(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112395957 B
(45)授权公告日 2024.06.04

(21)申请号 202011174089.5

(22)申请日 2020.10.28

(65)同一申请的已公布的文献号
申请公布号 CN 112395957 A

(43)申请公布日 2021.02.23

(73)专利权人 连云港杰瑞电子有限公司
地址 222000 江苏省连云港市高新区圣湖路18号

(72)发明人 张宇杰 项俊平 刘建华 张锋鑫 高超

(74)专利代理机构 连云港润知专利代理事务所 32255
专利代理师 刘喜莲

(51)Int.Cl.
G06V 10/778(2022.01)
G06V 20/40(2022.01)
G06V 10/762(2022.01)
G06V 10/82(2022.01)
G06V 10/25(2022.01)
G06N 3/0464(2023.01)

(56)对比文件
CN 110929560 A,2020.03.27
WO 2020206861 A1,2020.10.15

审查员 李纯菊

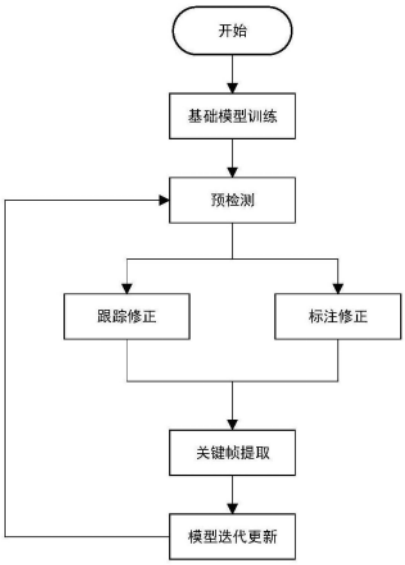
权利要求书2页 说明书5页 附图3页

(54)发明名称

一种针对视频目标检测的在线学习方法

(57)摘要

本发明公开了一种针对视频目标检测的在线学习方法,属于机器视觉领域。包括模型预训练、目标检测、跟踪修正、标注修正、关键帧提取和模型迭代更新。先利用开源或自标注数据集训练一个基础版本的当前模型;利用当前模型对视频序列进行预检测;利用改进的KCF跟踪算法和基于特征空间的k近邻算法对预检测结果分别进行方框修正和标注修正;利用基于特征空间相似度度量的关键帧提取方法,提取视频关键帧,去除重复图像;利用关键帧数据和修正检测结果对模型进行训练,实现模型的迭代更新。本发明该方法通过跟踪算法和聚类分析对检测和标注结果进行修正,利用修正后的结果重新训练目标检测模型,实现模型性能的不断改进,实现应用场景的自适应。



CN 112395957 B

1.一种针对视频目标检测的在线学习方法,其特征在于:该方法包括模型预训练、目标检测、跟踪修正、标注修正、关键帧提取和模型迭代更新,该方法具体包括如下步骤:

步骤1:利用开源或自标注数据集训练一个基础版本的改进YOLOv3目标检测模型,作为当前模型;

步骤2:利用当前模型对视频序列进行预检测,获取初始检测方框和目标类别;

步骤3:利用改进的KCF跟踪算法和基于特征空间的k近邻算法对预检测结果分别进行方框修正和标注修正;

步骤4:利用基于特征空间相似度度量的关键帧提取方法,提取视频关键帧,以压缩数据集大小,去除重复图像;

步骤5:利用关键帧数据和修正检测结果对模型进行训练,实现模型的迭代更新;

步骤6:回到步骤2,重复以上步骤2—步骤5操作;

步骤3所述方框修正的步骤包括:

步骤3.1:对于第n帧图像的所有预检测目标都初始化一个KCF跟踪器,分别进行正向和负向跟踪,得到邻近图像帧中的跟踪目标;

步骤3.2:对于临近帧k,计算其与前一帧的帧间差分图像,获取运动目标概率图;

步骤3.3:对于k帧中的每个跟踪目标,判断其是否静止,如果该目标处于运动状态,则根据运动目标概率图计算该目标的运动目标概率,如果运动目标概率值小于给定阈值,则认为该目标跟踪失败,停止跟踪;

步骤3.4:对每一帧都进行上述步骤3.1—步骤3.3处理,则可以得到所有图像中由跟踪器捕捉到的目标方框;

步骤3.5:将检测模型得到的目标方框与跟踪器捕捉到的目标方框进行融合,同时进行极大抑制算法NMS得到修正后的目标方框;

所述标注修正的步骤包括:

步骤3.6:获取ImageNet预训练VGG16网络模型;

步骤3.7:将检测得到的每个目标方框缩放到统一尺寸,然后传入VGG16网络,得到1000维的特征向量;

步骤3.8:计算不同目标特征之间的欧式距离作为目标相似度度量方式;

步骤3.9:对于每个检测目标,通过k近邻算法找出与其最近的k个目标,然后采用投票机制决定该检测目标的目标类别。

2.根据权利要求1所述的一种针对视频目标检测的在线学习方法,其特征在于:步骤1所述的改进的YOLOv3模型采用全尺度网络OSNet作为特征提取网络,特征金字塔中的上采样方法采用逆卷积神经网络实现,BBox的回归损失函数用GIOU替代MSE,anchors大小的选择仍采用聚类算法,得到9个聚类中心,形成改进的YOLOv3模型。

3.根据权利要求1所述的一种针对视频目标检测的在线学习方法,其特征在于:步骤3所述跟踪修正方法是:引入了帧间差分估计运动目标概率,首先利用跟踪算法对检测目标进行跟踪,判断目标是否静止,如果目标运动,则利用帧间差分获取方框目标概率,如果小于给定阈值,则认为跟踪失败,停止跟踪,利用跟踪结果进一步修正检测结果,即将跟踪到而未检测到的目标作为漏检目标添加到检测目标列表中。

4.根据权利要求1所述的一种针对视频目标检测的在线学习方法,其特征在于:步骤3

所述标注修正方法是,利用神经网络获取检测目标的特征,通过k近邻算法对目标标注进行投票更新,修正检测结果标注。

5.根据权利要求1所述的一种针对视频目标检测的在线学习方法,其特征在于:步骤4所述的关键帧提取方法是:利用目标检测网络特征提取层的输出作为图像特征,计算图像之间的距离来衡量图像的相似度,选取相似度的局部极大值作为视频关键帧。

6.根据权利要求1—5中任何一项所述的一种针对视频目标检测的在线学习方法,其特征在于:步骤1所述的模型训练的工作步骤包括:

步骤1.1:收集开源数据集,或采集特定场景下的视频数据,人工标注检测目标位置方框和目标类别,建立数据集;

步骤1.2:对数据集进行旋转、平移、缩放和镜像变换、添加随机白噪音、亮度、色度和饱和度变化,扩充数据集,同时将数据集随机分为训练集、验证集和测试集,其比例为6:2:2;

步骤1.3:用生成的扩充数据集,利用随机梯度法训练改进的YOLOv3模型,得到基础目标检测模型作为当前模型。

7.根据权利要求1—5中任何一项所述的一种针对视频目标检测的在线学习方法,其特征在于:步骤2所述预检测的步骤包括:

步骤2.1:将视频图像帧一帧一帧地送入当前模型,作为输入,进行前向推理得到模型输出;

步骤2.2:对模型输出进行解析,提取目标方框和目标标注;

步骤2.3:对得到的检测目标进行极大抑制算法NMS,剔除重复目标,得到最终检测目标,作为预检测结果。

8.根据权利要求1—5中任何一项所述的一种针对视频目标检测的在线学习方法,其特征在于:步骤4所述的关键帧提取方法步骤包括:

步骤4.1:将每帧图像通过目标检测网络的特征提取网络的输出作为图像特征提取出来;

步骤4.2:利用欧式距离计算图像特征之间的相似度;

步骤4.3:在时间轴上找出相似度的局部极大值作为视频关键帧提取出来。

9.根据权利要求1—5中任何一项所述的一种针对视频目标检测的在线学习方法,其特征在于:步骤5所述的模型迭代更新步骤包括:

步骤5.1:用提取的视频关键帧和其对于的修正后的目标方框和标注重构数据集,同时,对数据集进行旋转、平移、缩放和镜像变换、添加随机白噪音、亮度、色度和饱和度变化,扩充数据集;

步骤5.2:将新数据集划分为训练集、验证集和测试集,其比例为6:2:2;

步骤5.3:采用新数据集,利用随机梯度法训练改进的YOLOv3模型,得到改进模型,更新当前模型。

一种针对视频目标检测的在线学习方法

技术领域

[0001] 本发明属于深度学习、机器视觉领域,尤其涉及的是一种针对视频目标检测的在线学习方法。

背景技术

[0002] 目标检测即找出图像中所感兴趣的物体,包含物体定位和物体分类两个子任务,是机器视觉领域的基本任务之一,在智能交通、智能制造、安防监控、自动驾驶等领域有着广泛的应用。随着深度学习的发展,目标检测算法也逐步从基于手工特征的传统算法转向基于神经网络的深度学习算法。目前目标检测的研究主要侧重两个方向:基于图像的目标检测和基于视频的目标检测。

[0003] 基于图像的目标检测算法首先被提出,分为one-stage和two-stage两类方法。Two-stage方法沿用传统的目标检测流程,包含候选区域选取、特征提取和分类/回归等三部分。其中比较经典的算法是Region-based CNN(R-CNN)系列网络,经历了由R-CNN到Fast R-CNN到Faster R-CNN的不断改进。One-stage算法简化了two-stage的步骤,将目标检测问题转换为分类和回归问题,引入一个统一的框架,直接将输入图片像素映射为目标方框和目标类别,速度大大提升,比较经典的有Single Shot MultiBox Detector(SSD)和You Only Look Once(YOLO)系列。

[0004] 对于视频数据来说,尽管视频也可以被分解为一帧一帧的图像,借助图像目标检测算法进行检测,但是视频中还包含了时序上下文关系,例如目标在相邻视频帧中位置的连续性等,如果能充分利用这些特性,可以大大提高视频目标检测的精度和速度。这类算法一般是基于循环神经网路,比较经典的有Temporal Convolution Network(TCN)、Spatial-Temporal Memory Network(STMM)、Recurrent YOLO(ROLO)等。

[0005] 但是,不管是基于图像的还是基于视频的算法,以往的方法一般都是采用特定的数据集(开源或者自标注)进行模型训练。基于深度学习的算法存在一个很强的假设:测试数据集分布与训练数据集分布一致。所以,不管是基于图像还是基于视频,深度学习算法存在域适配问题,也就是说,很难通过单一的目标检测模型来实现全域的检测。当场景变化时,为了使检测器达到一定精度,往往需要重新采集数据,人工标注,然后重新训练模型,需要投入大量的人力和时间。

发明内容

[0006] 本发明所要解决的技术问题是针对现有技术的不足,提出一种针对视频目标检测的在线学习方法,该方法通过跟踪算法和聚类分析对检测和标注结果进行修正,然后,利用修正后的结果重新训练目标检测模型,实现模型性能的不断改进,实现应用场景的自适应。

[0007] 本发明为解决其技术问题所采用的技术方案是:提供了一种针对视频目标检测的在线学习方法,包括以下步骤:

[0008] 步骤1:准备基础数据集,该数据集可以是开源数据集或针对某一特定场景采集并

进行人工标注的数据集,训练改进的YOLOv3目标检测网络,获得基础目标检测模型作为当前模型;

[0009] 步骤2:利用当前模型对视频序列进行预检测,获取初始检测方框和目标类别;

[0010] 步骤3:利用跟踪算法和k近邻算法对预检测结果进行方框修正和标注修正;

[0011] 步骤4:提取视频关键帧,以压缩数据集大小,取出重复图像;

[0012] 步骤5:利用关键帧数据和修正检测结果对模型进行训练,实现模型的迭代更新;

[0013] 步骤6:回到步骤2,重复以上操作。

[0014] 步骤1所述的改进的YOLOv3模型的优选技术方案为:将原有YOLOv3的特征提取网络Darknet53替换为OSNet,后续网络与原有网络一致,采用三层金字塔结构,进行不同尺度下的目标检测,特征金字塔中的上采样方法采用逆卷积神经网络实现,计算BBox的损失函数时用GIOU代替原来的MSE,形成改进的YOLOv3模型。

[0015] 本发明所述方法进一步的优选技术方案是:

[0016] 步骤1所述的模型训练的的工作步骤包括:

[0017] 步骤1.1:收集开源数据集,或采集特定场景下的视频数据,人工标注检测目标位置方框和目标类别,建立数据集;

[0018] 步骤1.2:对数据集进行旋转、平移、缩放和镜像变换、添加随机白噪音、亮度、色度和饱和度变化等,扩充数据集,同时将数据集随机分为训练集、验证集和测试集,其比例为6:2:2;

[0019] 步骤1.3:用生成的扩充数据集,利用随机梯度法训练改进的YOLOv3模型,得到基础目标检测模型作为当前模型。

[0020] 本发明所述方法进一步的优选技术方案是:

[0021] 步骤2所述的预检测步骤包括:

[0022] 步骤2.1:将视频图像帧一帧一帧地送入当前模型,作为输入,进行前向推理得到模型输出;

[0023] 步骤2.2:对模型输出进行解析,提取目标方框和目标标注;

[0024] 步骤2.3:对得到的检测目标进行极大抑制算法(NMS),剔除重复目标,得到最终检测目标,作为预检测结果。

[0025] 本发明所述方法进一步的优选技术方案是:

[0026] 步骤3所述的方框修正步骤包括:

[0027] 步骤3.1:对于第n帧图像的所有预检测目标都初始化一个KCF跟踪器,分别进行正向和负向跟踪,得到邻近图像帧中的跟踪目标。

[0028] 步骤3.2:对于临近帧k,计算其与前一帧的帧间差分图像,获取运动目标概率图;

[0029] 步骤3.3:对于k帧中的每个跟踪目标,判断其是否静止,如果该目标处于运动状态,则根据运动目标概率图计算该目标的运动目标概率,如果该值小于给定阈值,则认为该目标跟踪失败,停止跟踪;

[0030] 步骤3.4:对每一帧都进行上述处理,则可以得到所有图像中由跟踪器捕捉到的目标方框;

[0031] 步骤3.5:将检测模型得到的目标方框与跟踪器捕捉到的目标方框进行融合,同时进行极大抑制算法(NMS)得到修正后的目标方框。

[0032] 本发明所述方法进一步的优选技术方案是：

[0033] 步骤3所述的标注修正步骤包括：

[0034] 步骤3.6:获取ImageNet预训练VGG16网络模型；

[0035] 步骤3.7:将检测得到的每个目标方框缩放到统一尺寸(224*224),然后传入VGG16网络,得到1000维的特征向量；

[0036] 步骤3.8:计算不同目标特征之间的欧式距离作为目标相似度度量方式；

[0037] 步骤3.9:对于每个检测目标,通过k近邻算法找出与其最近的k个目标,然后采用投票机制决定该检测目标的目标类别。

[0038] 本发明所述方法进一步的优选技术方案是：

[0039] 步骤4所述的关键帧提取方法步骤包括：

[0040] 步骤4.1:将每帧图像通过目标检测网络的特征提取网络的输出作为图像特征提取出来；

[0041] 步骤4.2:利用欧式距离计算图像特征之间的相似度；

[0042] 步骤4.3:在时间轴上找出相似度的的局部极大值作为视频关键帧提取出来。

[0043] 本发明所述方法进一步的优选技术方案是：

[0044] 步骤5所述的模型迭代更新步骤包括：

[0045] 步骤5.1:用提取的视频关键帧和其对于的修正后的目标方框和标注重构数据集,同时,对数据集进行旋转、平移、缩放和镜像变换、添加随机白噪音、亮度、色度和饱和度变化等,扩充数据集；

[0046] 步骤5.2:将新数据集划分为训练集、验证集和测试集,其比例为6:2:2；

[0047] 步骤5.3:采用新数据集,利用随机梯度法训练改进的YOLOv3模型,得到改进模型,更新当前模型。

[0048] 与现有技术相比,本发明的有益效果在于：

[0049] 1、本发明一种针对视频目标检测的在线学习方法,能够针对同类不同场景(例如交通监控中的不同路口)实现精准的目标检测,而不需要针对不同的场景专门收集数据,进行人工标注。利用开源数据集或者针对特定场景的自标注数据集训练的深度学习目标检测模型,受泛化能力制约,这种模型在新场景下的检测效果会下降,会出现漏检情况,为了提升模型在新场景下的检测精度。本发明方法使得模型具有场景适应能力。首先,利用该基础模型对目标视频序列进行检测,获取初步检测结果,然后利用目标跟踪算法,对检测到的目标进行前向和后向跟踪,获取该目标在临近视频帧中的位置,进而对检测结果进行修正,同时通过提取检测目标的特征,利用k近邻算法对目标标注进行修正,获取更加精准的检测结果。最后,提取视频关键帧,重新训练目标检测模型,该过程不断进行,不断适应场景的变化。

[0050] 2、本发明方法将YOLOv3的特征提取网络用OSNet网络替代,能够在不降低精度的情况下,大大降低网络的参数个数,降低GPU内存消耗,同时提高计算效率;引入边框修正和标注修正,能够使网路不断更新,适应新场景变化,而不降低检测精度;关键帧提取技术的引入能够大大降低视频数据中的冗余信息,减少训练数据集大小,提升模型训练效率。

附图说明

[0051] 图1为一种针对视频目标检测的在线学习方法的流程图；

[0052] 图2为改进YOLOv3的网络结构图；

[0053] 图3为跟踪算法流程图；

[0054] 图4为基于k近邻算法的标注修正示意图；

[0055] 图5为关键帧提取方法流程图。

具体实施方式

[0056] 以下进一步描述本发明的具体技术方案,以便于本领域的技术人员进一步地理解本发明,而不构成对其权利的限制。

[0057] 实施例1,一种针对视频目标检测的在线学习方法,借助方框修正和标注修正不断改进现有模型,实现场景自适应。如图1所示,该方法包括以下步骤:

[0058] 步骤1:准备基础数据集,训练基础网络模型

[0059] 基础数据集可以采用开源数据集,或者针对某一特定场景采集视频数据,人工标注出检测目标位置方框和目标类别,建立数据集,然后,对数据集进行旋转、平移、缩放和镜像变换、添加随机白噪音、亮度、色度和饱和度变化等,扩充数据集,最后,将扩充后的数据集随机分为训练集、验证集和测试集,其比例可根据需求自行决定,一般须满足训练集的数据量大于验证集和测试集,推荐选为6:2:2。

[0060] 目标检测网络采用改进的YOLOv3,其结构如图2所示,由特征提取层和目标检测层组成。其中,特征提取层由全尺度网络(OSNet)构建,目标检测层采用金字塔三层结构实现多尺度检测。

[0061] 特征提取层的前两层分别为卷积层和池化层,其中stride大小均为2,后面紧跟3个BLOCK结构,每个BLOCK结构由两个OSNet的bottleneck结构、一层卷积层和一侧池化层组成,其中池化层的stride为2,卷积层的stride为1。目标检测层中的每层检测网络由卷积序列层(Conv Set)、卷积层和YOLO层组成,其中卷积层的stride均为1。金字塔结构的不同层之间的连接通过一层卷积层和一层反卷积层组成,这里采用反卷积层实现上采样过程。同时,第二层Conv Set的输入与第二层BLOCK的输出进行融合,通过shortcut方式实现,第三层Conv Set的输入与第一层BLOCK的输出进行融合,通过shortcut方式实现。YOLO层anchors的大小由聚类算法给出,得到9个聚类中心,每一层分配3个anchors。

[0062] 模型训练中采用的BBox回归损失函数为GIOU,利用随机梯度法进行模型训练。

[0063] 步骤2:利用当前模型对视频序列进行预检测,获取初始检测方框和目标类别

[0064] 首先,将视频图像帧一帧一帧地送入当前模型,作为输入,进行前向推理得到模型输出;然后,对模型输出进行解析,提取目标方框和目标标注;最后,对得到的检测目标进行极大抑制算法(NMS),剔除重复目标,得到最终检测目标,作为预检测结果。

[0065] 步骤3:方框修正和标注修正

[0066] 该步骤利用跟踪算法和k近邻算法对预检测结果进行方框修正和标注修正。

[0067] 方框修正的具体流程如图3所示。首先,计算帧间差分图,进行阈值化出来,没有变化的像素单元格用0表示,有变化的单元格用1表示,为后续的运动目标概率计算做准备;然后,遍历前一步骤得到的所有检测结果,为每一个检测目标建立KCF跟踪器。同时对目标进

行前向和后向跟踪,直到跟踪失败,停止跟踪。跟踪失败的判定有以下步骤给出,首先判断目标是否处于静止状态,如果目标静止,则认为目标跟踪成功,否则,利用帧间差分图计算运动目标概率,即检测目标范围内的帧间差分图的像素均值,如果阈值大于给定阈值,认为该区域存在运动目标,跟踪成功,否则,认为跟踪失败;最后,更新目标方框。

[0068] 标注修正的过程由图4给出。首先,将所有检测目标的图像区域,缩放到统一固定尺寸(224*224),传入ImageNet预训练的特征提取网络VGG16,得到1000维的特征向量;然后,计算不同目标特征之间的欧式距离作为目标相似度度量,对于每个检测目标,通过k近邻算法找出与其最近的k个目标,采用投票机制决定该目标的目标类别;最后更新所有目标的类别。

[0069] 步骤4:关键帧提取

[0070] 该步骤的具体过程如图5所示。首先,将目标检测网络的特征提取层的输出(即图2中第三个BLOCK的输出)作为图像特征提取出来;然后,计算特征之间的欧式距离来衡量图像之间的相似程度,数值越大,相似度越低;最后,在时间轴上找出相似度的局部极大值(图像差距大)作为视频关键帧提取出来。

[0071] 步骤5:模型迭代更新

[0072] 用提取的视频关键帧和其对应的修正后的目标方框和标注重构数据集,同时,对数据集进行旋转、平移、缩放和镜像变换、添加随机白噪音、亮度、色度和饱和度变化等,扩充数据集。将新数据集划分为训练集、验证集和测试集,其比例可选为6:2:2。采用新数据集,利用随机梯度法训练改进的YOLOv3模型,得到改进模型,更新当前模型。

[0073] 采用上述本发明实施例所提供的在线学习方法能够提高现有目标检测模型的场景适应能力和泛化能力,能够使得利用特定场景训练的目标检测模型迁移到同类型的不同场景中,大大降低了模型对数据的依赖,降低了数据标注所需的人力和时间成本。

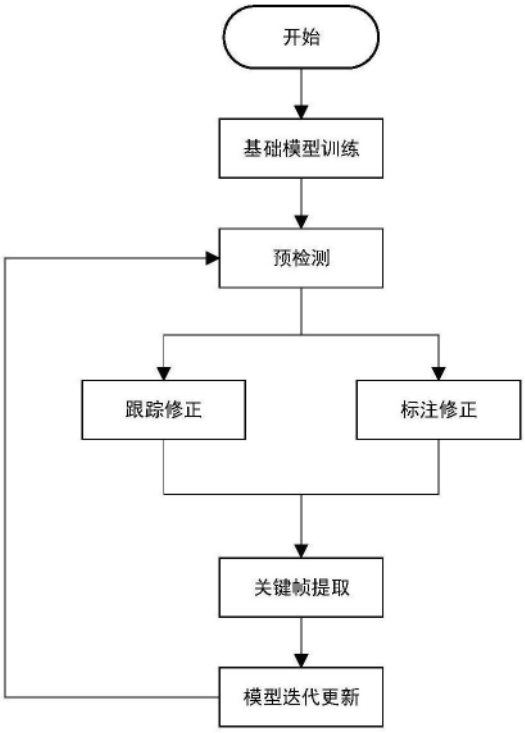

开始
基础模型训练
预检测
跟踪修正
标注修正
关键帧提取
模型迭代更新

图1

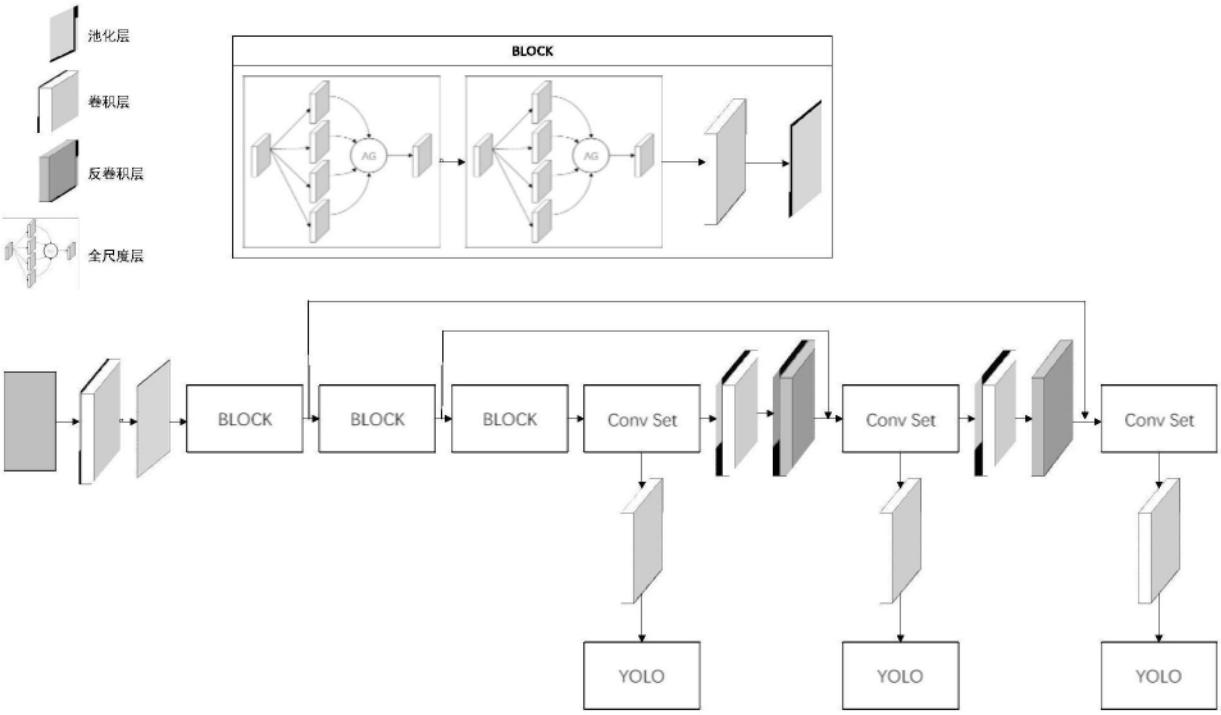

池化层
卷积层
反卷积层
全尺度层
BLOCK
AG
AG
BLOCK
BLOCK
BLOCK
Conv Set
Conv Set
Conv Set
YOLO
YOLO
YOLO

图2

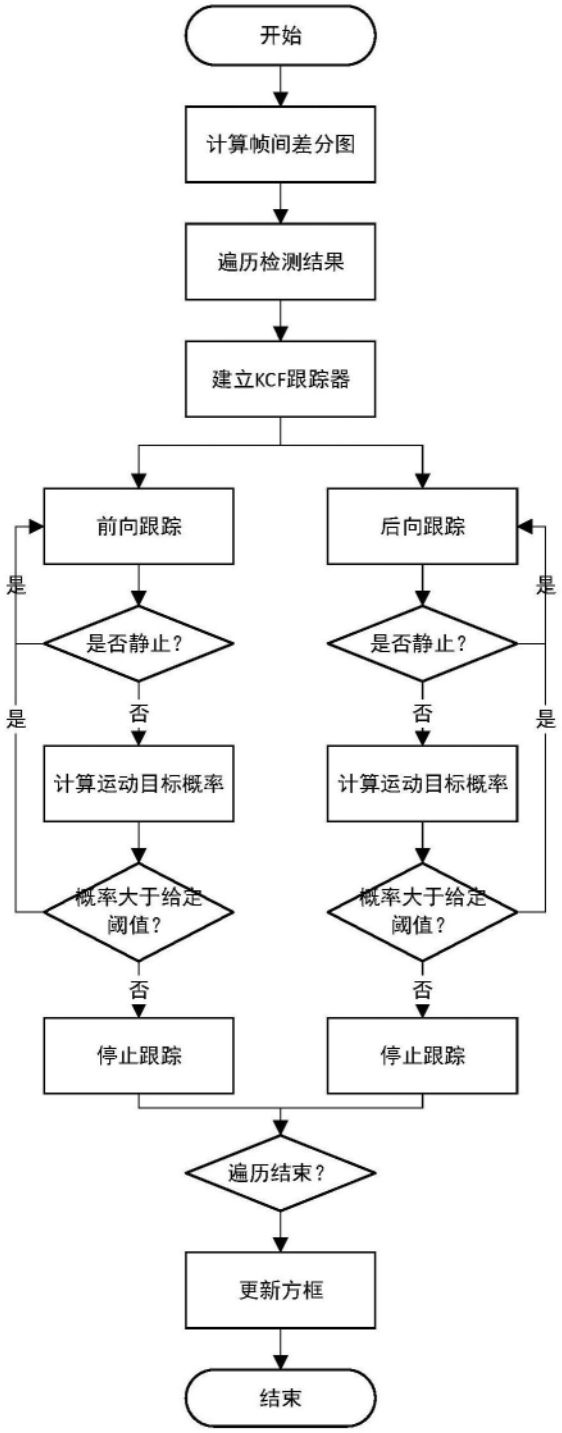

开始
计算帧间差分图
遍历检测结果
建立KCF跟踪器
前向跟踪
后向跟踪
是
是否静止？
是
否
否
是
计算运动目标概率
计算运动目标概率
概率大于给定阈值？
概率大于给定阈值？
否
否
停止跟踪
停止跟踪
遍历结束？
更新方框
结束

图3

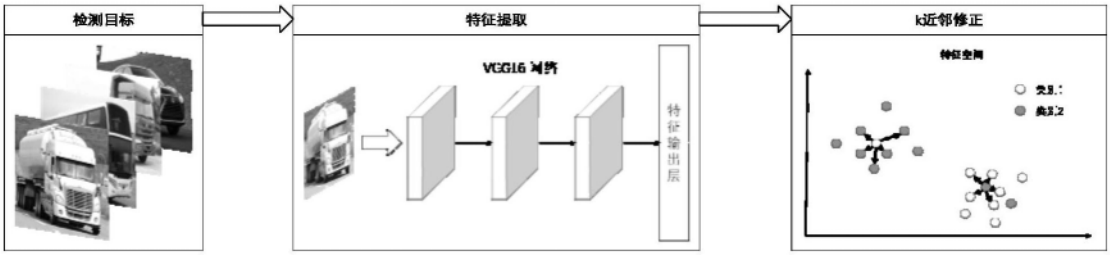

检测目标
特征提取
VGG16 网络
特征输出层
k近邻修正
特征空间

图4


开始
提取目标检测网络的特征网络输出
计算图像特征相似度
找出相似度在时间轴方向的局部极大值点
提取关键帧
结束

图5