



(12) 发明专利申请

(10) 申请公布号 CN 117237599 A

(43) 申请公布日 2023.12.15

(21) 申请号 202311084785.0

G06N 3/082 (2023.01)

(22) 申请日 2023.08.25

(71) 申请人 中银金融科技有限公司

地址 200120 上海市浦东新区(上海)自由贸易试验区银城中路200号4楼408室

(72) 发明人 张惠婷

(74) 专利代理机构 北京路浩知识产权代理有限公司 11002

专利代理师 李鑫

(51) Int. Cl.

G06V 10/25 (2022.01)

G06V 10/774 (2022.01)

G06V 10/82 (2022.01)

G06N 3/0464 (2023.01)

G06N 3/045 (2023.01)

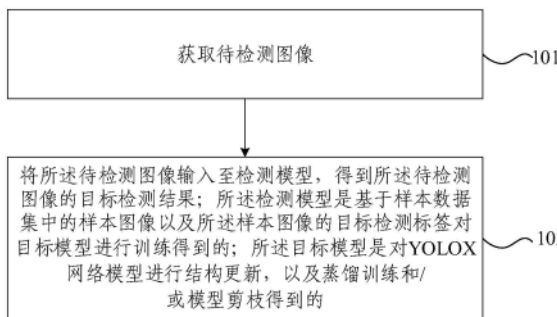
权利要求书2页 说明书13页 附图3页

(54) 发明名称

图像目标检测方法及装置

(57) 摘要

本发明提供一种图像目标检测方法及装置,涉及数据处理技术领域,该方法包括:获取待检测图像;将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果;其中,所述检测模型是基于样本数据集中的样本图像以及所述样本图像的目标检测标签对目标模型进行训练得到的;所述目标模型是对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝得到的。本发明实现通过对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝之后,采用多场景的样本图像进行微调训练,得到融合多种场景的检测模型,基于检测模型进行图像目标检测可实现在最大限度地压缩模型规模、降低运算成本的同时,可以最大限度地提高目标检测精度。



1. 一种图像目标检测方法,其特征在于,包括:

获取待检测图像;

将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果;其中,所述检测模型是基于样本数据集中的样本图像以及所述样本图像的目标检测标签对目标模型进行训练得到的;所述目标模型是对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝得到的。

2. 根据权利要求1所述的图像目标检测方法,其特征在于,所述检测模型是基于如下步骤训练得到的:

根据所述样本数据集,确定目标训练数据集;

根据所述目标训练数据集对所述YOLOX网络模型进行剪枝训练和/或蒸馏训练,得到轻量化神经网络模型;

对所述轻量化神经网络模型进行模型结构更新,得到所述目标模型;

将所述目标训练数据集中的样本图像输入至所述目标模型,得到所述目标训练数据集中样本图像的目标检测结果;

根据所述目标检测结果和所述目标检测标签,确定目标损失函数值;

基于所述目标损失函数值,对所述目标模型进行迭代训练,得到所述检测模型。

3. 根据权利要求2所述的图像目标检测方法,其特征在于,所述根据所述样本数据集,确定目标训练数据集,包括:

对所述样本数据集进行数据集划分,得到初始训练数据集;

确定所述初始训练数据集中样本图像的目标的类别信息、位置信息,以及尺寸信息;

根据所述类别信息、所述位置信息,以及所述尺寸信息,确定所述初始训练数据集中样本图像的均衡性;

在根据所述均衡性,确定所述初始训练数据集中的样本图像存在分布不均衡的情况下,对所述初始训练数据集中的样本图像进行数据增强;

对增强后的初始训练数据集进行预处理,得到所述目标训练数据集;

其中,所述预处理包括归一化处理和/或多尺度变换处理。

4. 根据权利要求2所述的图像目标检测方法,其特征在于,所述对所述轻量化神经网络模型进行模型结构更新,得到所述目标模型,包括:

将所述轻量化神经网络模型中特征提取网络的跨阶段局部网络更新为基于重参数化机器视觉几何模型构建的模块化网络结构;

在更新后的轻量化神经网络模型的所述模块化网络结构的最后一层增加卷积块注意力机制模块,得到所述目标模型。

5. 根据权利要求2所述的图像目标检测方法,其特征在于,所述根据所述目标检测结果和所述目标检测标签,确定目标损失函数值,包括:

根据所述目标检测结果和所述目标检测标签,获取至少一个分支损失函数值;

根据所述至少一个分支损失函数值,确定所述目标损失函数值;所述至少一个分支损失函数值包括位置损失函数值、置信度损失函数值、分类损失函数值以及正则化损失函数值;

其中,所述位置损失函数值是基于如下步骤确定的:

基于所述目标检测结果和所述目标检测标签,获取角度损失函数值、检测框重叠度损失函数值和形状损失函数值;

将所述角度损失函数值与预设值相加,根据相加结果,确定所述检测模型对应的距离损失函数值;

根据所述检测框重叠度损失函数值和所述形状损失函数值,以及所述距离损失函数值,确定所述位置损失函数值。

6. 根据权利要求1-4任一项所述的图像目标检测方法,其特征在于,所述将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果,包括:

将所述待检测图像输入至所述检测模型,得到所述待检测图像中目标的置信度信息、检测框信息和位置信息;

将所述置信度信息与置信度阈值进行比较,根据比较结果确定所述待检测图像中目标的类别预测值;

根据所述类别预测值、所述检测框信息和所述位置信息,确定所述待检测图像的目标检测结果。

7. 根据权利要求6所述的图像目标检测方法,其特征在于,所述置信度阈值是基于如下步骤确定的:

对所述样本数据集进行数据集划分,得到验证数据集;

将所述验证数据集中的样本图像输入至所述检测模型,得到所述验证数据集中的样本图像的目标检测结果;

根据所述验证数据集中的样本图像的目标检测结果以及目标检测标签,获取所述检测模型对应的评价指标;

根据所述评价指标,获取所述置信度阈值;

其中,所述评价指标包括精确率、召回率、平均精确度,以及精确率和召回率之间的调和平均分数中的至少一项。

8. 一种图像目标检测装置,其特征在于,包括:

获取单元,用于获取待检测图像;

检测单元,用于将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果;

其中,所述检测模型是基于样本数据集中的样本图像以及所述样本图像的目标检测标签对目标模型进行训练得到的;所述目标模型是对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝得到的。

9. 一种电子设备,包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,其特征在于,所述处理器执行所述程序时实现如权利要求1至7任一项所述图像目标检测方法。

10. 一种非暂态计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至7任一项所述图像目标检测方法。

图像目标检测方法及装置

技术领域

[0001] 本发明涉及数据处理技术领域,尤其涉及一种图像目标检测方法及装置。

背景技术

[0002] 图像目标检测的任务是找出图像中所有感兴趣的目标,并确定目标的位置和类别,这是计算机视觉中的核心问题之一。

[0003] 在实际应用中为了提高目标检测模型的检测精度,往往在搭建目标检测模型时,通过增加神经网络模型的层数或分支结构复杂度来提升图像目标检测的性能,导致神经网络模型的网络参数较多,调整成本高,且在复杂的场景下很容易因网络参数调整不准确的问题,导致检测准确性差。

发明内容

[0004] 本发明提供一种图像目标检测方法及装置,用以解决现有技术中增加神经网络模型的层数或分支结构复杂度来提升图像目标检测的性能,参数调整成本高,检测准确性差的缺陷,实现在降低成本的同时,确保检测精度。

[0005] 本发明提供一种图像目标检测方法,包括:

[0006] 获取待检测图像;

[0007] 将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果;

[0008] 其中,所述检测模型是基于样本数据集中的样本图像以及所述样本图像的目标检测标签对目标模型进行训练得到的;所述目标模型是对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝得到的。

[0009] 根据本发明提供的一种图像目标检测方法,所述检测模型是基于如下步骤训练得到的:

[0010] 根据所述样本数据集,确定目标训练数据集;

[0011] 根据所述目标训练数据集对所述YOLOX网络模型进行剪枝训练和/或蒸馏训练,得到轻量化神经网络模型;

[0012] 对所述轻量化神经网络模型进行模型结构更新,得到所述目标模型;

[0013] 将所述目标训练数据集中的样本图像输入至所述目标模型,得到所述目标训练数据集中样本图像的目标检测结果;

[0014] 根据所述目标检测结果和所述目标检测标签,确定目标损失函数值;

[0015] 基于所述目标损失函数值,对所述目标模型进行迭代训练,得到所述检测模型。

[0016] 根据本发明提供的一种图像目标检测方法,所述根据所述样本数据集,确定目标训练数据集,包括:

[0017] 对所述样本数据集进行数据集划分,得到初始训练数据集;

[0018] 确定所述初始训练数据集中样本图像的目标的类别信息、位置信息,以及尺寸信息;

- [0019] 根据所述类别信息、所述位置信息,以及所述尺寸信息,确定所述初始训练数据集中样本图像的均衡性;
- [0020] 在根据所述均衡性,确定所述初始训练数据集中的样本图像存在分布不均衡的情况下,对所述初始训练数据集中的样本图像进行数据增强;
- [0021] 对增强后的初始训练数据集进行预处理,得到所述目标训练数据集;
- [0022] 其中,所述预处理包括归一化处理和/或降噪处理。
- [0023] 根据本发明提供一种图像目标检测方法,所述对所述轻量化神经网络模型进行模型结构更新,得到所述目标模型,包括:
- [0024] 将所述轻量化神经网络模型中特征提取网络的跨阶段局部网络更新为基于重参数化机器视觉几何模型构建的模块化网络结构;
- [0025] 在更新后的轻量化神经网络模型的所述模块化网络结构的最后一层增加卷积块注意力机制模块,得到所述目标模型。
- [0026] 根据本发明提供一种图像目标检测方法,所述根据所述目标检测结果和所述目标检测标签,确定目标损失函数值,包括:
- [0027] 根据所述目标检测结果和所述目标检测标签,获取至少一个分支损失函数值;
- [0028] 根据所述至少一个分支损失函数值,确定所述目标损失函数值;所述至少一个分支损失函数值包括位置损失函数值、置信度损失函数值、分类损失函数值以及正则化损失函数值;
- [0029] 其中,所述位置损失函数值是基于如下步骤确定的:
- [0030] 基于所述目标检测结果和所述目标检测标签,获取角度损失函数值、检测框重叠度损失函数值和形状损失函数值;
- [0031] 将所述角度损失函数值与预设值相加,根据相加结果,确定所述检测模型对应的距离损失函数值;
- [0032] 根据所述检测框重叠度损失函数值和所述形状损失函数值,以及所述距离损失函数值,确定所述位置损失函数值。
- [0033] 根据本发明提供一种图像目标检测方法,所述将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果,包括:
- [0034] 将所述待检测图像输入至所述检测模型,得到所述待检测图像中目标的置信度信息、检测框信息和位置信息;
- [0035] 将所述置信度信息与置信度阈值进行比较,根据比较结果确定所述待检测图像中目标的类别预测值;
- [0036] 根据所述类别预测值、所述检测框信息和所述位置信息,确定所述待检测图像的目标检测结果。
- [0037] 根据本发明提供一种图像目标检测方法,所述置信度阈值是基于如下步骤确定的:
- [0038] 对所述样本数据集进行数据集划分,得到验证数据集;
- [0039] 将所述验证数据集中的样本图像输入至所述检测模型,得到所述验证数据集中的样本图像的目标检测结果;
- [0040] 根据所述验证数据集中的样本图像的目标检测结果以及目标检测标签,获取所述

检测模型对应的评价指标;

[0041] 根据所述评价指标,获取所述置信度阈值;

[0042] 其中,所述评价指标包括精确率、召回率、平均精确度,以及精确率和召回率之间的调和平均分数中的至少一项。

[0043] 本发明还提供一种图像目标检测装置,包括:

[0044] 获取单元,用于获取待检测图像;

[0045] 检测单元,用于将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果;

[0046] 其中,所述检测模型是基于样本数据集中的样本图像以及所述样本图像的目标检测标签对目标模型进行训练得到的;所述目标模型是对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝得到的。

[0047] 本发明还提供一种电子设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现如上述任一种所述图像目标检测方法。

[0048] 本发明还提供一种非暂态计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现如上述任一种所述图像目标检测方法。

[0049] 本发明还提供一种计算机程序产品,包括计算机程序,所述计算机程序被处理器执行时实现如上述任一种所述图像目标检测方法。

[0050] 本发明提供的图像目标检测方法及装置,通过对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝之后,采用多场景的样本图像进行微调训练,得到融合多种场景的检测模型,基于检测模型进行图像目标检测可实现在最大限度地压缩模型规模、降低运算成本的同时,可以最大限度地提高目标检测精度。

附图说明

[0051] 为了更清楚地说明本发明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一简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0052] 图1是本发明提供的图像目标检测方法的流程示意图之一;

[0053] 图2是本发明提供的图像目标检测方法的流程示意图之二;

[0054] 图3是本发明提供的检测模型的结构示意图;

[0055] 图4是本发明提供的RepBlock模块的结构示意图;

[0056] 图5是本发明提供的spp模块的结构示意图;

[0057] 图6是本发明提供的图像目标检测装置的结构示意图;

[0058] 图7是本发明提供的电子设备的结构示意图。

具体实施方式

[0059] 为使本发明的目的、技术方案和优点更加清楚,下面将结合本发明中的附图,对本发明中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,

而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0060] 需要说明的是,本实施例提供的图像目标检测方法及装置可以应用于各种所需进行目标检测的场景,如需要进行安全帽检测的场景、需要进行业务办理人员检测的场景、需要进行行人检测的场景等,本实施例对此不作具体地限定。以下以安全帽检测场景为例,对本实施例提供的图像目标检测方法展开描述,对于其它场景的图像目标检测,本实施例提供的方案同样适用。

[0061] 安全帽是现代工业制造、生产中对工人的安全保障的必备装备,通过对安全帽是否佩戴、以及佩戴是否合理进行检测,可以对施工区域内的人的安全进行预警。因此,如何对安全帽进行高效准确地检测是目前业界亟待解决的重要课题。

[0062] 而在实际应用中为了提高目标检测模型的检测精度,往往在搭建目标检测模型时,通过增加神经网络模型的层数或分支结构复杂度来提升图像目标检测的性能,导致神经网络模型的网络参数较多,而模型参数过大在边缘计算设备上的帧率较低,调整成本高,且在复杂的工地场景下很容易因网络参数调整不准确的问题,导致漏检和误检,严重影响检测准确性。

[0063] 针对上述问题,本实施例提供一种图像目标检测方法及装置,通过对目标检测模型YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝之后,再利用样本数据集对微调,由此实现在最大限度地压缩模型规模、降低运算量的同时,可以最大限度地提高图像目标检测精度,由此解决现有技术中安全帽在复杂场景下的误检率、漏检率都较高的问题,并且平衡耗时的的问题。

[0064] 需要说明的是,该方法的执行主体可以是电子设备、电子设备中的部件、集成电路、或芯片。该电子设备可以是移动电子设备,也可以为非移动电子设备。示例性的,移动电子设备可以为手机、平板电脑、笔记本电脑、掌上电脑、车载电子设备、可穿戴设备、超级移动个人计算机(ultra-mobile personal computer,UMPC)、上网本或者个人数字助理(personal digital assistant,PDA)等,非移动电子设备可以为服务器、网络附属存储器(Network Attached Storage,NAS)、个人计算机(personal computer,PC)、电视机(television,TV)、柜员机或者自助机等,本发明不作具体限定。

[0065] 下面结合图1-图6描述本发明的图像目标检测方法及装置,该方法和装置可以基于各种程序语言实现,如python语言或C++语言实现。

[0066] 图1为本实施例提供的图像目标检测方法的流程示意图之一,如图1所示,该方法包括如下步骤:

[0067] 步骤101,获取待检测图像;

[0068] 需要说明的是,在安全帽检测场景,待检测图像为在工地采集的所需进行安全帽检测的图像。

[0069] 可选地,待检测图像可以是通过相机或者带有相机的智能终端实时拍摄得到,也可以是扫描得到,还可以是通过互联网传输或者下载得到;本实施例对待检测图像的获取方式不作具体限定。

[0070] 可以理解的是,在获取到待检测图像后,可以对待检测图像进行预处理,包括但不限于,尺度归一化处理、图像对齐处理和滤波处理等,以提高图像目标检测的效率和精度。

[0071] 步骤102,将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果;其中,所述检测模型是基于样本数据集中的样本图像以及所述样本图像的目标检测标签对目标模型进行训练得到的;所述目标模型是对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝得到的。

[0072] 其中,检测模型是用于对图像进行目标检测,以输出图像的目标检测结果,所称的目标检测结果包括但不限于图像中目标的类别预测值、检测框信息和位置信息中的至少一项。

[0073] 需要说明的是,在安全帽检测场景,所称的类别预测值用于表征图像中的人是否佩戴安全帽,检测框信息用于标识安全帽对应的检测框,或者标识安全帽对应的检测框以及人对应的检测框;位置信息用于标识安全帽的位置信息,或者用于标识人对应的位置和安全帽的位置信息。

[0074] 检测模型是通过对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝后,再基于样本数据集进行微调训练得到的;所称的结构更新包括但不限于进行网络层替换和/或新增网络层。

[0075] 如图2所示,在执行步骤102之前,需要预先训练得到可精准进行图像目标检测的轻量级的检测模型,该检测模型是基于如下步骤训练的得到:

[0076] 首先,为了提高图像识别精度,需要采集多种不同类别的样本图像,且每一类别下对应多种不同场景的样本图像,以使样本数据集具备足够的深度和广度,以使得训练得到检测模型可在不同场景下进行精准地目标检测。

[0077] 所称的样本数据集可以是历史采集获取的,如将历史采集的数据进行抽帧、去重、标注,制作为样本数据集;也可以是在开源数据库中加载获取的,如从SHWD(Safety Helmet Wearing Detect Dataset,安全帽佩戴检测数据集)数据集中筛选获取。

[0078] 示例性地,由于非工地场景不符合安全帽检测应用场景,将SHWD数据集中的非工地场景数据进行剔除,并对将实际应用场景的样本图像进行标签标注,如图像中的人的类别标签可以为“Person”、安全帽的类别标注为“Hat”、其他的帽子的类别标注为“Ehat”、未佩戴带帽子的类别标注为“Nhat”进行标注,以形成包括2万张样本图像的样本数据集;所称的2万张样本图像中可以是包括4000张公开数据,以及16000张为私有数据。

[0079] 接着,基于样本数据集,对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝得到的目标模型。

[0080] 此处,确定目标模型的方式,可以先对YOLOX网络模型进行蒸馏训练和/或模型剪枝,得到轻量级神经网络模型,再对轻量级神经网络模型进行结构更新;或者先对YOLOX网络模型进行结构更新,再对更新后的YOLOX网络模型进行蒸馏训练和/或模型剪枝,本实施例对此不作具体地限定。

[0081] 接着,基于样本数据集对目标模型迭代进行微调训练,从而得到即具备高检测精确、又具有轻量级结构的检测模型。

[0082] 同时为了进一步验证检测模型的有效性,还可以基于精确率P、召回率R、平均精确度Map,以及精确率和召回率之间的调和平均分数FI中的至少一项对检测模型进行性能评价;还可以基于P、召回率R、平均精确度Map,以及精确率和召回率之间的调和平均分数FI中的至少一项确定检测模型对应的最佳置信度阈值。

[0083] 可选地,在通过迭代训练得到检测模型之后,可以将检测模型部署至边缘设备(或嵌入式设备端),以便在需要对待检测图像进行目标检测时,调用边缘设备中的检测模型,即可检测得到待检测图像的目标检测结果。

[0084] 需要说明的是,在获取到待检测图像的目标检测结果之后可以是在边缘设备中对目标检测结果进行显示、存储以及异常告警等。

[0085] 本实施例提供的图像目标检测方法,通过对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝之后,采用多场景的样本图像进行微调训练,得到融合多种场景的检测模型,基于检测模型进行图像目标检测可实现在最大限度地压缩模型规模、降低运算成本的同时,可以最大限度地提高目标检测精度。

[0086] 在一些实施例中,所述检测模型是基于如下步骤训练得到的:

[0087] 根据所述样本数据集,确定目标训练数据集;

[0088] 根据所述目标训练数据集对所述YOLOX网络模型进行剪枝训练和/或蒸馏训练,得到轻量化神经网络模型;

[0089] 对所述轻量化神经网络模型进行模型结构更新,得到所述目标模型;

[0090] 将所述目标训练数据集中的样本图像输入至所述目标模型,得到所述目标训练数据集中样本图像的目标检测结果;

[0091] 根据所述目标检测结果和所述目标检测标签,确定目标损失函数值;

[0092] 基于所述目标损失函数值,对所述目标模型进行迭代训练,得到所述检测模型。

[0093] 如图2所示,检测模型的训练步骤进一步包括:

[0094] 首先,将样本数据集按照一定的比例,如8:1:1的比例进行数据集划分,以得到初始训练样本集、初始验证数据集以及初始测试数据集。

[0095] 需要说明的是,在进行数据集划分的过程中,为了保证数据分布的一致性,使得各数据集的方差更小,模型的泛化能力更强,可以对样本数据集中所包括的公开数据集和私有数据集进行分层划分,具体可以按照一定比例从公开数据集提取一定比例的训练样本、测试样本以及验证样本;再从私有数据集提取一定比例的训练样本、测试样本以及验证样本;再将从公开数据集以及私有数据集中提取的训练样本形成训练样本集,将从公开数据集以及私有数据集中提取的测试样本形成测试样本集,将从公开数据集以及私有数据集中提取的验证样本形成验证样本集,使得训练样本集、测试样本集以验证样本集从不同数据来源的样本数据的比例保持一致。

[0096] 接着,可以直接将初始训练数据集作为目标训练数据集,或者对初始训练数据集进行处理后,得到目标训练数据集,本实施例对此不作具体地限定。

[0097] 在一些实施例中,根据所述样本数据集,确定目标训练数据集的步骤进一步包括:

[0098] 对所述样本数据集进行数据集划分,得到初始训练数据集;

[0099] 确定所述初始训练数据集中样本图像的目标的类别信息、位置信息,以及尺寸信息;

[0100] 根据所述类别信息、所述位置信息,以及所述尺寸信息,确定所述初始训练数据集中样本图像的均衡性;

[0101] 在根据所述均衡性,确定所述初始训练数据集中的样本图像存在分布不均衡的情况下,对所述初始训练数据集中的样本图像进行数据增强;

[0102] 对增强后的初始训练数据集进行预处理,得到所述目标训练数据集;

[0103] 其中,所述预处理包括归一化处理 and/或多尺度变换处理。

[0104] 可选地,在将样本数据集按照一定的比例划分得到初始训练样本集之后,为了使得训练得到的检测模型的泛化性更强,需要确保训练数据集中样本图像分布均衡,为此还可以对初始训练数据集执行如下数据增强步骤:

[0105] 确定初始训练数据集中样本图像的目标的类别信息、位置信息、以及尺寸信息,通过对类别信息、位置信息、以及尺寸信息进行分析,得到初始训练数据集中样本图像的目标的类别的个数比例、位置分布、以及尺寸分布,进而确定初始训练数据集中样本图像的均衡性;在根据均衡性确定初始训练数据集中样本图像分布不均衡的情况下,对初始训练数据集中类别分布不均衡、位置分布不均衡以及尺寸分布不均衡的样本图像进行数据增强,以确保初始训练数据集中样本图像分布的均衡性,由此提高检测模型的泛化性,以适应各种场景的图像目标检测。所称的数据增强包括但不限于用于将图片裁剪成很多部分以进行数据增强的CROP (Common, Random, Preprocessing Data Augmentation Methods, 常规、随机、预处理数据增强方法)、仿射变换等数据增强手段对进行有效扩增,以有效提升各场景下图像的检测精度。

[0106] 为了进一步提高模型训练的收敛速度以及泛化性,还可以对增强后的初始训练数据集进行预处理,包括但不限于归一化处理和/或多尺度变换,以得到目标训练数据集。

[0107] 其中,所称的归一化处理可以是对样本图像采用基于ImageNet的均值和方差进行归一化;

[0108] 所称的样本图像的尺寸可以根据实际需求进行设置,如640*640所述称的多尺度变换,可以是在预先设定的随机变化系数区间,如[0.1, 2]对样本图像进行多尺度变换。

[0109] 进一步地,在获取到目标训练数据集之后,可以是基于目标训练数据集对YOLOX网络模型进行剪枝训练和/或蒸馏训练,得到轻量化神经网络模型。

[0110] 可选地,基于YOLOX网络构建可用于图像目标检测的教师模型和学生模型;所称的教师模型可以是大规模的YOLOX-x (带有额外增强的目标检测模型) 网络,所称的学生模型可以是小规模无anchor (锚点, 也即预定义边界框) 和解耦头的YOLOX-s (具有小型网络结构的目标检测模型) 网络。通过采用无anchor的YOLOX-s网络可以有效减少超参数的设定,降低了网络的检测头的输出结果部分,进而提高模型检测效率。

[0111] 接着,基于目标训练数据集对YOLOX-x网络进行初次训练,对小规模的YOLOX-s网络进行剪枝训练;然后,基于训练后的YOLOX-x网络输出的样本图像的目标检测结果对剪枝训练后的小规模的YOLOX-s网络进行蒸馏训练,得到轻量化神经网络模型。

[0112] 所称的剪枝可以是采用剪枝算法,如基于BN (Batch Normalization) 层的通道剪枝技术,对骨干网络的BN层的权重加入L1正则化,以使得在检测模型的精度不造成的影响的前提下,对检测模型的网络参数进行稀疏化,至少可消除50%左右的冗余参数,提高了网络的推理速度。

[0113] 所称的剪枝可以是采用递进式剪枝,以防止检测模型在一次幅度剪枝导致精度下降较多,防止进行多次地缩减参数,多次训练微调,使得同等参数量的情况下,模型的精度较高。

[0114] 接着,在获取到轻量化神经网络模型之后,对轻量化神经网络模型进行模型结构

更新,以得到目标模型;具体可以是对部分网络结构进行替换和/或增加新的网络结构以实现模型结构更新。

[0115] 在一些实施例中,对所述轻量化神经网络模型进行模型结构更新,得到所述目标模型,包括:

[0116] 将所述轻量化神经网络模型中特征提取网络的跨阶段局部网络更新为基于重参数化机器视觉几何模型构建的模块化网络结构;

[0117] 在更新后的轻量化神经网络模型的所述模块化网络结构的最后一层增加卷积块注意力机制模块,得到所述目标模型。

[0118] 图3为本实施例提供的检测模型的结构示意图;如图3所示,将轻量化神经网络模型中特征提取网络的跨阶段局部网络(Cross-Stage Partial Connections Darknet, Cspdarknet)替换为高效的重参数化机器视觉几何模型(Re-param Visual Geometry Group, RepVGG Block)构建的模块化网络结构,以使得检测模型,在训练时采取类似残差的跨层连接的结构,通过多个不同尺度的卷积核来获取特征信息,将多个特征再进行线性融合,这样增强特征信息的提取,检测模型再推理时,对并行分支进行线性叠加后的对网络结构进行重建为结构整齐,运行效率高,易进行剪枝的单路结构,并且嵌入式板端对3*3的卷积进行了优化,提高模型在板端的运行速度、降低耗时,以及提高检测模型的精度。

[0119] 接着,对更新后的轻量化神经网络模型的Repvgg block的模块化网络结构的最后一层加入CBAM(Convolutional Block Attention Module,卷积块注意力模块)注意力机制,得到目标模型,实现增强了网络学习特征的表达能力,且对特征提取网络提取的特征的通道和空间进行了增强,弥补了网络的剪枝带来的影响。

[0120] 另外,通过消融对比试验,确定本实施例提供的方法通过在网络结构的SPP(Spatial Pyramid Pooling,空间金字塔池化)层,也即Repvgg block的模块化网络结构的最后一层后加入注意力机制模块可达到较佳的检测效果。

[0121] 如图3所示,检测模型的具体结构包括但不限于STEM层,用于表示特征提取网络的初始层,主要是由RepVGG Block构建生成的、RepVGG Block层、Repblock层、SPP(Spatial Pyramid Pooling,空间金字塔池化)层、CBAM层、Upsp(Upsample,上采样操作)层、Ccat(Concat,拼接操作)层、C3(Csp Bottleneck With 3Convs,包括3个卷积层以及一层瓶颈层的网络层),以及CBS(包括conv层、BN层以及silu(Sigmoid Linear Unit Layer,Sigmoid非线性激活单元)的网络层);

[0122] 如图4所示,Repblock是多个RepVGG Block堆叠形成的,RepVGG Block是由多个conv(卷积)层以及BN(Batch Normalization,批归一化)层。

[0123] 如图5所示,SPP(Spatial Pyramid Pooling,空间金字塔池化)包括多层CBS层、多层Max pool(最大池化)层以及Ccat;其中,CBS层包括conv层、BN层以及silu的网络层。

[0124] 进一步地,在获取到目标模型之后,可以将目标训练数据集的样本图像输入至目标模型,得到目标模型输出的样本图像的目标检测结果;所称的目标检测结果可以是经过标签平滑处理的检测结果。

[0125] 接着,根据目标检测结果和目标检测标签,确定目标损失函数值。

[0126] 在一些实施例中,根据所述目标检测结果和所述目标检测标签,确定目标损失函数值,包括:

[0127] 根据所述目标检测结果和所述目标检测标签,获取至少一个分支损失函数值;

[0128] 根据所述至少一个分支损失函数值,确定所述目标损失函数值;所述至少一个分支损失函数值包括位置损失函数值、置信度损失函数值、分类损失函数值以及正则化损失函数值;

[0129] 其中,所述位置损失函数值是基于如下步骤确定的:

[0130] 基于所述目标检测结果和所述目标检测标签,获取角度损失函数值、检测框重叠度损失函数值和形状损失函数值;

[0131] 将所述角度损失函数值与预设值相加,根据相加结果,确定所述检测模型对应的距离损失函数值;

[0132] 根据所述检测框重叠度损失函数值和所述形状损失函数值,以及所述距离损失函数值,确定所述位置损失函数值。

[0133] 可选地,位置损失函数值的计算公式如下:

$$[0134] \quad \text{Loss}_{\text{sioU}} = 1 - \text{Iou} + \frac{\Delta + \Omega}{2};$$

[0135] 其中, $\text{Loss}_{\text{sioU}}$ 为位置损失函数值,Iou为检测框重叠度损失函数值,也即样本图像的目标检测结果中的预测框与目标检测标签中的真实框之间的重合度的度量值; Δ 为距离损失函数值、 Ω 为形状损失函数值。

[0136] 其中,距离损失函数值的计算公式如下:

$$[0137] \quad \Delta = \sum_{t=x,y} (1 - e^{-m_t});$$

$$[0138] \quad \rho_x = \left(\frac{b_{cx}^{gt} - b_{cx}}{c_{w2}} \right)^2;$$

$$[0139] \quad \rho_y = \left(\frac{b_{cy}^{gt} - b_{cy}}{c_{h2}} \right)^2;$$

$$[0140] \quad \gamma = 1 + \Lambda;$$

$$[0141] \quad \Lambda = 1 - 2\sin^2\left(\arcsin\left(\frac{c_{h1}}{\sigma}\right) - \frac{\pi}{4}\right);$$

$$[0142] \quad \sin \alpha = \frac{c_{h1}}{\sigma};$$

[0143] 其中, σ 是预测框和真实框之间的中心点距离, $\alpha \in \left[0, \frac{\pi}{4}\right]$ 为角度系数, $\Lambda \in [0, 1]$

为角度损失函数值,其在 α 的区间内是单调递增的; x 和 y 分别为横坐标和纵坐标, b_{cx}^{gt} 和 b_{cy}^{gt} 分别为真实框的中心点的横坐标和纵坐标, b_{cx} 和 b_{cy} 分别为预测框的中心点的横坐标和纵坐标, c_{w2} 和 c_{h2} 分别为预测框和真实框之间的最小外接矩形框的宽和高, c_{h1} 为预测框和真实框的中心点之间的高度, ρ_x 为真实框和预测框的 x 方向的归一化距离, ρ_y 为真实框和预测框的 y 方向的归一化距离, γ 是角度损失对距离损失的一个惩罚因子,角度损失越大, γ 越大,距离损失越大。

[0144] 同时根据目标检测结果和目标检测标签,获取检测模型对应的置信度损失函数值、分类损失函数值以及正则化损失函数值。

[0145] 进一步地,联合位置损失函数值、置信度损失函数值、分类损失函数值以及正则化损失函数值确定检测模型的损失函数值,由此根据检测模型的损失函数值进行迭代训练,得到可对图像进行精准目标检测的检测模型。具体确定方式可以是加权相加或直接相加,本实施例对此不作具体地限定。

[0146] 所称的置信度损失函数值和分类损失函数值可以是采用BCE Loss(Binary Cross Entropy Loss,二值交叉熵损失函数值)等,本实施例对此不作具体地限定。

[0147] 正则化损失函数值用于对网络进行稀疏化训练,由此降低检测模型训练所需的参数,降低检测模型过拟合的风险。

[0148] 本实施例提供的方法,通过将距离损失函数值 Δ 中的 γ 的单调性与角度损失函数值 Δ 保持一致,使得距离损失函数值 Δ 相对于 γ 为单调递增,由此使得检测网络在快速降低角度损失函数值的同时,使得预测框与真实框在同一水平上,有效提升了检测网络的收敛速,并且快速获取到局部最优解,由此检测网络的检测精度。

[0149] 进一步地,采用目标损失函数值,基于优化器,对目标模型进行微调训练,直到达到最大迭代次数/或模型收敛,即可得到检测模型。

[0150] 需要说明的是,所称的优化器可以采用SGD(Stochastic Gradient Descent,随机梯度下降)作为优化器对梯度进行更新,并将目标损失函数值进行反向传播对目标模型的参数进行迭代优化。

[0151] 所称的最大迭代次数可以是根据实际需求进行预先设定的,如200次,或者通过可视化工具tensorboard观测写入到日志中的网络损失函数值的变化状态,观察网络损失函数值在某一迭代周期内趋于稳定,如110次到150次的情况下,重新更新最大迭代次数,再次基于更新的最大迭代次数进行目标模型重训练,直到得到最优的最大迭代次数,由此动态降低训练的总迭代次数,防止模型过训练,提升模型的鲁棒性。

[0152] 另外,为了提高正样本的质量(也即配带安全帽的样本图像)以及小目标的检测精度,在模型迭代训练过程中,还可以实时调整目标次迭代过程中Mosaic(裁剪拼接增强)的概率以及Mixup(混合增强)的概率,如在最后25个迭代训练过程中,将Mosaic的概率设置0.3,将Mixup的概率将为0。

[0153] 其中,Mosaic、Mixup是网络模型训练过程中为了增强数据的多样性常用的数据增强的方式。Mosaic用于对多张图像进行随机裁剪,再拼接到一张图作为训练数据;Mixup用于将随机的两张样本图像按比例混合,分类的结果按比例分配。

[0154] 此外,在训练过程中,还采用SimOTA(Similarity Optimized Training Algorithm,相似性优化训练算法)动态匹配策略,一方面通过引入中心半径和区域内部点的概念,对正样本进行扩充,从而增加正样本的数量并提高匹配的准确性,另一方面通过合理选择中心半径和区域内部点的个数,可以在一定程度上控制正样本的数量,以使得正样本的数量控制在[1,10]区间,避免正样本数量过多或过少,从而使模型能够更好地学习目标特征,并减少对负样本的误判,由此改进了传统分配算法前期带来的低质量样本,将中心半径以及区域内部的点合并扩充正样本的数量和提高了网络的收敛速度。

[0155] 本实施例提供的方法,通过针对YOLOX网络模型的模型结构的改进、注意力机制的

添加、模型的剪枝和蒸馏,得到目标模型,并基于改进的损失函数值(也即目标损失函数值)对目标模型进行微调训练,由此得到泛化性、鲁棒性以及检测精确性更强、更轻量化的检测模型,由此提高目标检测的计算耗时、提高检测的准确率以及鲁棒性。

[0156] 在一些实施例中,步骤102进一步包括:

[0157] 将所述待检测图像输入至所述检测模型,得到所述待检测图像中目标的置信度信息、检测框信息和位置信息;

[0158] 将所述置信度信息与置信度阈值进行比较,根据比较结果确定所述待检测图像中目标的类别预测值;

[0159] 根据所述类别预测值、所述检测框信息和所述位置信息,确定所述待检测图像的目标检测结果。

[0160] 可选地,将待检测图像输入至检测模型,由检测模型输出的待检测图像中目标的置信度信息、检测框信息和位置信息;

[0161] 接着,将目标的置信度信息与置信度阈值进行比较,以根据比较结果确定目标的类别;示例性地,若目标的置信度信息大于或等于置信度阈值,则确定目标的类别为佩戴头盔;若小于置信度阈值,则确定目标的类别为未佩戴头盔。

[0162] 此处的置信度阈值可以根据实际需求进行设置,也可以是根据验证数据集进行验证确定的,本实施例对此不作具体地限定。

[0163] 在一些实施例中,所述置信度阈值是基于如下步骤确定的:

[0164] 对所述样本数据集进行数据集划分,得到验证数据集;

[0165] 将所述验证数据集中的样本图像输入至所述检测模型,得到所述验证数据集中的样本图像的目标检测结果;

[0166] 根据所述验证数据集中的样本图像的目标检测结果以及目标检测标签,获取所述检测模型对应的评价指标;

[0167] 根据所述评价指标,获取所述置信度阈值;

[0168] 其中,所述评价指标包括精确率、召回率、平均精确度,以及精确率和召回率之间的调和平均分数中的至少一项。

[0169] 可选地,置信度阈值的确定步骤包括:

[0170] 首先,按照一定比例对样本数据集进行划分得到验证数据集;将验证数据集的样本图像输入至检测模型中,得到检测模型输出的样本图像的目标检测结果。

[0171] 接着,将样本图像的目标检测结果与样本图像的目标检测标签进行对比,以获取检测模型对应的精确率、召回率、平均精确度,以及精确率和召回率之间的调和平均分数;

[0172] 对检测模型对应的精确率、召回率、平均精确度,以及精确率和召回率之间的调和平均分数进行加权相加,得到检测模型的总评价值,通过置信度阈值进行分割,得到总评价值曲线,将最大总评价值对应的置信度阈值作为检测模型对应的置信度阈值。

[0173] 接着,将类别预测值、检测框信息和位置信息整合作为待检测图像的目标检测结果。

[0174] 本实施例提供的方法,通过检测模型对应的评价指标适应性地确定目标分类所需参照的置信度阈值,由此提高目标检测的准确性和泛化性。

[0175] 下面对本发明提供的图像目标检测装置进行描述,下文描述的图像目标检测装置

与上文描述的图像目标检测方法可相互对应参照。

[0176] 如图6所示,本实施例提供一种图像目标检测装置,该装置包括:

[0177] 获取单元601用于获取待检测图像;

[0178] 检测单元602用于将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果;

[0179] 其中,所述检测模型是基于样本数据集中的样本图像以及所述样本图像的目标检测标签对目标模型进行训练得到的;所述目标模型是对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝得到的。

[0180] 本实施例提供的图像目标检测装置,对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝之后,采用多场景的样本图像进行微调训练,得到融合多种场景的检测模型,基于检测模型进行图像目标检测可实现在最大限度地压缩模型规模、降低运算成本的同时,可以最大限度地提高目标检测精度。

[0181] 图7示例了一种电子设备的实体结构示意图,如图7所示,该电子设备可以包括:处理器(processor)701、通信接口(Communications Interface)702、存储器(memory)703和通信总线704,其中,处理器701,通信接口702,存储器703通过通信总线704完成相互间的通信。处理器701可以调用存储器703中的逻辑指令,以执行上述任一实施例提供的图像目标检测方法,该方法包括:获取待检测图像;将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果;其中,所述检测模型是基于样本数据集中的样本图像以及所述样本图像的目标检测标签对目标模型进行训练得到的;所述目标模型是对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝得到的。

[0182] 此外,上述的存储器703中的逻辑指令可以通过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,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品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,服务器,或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0183] 另一方面,本发明还提供一种计算机程序产品,所述计算机程序产品包括计算机程序,计算机程序可存储在非暂态计算机可读存储介质上,所述计算机程序被处理器执行时,计算机能够执行上述各方法所提供的图像目标检测方法,该方法包括:获取待检测图像;将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果;其中,所述检测模型是基于样本数据集中的样本图像以及所述样本图像的目标检测标签对目标模型进行训练得到的;所述目标模型是对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝得到的。

[0184] 又一方面,本发明还提供一种非暂态计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现以执行上述各方法提供的图像目标检测方法,该方法包括:获取待检测图像;将所述待检测图像输入至检测模型,得到所述待检测图像的目标检测结果;其中,所述检测模型是基于样本数据集中的样本图像以及所述样本图像的目标

检测标签对目标模型进行训练得到的;所述目标模型是对YOLOX网络模型进行结构更新,以及蒸馏训练和/或模型剪枝得到的。

[0185] 以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0186] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件。基于这样的理解,上述技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在计算机可读存储介质中,如ROM/RAM、磁碟、光盘等,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,服务器,或者网络设备等)执行各个实施例或者实施例的某些部分所述的方法。

[0187] 最后应说明的是:以上实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

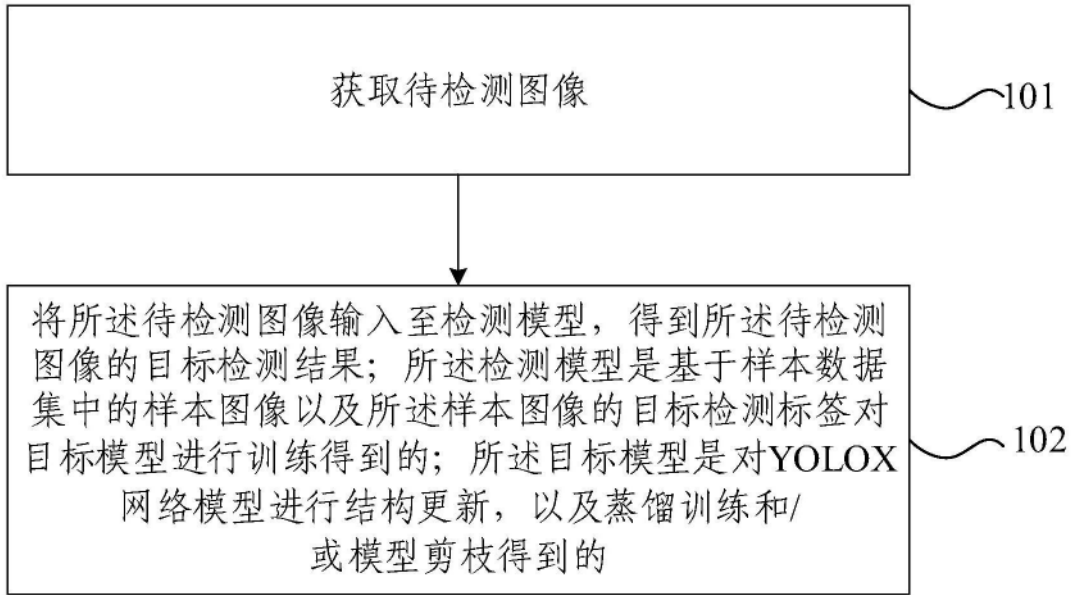


图1

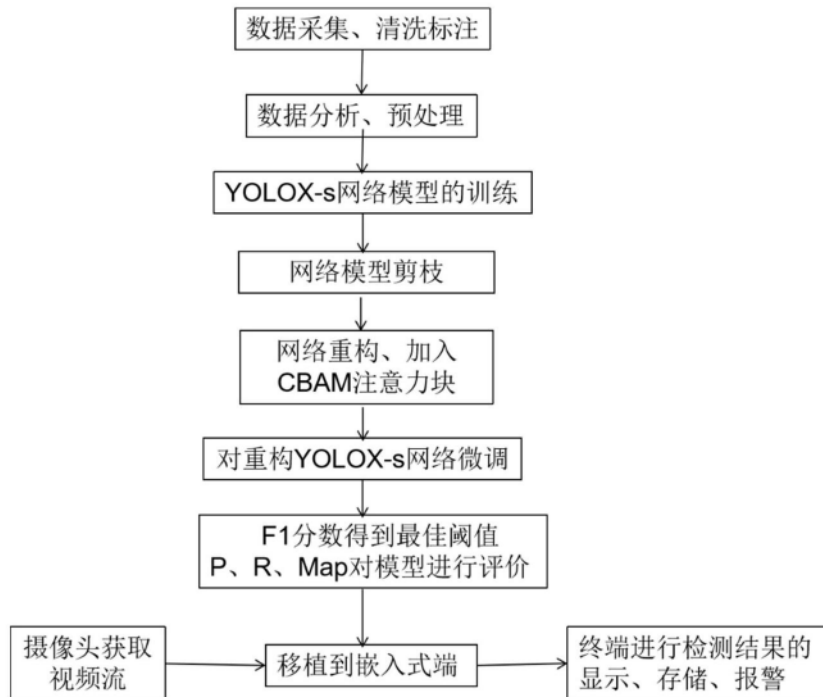


图2

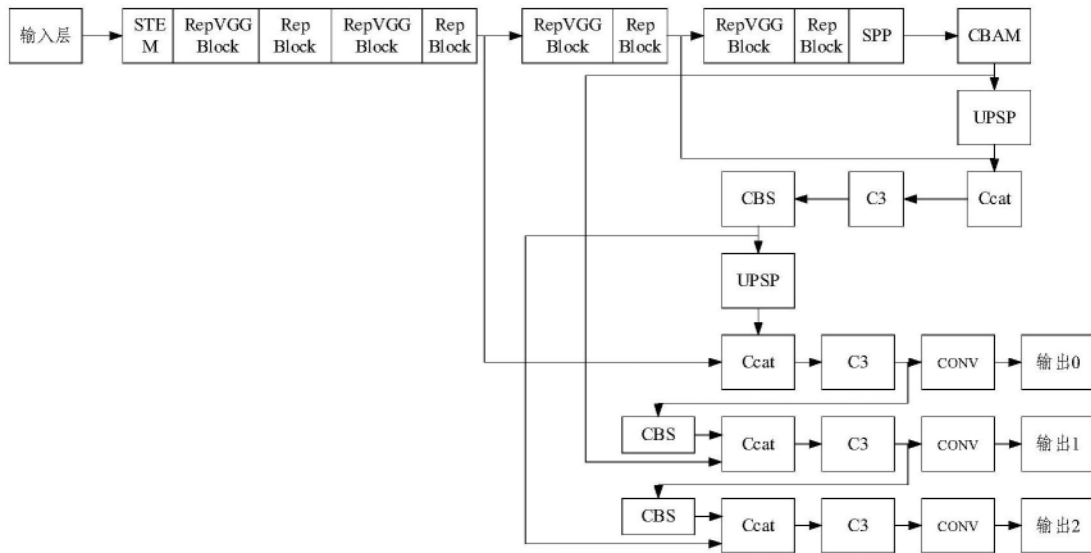


图3

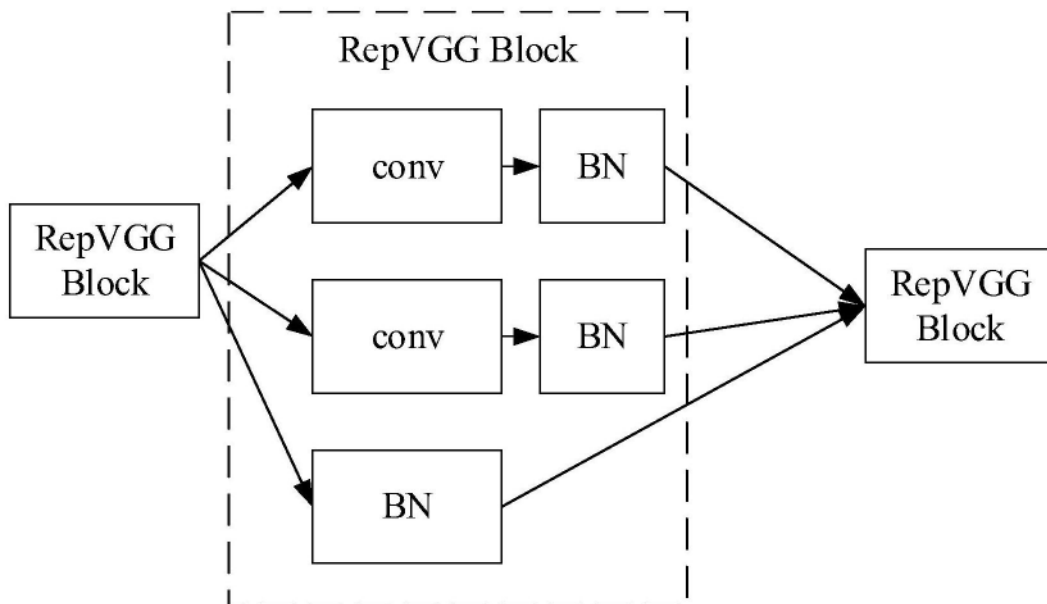


图4

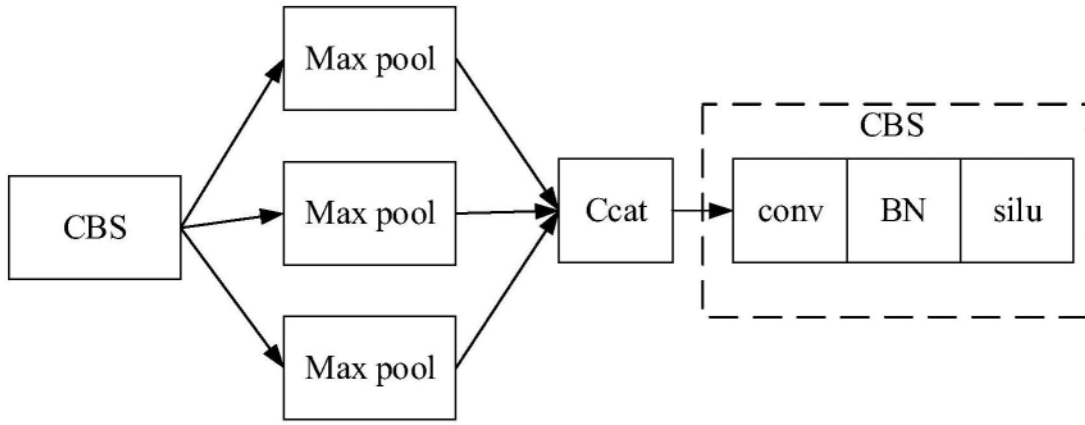


图5

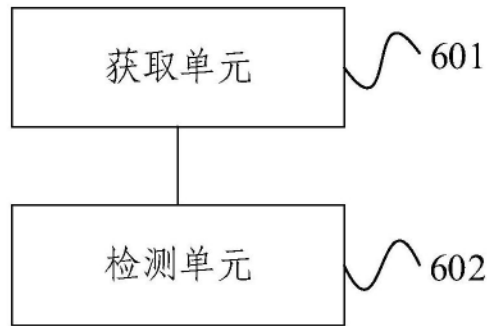


图6

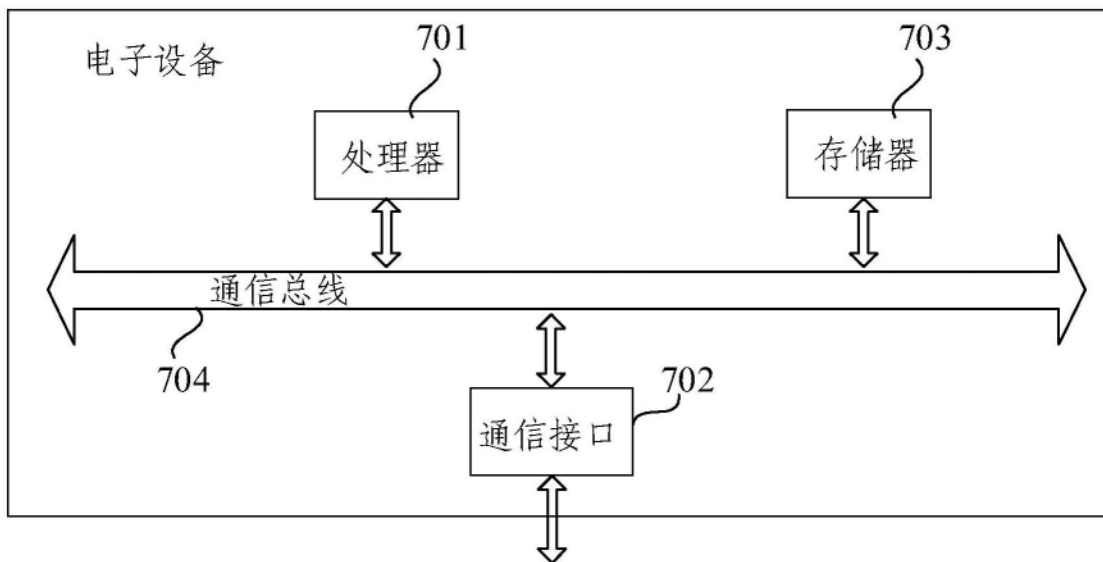


图7