



(10) 授权公告号 CN 109359552 B

(45) 授权公告日 2020.11.13

(21) 申请号 201811105659.8

CN 103279791 A, 2013.09.04

(22) 申请日 2018.09.21

WO 2016131300 A1, 2016.08.25

(65) 同一申请的已公布的文献号

CN 106295594 A, 2017.01.04

申请公布号 CN 109359552 A

US 2016328613 A1, 2016.11.10

(43) 申请公布日 2019.02.19

CN 107341517 A, 2017.11.10

(73) 专利权人 中山大学

Jian-Fang Hu等. "Jointly Learning Heterogeneous Features for RGB-D Activity Recognition".《IEEE Transactions on Pattern Analysis and Machine Intelligence》.2017, 第39卷(第11期), 第2186-2200页.

地址 510275 广东省广州市海珠区新港西路135号

郑伟诗等. "非对称行人重识别:跨摄像机持续行人追踪".《中国科学:信息科学》.2018, 第48卷(第5期), 第545-563页. (续)

(72) 发明人 赖剑煌 程海杰 张权

审查员 张博

(74) 专利代理机构 广州市华学知识产权代理有限公司 44245

代理人 刘巧霞

(51) Int.Cl.

G06K 9/00 (2006.01)

(续)

(56) 对比文件

CN 105354548 A, 2016.02.24

CN 105354548 A, 2016.02.24

CN 106530340 A, 2017.03.22

CN 105631413 A, 2016.06.01

CN 106709436 A, 2017.05.24

权利要求书2页 说明书5页 附图2页

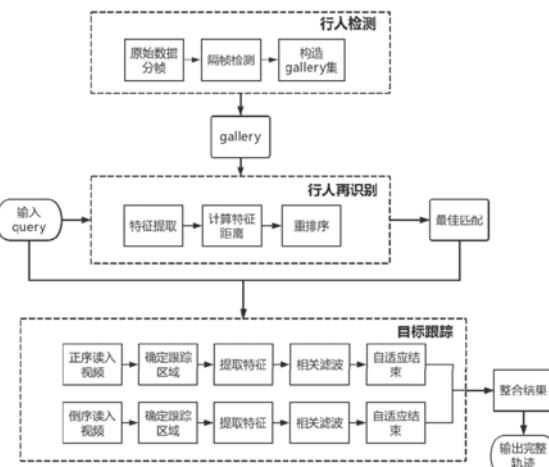
(54) 发明名称

点, 具有很强的工程意义。

一种高效的跨摄像头行人双向跟踪方法

(57) 摘要

本发明公开了一种高效的跨摄像头行人双向跟踪方法, 用于获取行人在已出现过摄像头中的完整轨迹。包括步骤: (1) 从监控视频中获取图像, 对图像中的行人进行检测, 根据检测结果建立候选行人库; (2) 构建行人再识别模型, 提取待查行人和候选行人库中所有图像的特征, 一一计算待查行人特征与候选行人库中所有图像特征的距离特征, 获取待查行人在其他摄像头下的最佳匹配块, 并将其作为跟踪的起始位置; (3) 进行目标跟踪, 跟踪过程中通过正反向处理视频完成对行人的双向跟踪; (4) 将不同摄像头下的轨迹进行整合, 得到行人最终的轨迹输出。本发明对真实场景下的跨境跟踪具有速度快、精度高的优



[转续页]

[接上页]

(51) Int.Cl.

G06K 9/62 (2006.01)

G06T 7/246 (2017.01)

(56) 对比文件

李孟庭.“双向多轨迹判定方法在目标跟踪中的应用研究”.《计算机应用与软件》.2017,第

34卷(第7期),第137-142页.

Gwo Giun Lee等.“Bi-Directional Trajectory Tracking With Variable Block-Size Motion Estimation for Frame Rate Up-Convertor”.《IEEE Journal on Emerging and Selected Topics in Circuits and Systems》.2014,第4卷(第1期),第29-42页.

1. 一种高效的跨摄像头行人双向跟踪方法,其特征在于,包括步骤:

(1) 从监控视频中获取图像,对图像中的行人进行检测,根据检测结果建立候选行人库;

(2) 构建行人再识别模型,提取待查行人和候选行人库中所有图像的特征,一一计算待查行人特征与候选行人库中所有图像特征的距离特征,获取待查行人在其他摄像头下的最佳匹配块,并将其作为跟踪的起始位置;

(3) 进行目标跟踪,跟踪过程中通过正反向处理视频完成对行人的双向跟踪;

(4) 将不同摄像头下的轨迹进行整合,得到行人最终的轨迹输出;

所述步骤(3)中,通过构建目标跟踪模型进行目标跟踪,步骤是:

采用在ImageNet数据集上预训练的VGG模型作为提取行人特征的初始模型,然后利用摄像头视频中的行人和背景图片对模型参数进行精细调整,结合手工特征得到行人的完整表观特征和语义特征,基于上述特征训练得到目标跟踪模型;

所述步骤(3)中,进行目标跟踪的过程是:

对于给定的待跟踪视频帧,首先根据上一帧的跟踪结果确定跟踪区域;

接着对跟踪区域提取深度特征和手工特征,并进行融合;

然后让融合特征经过一个预训练好的相关滤波器计算特征的响应图,取响应最大点作为目标跟踪模型下一帧的预测位置,反复迭代此过程获得行人在该摄像头下的单向轨迹;

所述步骤(3)中,在目标跟踪过程中,根据跟踪的置信分自适应判断行人跟踪停止的位置,步骤是:

将相关滤波得到的响应图的峰值作为第*i*帧跟踪结果的置信分*S<sub>i</sub>*,指定一个跟踪的最大长度*L*,模型生成置信分序列  $\vec{S} = (S_0, S_1, \dots, S_i, \dots, S_{L-1})$ ,并在置信分序列的全局最小值处判定为跟踪停止,其具体的停止帧号可描述为  $\text{index} = \arg \min_i \vec{S}$ 。

2. 根据权利要求1所述的跨摄像头行人双向跟踪方法,其特征在于,所述步骤(1)中,获取监控视频后,对视频进行分帧,采用隔帧检测的方式构建候选行人库。

3. 根据权利要求1所述的跨摄像头行人双向跟踪方法,其特征在于,所述步骤(1)中,对图像中的行人进行检测,步骤是:采用在COCO数据集上预训练好的YOLOv3模型作为行人检测初始模型,接着将YOLOv3模型的最后一个全连接层替换为一个二分类的softmax分类器,然后在公开行人检测数据集上对模型参数进行微调,最终得到行人检测模型;根据该行人检测模型对视频图像中的行人进行检测。

4. 根据权利要求1所述的跨摄像头行人双向跟踪方法,其特征在于,所述步骤(2)中,构建行人再识别模型,步骤是:

使用在ImageNet数据集上预训练的ResNet50分类模型作为主干网络,将主干网络分为两个分支,一个分支提取原网络softmax分类层的输入特征作为全局特征;另一个分支提取分类网络的高层特征并将其水平均匀分为六块,通过池化和降维的操作得到固定大小的六个局部特征;

在训练阶段,每个特征独立经过一个softmax分类器进行训练;

在测试阶段,将全局特征和六个局部特征拼接得到行人的完整特征表达;然后再用公开的行人再识别数据集对模型进行微调,最终得到行人再识别模型。

5.根据权利要求1所述的跨摄像头行人双向跟踪方法,其特征在于,所述步骤(3)中,进行目标跟踪时,通过正反向读入视频分别进行跟踪,获得两条单向轨迹,通过时序拼接得到该行人在此摄像头下的完整轨迹。

## 一种高效的跨摄像头行人双向跟踪方法

### 技术领域

[0001] 本发明涉及计算机视觉领域,更具体地,涉及一种高效的跨摄像头行人双向跟踪方法,用于获取行人在已出现过摄像头中的完整轨迹。

### 背景技术

[0002] 随着科技的进步和社会的发展,城市人口密度越来越高,安全问题也越来越引起人们的重视。为了及时预防和避免安全事件的发生,大量的监控摄像头被安装和应用在公共场所。目前视频监控的布设已十分广泛,如何从多个视频监控的数据中有效获取行人的轨迹信息,对于社会安防体系具有非常重要的价值。因此,跨摄像头行人跟踪已成为计算机视觉领域的一个热点研究内容。

[0003] 跨摄像头行人跟踪是对行人不仅在当前摄像头进行跟踪,还要在其他摄像头找到此人并继续跟踪,直到其完全消失在所有摄像头为止。可以看出跨摄像头行人跟踪主要包括三个子任务,即行人检测、行人再识别和目标跟踪。因此,在监控系统中对行人进行跨摄像头跟踪时,一般有以下几个问题需要解决:首先需要在监控场景的图像中,用行人检测技术,把所有出现的行人检测出来;获得行人的位置后,当行人在同一个摄像头区域中移动时,需要用行人跟踪的技术不断更新行人当前的位置;当行人离开了一个摄像头区域,进入另外一个无交叠的摄像头区域时,需要将不同摄像头中的行人用行人再识别匹配起来。随着计算机视觉和深度学习的快速发展,行人检测、行人再识别和目标跟踪已作为热点课题被研究多年,也取得了很大进展,这也为跨摄像头行人跟踪提供了坚强的理论和技术支撑。

[0004] 目前,跨摄像头行人跟踪普遍是基于轨迹匹配的跨摄像头行人跟踪方法。其主要思想是:首先对于指定的待查行人,使用单目标跟踪方法获取待查行人在当前摄像头下的完整轨迹,接着使用多目标跟踪算法对各个摄像头下的所有行人进行跟踪获取其在单个摄像头下的轨迹;最后使用行人再识别算法对第一步中行人的跟踪轨迹与第二步中其他摄像头下的所有行人轨迹进行匹配获取该行人在其他摄像头下的跟踪轨迹。然而,这种方法由于是对所有摄像头下所有行人进行跟踪,这本身在时间上就需要极大的代价,且精确度也不高;同时,在做行人再识别时,由于是对行人的轨迹进行特征匹配,会额外引入一个时序的模块和一些时空约束,这样既增加了模型的复杂度,也增加了行人匹配的处理时间。

### 发明内容

[0005] 为了克服传统基于轨迹匹配的跨摄像头行人跟踪方法耗时长、精度不高的缺点,本发明提出了一种高效的跨摄像头行人双向跟踪方法,该方法可快速精确获取行人在已出现过摄像头中的完整轨迹。

[0006] 本发明的目的通过以下的技术方案实现:一种高效的跨摄像头行人双向跟踪方法,包括步骤:

[0007] (1)从监控视频中获取图像,对图像中的行人进行检测,根据检测结果建立候选行人库;

[0008] (2) 构建行人再识别模型,提取待查行人和候选行人库中所有图像的特征,一一计算待查行人特征与候选行人库中所有图像特征的距离特征,获取待查行人在其他摄像头下的最佳匹配块,并将其作为跟踪的起始位置;

[0009] (3) 进行目标跟踪,跟踪过程中通过正反向处理视频完成对行人的双向跟踪;

[0010] (4) 将不同摄像头下的轨迹进行整合,得到行人最终的轨迹输出。

[0011] 本发明通过上述行人检测、行人再识别、目标跟踪步骤,只要选择行人在任意摄像头下出现过的一个位置,就可以输出该行人在所有摄像头下出现过的完整轨迹,其具有准确度高、速度快的优点。

[0012] 优选的,所述步骤(1)中,获取监控视频后,对视频进行分帧,采用隔帧检测的方式构建候选行人库。隔帧检测即指定间隔帧数,仅对帧号F满足 $F \bmod K = 0$ 的视频帧进行行人检测,从而可提高检测速度。

[0013] 优选的,所述步骤(1)中,对图像中的行人进行检测,步骤是:采用在COCO数据集上预训练好的YOLOv3模型作为行人检测初始模型,接着将YOLOv3模型的最后一个全连接层替换为一个二分类的softmax分类器,然后在公开行人检测数据集上对模型参数进行微调,最终得到行人检测模型;根据该行人检测模型对视频图像中的行人进行检测。

[0014] 优选的,所述步骤(2)中,构建行人再识别模型,步骤是:

[0015] 使用在ImageNet数据集上预训练的ResNet50分类模型作为主干网络,将主干网络分为两个分支,一个分支提取原网络softmax分类层的输入特征作为全局特征;另一个分支提取分类网络的高层特征并将其水平均匀分为六块,通过池化和降维的操作得到固定大小的六个局部特征;

[0016] 在训练阶段,每个特征独立经过一个softmax分类器进行训练;

[0017] 在测试阶段,将全局特征和六个局部特征拼接得到行人的完整特征表达;然后再用公开的行人再识别数据集对模型进行微调,最终得到行人再识别模型。

[0018] 优选的,所述步骤(3)中,通过构建目标跟踪模型进行目标跟踪,步骤是:

[0019] 采用在ImageNet数据集上预训练的VGG模型作为提取行人特征的初始模型,然后利用摄像头视频中的行人和背景图片对模型参数进行精细调整,结合传统的手工特征(如HOG特征、SIFT特征)得到行人的完整表观特征和语义特征,基于上述特征训练得到目标跟踪模型。

[0020] 更进一步的,所述步骤(3)中,进行目标跟踪的过程是:

[0021] 对于给定的待跟踪视频帧,首先根据上一帧的跟踪结果确定跟踪区域;

[0022] 接着对跟踪区域提取深度特征和手工特征,并进行融合;

[0023] 然后让融合特征经过一个预训练好的相关滤波器计算特征的响应图,取响应最大点作为目标跟踪模型下一帧的预测位置,反复迭代此过程获得行人在该摄像头下的单向轨迹。

[0024] 更进一步的,所述步骤(3)中,在目标跟踪过程中,根据跟踪的置信分自适应判断行人跟踪停止的位置,步骤是:

[0025] 将相关滤波得到的响应图的峰值作为第i帧跟踪结果的置信分 $S_i$ ,指定一个跟踪的最大长度L,模型生成置信分序列 $\vec{S} = (S_0, S_1, \dots, S_i, \dots, S_{L-1})$ ,并在置信分的全

局最小值处判定为跟踪停止,其具体的停止帧号可描述为  $\text{index} = \arg \min_i \hat{S}$ 。

[0026] 更进一步的,所述步骤(3)中,进行目标跟踪时,通过正反向读入视频分别进行跟踪,获得两条单向轨迹,通过时序拼接得到该行人在此摄像头下的完整轨迹。

[0027] 与现有技术相比,本发明技术方案的有益效果是:

[0028] 1、由于监控摄像头具有分辨率低、数据量大、场景复杂的特点,跨摄像头行人跟踪的瓶颈在于处理时间长、跟踪效果差。本发明中在进行目标跟踪时,提出对行人进行双向跟踪,这样在进行行人再识别时获取的最佳匹配位置无论是位于行人刚出现,还是行人即将消失,只要确保身份一致,本发明都能得到该行人在摄像头下的完整轨迹,不需要对摄像头视频中的所有行人进行检测和跟踪,极大地提升了跟踪的速度和精度。

[0029] 2、本发明提出采取隔帧检测去构建候选用人库,可大大地提升检测的速度。

[0030] 3、本发明目标跟踪过程中根据跟踪置信分自适应判断跟踪停止,有效地解决了人为设定跟踪结束条件(如设定阈值帧数和边界条件)在时间效率和精确度上的缺点。

[0031] 4、本发明行人再识别模型为基于单帧图像的匹配模型,与传统基于轨迹特征的匹配模型相比,本发明模型更为简洁,并且候选用人库的容量也较为合理,故本发明的时间效率会大幅提高;除此之外,本发明在模型中使用了全局特征和多个局部特征相结合,相较于单一特征,本发明得到了行人更完整的特征表达,故在精度上也有提高。

## 附图说明

[0032] 图1是本实施例方法的总体流程图。

[0033] 图2是本实施例方法中行人再识别步骤的网络流程图。

[0034] 图3是本实施例方法中目标跟踪步骤的流程图。

## 具体实施方式

[0035] 附图仅用于示例性说明,不能理解为对本专利的限制;对于本领域技术人员来说,附图中某些公知结构及其说明可能省略是可以理解的。下面结合附图和实施例对本发明的技术方案做进一步的说明。

[0036] 如图1所示,本实施例一种高效的跨摄像头行人双向跟踪方法,主要包括行人检测、行人再识别、目标跟踪三个部分,下面对各个部分进行具体说明。

### 一、行人检测

[0038] 要进行行人检测,首先建立行人检测模型,该模型的建立步骤是:采用在COCO数据集上预训练好的YOLOv3模型作为行人检测初始模型,接着将YOLOv3模型的最后一个全连接层替换为一个二分类的softmax分类器,然后在多个公开行人检测数据集包括Caltech、Cityperson等上对模型参数进行精细调节,将此模型作为行人检测模型。

[0039] 在检测过程中对监控视频进行分帧,然后进行隔帧检测,即指定间隔帧数K(K=10),对于帧号F满足  $F \bmod K = 0$  的视频帧代入上述的行人检测模型进行行人检测,并将检测结果保存,作为行人再识别步骤中的候选用人库。

[0040] 本发明中采用隔帧检测的方法构建候选用人库,因此大大地提升了检测的速度,理论上候选用人库中只需要有每个行人在摄像头下的一张图片即可,但考虑到某些摄像头下某段时间内可能有大量行人同时出现,极易出现行人严重遮挡的情况,为了给目标跟踪

提供一个良好的初始跟踪位置,本发明主要是每隔K(K=10)帧对摄像头视频中行人进行检测一次,这意味着在行人检测步骤中构造得到的候选行人库容量大幅度减少,并且构造候选人库的时间仅仅需要传统基于轨迹匹配的跨摄像头行人跟踪方法在行人检测用时的K分之一,大幅度提高了检测的速度。

[0041] 二、行人再识别

[0042] 要进行行人再识别,首先建立行人再识别模型,参见图2,该模型的建立步骤是:

[0043] 首先,使用在ImageNet数据集上预训练的ResNet50分类模型作为主干网络,将主干网络分为两个分支,一个分支提取原网络softmax分类层的输入特征作为全局特征;另一个分支提取分类网络的高层特征并将其水平均匀分为六块,通过池化和降维的操作得到固定大小的六个局部特征。在训练阶段,每个特征独立经过一个softmax分类器进行训练;在测试阶段,将全局特征和六个局部特征拼接得到行人的完整特征表达。然后再用多个公开的行人再识别数据集包括MSMT17、DukeMTMC-reID、Market1501等对模型进行精细调整,调整后的最终模型作为行人再识别模型。

[0044] 在行人再识别过程中,利用上述模型同时提取指定待查行人和候选行人库中所有图像的特征,一一计算待查行人特征和候选行人库中所有图片的特征的余弦距离,经过一个重排序的操作,获取行人在其他摄像头下的最佳匹配块,并将其作为跟踪的起始位置用于下面的目标跟踪。

[0045] 本实施例行人再识别模型为基于单帧图像的匹配模型,模型简洁,时间效率高。另外,本发明在模型中使用了全局特征和多个局部特征相结合,相较于单一特征,本发明得到了行人更完整的特征表达,因而精度更高。

[0046] 三、目标跟踪

[0047] 要进行目标跟踪,首先建立目标跟踪模型,该模型的建立步骤是:

[0048] 采用在ImageNet数据集上预训练的VGG模型作为提取行人特征的初始模型,然后利用摄像头视频中的行人和背景图片对模型参数进行精细调整,增强模型对行人特征的响应,抑制背景噪声的干扰,同时结合传统手工特征,如HOG特征、灰度特征、CN特征等辅助深度特征,得到行人的完整表观特征和语义特征,依旧上述特征构建目标跟踪模型。

[0049] 本实施例跟踪过程中通过正反向处理视频完成对行人的双向跟踪,即通过正反向读入视频分别进行跟踪,获得两条单向轨迹,通过时序拼接得到该行人在此摄像头下的完整轨迹。参见图3,以单向跟踪为例,说明其具体的流程:

[0050] 对于给定的待跟踪视频帧,首先根据上一帧的跟踪结果确定跟踪区域;接着对跟踪区域提取深度特征和手工特征并进行融合;然后让融合特征经过一个预训练好的相关滤波器计算特征的响应图,取响应最大点作为目标跟踪模型下一帧的预测位置,反复迭代此过程可获得行人在该摄像头下的单向轨迹。

[0051] 另外,本实施例中提出在目标跟踪过程中,根据跟踪的置信分自适应判断行人跟踪停止的位置,步骤是:

[0052] 将相关滤波得到的响应图的峰值作为第i帧跟踪结果的置信分 $S_i$ ,指定一个跟踪的最大长度L,模型生成置信分序列 $\vec{S} = (S_0, S_1, \dots, S_i, \dots, S_{L-1})$ ,并在置信分的全局最小值处判定为跟踪停止,其具体的停止帧号可描述为 $\text{index} = \arg \min_i \vec{S}$ 。

[0053] 通过上述3个步骤,只要将行人在任意摄像头下出现过的一个位置作为初始值输入到行人再识别模型,就可以通过目标跟踪模型输出该行人在所有摄像头下出现过的完整轨迹。通过实验表明:在给定三个摄像头视频,每个视频20分钟,帧速20FPS,分辨率320\*240的条件下,本发明在检测阶段其检测速度可达45FPS,准确度为90%;在行人再识别阶段首位命中率为92%;在跟踪阶段跟踪精度为95%。在测试阶段系统的跟踪速度可达18FPS。总体而言,利用本发明去对行人进行跨摄像头跟踪不仅会大大缩短跟踪时间和减少成本,在精确度上也更加准确,具有很强的工程意义。

[0054] 显然,可通过各种手段实施本发明描述的技术。举例来说,这些技术可实施在硬件、固件、软件或其组合中。对于硬件实施方案,处理模块可实施在一个或一个以上专用集成电路(ASIC)、数字信号处理器(DSP)、可编程逻辑装置(PLD)、现场可编辑逻辑门阵列(FPGA)、处理器、控制器、微控制器、电子装置、其他经设计以执行本发明所描述的功能的电子单元或其组合内。

[0055] 对于固件和/或软件实施方案,可用执行本文描述的功能的模块(例如,过程、步骤、流程等)来实施所述技术。固件和/或软件代码可存储在存储器中并由处理器执行。存储器可实施在处理器内或处理器外部。

[0056] 本领域普通技术人员可以理解:实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成,前述的程序可以存储在一计算机可读取存储介质中,该程序在执行时,执行包括上述方法实施例的步骤;而前述的存储介质包括:ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0057] 显然,本发明的上述实施例仅仅是为清楚地说明本发明所作的举例,而并非是对本发明的实施方式的限定。对于所属领域的普通技术人员来说,在上述说明的基础上还可以做出其它不同形式的变化或变动。这里无需也无法对所有的实施方式予以穷举。凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明权利要求的保护范围之内。

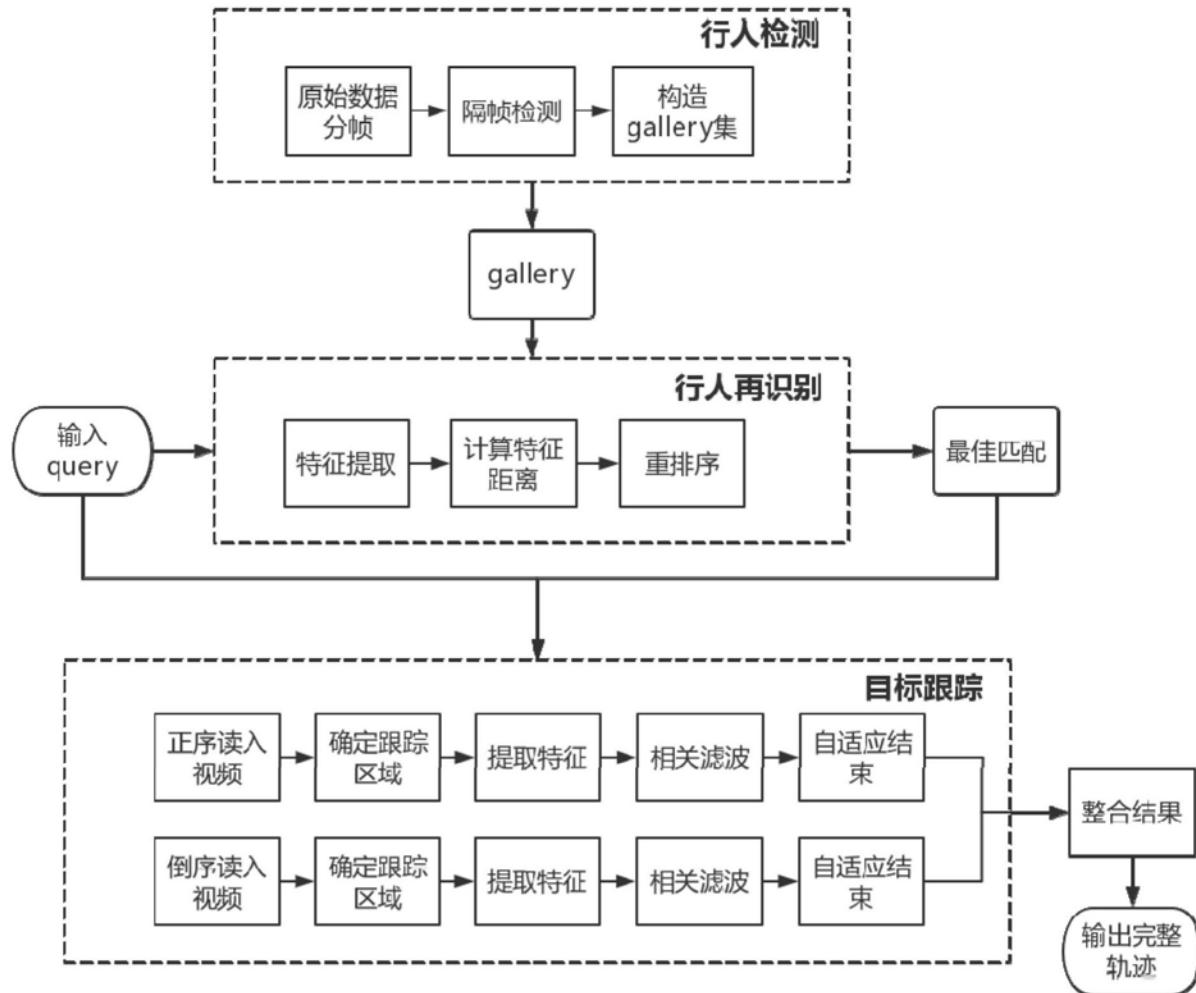


图1

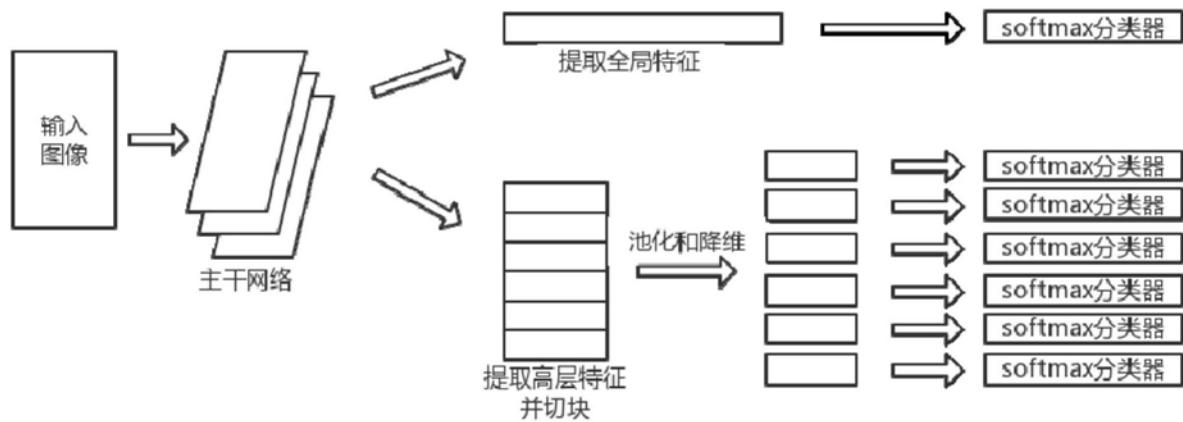


图2

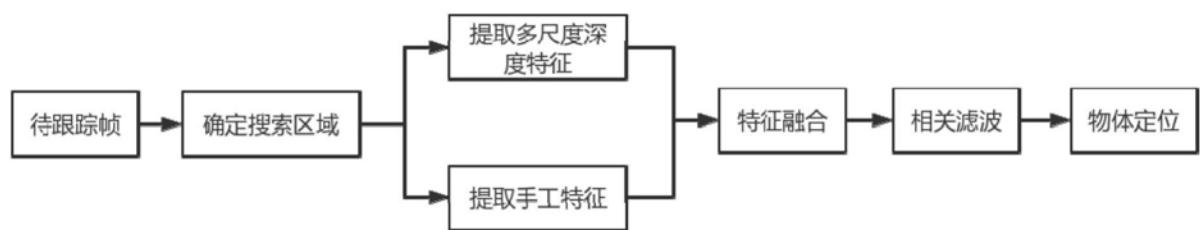


图3