设的anchor上进行回归和分类,并且分类和

回归都是对feature map上的单个点上进行的。图1C为feature map的点和输入图像的关系示意图,如图1C所示(为了简化讨论,此处做一维的展示,二维情况类似)。如果在该特征图层上预测一个远小于感受野的目标(图1C中物体121),则feature map上的这个点包含了过多背景信息,会严重影响分类结果;反之如果在该特征图层上预测一个远大于感受野的目标(图1C中物体122),则feature map上的这个点只包含了物体的局部,分类信息不完整,也会影响分类结果。

[0087] feature map的每个点映射回输入图,实际上都是一个固定小区域s,图1D为不同层feature map中感受野示意图,如图1D所示,区域131为最高层feature map空间上一个点在原图中的实际感受野,区域132则为中层feature map上的点的感受野,区域133则为最底层的特征图上点的感受野。对于检测中的分类任务,则完全依靠固定小区域s的信息来进行预测。图1D中,如果要检测猫,则最优的策略是先验的anchor放在最高层的feature map(因为最高层的feature map的感受野可以覆盖到整个猫),如果要检测猫的眼睛,最优策略是先验的 anchor放在对底层的feature map(因为最底层的feature map的感受野很小,与猫眼大小相当)。如果反过来,则用整只猫来预测猫眼,和用猫眼的信息来预测猫,都是不合理的。虽然Yolo系列中,先验anchor是通过排序后从大到小的均匀分配到由顶至底的feature map的,但是这种分配是固定的,并没有结合具体的实际情况(训练集的先验信息)来调整的,会在实际应用中出现不合理的情况:大的感受野信息来预测小的物体,或者小的感受野来预测大的物体。比如在台标检测中,所有的台标相对于整图来说都是占比非常小的目标(可以参考图1D中的猫眼睛的比例)。但是现有方法会把anchor均匀分配到三层特征图层,这样非常影响检测中的分类效果,导致检测小目标精度很低。

[0088] 基于上述问题,本申请实施例提供一种目标检测方法,通过anchor的分配与feature map的感受野匹配的策略,实现anchor与感受野最接近的feature map相匹配,从而提高检测性能。

[0089] 下面说明本申请实施例提供的目标检测设备的示例性应用,本申请实施例提供的目标检测设备设备可以实施为笔记本电脑,平板电脑,台式计算机,机顶盒,移动设备(例如,移动电话,便携式音乐播放器,个人数字助理,专用消息设备,便携式游戏设备)等各种类型的用户终端,也可以实施为服务器。下面,将说明设备实施为服务器时示例性应用。

[0090] 参见图2,图2为本申请实施例提供的目标检测系统20网络架构示意图,如图2所示,该目标检测系统20包括:终端100通过网络200连接服务器300,网络200可以是广域网或者局域网,又或者是二者的组合。

[0091] 在该目标检测系统20中,可以是终端100向服务器300发送视频发布请求,该视频发布请求中携带有待发布的视频,该待发布的视频可以是用户通过终端100录制得到的,还可以是终端100下载到本地的,服务器300在接收到该视频发布请求后,利用训练好的神经网络模型对待发布的视频利用本申请实施例提供的方法进行目标检测,以检测待发布的视频中是否包括预设的网站水印、预设的台标等目标对象,得到检测结果,当基于检测结果确定待发布的视频满足发布要求时,将该视频进行发布,并向终端100返回发布成功的响应消息;当检测结果确定待发布的视频不满足发布要求时,向终端100返回发布失败的响应消息。

[0092] 服务器300可以是专门进行目标检测的服务器,还可以是与业务服务器为同一服

务器,例如可以是视频服务器、即时通讯服务器等等。当服务器300为专门进行目标检测的服务器时,终端100在发布视频文件时,可以先将携带有待发布的视频文件的视频发布请求发送至服务器300,由服务器300对视频文件的各个视频帧图像进行目标检测,从而确定视频文件的检测结果,当基于该检测结果确定满足发布条件时,服务器300再将待发布的视频文件发送至业务服务器,以由业务服务器发布该视频文件。

[0093] 在一些实施例中,服务器300可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN、以及大数据和人工智能平台等基础云计算服务的云服务器。终端100可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表等,但并不局限于此。终端以及服务器可以通过有线或无线通信方式进行直接或间接地连接,本发明实施例中不做限制。

[0094] 参见图3,图3为本申请实施例提供的服务器300的结构示意图,图3所示的服务器300包括:至少一个处理器310、存储器350、至少一个网络接口320和用户接口330。服务器300中的各个组件通过总线系统340耦合在一起。可理解,总线系统340用于实现这些组件之间的连接通信。总线系统340除包括数据总线之外,还包括电源总线、控制总线和状态信号总线。但是为了清楚说明起见,在图3中将各种总线都标为总线系统340。

[0095] 处理器310可以是一种集成电路芯片,具有信号的处理能力,例如通用处理器、数字信号处理器(DSP, Digital Signal Processor),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,其中,通用处理器可以是微处理器或者任何常规的处理器等。

[0096] 用户接口330包括使得能够呈现媒体内容的一个或多个输出装置331,包括一个或多个扬声器和/或一个或多个视觉显示屏。用户接口330还包括一个或多个输入装置332,包括有助于用户输入的用户接口部件,比如键盘、鼠标、麦克风、触屏显示屏、摄像头、其他输入按钮和控件。

[0097] 存储器350可以是可移除的,不可移除的或其组合。示例性的硬件设备包括固态存储器,硬盘驱动器,光盘驱动器等。存储器350可选地包括在物理位置上远离处理器310的一个或多个存储设备。存储器350包括易失性存储器或非易失性存储器,也可包括易失性和非易失性存储器两者。非易失性存储器可以是只读存储器(ROM, Read Only Memory),易失性存储器可以是随机存取存储器(RAM, Random Access Memory)。本申请实施例描述的存储器350旨在包括任意适合类型的存储器。在一些实施例中,存储器350能够存储数据以支持各种操作,这些数据的示例包括程序、模块和数据结构或者其子集或超集,下面示例性说明。

[0098] 操作系统351,包括用于处理各种基本系统服务和执行硬件相关任务的系统程序,例如框架层、核心库层、驱动层等,用于实现各种基础业务以及处理基于硬件的任务;

[0099] 网络通信模块352,用于经由一个或多个(有线或无线)网络接口320到达其他计算设备,示例性的网络接口320包括:蓝牙、无线相容性认证(WiFi)、和通用串行总线(USB, Universal Serial Bus)等;

[0100] 输入处理模块353,用于对一个或多个来自一个或多个输入装置332之一的一个或多个用户输入或互动进行检测以及翻译所检测的输入或互动。

[0101] 在一些实施例中,本申请实施例提供的装置可以采用软件方式实现,图3示出了

存储在存储器350中的一种目标检测装置354,该目标检测装置354可以是服务器300中的目标检测装置,其可以是程序和插件等形式的软件,包括以下软件模块:第一获取模块3541、第一确定模块3542、目标检测模块3543 和输出模块3544,这些模块是逻辑上的,因此根据所实现的功能可以进行任意的组合或进一步拆分。将在下文中说明各个模块的功能。

[0102] 在另一些实施例中,本申请实施例提供的装置可以采用硬件方式实现,作为示例,本申请实施例提供的装置可以是采用硬件译码处理器形式的处理器,其被编程以执行本申请实施例提供的目标检测方法,例如,硬件译码处理器形式的处理器可以采用一个或多个应用专用集成电路(ASIC,Application Specific Integrated Circuit)、DSP、可编程逻辑器件(PLD,Programmable Logic Device)、复杂可编程逻辑器件(CPLD,Complex Programmable Logic Device)、现场可编程门阵列(FPGA,Field-Programmable Gate Array)或其他电子元件。

[0103] 将结合本申请实施例提供的服务器的示例性应用和实施,说明本申请实施例提供的目标检测方法。

[0104] 参见图4,图4为本申请实施例提供的目标检测方法的一种实现流程示意图,将结合图4示出的步骤进行说明。

[0105] 步骤S101,获取待检测图像,并对该待检测图像进行预处理,得到预处理后的待检测图像。

[0106] 这里,步骤S101在实现时,可以是接收到终端发送的目标检测请求后,从目标检测请求中解析得到待检测图像,还可以是基于目标检测请求中携带的待检测图像的标识,获取待检测图像。待检测图像可以是一个个单独的图像,还可以是从视频中解析出来的多个视频帧图像。在获取到待检测图像之后,可以获取待检测图像的属性信息,然后基于待检测图像的属性信息对该待检测图像进行预处理,得到预处理后的待检测图像。在实际实现时,对待检测图像进行预处理,可以是将待检测图像的尺寸处理为神经网络模型所要求的尺寸。例如,神经网络模型所要求的图像尺寸为416*416,如果待检测图像的尺寸为1000*1000,那么此时需要对待检测图像进行下采样,从而将待检测图像的尺寸处理为 416*416。在一些实施例中,还可以是将待检测图像处理为神经网络模型所要求的色彩空间,例如,待检测图像为YUV彩色图像,而神经网络模型所要求的色彩空间为RGB,那么此时需要对待检测图像进行预处理,将待检测图像处理为RGB彩色图像。

[0107] 步骤S102,确定锚点框的尺寸信息和神经网络模型对应的各个特征图层的感受野。

[0108] 步骤S103,将该预处理后的待检测图像输入至训练好的神经网络模型,得到该待检测图像的目标检测结果。

[0109] 这里,训练好的神经网络模型对应的各个特征图层的感受野与锚点框的尺寸信息满足匹配条件,也就是说,在该训练好的神经网络模型中,并不是将anchor均匀分配到各个特征图层,而是根据各个特征图层的感受野大小来分配适合的anchor,使得后续的分类和回归都能在跟目标对象大小一致的感受野上进行,从而提高目标检测的性能,尤其能够提高对小目标检测任务的检测准确率和召回率。

[0110] 目标检测结果至少包括待检测图像中包括目标对象的概率值,以表征待检测图像中是否包括目标对象,当目标检测结果表征待检测图像中包括目标对象时,目标检测结果

中还可以包括目标对象在待检测图像中的位置信息。

[0111] 步骤S104,输出该目标检测结果。

[0112] 这里,当目标检测结果为待检测图像中不包括目标对象时,可以是直接输出目标检测结果,当目标检测结果为待检测图像中包括目标对象时,输出该目标检测结果可以是基于目标对象的位置信息对待检测图像进行标记,然后输出标记后的待检测图像,从而能够使得能够直观的了解目标对象在待检测图像中的位置。

[0113] 在一些实施例中,输出目标检测结果可以是直接在目标检测设备的显示装置中显示目标检测结果,还可以是将目标检测结果发送至请求进行目标检测的终端。

[0114] 在本申请实施例提供的目标检测方法中,在获取到待检测图像之后,首先对该待检测图像进行预处理,从而将待检测图像处理为训练好的神经网络模型所要求的尺寸和格式,得到预处理后的待检测图像,然后将预处理后的待检测图像输入至训练好的神经网络模型,得到待检测图像的目标检测结果,并输出该目标检测结果,在本申请实施例中,该训练好的神经网络模型对应的各个特征图层的感受野与锚点框的尺寸信息满足匹配条件,使得在目标检测时所进行的分类和回归都能在跟目标对象大小一致的感受野上进行,从而极大的提高检测性能和检测准确度。

[0115] 在本申请实施例提供的目标检测方法中,在获取到待检测图像之后,首先对该待检测图像进行预处理,从而将待检测图像处理为训练好的神经网络模型所要求的尺寸和格式,得到预处理后的待检测图像,然后将预处理后的待检测图像输入至训练好的神经网络模型,得到待检测图像的目标检测结果,并输出该目标检测结果,在本申请实施例中,该训练好的神经网络模型对应的各个特征图层的感受野与锚点框的尺寸信息满足匹配条件,使得在目标检测时所进行的分类和回归都能在跟目标对象大小一致的感受野上进行,从而极大的提高检测性能和检测准确度。

[0116] 在一些实施例中,当需要检测的是单张图像时,通过上述的步骤S101至步骤S103即可得到目标检测结果;当需要检测的是视频文件时,步骤S101中的“获取待检测图像”可以通过以下步骤实现:

[0117] 步骤S1011,获取待检测的视频文件。

[0118] 这里,步骤S1011在实现时,可以从自身的存储空间中获取待检测的视频文件,还可以是在接收到终端发送的视频发布请求时,将视频发布请求中携带的视频文件确定为待检测的视频文件。由于在对视频进行发布之前需要对视频进行审核,因此可以对待发布的视频文件进行目标检测,以确定是否审核通过。

[0119] 步骤S1012,对该待检测的视频文件进行解析,得到各个视频帧图像。

[0120] 这里,由于视频文件时将一帧一帧的图像进行编码得到的,当需要得到视频文件中的各个视频帧图像时,需要对视频文件进行解码,从而得到各个视频帧图像。

[0121] 步骤S1013,将该各个视频帧图像确定为各个待检测图像。

[0122] 在得到各个待检测图像之后,同样也需要对各个待检测图像进行预处理,从而得到预处理后的待检测图像。

[0123] 对应地,在步骤S102之后,该方法还包括:

[0124] 步骤S21,获取该各个视频帧图像的目标检测结果。

[0125] 这里,各个视频帧图像的目标检测结果表征各个视频帧图像中是否包括目标对

象,当目标检测结果为视频帧图像包括目标对象时,目标检测结果还包括目标对象在视频帧图像中的位置信息。

[0126] 步骤S22,基于各个视频帧图像的目标检测结果,确定该待检测的视频文件中目标检测结果为包括目标对象的视频帧图像的图像个数。

[0127] 步骤S23,基于该图像个数确定该待检测的视频文件的检测结果。

[0128] 这里,当该图像个数大于或者等于预设的个数阈值时,确定该待检测的视频文件的检测结果为包括目标对象,也就是说,当待检测的视频文件中包括目标对象的视频帧图像的个数超过个数阈值时,才认为视频文件中包括目标对象,其中,该个数阈值可以是预先设置的正整数,例如该个数阈值可以为1,那么也即只要视频文件中存在一个视频帧图像中包括目标对象,即认为视频文件中包括目标对象;再例如该个数阈值可以为3,那么也即当视频文件中有3个或者3个以上视频帧图像中包括目标对象时,才认为视频文件中包括目标对象。

[0129] 通过上述步骤S21至步骤S23,能够根据视频文件中各个视频帧图像的目标检测结果确定视频文件的检测结果,从而能够基于视频文件的检测结果进行后续的审核或者推荐流程。

[0130] 在一些实施例中,进行目标检测的目的可以是对视频文件进行审核,还可以是确定推荐视频,对应地,在确定出待检测的视频文件的检测结果之后,可以基于该检测结果删除不合规的视频文件,或者基于视频文件的检测结果确定视频文件是否为待推荐的目标文件。当进行目标检测的目的是对视频文件进行审核,以检测不合规视频时,那么,在步骤S23之后,还可以执行以下步骤:

[0131] 步骤S24A,当该视频文件的检测结果为包括目标对象时,确定该视频文件为目标视频文件。

[0132] 这里,当进行目标检测的目的是对视频文件进行审核时,目标对象可以是预设水印或者预设台标,当确定视频文件的检测结果为包括目标对象时,说明该视频文件中包括预设水印或者预设台标,此时将该视频文件确定为目标视频文件。

[0133] 步骤S25A,在确定达到预设的删除时机时,删除该目标视频文件。

[0134] 这里,可以是一旦确定出视频文件为目标视频文件即认为达到删除时机,还可以是预先设置好每间隔一定时长进行一次目标视频文件的删除,此时在达到预设的删除时刻时认为达到删除时机,删除检测出的目标视频文件。

[0135] 通过上述的步骤S24A至步骤S25A,在确定出视频文件的检测结果后,如果该视频文件的检测结果表明该视频文件的审核不通过或者表明该视频文件满足删除条件,确定该视频文件为目标视频文件,并在达到预设的删除时机时,将目标视频文件进行删除,从而能够实现视频文件的自动智能审核和删除,从而提高审核速度和准确度。

[0136] 在一些实施例中,当进行目标检测的目的是确定推荐视频时,那么在步骤 S23之后,还可以执行以下步骤:

[0137] 步骤S24B,当该视频文件的检测结果为包括目标对象时,确定该视频文件为目标视频文件。

[0138] 这里,目标对象可以为预设网站水印、预设台标等。当视频文件的检测结果为包括目标对象时,说明该视频文件可以是预设网站所发布的视频,或者是预设电视台的节目视

频。

[0139] 步骤S25B,在确定达到预设的推荐时机时,确定目标推荐对象。

[0140] 这里,可以预先设置推荐时刻,当达到预设的推荐时刻时,认为达到推荐时机,此时确定目标推荐对象,其中目标推荐对象为历史观看过具有目标对象的视频的终端,或者当前正在观看具有目标对象的视频的终端。

[0141] 步骤S26B,向该目标推荐对象推荐该目标视频文件。

[0142] 这里,步骤S26B在实现时,可以是将该目标视频文件的标识和/或该目标视频文件的封面图像发送至目标推荐对象,以在目标推荐对象的显示屏上显示该目标视频文件的标识和/或封面图像,从而实现推荐目的。

[0143] 在一些实施例中,进行目标检测的目的还可以是待发布的视频进行审核,此时目标对象可以是预设水印或预设台标,在确定出视频文件的检测结果后,可以基于该检测结果确定审核结果,以确定视频文件是否通过审核。例如当目标对象为经常发布预设类型网站的网站水印时,如果视频文件的检测结果为包括目标对象,那么确定该视频文件的审核结果为审核不通过;如果视频文件的检测结果为不包括目标对象,那么确定该视频文件的审核结果为审核通过。

[0144] 在一些实施例中,在进行目标检测之前,首选需要得到训练好的神经网络模型,在本申请实施例中,在对神经网络模型进行训练之前,首先要通过步骤 S102确定锚点框的尺寸信息,并且还要确定出神经网络模型各个特征图层的感受野大小,进而才能基于锚点框的尺寸信息和感受野大小对锚点框进行自适应分配。在实现时,可以通过如图5所示的步骤S31至步骤S32实现步骤S102中的“确定锚点框的尺寸信息”:

[0145] 步骤S31,获取多个样本图像和该多个样本图像的标注信息。

[0146] 这里,样本图像的标注信息至少包括标注对象对应的标注框的尺寸信息。该标注框的形状一般为长方形或正方形,标注框的尺寸信息可以包括标注框的长度和宽度,还可以包括标注框的长度与样本图像的长度的比值以及标注框的宽度与样本图像的宽度的比值。

[0147] 在一些实施例中,该样本图像可以为经过预处理之后得到的尺寸、格式都符合神经网络模型输入要求的图像,此时标注框的尺寸信息可以仅包括标注框的长度和宽度。

[0148] 步骤S32,对多个标注框的尺寸信息进行聚类处理,得到N个锚点框的尺寸信息。

[0149] 这里,N为大于1的整数,例如N可以为9,还可以为16。

[0150] 步骤S32在实现时,可以是利用K-Mean聚类方法,确定出N个聚类中心,从而得到N个锚点框的尺寸信息,还可以利用均值漂移聚类算法、基于高斯模型的最大期望聚类算法来确定N个聚类中心,从而得到N个锚点框的尺寸信息。在本申请实施例中,锚点框的尺寸信息包括锚点框的长度和宽度,还可以包括长宽比和对角线长度。在一些实施例中,还可以对锚点框的尺寸信息进行归一化,在实现时,可以将锚点框的长度除以样本图像的长度,得到归一化的锚点框长度,将锚点框的宽度除以样本图像的宽度,得到归一化的锚点框宽度。举例来说,锚点框的原始尺寸信息为30*40,样本图像的尺寸信息为416*416,那么归一化后的锚点框的尺寸信息为0.072*0.096,对角线长度为0.12。

[0151] 在确定出锚点框的尺寸信息后,可以通过如图5所示的步骤S33至步骤S34实现步骤S102中的“确定神经网络模型对应的各个特征图层的感受野”:

[0152] 步骤S33,获取预设的神经网络模型。

[0153] 这里,该神经网络模型可以为卷积神经网络模型。

[0154] 步骤S34,基于该神经网络模型的参数信息确定该神经网络模型中各个特征图层对应的感受野大小。

[0155] 这里,在实现时,可以根据公式(1-1)计算各个特征图层对应的感受野大小:

$$[0156] \quad r_0 = \frac{\left(\sum_{l=1}^L ((k_l - 1) \prod_{i=1}^{l-1} s_i) + 1 \right)}{\sqrt{L}} \quad (1-1);$$

[0157] 其中,在公式(1-1)中, r_0 表示当前特征图层每个点的感受野, L 代表当前特征图层之前的网络的层数, k_1 表示第1层的卷积核大小, s_i 表示*i*层卷积核的步长(stride)。

[0158] 例如,最顶层的特征图层的感受野大小为101,中间层的特征图层的感受野大小为60,最底层的特征图层的感受野大小为30。

[0159] 在确定出锚点框的尺寸信息以及各个特征图层的感受野之后,可以通过下述的步骤S35至步骤S36完成锚点框的自适应分配:

[0160] 步骤S35,基于该各个特征图层对应的感受野大小和*N*个锚点框的尺寸信息,确定各个锚点框对应的特征图层。

[0161] 这里,步骤S35可以通过以下步骤实现:

[0162] 步骤S351,基于各个特征图层对应的感受野大小和第*i*个锚点框的尺寸信息,确定第*i*个锚点框与各个感受野大小的各个差值。

[0163] 这里, $i=1,2,\dots,N$ 。当锚点框的尺寸信息是没有进行归一化的,那么步骤S351在实现时,直接将第*i*个锚点框的尺寸信息与各个感受野大小相减,并求绝对值,得到第*i*个锚点框与各个感受野大小的各个差值。进一步地,可以是将第*i*个锚点框的对角线长度与各个感受野大小相减,并求绝对值,得到第*i*个锚点框与各个感受野大小的各个差值。举例来说,第*i*个锚点框的对角线为50,三个感受野的大小从上到下分别对101、60和30,那么第*i*个锚点框与各个感受野的差值分别为61、10和20。

[0164] 在一些实施例中,当锚点框的尺寸信息是归一化后的,那么步骤S351在实现时,首先要将归一化后的第*i*个锚点框的尺寸信息乘以样本图像的尺寸,得到归一化之前的第*i*个锚点框的尺寸信息,然后再将归一化之前的第*i*个锚点框的尺寸信息与各个感受野大小相减,并求绝对值,得到第*i*个锚点框与各个感受野大小的各个差值。举例来说,归一化后的第*i*个锚点框的尺寸信息中对角线长度为0.12,样本图像的尺寸为416*416,那么归一化之前的对角线长度为 $0.12*416=50$,三个感受野的大小从上到下分别对101、60和30,第*i*个锚点框与各个感受野的差值分别为61、10和20。

[0165] 步骤S352,将各个差值中的最小差值对应的特征图层确定为第*i*目标特征图层。

[0166] 承接上述的举例,第一个锚点框与各个感受野的差值分别为61、10、20,也即与中间层的特征图层的感受野的差别最小,因此将中间的特征图层确定为第*i*个目标特征图层。

[0167] 步骤S353,确定第*i*目标层为第*i*个锚点框对应的特征图层,此时确定该第*i*目标特征图层的感受野与该第*i*个锚点框的尺寸信息满足匹配条件。

[0168] 步骤S36,将各个锚点框分配至对应的特征图层。

[0169] 假设有9个锚点框,其中锚点框1-6对应中间的特征图层,锚点框7-9对应最下面的

特征图层,而最上面的特征图层没有对应的锚点框。

[0170] 通过上述的步骤S31至步骤S36即可完成锚点框和神经网络特征图层的自适应分配,从而实现对神经网络模型架构的改变,在此之后,可以通过以下步骤得到训练好的神经网络模型:

[0171] 步骤S41,将各个样本图像分别输入该神经网络模型,得到各个样本图像对应的预测结果。

[0172] 这里,各个样本图像为彩色图像,各个样本图像输入该神经网络模型后,得到三个特征图层,并且各个锚点框对应与感受野大小最接近的特征图层,以保证在于锚点框最接近的特征图层中进行分类和回归。

[0173] 预测结果至少包括样本图像中具有目标对象的概率值,以及目标对象在样本图像中的位置信息,当目标对象有多个时,那么预测结果中包括样本图像中具有各个目标对象的各个概率值,以及各个目标对象在样本图像中的各个位置信息。

[0174] 步骤S42,利用各个样本图像对应的预测结果和各个样本图像的标注信息对该神经网络模型进行反向传播训练,以对该神经网络模型的参数进行调整,以得到训练好的神经网络模型。

[0175] 在该训练好的神经网络模型中,每个特征图层是否分配anchor,或者是分配哪些anchor是根据每个特征图层的感受野大小和各个anchor的尺寸大小的差值确定的,也就是说,在该训练好的神经网络模型中,anchor是自适应地分配到各个特征图层的,因此在进行目标检测时,分类和回归都是在于目标对象大小最接近的感受野上进行,从而提高目标检测的精度。

[0176] 基于前述的实施例,本申请实施例再提供一种目标检测方法,应用于图2所示的网络架构,图6为本申请实施例提供的目标检测方法的再一种实现流程示意图,如图6所示,该流程包括:

[0177] 步骤S601,终端基于接收到的视频录制操作,进行视频录制。

[0178] 这里,该视频录制操作可以通过相机应用启动摄像头以开始录制视频的操作,还可以是通过即时通讯应用或者短视频应用启动摄像头以开始录制视频的操作,还可以是启动录屏的操作。

[0179] 步骤S602,终端基于接收到的录制完成操作,获取录制好的视频文件。

[0180] 步骤S603,终端基于接收到水印编辑操作,对录制好的视频文件增加水印。

[0181] 这里,所增加的水印可以是用户自行生成的,例如可以是基于用户名生成的水印,还可以是终端预设的水印,还可以是业务应用预设的水印,例如可以是利用短视频应用的logo而生成的水印。

[0182] 步骤S604,终端响应于发布视频文件的操作,向服务器发送视频发布请求。

[0183] 这里,该视频发布请求中携带有已增加有水印的视频文件。

[0184] 步骤S605,服务器解析该视频发布请求,以获取该视频文件。

[0185] 步骤S606,服务器对该视频文件进行解码,得到各个视频帧图像。

[0186] 步骤S607,服务器将各个视频帧图像分别输入训练好的神经网络模型,得到各个视频帧图像的目标检测结果。

[0187] 这里,训练好的神经网络模型的特征图层的感受野大小与锚点框的尺寸信息满足

匹配条件,这样在对各个视频帧图像进行目标检测时,能够在于目标对象最接近的感受野上进行分类和回归,从而能够提高检测精度。该目标检测结果中包括视频帧图像中包括目标对象的概率值,以及目标对象在视频帧图像中的位置信息。

[0188] 步骤S608,服务器根据各个视频帧图像的目标检测结果确定该视频文件的检测结果。

[0189] 这里,步骤S608在实现时,可以是首先基于各个视频帧图像的目标检测结果,确定该待检测的视频文件中目标检测结果为包括目标对象的视频帧图像的图像个数;然后基于该图像个数确定该待检测的视频文件的检测结果,其中,当该图像个数大于预设的个数阈值时,确定该待检测的视频文件的检测结果为包括目标对象。目标对象可以是预设网站的水印,还可以是预设的水印或者台标等。

[0190] 步骤S609,服务器根据该视频文件的检测结果确定该视频文件是否满足发布条件。

[0191] 这里,当视频文件的检测结果为不包括目标对象时,确定该视频文件满足发布条件时,进入步骤S610;当视频文件的检测结果为包括目标对象时,确定该视频文件不满足发布条件时,进入步骤S611。

[0192] 步骤S610,服务器发布该视频文件,并向终端返回发布成功的响应消息。

[0193] 步骤S611,服务器向终端返回发布失败的响应消息。

[0194] 这里,该发布失败的响应消息中可以携带有发布失败的原因,以使得终端用户能够及时了解发布失败原因,并给予发布失败原因对视频文件进行更新。

[0195] 步骤S612,服务器将该视频文件确定为目标视频文件。

[0196] 步骤S613,服务器在确定达到删除时机时,删除该目标视频文件。

[0197] 在本申请实施例提供的目标检测方法中,在终端获取到待发布的视频文件之后,将待发布的视频文件携带于视频发布请求中发送至服务器,服务器利用训练好的神经网络模型对视频文件中的各个视频帧图像进行目标检测,该训练好的神经网络模型对应的各个特征图层的感受野与锚点框的尺寸信息满足匹配条件,从而能够保证在进行目标检测时,分类和回归都在与目标对象尺寸最接近的感受野上进行,从而能够提高检测精度,在得到各个视频帧图像的目标检测结果后,服务器再基于各个视频帧图像的目标检测结果确定视频文件的检测结果,然后在基于该检测结果确定该视频文件满足发布条件时,将该视频文件进行发布;当确定该视频文件不满足发布条件时,对该视频文件不予进行发布,并且在达到删除时机时,将该视频文件进行删除,从而能够保证发布视频文件的合法合规性,为用户提供安全纯洁的网络环境。

[0198] 下面,将说明本申请实施例在一个实际的应用场景中的示例性应用。

[0199] 本申请实施例提供一种自适应的anchor based目标检测方法,对于检测中的位置回归和分类能自适应的选择尺寸合适的anchor(先验尺寸)来计算,能够显著提高检测性能。

[0200] 图7为本申请实施例提供的目标检测方法的检测框架示意图,如图7所示,该检测框架包括:特征提取网络701、特征金字塔网络702、Yolo Block 703,其中Yolo Block703用作特征转换、降维等,每一个特征图层之后都接一个Yo lo Block,所以每个Yolo Block最后输出的通道数依次减少。

[0201] 在图7中,仅示例性示出了中间一层的特征图层的检测头704,并且anchor的分配方式由均匀分配变为自适应的分配,以此提高检测性能。

[0202] 以下说明本申请实施例提供的目标检测方法的实现过程。

[0203] 第一步、计算先验anchor信息。

[0204] Anchor是带有尺寸信息的先验框。在计算anchor的值时,先要获取样本图像的标注信息中的目标标注框的尺寸信息(bb_w,bb_h)和长宽比信息(bb_w/im_w,bb_h/im_h),其中bb_w和bb_h为目标标注框的宽和高,im_w和im_h为图片宽和高;再利用聚类算法(例如可以是k-means),对获取到的目标标注框的尺寸信息和长宽比进行聚类,得到k(一般k=9)个聚类中心的值,得到k个聚类中心的值,也就能够获得k个先验anchor的值。在本申请实施例中,聚类中心的值可以包括先验框的长和宽,此时可以根据先验框的长和宽,确定出先验框的对角线长度,并将该对角线长度进行归一化得到anchor的值。其中,将该对角线长度进行归一化得到anchor的值可以是将对角线长度除以图像的长度,还可以是将对角线长度除以图像的宽度。

[0205] 在一些实施例中,还可以是直接将先验框的对角线长度确定为anchor的值。

[0206] 第二步、获取预设的初始化神经网络模型,并计算feature map的实际感受野。

[0207] 在实现时,可以根据公式(1-1)计算当前feature map中空间每个点的感受野:

$$r_0 = \frac{\left(\sum_{l=1}^L ((k_l - 1) \prod_{i=1}^{l-1} s_i) + 1 \right)}{\sqrt{L}} \quad (1-1);$$

[0209] 其中,在公式(1-1)中, r_0 表示当前feature map中空间每个点的感受野,L代表当前feature map之前的网络的层数, k_l 表示第l层的卷积核大小, s_i 表示i层卷积核的stride。

[0210] 利用公式(1-1)可以获得每一层feature map的每个空间点的感受野。如在Yolo-v3中,feature_map_top,feature_map_middle,feature_map_bottom对应的感受野分别为feature_map_top_s=101,feature_map_middle_s=60,feature_map_bottom_s=30。

[0211] 第三步、自适应地将anchor分配给feature map。

[0212] 在本申请实施例中,anchor分配给feature map的准则为:anchor尺寸和feature map感受野最接近。在实现时可以通过以下步骤:

[0213] 步骤S801,计算anchor相对于模型输入图像的尺寸anchor*input_size。

[0214] 在实现是,可以是将对角线长度乘以输入图像的尺寸得到anchor*input_size,比如模型输入的图像的大小是416*416(即输入图片会被调整为416*416),则anchor*input_size为anchor*416。

[0215] 需要说明的是,如果在第一步计算先验anchor时,是在得到K个聚类中心的值从而得到先验框的对角线长度后,anchor的值是将对角线长度进行归一化得到的,那么需要根据步骤①计算anchor相对于模型输入图像的尺寸anchor*input_size;如果在第一步计算先验anchor时,是在得到K个聚类中心的值从而得到先验框的对角线长度后,直接将对角线长度确定为anchor的值,那么此时在该步骤①中直接将anchor的值作为anchor*input_size。

[0216] 步骤S802,计算anchor*input_size与不同feature map感受野的差值。

[0217] 在实现时,可以通过公式(1-2)计算 $\text{anchor} * \text{input_size}$ 与不同feature map感受野的差值 $\text{dis_anchor_feature_i}$:

[0218] $\text{dis_anchor_feature_i} = \text{abs}(\max(\text{anchor} * \text{input_size}) - \text{feature_map_i_s})(1-2)$;

[0219] 其中, feature_map_i_s 为不同feature map的感受野。

[0220] 步骤S803,确定各个 $\text{anchor} * \text{input_size}$ 对应的最接近的feature map。

[0221] 这里,将与 $\text{anchor} * \text{input_size}$ 之间差值最小的feature map确定为与该 $\text{anchor} * \text{input_size}$ 最接近的feature map,并把该anchor分配给该feature map。

[0222] 通过上述的步骤S801至步骤S803,就自适应地将anchor分配给感受野与anchorsize最匹配的feature map。

[0223] 比如对于台标检测的任务,因为台标都是很小的目标,anchor中较小的分布在(0.06,0.06)附近(比如 $\text{anchor}[1:6]$),此时 $\text{anchor} * \text{input_size}$ 为 $0.06 * 416 = 24.96$,假设 $\text{feature_map_top_s} = 101$, $\text{feature_map_middle_s} = 60$, $\text{feature_map_bottom_s} = 30$,由于24.96与101、60、30相比,最接近30,因此 $\text{anchor}[1:6]$ 都分配到 $\text{feature_map_bottom}$,再例如较大的anchor为分布在(0.18,0.18)附近(比如 $\text{anchor}[7,9]$),此时 $\text{anchor} * \text{input_size}$ 为 $0.18 * 416 = 74.88$,由于74.88与101、60、30相比,最接近60,因此 $\text{anchor}[7,9]$ 分配到 $\text{feature_map_middle}$,. 则最终的目标检测网络可以如图8所示, $\text{anchor}[1,6]$ 801分配给了最底层的feature map 802, $\text{anchor}[7,9]$ 803分配给了中间层的feature map 804,而最高层的feature map没有分配anchor,从而利用基于感受野适配的anchor自适应分配原则,将anchor分配给感受野一致的feature map进行后续的分类和回归,能够极大的提高检测性能。

[0224] 利用样本图像和样本图像对应的标注信息,对图8所示的目标检测网络进行训练,得到训练好的神经网络模型。在进行推理测试时,利用训练好的神经网络模型对待检测图像进行目标检测,从而得到检测结果。利用本申请实施例提供的目标检测方法在进行小目标的检测任务时,如水印和台标检测中,准确率和召回率都显著优于其它已有参考系统。

[0225] 需要说明的是,在图8中是以Yolo系列来进行示例性说明,本申请实施例提供的目标检测方法,自适应地分配anchor给最佳的feature map来进行分类和回归适用于任何anchor based检测方法。

[0226] 本申请实施例提供的目标检测方法至少能够应用到智能视频审核、视频推荐的多个项目和产品中,例如在智能视频审核中,能够识别视频中是否具有预设的水印和台标,能够实现全自动、高精度的对目标台标和水印的检测和识别,在保障视频安全的同时,还能够提高用户体验。

[0227] 下面继续说明本申请实施例提供的目标检测装置354的实施为软件模块的示例性结构,在一些实施例中,如图3所示,存储在存储器350的目标检测装置354中的软件模块可以包括:

[0228] 第一获取模块3541,用于获取待检测图像,并对该待检测图像进行预处理,得到预处理后的待检测图像;

[0229] 第一确定模块3542,用于确定锚点框的尺寸信息和神经网络模型对应的各个特征图层的感受野;

- [0230] 目标检测模块3543,用于将该预处理后的待检测图像输入至训练好的神经网络模型,得到该待检测图像的目标检测结果,
- [0231] 该训练好的神经网络模型对应的各个特征图层的感受野与锚点框的尺寸信息满足匹配条件;
- [0232] 输出模块3544,用于输出该目标检测结果。
- [0233] 在一些实施例中,该第一确定模块3542,还用于:
- [0234] 第二获取模块,用于获取多个样本图像和该多个样本图像的标注信息,该标注信息至少包括标注对象对应的标注框的尺寸信息;
- [0235] 聚类模块,用于对多个标注框的尺寸信息进行聚类处理,得到N个锚点框的尺寸信息,N为大于1的整数。
- [0236] 在一些实施例中,该第一确定模块3542,还用于:
- [0237] 获取预设的神经网络模型;
- [0238] 基于该神经网络模型的参数信息确定该神经网络模型中各个特征图层对应的感受野大小;
- [0239] 在一些实施例中,该装置还包括:
- [0240] 第二确定模块,用于基于该各个特征图层对应的感受野大小和N个锚点框的尺寸信息,确定各个锚点框对应的特征图层;
- [0241] 锚点分配模块,用于将各个锚点框分配至对应的特征图层。
- [0242] 在一些实施例中,该第二确定模块,还用于:
- [0243] 基于各个特征图层对应的感受野大小和第i个锚点框的尺寸信息,确定第i个锚点框与各个感受野大小的各个差值, $i=1,2,\dots,N$;
- [0244] 将各个差值中的最小差值对应的特征图层确定为第i目标特征图层;
- [0245] 确定第i目标层为第i个锚点框对应的特征图层,并且确定该第i目标特征图层的感受野与该第i个锚点框的尺寸信息满足匹配条件。
- [0246] 在一些实施例中,该装置还包括:
- [0247] 输入模块,用于将各个样本图像分别输入该神经网络模型,得到各个样本图像对应的预测结果;
- [0248] 训练模块,用于利用各个样本图像对应的预测结果和各个样本图像的标注信息对该神经网络模型进行反向传播训练,以对该神经网络模型的参数进行调整,以得到训练好的神经网络模型。
- [0249] 在一些实施例中,该第一获取模块还用于:
- [0250] 获取待检测的视频文件;
- [0251] 对该待检测的视频文件进行解析,得到各个视频帧图像;
- [0252] 将该各个视频帧图像确定为各个待检测图像。
- [0253] 在一些实施例中,该装置还包括:
- [0254] 获取模块,用于获取该各个视频帧图像的目标检测结果;
- [0255] 第三确定模块,用于基于各个视频帧图像的目标检测结果,确定该待检测的视频文件中目标检测结果为包括目标对象的视频帧图像的图像个数;
- [0256] 第四确定模块,用于基于该图像个数确定该待检测的视频文件的检测结果,其中,

当该图像个数大于预设的个数阈值时,确定该待检测的视频文件的检测结果为包括目标对象。

[0257] 在一些实施例中,该装置还包括:

[0258] 第五确定模块,用于当该视频文件的检测结果为包括目标对象时,确定该视频文件为目标视频文件;

[0259] 删除模块,用于在确定达到预设的删除时机时,删除该目标视频文件。

[0260] 在一些实施例中,该装置还包括:

[0261] 第六确定模块,用于当该视频文件的检测结果为包括目标对象时,确定该视频文件为目标视频文件;

[0262] 第七确定模块,用于在确定达到预设的推荐时机时,确定目标推荐对象;

[0263] 推荐模块,用于向该目标推荐对象推荐该目标视频文件。

[0264] 本申请实施例提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行本申请实施例上述的目标检测方法。

[0265] 本申请实施例提供一种存储有可执行指令的计算机可读存储介质,其中存储有可执行指令,当可执行指令被处理器执行时,将引起处理器执行本申请实施例提供的方法,例如,如图4、图5、图6示出的方法。

[0266] 在一些实施例中,计算机可读存储介质可以是FRAM、ROM、PROM、EP ROM、EEPROM、闪存、磁表面存储器、光盘、或CD-ROM等存储器;也可以是包括上述存储器之一或任意组合的各种设备。

[0267] 在一些实施例中,可执行指令可以采用程序、软件、软件模块、脚本或代码的形式,按任意形式的编程语言(包括编译或解释语言,或者声明性或过程性语言)来编写,并且其可按任意形式部署,包括被部署为独立的程序或者被部署为模块、组件、子例程或者适合在计算环境中使用的其它单元。

[0268] 作为示例,可执行指令可以但不一定对应于文件系统中的文件,可以可被存储在保存其它程序或数据的文件的一部分,例如,存储在超文本标记语言(H TML,Hyper Text Markup Language)文档中的一个或多个脚本中,存储在专用于所讨论的程序的单个文件中,或者,存储在多个协同文件(例如,存储一个或多个模块、子程序或代码部分的文件)中。

[0269] 作为示例,可执行指令可被部署为在一个计算设备上执行,或者在位于一个地点的多个计算设备上执行,又或者,在分布在多个地点且通过通信网络互连的多个计算设备上执行。

[0270] 综上所述,通过本申请实施例利用神经网络模型的感受野来分配适合的an chor,使得分类和回归都能在跟目标大小一致的感受野上进行,能够极大的提高检测性能。

[0271] 以上所述,仅为本申请的实施例而已,并非用于限定本申请的保护范围。凡在本申请的精神和范围之内所作的任何修改、等同替换和改进等,均包含在本申请的保护范围之内。

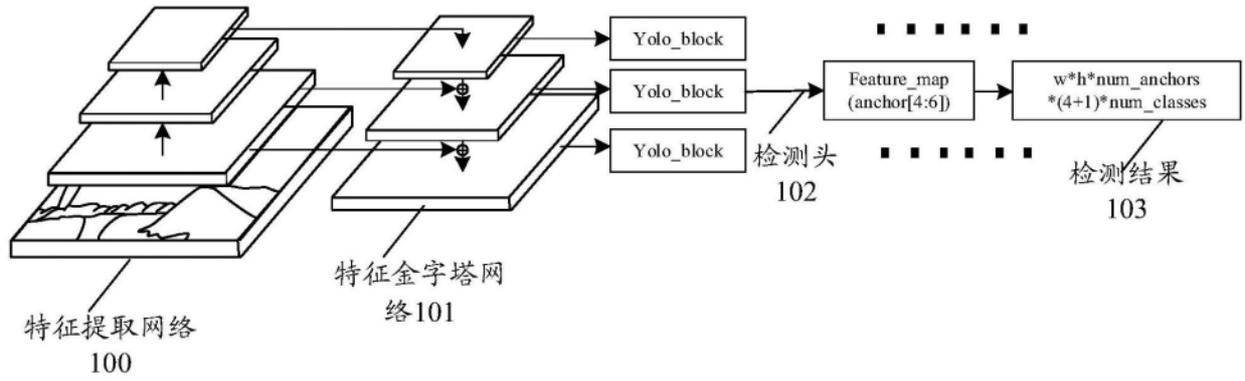


图1A

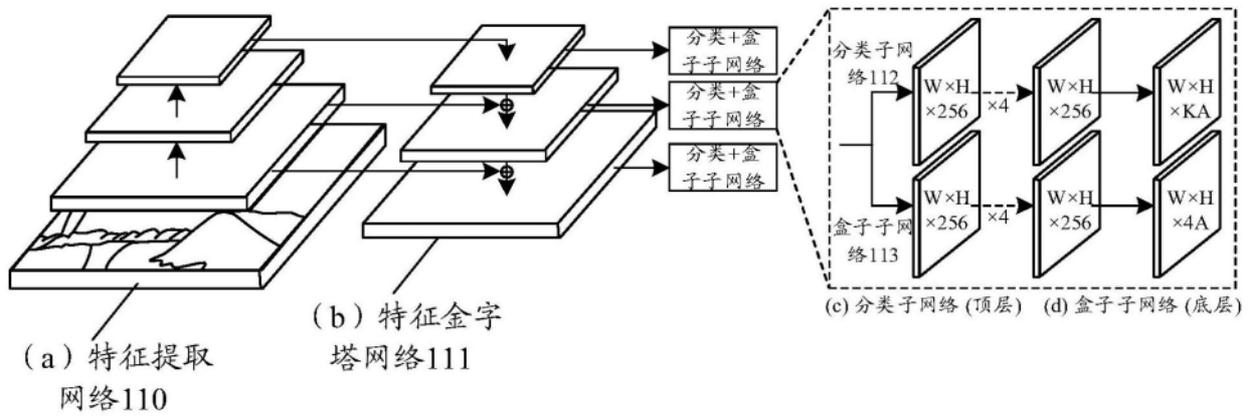


图1B

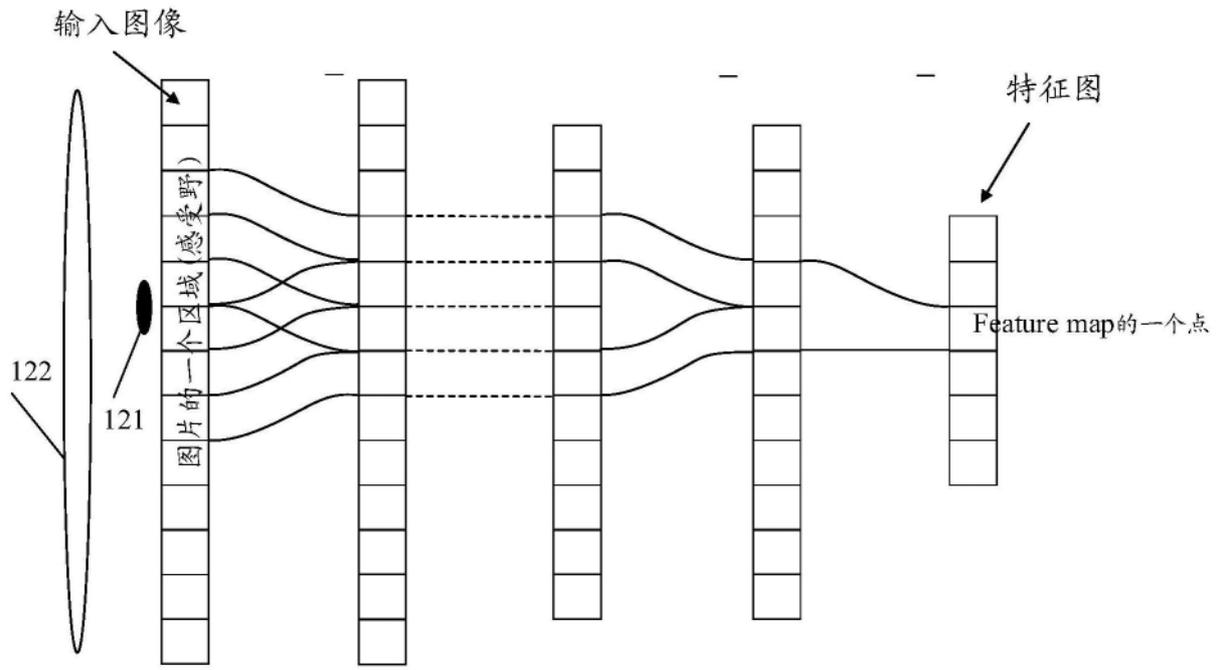


图1C

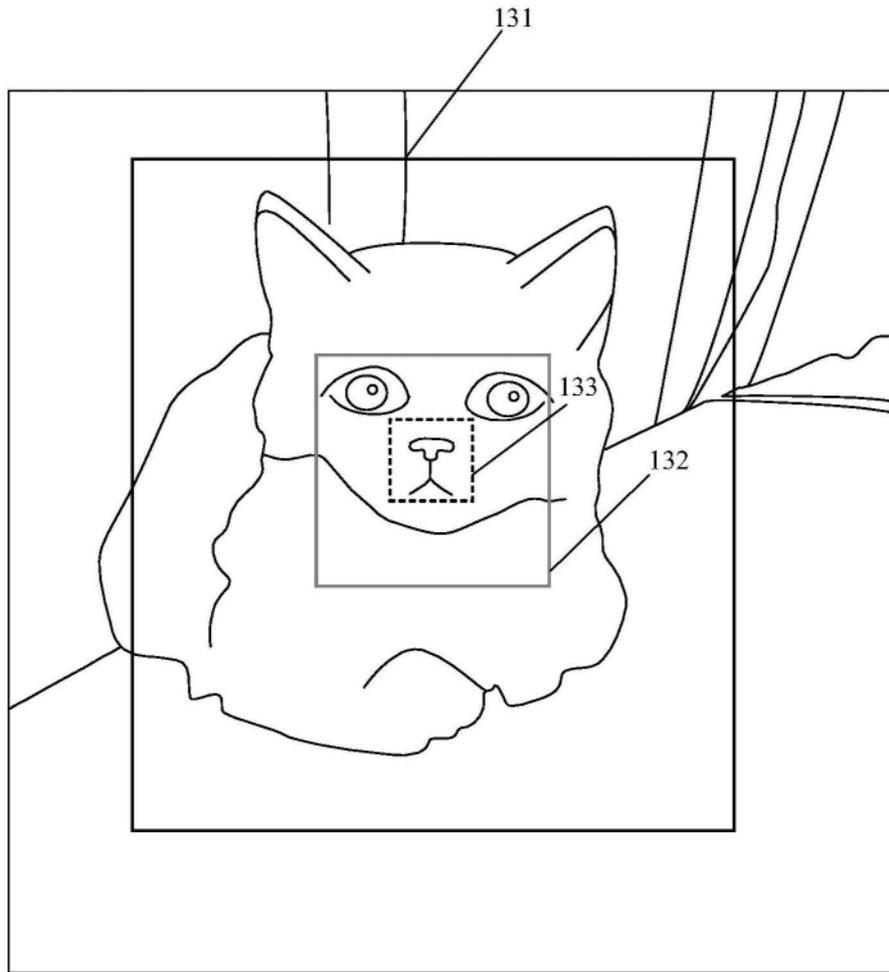


图1D

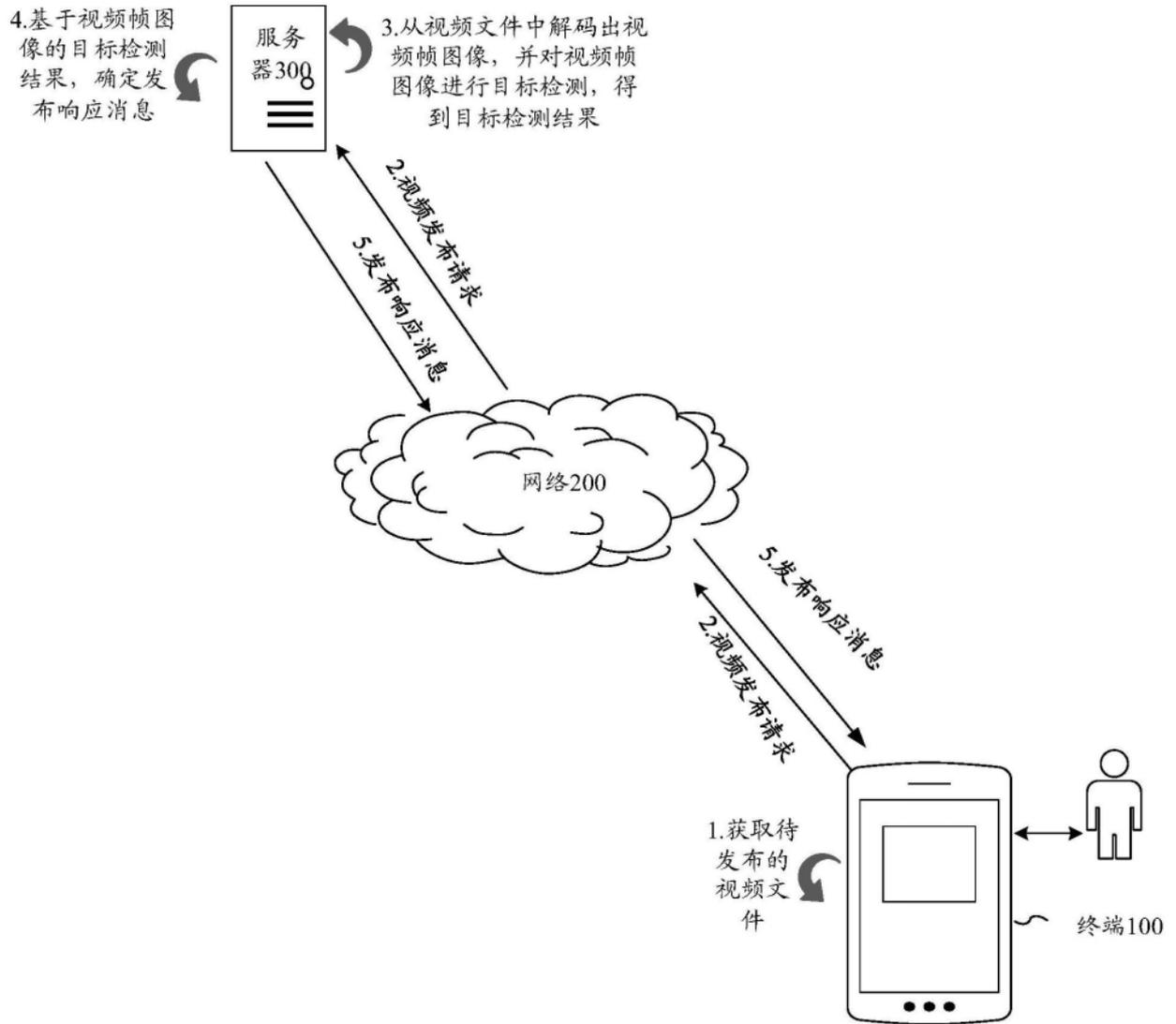


图2

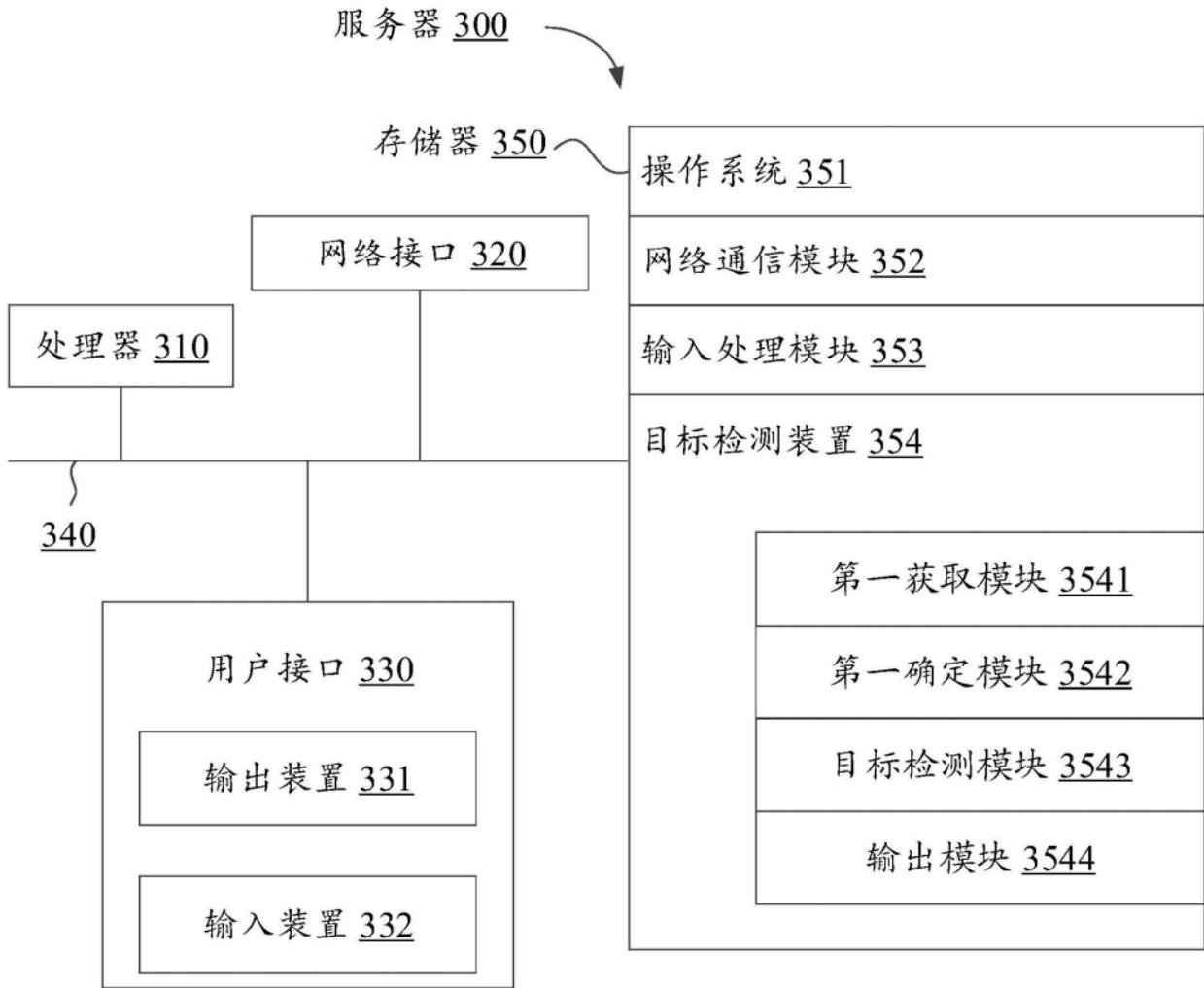


图3

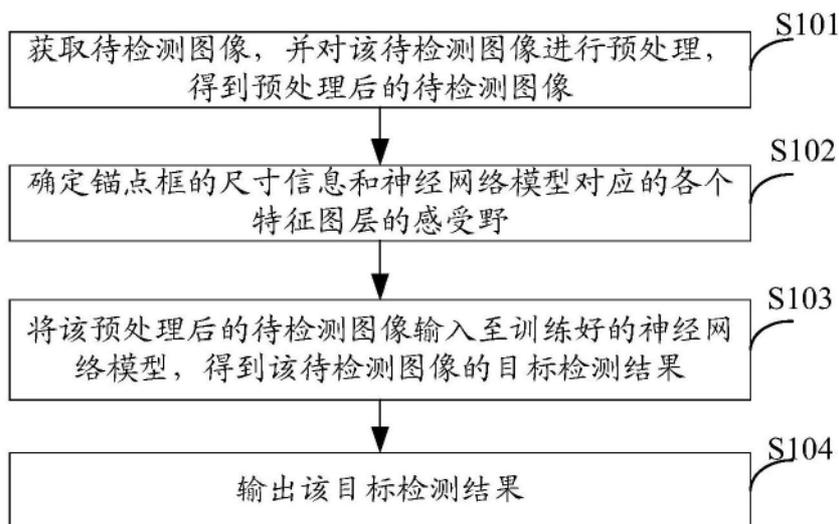


图4

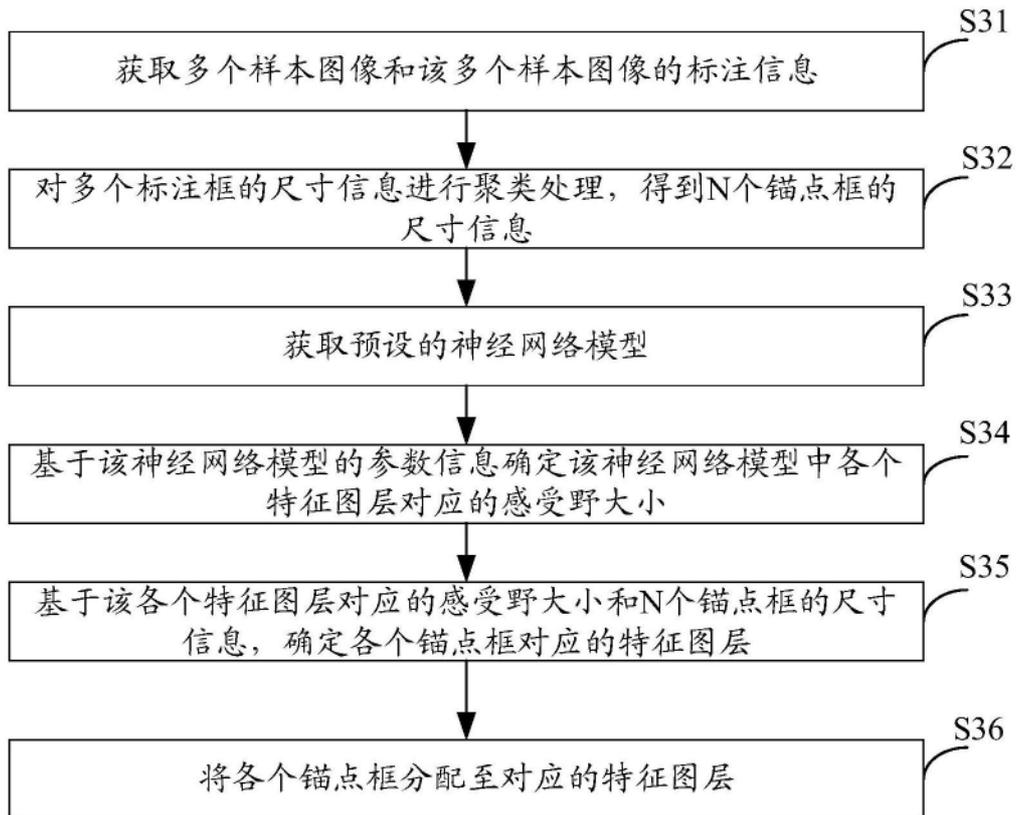


图5

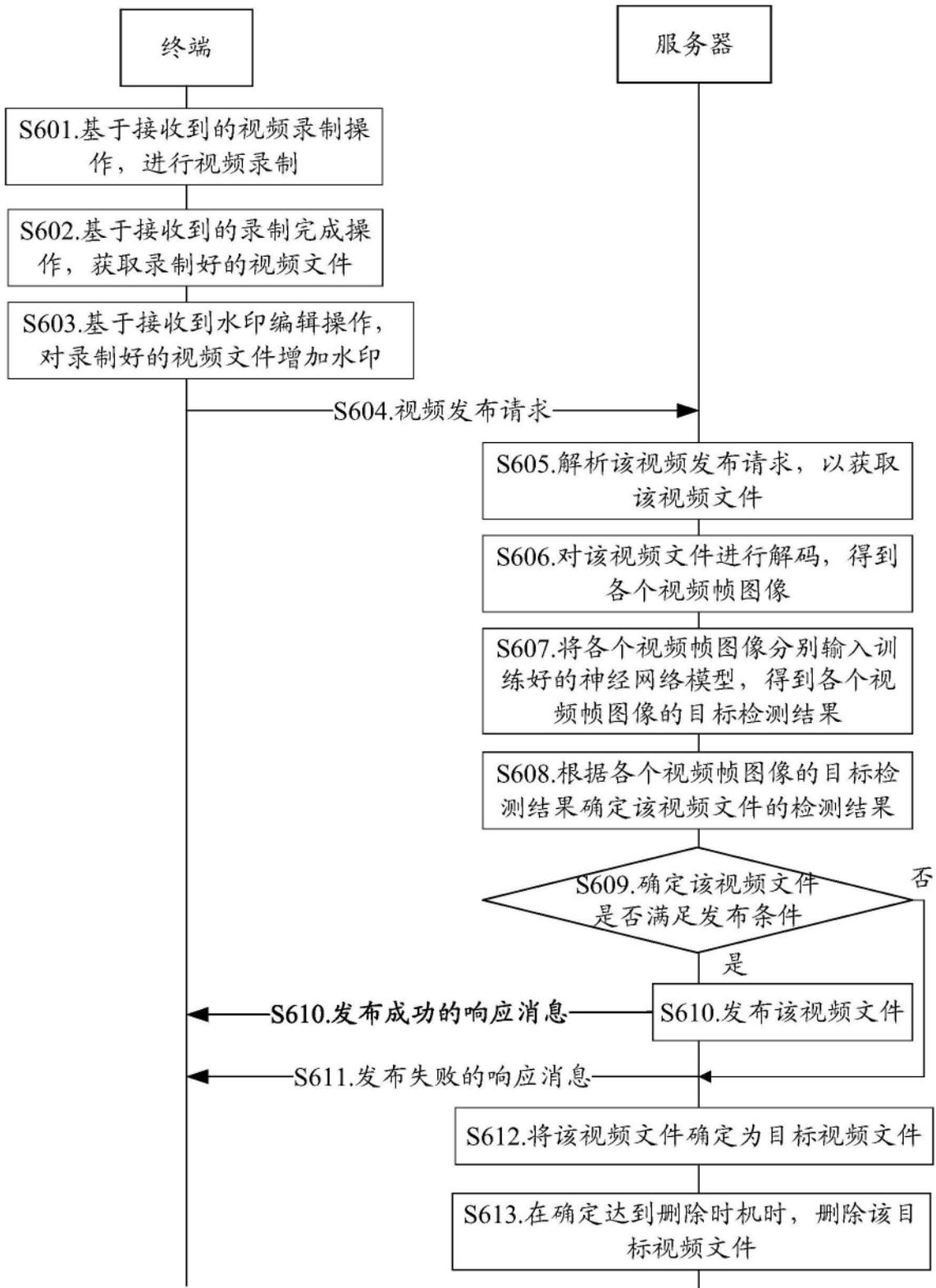


图6

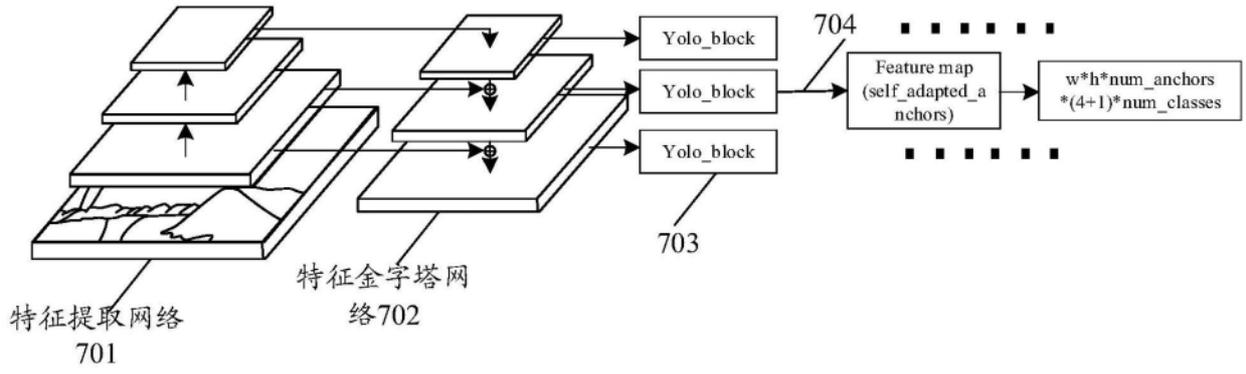


图7

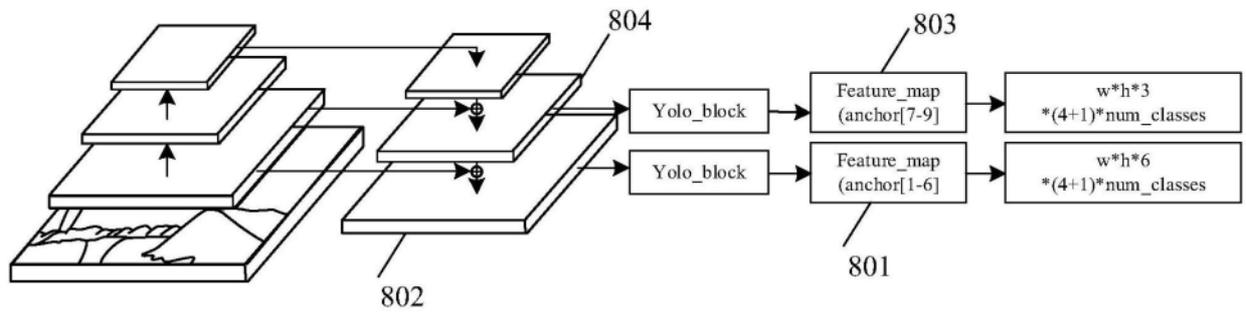


图8