(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 110651310 A
(43)申请公布日 2020.01.03

(21)申请号 201880033667.8

(22)申请日 2018.04.05

(30)优先权数据
62/601,953 2017.04.05 US

(85)PCT国际申请进入国家阶段日
2019.11.21

(86)PCT国际申请的申请数据
PCT/US2018/026341 2018.04.05

(87)PCT国际申请的公布数据
WO2018/187632 EN 2018.10.11

(71)申请人 卡内基梅隆大学
地址 美国宾夕法尼亚州
申请人 里斯本高等理工学院

(72)发明人 何塞·M·F·莫拉
若昂·保罗·科斯泰拉 张商行
叶夫根尼·托罗波夫

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262
代理人 陆建萍 杨明钊

(51)Int.Cl.
G08G 1/01(2006.01)
G06N 3/08(2006.01)
G06K 9/46(2006.01)

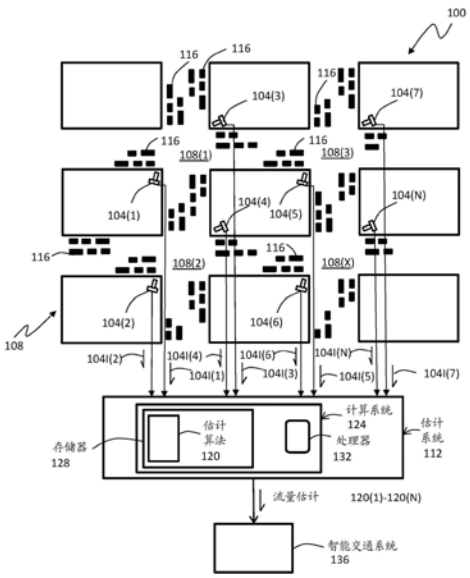
权利要求书6页 说明书23页 附图11页

(54)发明名称

估计对象密度和/或流量的深度学习方法及相关方法和软件

(57)摘要

利用人工神经网络(ANN)估计一个或更多个场景中对象的密度和/或流量(速度)的方法和软件,每个场景在一个或更多个图像中捕获。在一些实施例中,ANN及其训练被配置成提供可靠的估计,尽管存在一个或更多个挑战,这些挑战包括但不限于低分辨率图像、低帧率图像采集、高对象遮挡率、大摄像机视角、大范围变化的照明条件和大范围变化的天气条件。在一些实施例中,在ANN中使用完全卷积网络(FCN)。在一些实施例中,长短期记忆网络(LSTM)与FCN一起使用。在这样的实施例中,LSTM可以以残差学习方式或直接连接方式连接到FCN。还公开了生成用于训练基于ANN的估计算法的训练图像的方法,该方法使得估计算法的训练成本更低。



CN 110651310 A

1.一种向智能交通系统提供交通密度和/或交通流量数据的方法,其中,所述交通密度和/或交通流量数据用于道路网络的多个区域,每个区域具有与其相关联的交通摄像机,所述交通摄像机捕获相应区域的至少一个相应交通图像,所述方法包括:

使用所述至少一个相应交通图像,在每个交通摄像机本地自动估计所述道路网络的相应区域中的交通密度和/或交通流量,对每个相应区域进行估计包括:

使用训练标注图像来训练基于人工神经网络(ANN)的密度估计算法,其中,所述基于ANN的密度估计算法被配置成在训练时处理一个或更多个输入图像,以自动确定所述输入图像中存在的交通的交通密度和/或交通流量;

接收所述至少一个相应交通图像;以及

在所述基于ANN的密度估计算法的训练之后,使所述基于ANN的密度估计算法处理所述至少一个相应交通图像,以确定对于所述相应区域的交通密度和/或交通流量;以及

从每个交通摄像机本地向所述智能交通系统传输与所述道路网络的相应区域中的所述交通密度和/或交通流量相对应的交通密度和/或交通流量数据。

2.根据权利要求1所述的方法,还包括,连同所述交通密度和/或交通流量的传输,传输对于所述道路网络的相应区域的位置信息。

3.根据权利要求1所述的方法,其中,每个交通摄像机具有摄像机视角,所述方法还包括:接收所述训练图像,其中所述训练图像具有与所述交通摄像机的摄像机视角基本相同的训练视角,并且所述训练图像中每一个是包含多个车辆像素级掩模的交通图像,其中所述车辆像素级掩模已经使用分割器被放置在来自所述交通摄像机的交通图像中,所述分割器在已用边界框标注的、来自所述交通摄像机的所述交通图像上操作。

4.根据权利要求3所述的方法,其中,标注的真实图像已由检测器自动标注。

5.根据权利要求3所述的方法,还包括使用从不同类型的不同车辆的3D CAD模型再现的图像来训练所述分割器。

6.根据权利要求1所述的方法,其中,每个交通摄像机具有摄像机视角,所述方法还包括:

接收训练图像,其中,所述训练图像具有与所述交通摄像机的摄像机视角基本相同的训练视角,所述训练图像中每一个是混合图像,所述混合图像具有所述道路的区域的实际背景,并且包含覆盖在所述背景上的不同类型的多个再现车辆,其中所述再现车辆:

基于不同车辆类型的不同车辆的3D CAD模型;

使用基于所述摄像机视角的实际视角和实际遮挡被放置在所述混合图像中;并且

被自动标记用于基于所述3D CAD模型进行归类。

7.根据权利要求1所述的方法,其中,所述基于ANN的估计算法被配置成从交通图像的时序集合中确定交通流量,并且所述方法还包括向所述智能交通系统提供交通流量数据,其中提供所述交通流量数据包括:

接收所述交通图像的时序集合;

使所述基于ANN的密度估计算法处理所述交通图像的集合,以确定对于所述交通图像的集合的交通流量;以及

向所述智能交通系统传输对应于所述交通流量的交通流量数据。

8.根据权利要求1所述的方法,其中,所述基于ANN的估计算法包括完全卷积网络

(FCN)。

9.根据权利要求7所述的方法,其中,所述基于ANN的估计算法包括长短期记忆网络(LSTM)。

10.根据权利要求9所述的方法,其中,所述LSTM直接连接到所述FCN。

11.根据权利要求9所述的方法,其中,所述LSTM以残差学习的方式连接到所述FCN。

12.根据权利要求11所述的方法,其中,每个交通摄像机具有摄像机视角,所述方法还包括:接收所述训练图像,其中所述训练图像具有与所述交通摄像机的摄像机视角基本相同的训练视角,并且所述训练图像中的每一个是包含多个车辆像素级掩模的交通图像,其中所述车辆像素级掩模使用分割器被放置在来自所述交通摄像机的交通图像中,所述分割器在用边界框标注的、来自所述交通摄像机的所述交通图像上操作。

13.根据权利要求12所述的方法,其中,标注的真实图像已由检测器自动标注。

14.根据权利要求12所述的方法,还包括使用从不同类型的不同车辆的3D CAD模型再现的图像来训练所述分割器。

15.根据权利要求11所述的方法,其中,每个交通摄像机具有摄像机视角,所述方法还包括:

接收训练图像,其中,所述训练图像具有与所述交通摄像机的摄像机视角基本相同的训练视角,所述训练图像中的每一个是混合图像,所述混合图像具有所述道路的区域的实际背景,并且包含覆盖在所述背景上的不同类型的多个再现车辆,其中所述再现车辆:

基于不同车辆类型的不同车辆的3D CAD模型;

使用基于所述摄像机视角的实际视角和实际遮挡被放置在所述混合图像中。

16.根据权利要求7所述的方法,其中,所述基于ANN的算法还包括连接到所述FCN的多任务网络。

17.一种生成用于基于人工神经网络(ANN)的密度估计算法的训练图像集合的方法,其中,所述基于ANN的密度估计算法用于检测由固定摄像机捕获的场景内的一种或更多种目标类型的对象的密度,所述方法包括:

接收由所述固定摄像机捕获的真实图像的第一集合,其中,每个真实图像包括具有所述目标类型内的类型的至少一个真实对象;

使用基于ANN的检测器处理每个真实图像,以便用相应的各自的第一边界框区域标注每个真实图像内所述一种或更多种目标类型的对象的出现,其中所述处理产生标注的真实图像的集合;以及

使用基于ANN的分割器处理每个标注的真实图像,以便生成相应的像素级背景掩模,所述背景掩模进一步用于生成用于训练的密度图。

18.根据权利要求17所述的方法,其中,生成训练图像的集合还包括将所述场景的在所述边界框区域之外的区域分割为背景。

19.根据权利要求17所述的方法,还包括在处理所述第一集合中的每个真实图像之前训练所述基于ANN的检测器,其中所述基于ANN的检测器的训练包括:

接收由所述固定摄像机捕获的真实图像的第二集合,其中所述第二集合中的每个真实图像包括具有所述目标类型内的类型的至少一个真实对象;

接收对应于真实图像的所述第二集合的手动标记的真实图像的集合,其中,每个手动

标记的真实图像包括所述至少一个真实对象,并且还包括用于所述至少一个真实对象中的每一个的手动插入的边界框;以及

使用真实图像的所述第二集合和相应的手动标记的真实图像的集合来训练所述基于ANN的检测器。

20.根据权利要求19所述的方法,其中,每个交通摄像机具有摄像机视角,所述方法还包括:接收所述训练图像,其中所述训练图像具有与所述交通摄像机的摄像机视角基本相同的训练视角,并且所述训练图像中的每一个是包含多个车辆像素级掩模的交通图像,其中所述车辆像素级掩模使用分割器被放置在来自所述交通摄像机的交通图像中,所述分割器在用边界框标注的、来自所述交通摄像机的交通图像上操作。

21.一种估计由摄像机捕获的场景的图像中的不同目标类型的对象的密度的方法,所述方法包括:

接收训练数据集,其中所述训练数据集包括多个合成图像,所述多个合成图像具有与所述摄像机的摄像机视角基本相同的训练视角,所述合成图像中的每一个包含不同类型的多个再现对象,其中所述再现对象:

基于不同类型的不同对象的3D CAD模型;

使用基于所述摄像机视角的实际视角和实际遮挡被放置在所述合成图像中;并且

被自动标记用于基于所述3D CAD模型进行归类;使用所述合成图像来训练基于人工神经网络(ANN)的密度估计算法,其中所述基于ANN的密度估计算法被配置成在训练时处理输入图像以自动估计存在于所述输入图像中的不同目标类型的对象的密度;接收所述图像;以及

在训练所述基于ANN的密度估计算法之后,使用所述基于ANN的密度估计算法处理所述图像中每一个以估计对于所述图像的密度。

22.一种向智能交通系统提供交通密度和/或交通流量数据的方法,其中,所述交通密度和/或交通流量数据用于道路网络的多个区域,每个区域具有与其相关联的交通摄像机,所述交通摄像机捕获相应区域的至少一个相应交通图像,所述方法包括:

在集中式估计系统处,从每个交通摄像机接收对应于所述至少一个相应交通图像中的每一个的相应的低分辨率图像;

在所述集中式估计系统处,使用所述相应的低分辨率图像自动估计所述道路网络的相应区域中的交通密度和/或交通流量,在每个相应区域中进行所述估计包括:

使用训练标注图像来训练基于人工神经网络(ANN)的密度估计算法,其中所述基于ANN的密度估计算法被配置成在训练时处理一个或更多个输入图像,以自动确定所述输入图像中存在的交通的交通密度和/或交通流量;以及

在所述基于ANN的密度估计算法的训练之后,使所述基于ANN的密度估计算法处理对于所述区域中的每一个区域的相应的低分辨率图像,以确定对于所述相应区域的交通密度和/或交通流量;以及从所述集中式估计系统向所述智能交通系统传输与所述道路网络的区域的交通密度和/或交通流量相对应的交通密度和/或交通流量数据。

23.根据权利要求22所述的方法,其中,接收对应于所述至少一个相应交通图像中的每一个的相应的低分辨率图像包括接收视频。

24.根据权利要求23所述的方法,其中,所述视频具有小于1帧每秒的帧率。

25.根据权利要求22所述的方法，还包括，连同所述交通密度和/或交通流量的传输，传输对于所述道路网络的相应区域的位置信息。

26.根据权利要求22所述的方法，其中，每个交通摄像机具有摄像机视角，所述方法还包括：接收训练图像，其中所述训练图像具有与所述交通摄像机的摄像机视角基本相同的训练视角，并且所述训练图像中的每一个是包含多个车辆像素级掩模的交通图像，其中所述车辆像素级掩模使用分割器被放置在来自所述交通摄像机的交通图像中，所述分割器在用边界框标注的、来自所述交通摄像机的交通图像上操作。

27.根据权利要求26所述的方法，其中，标注的真实图像已由检测器自动标注。

28.根据权利要求22所述的方法，其中，每个交通摄像机具有摄像机视角，所述方法还包括：

接收所述训练图像，其中所述训练图像具有与所述交通摄像机的摄像机视角基本相同的训练视角，所述训练图像中每一个是混合图像，所述混合图像具有所述道路的区域的实际背景，并且包含覆盖在所述背景上的不同类型的多个再现车辆，其中所述再现车辆：

基于不同车辆类型的不同车辆的3D CAD模型；

使用基于所述摄像机视角的实际视角和实际遮挡被放置在所述混合图像中；并且

被自动标记用于基于所述3D CAD模型进行归类。

29.根据权利要求22所述的方法，其中，所述基于ANN的估计算法被配置成从交通图像的时序集合中确定交通流量，并且所述方法还包括向所述智能交通系统提供交通流量数据，其中提供所述交通流量数据包括：

接收交通图像的时序集合；

使所述基于ANN的密度估计算法处理所述交通图像的集合，以确定对于所述交通图像的集合的交通流量；以及

向所述智能交通系统传输对应于所述交通流量的交通流量数据。

30.根据权利要求22所述的方法，其中，所述基于ANN的估计算法包括完全卷积网络(FCN)。

31.根据权利要求30所述的方法，其中，所述基于ANN的估计算法包括长短期记忆网络(LSTM)。

32.根据权利要求31所述的方法，其中，所述LSTM直接连接到所述FCN。

33.根据权利要求31所述的方法，其中，所述LSTM以残差学习的方式连接到所述FCN。

34.根据权利要求33所述的方法，其中，每个交通摄像机具有摄像机视角，所述方法还包括：接收所述训练图像，其中所述训练图像具有与所述交通摄像机的摄像机视角基本相同的训练视角，并且所述训练图像中每一个是包含多个车辆像素级掩模的交通图像，其中所述车辆像素级掩模使用分割器被放置在来自所述交通摄像机的交通图像中，所述分割器在用边界框标注的、来自所述交通摄像机的交通图像上操作。

35.根据权利要求34所述的方法，其中，标注的真实图像已由检测器自动标注。

36.根据权利要求33所述的方法，其中，每个交通摄像机具有摄像机视角，所述方法还包括：

接收所述训练图像，其中所述训练图像具有与所述交通摄像机的摄像机视角基本相同的训练视角，所述训练图像中每一个是混合图像，所述混合图像具有所述道路的区域的实

际背景,并且包含覆盖在所述背景上的不同类型的多个再现车辆,其中所述再现车辆:

基于不同车辆类型的不同车辆的3D CAD模型;

使用基于所述摄像机视角的实际视角和实际遮挡被放置在所述混合图像中;并且

被自动标记用于基于所述3D CAD模型进行归类。

37.根据权利要求30所述的方法,其中,所述基于ANN的算法还包括连接到所述FCN的多任务网络。

38.一种从摄像机获取的图像确定对象计数的方法,所述方法在机器中执行,并且包括:

使用完全卷积网络(FCN)联合估计所述对象计数和对象密度,其中联合估计包括使用所述FCN将所述图像内的密集特征映射到所述对象密度;

使用残差学习框架,参照所述图像中的所述对象密度之和来学习残差函数;

组合来自所述FCN的较浅层的外观特征和来自所述FCN的较深层的语义特征,以产生更密集的特征图;以及

使用所述更密集的特征图确定所述对象计数。

39.一种从摄像机获取的视频的帧确定对象计数的方法,所述方法在机器中执行,并且包括:

使用完全卷积网络(FCN)联合估计所述对象计数和对象密度,其中联合估计包括使用所述FCN创建对于所述视频的帧的对象密度图;

向长短期记忆(LSTM)网络提供所述对象密度图,其中所述LSTM网络被配置成参照所述帧的每一个中的密度之和来学习残差函数;以及

将所述残差函数和所述密度求和,以确定所述对象计数。

40.根据权利要求38所述的方法,还包括进行超空洞组合,以对所述FCN中的空洞卷积进行积分,并组合所述FCN内不同卷积层的特征图。

41.一种从摄像机获取的视频的帧确定对象计数的方法,所述方法在机器中执行,并且包括:

使用完全卷积网络(FCN)联合估计所述对象计数和对象密度,其中联合估计包括使用所述FCN创建对于所述视频的帧的对象密度图;

向长短期记忆(LSTM)网络提供所述对象密度图,其中所述LSTM网络被配置成参照所述帧的每一个中的密度之和来学习残差函数;以及

将所述残差函数和所述密度求和,以确定所述对象计数。

42.一种从来自多个摄像机的图像确定对象计数的方法,所述方法包括:

从所述多个摄像机内的多个源摄像机接收标记图像的集合;

从所述多个摄像机内的多个目标摄像机接收未标记图像的集合;

使用特征提取器从所述标记图像和所述未标记图像提取特征;

使用多域分类器将提取的特征分类到某个源或目标中;

使用密度估计分支估计车辆密度;

使用所述多域分类器和所述特征提取器之间的梯度反转来执行基于反向传播的训练;以及

在所述基于反向传播的训练之后,从所述图像中确定所述对象计数。

43.一种机器可读存储介质,其包含用于执行根据权利要求1-42中任一项所述的方法的机器可执行指令。

估计对象密度和/或流量的深度学习方法及相关方法和软件

[0001] 相关申请数据

[0002] 本申请要求2017年4月5日提交的且题为“Extract Urban Traffic Information From Citywide Infrastructure Sensors to Augment Autonomous Intelligent System”的序列号为62/601,953的美国临时专利申请的优先权的权益,所述临时专利申请的全部内容以引用的方式并入本文。

发明领域

[0003] 本发明大体上涉及机器/计算机视觉领域。特别地,本发明涉及使用机器视觉来估计对象的密度和/或流量的深度学习方法,以及相关的方法和软件。

[0004] 背景

[0005] 随着公共部门中仪器的不断增加和自动化程度的不断提高,数据量也越来越大,人们希望将这些数据用于诸如提高自动化程度、并且更广泛地说提高对决策者可能有用的信息的实时意识的用途。例如,城市越来越多地安装了各种传感器,如磁环检测器、红外传感器、压力垫、路边雷达和网络摄像机。这些传感器中的许多传感器可以提供城市中有关交通流量的相关信息,如车辆速度、计数和类型。特别地,安装在城市街道或城市其他道路的交叉路口的摄像机越来越多,使得能够提取每种类型车辆的交通流量的实时估计,例如黄色出租车的流速。全市范围的网络摄像机全天候连续捕获交通视频,生成大规模的交通视频数据。这些摄像机要么是低质量的,要么它可能只是为了处理它们的视频的低质量版本。这排除了大多数现有的用于交通流量分析的技术。

[0006] 预计无人驾驶车辆将越来越多地出现在城市街道上,并在不久的将来成为交通流量的重要组成部分。虽然无人驾驶车辆传感提供了对其操作的本地条件的认识,但随着物联网的出现,越来越多的基础设施传感器有可能为无人驾驶汽车提供对整个城市或其他地区的交通条件的全局认识。在处理无人驾驶汽车的传感器套件所收集的传感数据方面已经付出了很大努力。然而,在同时处理来自城市交通摄像机的流式视频以构建多模型自主系统方面,工作却少得多。事实上,典型交通摄像机中固有的问题(如低帧速率和低分辨率)以及车辆交通本质中固有的问题(如变化的天气条件、每天变化的照明条件以及各种各样的车辆类型和型号)使得以任何有意义的方式使用来自这些摄像机的信息极具挑战性。

[0007] 公开概述

[0008] 在一个实施方式中,本公开涉及一种向智能交通系统提供交通密度和/或交通流量数据的方法,其中交通密度和/或交通流量数据用于道路网络的多个区域,每个区域具有与其相关联的交通摄像机,该交通摄像机捕获相应区域的至少一个相应交通图像。该方法包括使用至少一个相应交通图像,在每个交通摄像机本地自动估计道路网络的相应区域中的交通密度和/或交通流量,对于每个相应区域进行估计包括:使用训练标注图像(training annotated image)来训练基于人工神经网络(ANN)的密度估计算法,其中基于ANN的密度估计算法被配置成在训练时处理一个或更多个输入图像,以自动确定对于输入图像中存在的交通的交通密度和/或交通流量;接收至少一个相应交通图像;以及在基于

ANN的密度估计算法的训练之后，使基于ANN的密度估计算法处理至少一个相应交通图像，以确定对于相应区域的交通密度和/或交通流量；以及从每个交通摄像机本地向智能交通系统传输与道路网络的相应区域中的交通密度和/或交通流量相对应的交通密度和/或交通流量数据。

[0009] 在另一实施方式中，本公开涉及一种生成用于基于人工神经网络(ANN)的密度估计算法的训练图像的集合的方法，其中基于ANN的密度估计算法用于检测由固定摄像机捕获的场景内的一种或更多种目标类型的对象的密度。该方法包括，接收由固定摄像机捕获的真实图像的第一集合，其中每个真实图像包括具有目标类型内的类型的至少一个真实对象；使用基于ANN的检测器处理每个真实图像，以便用相应的各自的第一边界框区域标注每个真实图像内一种或更多种目标类型的对象的出现，其中该处理产生标注的真实图像的集合；以及使用基于ANN的分割器处理每个标注的真实图像，以便生成相应的像素级(pixel-wise)背景掩模，该背景掩模进一步用于生成用于训练的密度图。

[0010] 在又一实施方式中，本公开涉及一种估计由摄像机捕获的场景图像中不同目标类型的对象密度的方法。该方法包括，接收训练数据集，其中训练数据集包括具有的训练视角与摄像机的摄像机视角基本相同的多个合成图像，每个合成图像包含不同类型的多个再现的对象，其中这些再现的对象：基于不同类型的不同对象的3D CAD模型；已经使用基于摄像机视角的实际视角和实际遮挡被放置在合成图像中；并且被自动标记用于基于3D CAD模型进行归类；使用合成图像来训练基于人工神经网络(ANN)的密度估计算法，其中基于ANN的密度估计算法被配置成在训练时处理输入图像以自动估计存在于输入图像中的不同目标类型的对象密度；接收图像；以及在基于ANN的密度估计算法的训练之后，使用基于ANN的密度估计算法处理图像中每一个以估计图像的密度。

[0011] 在又一实施方式中，本公开涉及一种向智能交通系统提供交通密度和/或交通流量数据的方法，其中交通密度和/或交通流量数据用于道路网络的多个区域，每个区域具有与其相关联的交通摄像机，该交通摄像机捕获相应区域的至少一个相应交通图像。该方法包括，在集中式估计系统处从每个交通摄像机接收对应于至少一个相应交通图像中的每一个的相应的低分辨率图像；在集中式估计系统处，使用相应的低分辨率图像自动估计道路网络的相应区域中的交通密度和/或交通流量，在每个相应区域中进行估计包括：使用训练标注图像来训练基于人工神经网络(ANN)的密度估计算法，其中基于ANN的密度估计算法被配置成在训练时处理一个或更多个输入图像，以自动确定输入图像中存在的交通的交通密度和/或交通流量；以及在基于ANN的密度估计算法的训练之后，使基于ANN的密度估计算法处理对于区域中每一个的相应的低分辨率图像，以确定对于相应区域的交通密度和/或交通流量；以及从集中式估计系统向智能交通系统传输与道路网络的区域的交通密度和/或交通流量相对应的交通密度和/或交通流量数据。

[0012] 在又一实施方式中，本公开涉及一种从摄像机获取的图像确定对象计数的方法。该方法在机器中执行，并且包括：使用完全卷积网络(FCN)联合估计对象计数和对象密度，其中联合估计包括使用FCN将图像内的密集特征映射到对象密度；使用残差学习框架，参照图像中的对象密度之和来学习残差函数；组合来自FCN的较浅层的外观特征和来自FCN的深层的语义特征，以产生更密集的特征图；以及使用更密集的特征图确定对象计数。

[0013] 在进一步的实施方式中，本公开涉及一种从摄像机获取的视频的帧确定对象计数

的方法。该方法在机器中执行，并且包括：使用完全卷积网络(FCN)联合估计对象计数和对象密度，其中联合估计包括使用FCN创建对于视频的帧的对象密度图；向长短期记忆(LSTM)网络提供该对象密度图，其中LSTM网络被配置成参照帧的每一个中的密度之和来学习残差函数；以及将残差函数和密度求和，以确定对象计数。

[0014] 在另一个实施方式中，本公开涉及一种从摄像机获取的视频的帧确定对象计数的方法。该方法在机器中执行，并且包括：使用完全卷积网络(FCN)联合估计对象计数和对象密度，其中联合估计包括使用FCN创建对于视频的帧的对象密度图；向长短期记忆(LSTM)网络提供该对象密度图，其中LSTM网络被配置成参照帧的每一个中的密度之和来学习残差函数；以及将残差函数和密度求和，以确定对象计数。

[0015] 在又一实施方式中，本公开涉及一种从来自多个摄像机的图像确定对象计数的方法。该方法包括，从多个摄像机内的多个源摄像机接收标记图像的集合；从多个摄像机内的多个目标摄像机接收未标记图像的集合；使用特征提取器从标记图像和未标记图像提取特征；使用多域分类器将提取的特征分类到某个源或目标中；使用密度估计分支估计车辆密度；使用多域分类器和特征提取器之间的梯度反转来执行基于反向传播的训练；以及在基于反向传播的训练之后，从图像中确定对象计数。

[0016] 附图简述

[0017] 为了说明本发明的目的，附图示出本发明的一个或更多个实施例的多个方面。然而，应该理解，本发明不限于附图中所示的精确布置和工具，其中：

[0018] 图1是本公开的基于人工神经网络(ANN)的估计系统的示例车辆交通实现的高级图示，该估计系统根据由一个或更多个交通摄像机拍摄的图像自动估计交通密度和/或交通流量；

[0019] 图2是向智能交通系统提供交通密度和/或交通流量数据的示例方法的图示；

[0020] 图3是可以在本公开的估计算法(如图1的估计算法)中实现的示例FCN-rLSTM网络的高级示意图；

[0021] 图4是图3的FCN-rLSTM网络的体系结构的示意图；

[0022] 图5是示出直接连接的FCN-LSTM(FCN-dLSTM)网络和残差学习FCN-LSTM(FCN-rLSTM)网络之间差异的高级示意图；

[0023] 图6是生成用于训练基于ANN的网络的合成训练图像的集合的示例方法的图示；

[0024] 图7是可用于执行图6方法的示例合成训练图像生成系统的示意图；

[0025] 图8是生成用于训练基于ANN的网络的混合训练图像的视频的示例方法的示意图；

[0026] 图9是示例计算系统的高级示意图，该示例计算系统可用于实施用于执行本文公开的任何一种或更多种方法的软件；

[0027] 图10是可在本公开的估计算法(如图1的估计算法)中实施的示例多任务学习FCN(FCN-MT)网络的高级示意图；

[0028] 图11是用于利用对抗学习的多摄像机域自适应(MDA)的示例系统和方法的示意图，该示例系统和方法可以被实施用于训练与不同背景、照明和/或天气条件的图像一起使用的估计算法；以及

[0029] 图12是用于实施本公开的MDA方法(如图11的MDA方法)的示例网络体系结构的高级示意图。

[0030] 详细描述

[0031] 在一些方面，本公开涉及使用机器视觉来估计对象的密度和/或流量的方法。(如本文所使用的，术语“和/或”当用作两个项目之间的连接词时，表示项目中的一个、项目中的另一个或两个项目，这取决于实现的选择。)通常，对象本质上以多种方式中的任何一种或更多种视觉地变化。例如，在交通环境中，对象是车辆，车辆通常(除非存在诸如“仅公共汽车”或“不允许卡车”的限制)几乎可以是任何类型(例如，客车、SUV、小型货车、货车、箱式卡车、牵引车拖车、城市公共汽车、州际公共汽车、娱乐车辆、工作卡车等)，并且几乎可以是过去或现在的任何制造商品牌和型号。作为另一个示例，在人群环境中，对象通常是人，人可以具有任何性别、身材、年龄、站立、坐着、面向任何方向等。仅这些条件本身和低视频质量就使得可靠的对象检测和任何后续密度和/或流量估计具有挑战性。更不用说，根据摄像机视角和对象之间的接近程度，对象中一个或更多个被对象中另一个的部分遮挡进一步增加了重大挑战。本文公开的技术和方法还可以克服由于对象可变性而带来的挑战，并允许鲁棒和可靠的密度和/或流量估计。

[0032] 这种估计方法可以在室外实施，户外的天气条件一直在变化，由于每天和每年的周期以及不断变化的天气条件，照明条件也在变化。例如，如通过交通摄像机所见，由于诸如阴影、眩光、反射(例如，来自潮湿表面)、光强度等的因素，场景的变化的视觉特性会使城市街道、交叉路口或其他位置上的交通的密度和/或流量的估计变得非常复杂。类似地，户外空间中人群的密度和/或流量的估计同样会因场景的变化的视觉特性而变得复杂。本文公开的技术和方法可以克服由于照明和/或天气条件(取决于它们的存在)带来的挑战，并且允许鲁棒和可靠的密度和/或流量估计。

[0033] 本公开的估计方法可以使用任何合适类型的成像设备来实施，如交通摄像机，包括具有低分辨率(例如，350×250像素或更少)和低帧率(例如，1帧每秒(FPS)或更少)中的任何一种或更多种的交通摄像机。例如，城市越来越多地在其整个街道和道路网络中的许多位置处部署交通摄像机，包括在许多交叉路口、已知的交通瓶颈处和沿重要的交通走廊等，并且这些交通摄像机通常是低分辨率的并具有低帧率，或者，如果具有较高的质量，则只处理摄像机捕获的低分辨率版本的视频是令人感兴趣的。这些特性中的任一个和两个也会使合理和有效地估计对象的密度和/或流量的能力复杂化。本文公开的技术和方法也可以克服由于摄像机特性而带来的挑战，并允许鲁棒和可靠的密度和/或流量估计。

[0034] 在一些方面，本公开涉及使用估计的对象密度和/或流量来影响行动者的行为，该行动者可以是基于机器的系统或人类。在车辆交通环境中，对象密度和/或流量估计如何被用来影响行动者行为的示例比比皆是。例如，道路网络可以包括交通控制基础设施，如交通控制信号，其由被编程为自动控制交通控制信号的集中式或分布式控制器系统控制。这里，控制器系统可以被编程为使用交通密度和/或流量估计来以倾向于改善交通流量的方式控制单独的交通控制系统或交通控制系统的集合。作为另一个示例，行动者可以是自动驾驶汽车，其可以配备有控制系统，该控制系统被编程为使用交通密度和/或流量估计来计算适当的路线，该路线可以倾向于避免道路网络的拥挤区域，以便最小化行驶时间。作为进一步的示例，自动驾驶汽车行动者可以选择使用由监视摄像机提供的关于行动者传感器系统的盲点中的交通的信息，如在拐角周围。作为又一个示例，人类驾驶员可以访问导航设备(例如，内置在车辆中或者与车辆分开提供，如运行合适app的智能手机或者辅助全球定位系统

(GPS)导航设备等),这些导航设备被编程为以有用的方式使用交通密度和/或流量估计,如通过计算作为这种估计的函数的路线并将这些路线呈现给用户,或者基于估计向用户显示密度和/或流量信息,如叠加在显示给用户的地图图像上。这些只是交通密度和/或流量估计如何用于控制行动者行为的几个示例。

[0035] 在本公开内容中和在所附权利要求中,术语“智能交通系统”用于表示任何基于机器的系统,该系统利用出于某种目的而使用本文公开的方法确定的交通密度和/或流量估计。例如,智能交通系统可以是自动驾驶车辆、用于自动驾驶车辆的控制系统、交通控制基础设施控制系统(集中式或分布式)、导航设备和实时交通显示系统等。关于实时交通显示系统,这可以是例如移动或固定设备或基于估计以电子方式显示覆盖有交通密度和/或流量信息的地图的设备。

[0036] 同样在本公开中,术语“估计系统”用于表示使用本文公开的方法自动确定对象密度和/或流量估计的系统。这种估计系统可以由一个或更多个计算设备(例如,服务器、台式计算机、膝上型计算机、平板计算机、摄像机上的计算设备等)和一个或更多个合适的有线和/或无线通信系统组成,通信系统用于接收图像数据并将估计传送给估计的任何合适的用户,如一个或更多个智能交通系统。在一些实施例中,估计系统可以包括生成在本文公开的估计方法中使用的图像的成像设备。然而,例如,如果成像设备由另一实体提供,如城市、州或其他政府或非政府实体,则估计系统可以被认为排除了这种成像设备。本领域的技术人员将容易理解适用于所公开的估计系统和智能交通系统的计算设备、通信系统和成像设备的类型,使得对于本领域的技术人员来说,理解本发明的范围不需要对每种设备的详细描述。

[0037] 注意,以下示例涉及估计车辆交通特性,其确实是本公开的方法和系统的有用部署。然而,应当注意,估计车辆交通特性并不是这些方法和系统的唯一用途,因为相同或相似的技术可以用于其他目的,如估计人群、动物群体和其他无生命但运动的对象群体的特性,以及其他对象的运动特性。本领域的技术人员将很容易理解如何从这些基于车辆的示例修改技术和方法,并将其应用于其他场景,而无需过多的实验。事实上,如“智能交通系统”等方面可以根据需要修改为“智能人群系统”或“智能对象分组系统”,以适应特定的应用。

[0038] 还应注意,原始提交的所附权利要求是本公开的一部分,就好像它们逐字出现在该详细描述部分中一样。

[0039] 示例车辆交通实现

[0040] 考虑到前述内容并且现在参考附图,图1示出了本公开的基于ANN的估计系统的示例车辆交通实现100,其中多个交通摄像机104(1)至104(N)部署在道路网络108的交叉路口108(1)至108(X)处,道路网络108可以是例如城市内的街道网络。每个交通摄像机104(1)至104(N)具有相对于相应的交叉路口108(1)至108(X)的视角,并且提供场景(未示出)的一个或更多个图像104I(1)至104I(N)(例如,以静止图像或视频图像的形式),在该示例中,场景是每个交通摄像机独有的场景。在一些实施例中,交通摄像机104(1)至104(N)可以都是相同类型的。例如,每个交通摄像机104(1)至104(N)可以是低分辨率摄像机,如分辨率小于350×250像素的摄像机。附加地或替代地,每个交通摄像机104(1)至104(N)可以具有低帧率,如小于1FPS的帧率。在一些实施例中,每个交通摄像机可以具有高分辨率(例如,2000像

素×1000像素或更高)和/或高帧率,如10FPS或更高。在一些实施例中,交通摄像机104(1)至104(N)可以是不同类型的。基本上,除非另有说明,交通摄像机104(1)至104(N)的分辨率和帧率可以是任何分辨率和帧率。

[0041] 在一个示例中,交通摄像机104(1)至104(N)与集中式估计系统112通信,该集中式估计系统112使用来自每个交通摄像机的一个或更多个图像104I(1)至104I(N)来确定相应场景内车辆116(为了方便起见,仅标记了几个)的密度和/或流量的估计。在这点上,应当注意,本文和所附权利要求中使用的术语"集中式"并不意味着必须有单个计算机或执行估计的其他设备,尽管可能是这种情况。相反,在道路网络108和相应多个交通摄像机104(1)至104(N)的环境中,"集中式"意味着远离交通摄像机来执行处理。这种处理可以在单个计算设备中或者跨多个计算设备(例如,web服务器)执行。

[0042] 在一些实施例中,可能希望最小化需要从交通摄像机104(1)到104(N)传输的数据量。这可以通过例如部署具有低分辨率和/或低帧率的交通摄像机来实现。然而,这也可以通过将由交通摄像机104(1)至104(N)捕获的图像104I(1)至104I(N)下转换到比相应交通摄像机的捕获(as-captured)分辨率低的分辨率,并将较低分辨率图像传输到集中式估计系统112和/或传输少于每个摄像机所捕获的所有图像来实现。作为后者的示例,如果特定的交通摄像机具有20FPS的帧速率,则摄像机或与其进行数据通信的其他设备可以每秒只传输这些图像中的一个。如本文所述,本公开的基于ANN的估计算法特别适合用于处理低分辨率图像和/或低帧率视频,同时仍然提供有用的估计。

[0043] 如上所述,车辆116可以彼此不同,并且可以是各种类型、品牌、型号等中的任何一种。本领域技术人员将容易理解,交通摄像机104(1)至104(N)可以以本领域已知的任何合适的有线或无线方式与估计系统112通信。

[0044] 估计系统112包括一个或更多个基于人工神经网络(ANN)的估计算法120(例如,分布式处理体系结构中的多个算法),估计算法120使用由交通摄像机104(1)至104(N)获取的图像104I(1)至104I(N)来执行估计,以生成相应的各自的交通密度和/或流量估计120(1)至120(N),这取决于估计算法120是如何配置的。估计算法120可以是任何合适的基于ANN的算法,如包括一个或更多个卷积神经网络(CNN)的算法。合适的基于ANN的算法的示例包括但不限于包括基于区域的CNN(R-CNN)(例如,快速RCNN和更快的RCNN)和/或完全卷积网络(FCN)(例如,多任务学习FCN(FCN-MT)、具有残差学习的长短期记忆FCN(FCN-rLSTM))等的算法。基本上,对估计算法120中使用的ANN的类型没有限制,除了要求它足够鲁棒并且可训练以克服需要执行估计算法以提供有用估计的困难对象检测任务的挑战。一般来说,本文公开的非LSTM ANN适用于从单个图像估计密度,而基于LSTM的ANN由于其在时序图像的集合中"记住"从一帧到下一帧的信息的能力而特别适用于估计密度和流量。下面描述了可以在估计算法120中实施的基于ANN的技术的详细示例。

[0045] 每个估计算法120被适当地训练以在不同的照明和天气条件下以及不同的车辆位置处检测不同类型的不同车辆。下面详细描述用于训练估计算法120的示例训练方法,包括基于合成图像和基于混合图像的训练。

[0046] 估计系统112可以包括合适的计算系统124,计算系统124包括存储器128和执行估计算法的一个或更多个处理器132,其中存储器128包含估计算法120等。如本领域技术人员将容易理解的,存储器128可以由任何一种或更多种类型的存储器组成,包括本领域公知的

非易失性和易失性存储器。类似地，本领域技术人员将容易理解，一个或更多个处理器132可以是任何合适的类型(例如，通用处理器、现场可编程门阵列等)，并且如果提供了不止一个处理器，则处理器可以包含在可位于任何合适位置的一个或更多个计算设备(例如，服务器、台式计算机、膝上型计算机等)中。基本上，对计算系统124的类型和配置没有限制，除了它能够适当地执行估计算法120并实现相关功能，如与交通摄像机104(1)至104(N)通信以及根据需要与其他系统通信。

[0047] 估计系统112可以向一个或更多个智能交通系统136提供交通密度和/或流量估计120(1)至120(N)，智能交通系统136可以使用这些估计来影响行动者(未示出)的行为。如上所述，可作为智能交通系统136提供的智能交通系统的示例包括但不限于自驾驶车辆、用于自驾驶车辆的控制系统、交通控制基础设施控制系统(集中式或分布式)、导航设备和实时交通显示系统等，以及它们的任意组合和/或数量。

[0048] 图1的前述描述涉及一个实施例，其中估计由集中式估计系统112执行，其中用于确定交通密度和/或交通流量估计的处理远离交通摄像机104(1)到104(N)执行。在替代实施例中，集中式估计系统112可以由分布式处理系统代替，在分布式处理系统中，对每个交通摄像机104(1)至104(N)的图像104I(1)至104I(N)的估计处理在该交通摄像机本地执行。例如，每个交通摄像机104(1)至104(N)可以包括执行本地存储的估计算法的机载计算系统(未示出)。在这种环境中，术语“本地”和类似术语意味着估计在每个交通摄像机104(1)至104(N)上执行，或者由靠近每个交通摄像机的合适的计算系统执行，如位于靠近交通摄像机的保护外壳中的计算系统。本领域技术人员将容易理解，什么构成“本地”处理和“集中式”处理的对比。应当注意，本地处理可以包括处理来自彼此靠近定位的两个或更多个交通摄像机的图像。例如，如果一对摄像机位于同一交叉路口，但是在交叉路口的不同区域上进行训练，则该交叉路口可以配备有单个计算系统来执行针对两个摄像机的估计。在诸如这样的分布式估计系统中，每个摄像机或其他本地计算系统可以向智能交通系统136提供相应的交通密度和/或交通流量数据。

[0049] 图2示出了向智能交通系统(如图1的智能交通系统136)提供交通密度和/或交通流量数据的示例方法200。在该示例中，交通密度和/或交通流量数据是针对道路网络(如图1的道路网络108)的多个区域的，其中每个区域都与交通摄像机(如图1的交通摄像机104(1)至104(N)中的任何一个)相关联，该交通摄像机捕获道路网络的相应区域的至少一个相应交通图像。方法200几乎可以在任何车辆交通实现中实施，如具有多个交通摄像机104(1)至104(N)的图1的车辆交通实现100。参考图2，在块205处，使用基于ANN的估计算法(如图1的估计算法120)在道路或道路网络的相应区域中，对每个交通摄像机本地估计交通密度和/或交通流量。道路的每个区域可以是任何区域，如交叉路口、高速公路交换站、通向收费结构的通道、隧道入口、桥跨(bridge span)的一部分等。通常，每个区域都是对了解交通水平感兴趣的区域。每个交通摄像机104(1)至104(N)提供单个帧或图像的时序集合，通常是视频，估计算法120分析图像以确定相应区域中的交通密度和/或交通流量。在交通摄像机相当初级的一些实施例中，每个交通图像具有小于350×250像素的分辨率和小于例如1帧每秒的帧率。

[0050] 在块210(图2)处，估计包括使用训练标注图像针对每个区域训练估计算法。估计算法被配置成当被训练时处理一个或更多个输入图像，以自动确定一个或更多个输入图像

中存在的交通的交通密度和/或交通流量。本领域的技术人员将理解，估计算法通常将持续处理从网络中的每个交通摄像机获得的图像，以在整个期望时间段内提供实时的交通密度和/或交通流量估计，所述期望时间段可以是全天或一个或更多个特定时间段，如在早上、中午和晚上的“高峰时间”，可能是每天或仅在工作周期间，或者在任何其他期望的时间表上。本领域技术人员还将理解，如果需要进行交通流量估计，估计算法必须适当配置，并且它必须处理时序图像，以便检测每个场景中车辆的移动。

[0051] 在块215处，针对道路网络的每个区域的估计包括接收针对该区域的交通图像。在块220处，估计还包括，在估计算法的训练之后，用估计算法针对每个区域处理对于该区域的交通图像，以确定交通密度和/或交通流量。在块225处，针对每个区域所确定的交通密度和/或交通流量(可选地与关于道路的区域位置的位置信息一起)从每个交通摄像机的本地传输到智能交通系统。

[0052] 应当注意，如本文中所使用的，术语“传输”和类似术语意味着传送相应信息或数据的任何方式，包括但不限于经由任何合适协议的有线传输、经由任何合适协议的无线传输及其任意组合，这取决于特定实施方式中物理部件的关系的布置。位置信息可以是指示道路区域位置的任何信息，如GPS坐标、交叉路口名称、道路名称和沿着道路的位置、摄像机标识符等、以及它们的任意组合。基本上，位置信息可以是单独或与其他信息(如关系表)组合来指示道路区域位置的任何信息。

[0053] 本领域技术人员将容易理解，块205至220可以适用于估计除车辆交通之外的对象的密度和/或流量，并且不需要在智能交通系统的环境中执行。实际上，自动确定的对象密度和/或对象流量数据可以用于任何合适的目的，包括具有与上面参考图1提到的示例智能交通系统相似功能的系统。

[0054] 估计算法

[0055] 如上所述，许多估计算法可以用于本公开的密度和/或流量估计算法，如图1的估计系统112的估计算法120。通过将FCN网络与多任务学习(MT)网络相结合，或者在时序图像的情况下与LSTM网络相结合，可以改善结果。此外，通过利用残差学习来增强FCN-LSTM网络以创建FCN-rLSTM网络，可以进一步改善LSTM增强的FCN。下面将马上详细描述适用于本公开的估计算法的FCN-MT网络和FCN-rLSTM网络的示例。然而，本领域的技术人员将容易理解，这些示例仅仅是说明性的，并且可以做出许多变化，包括缩略(例如，通过去除残差学习)，并且存在替代方案，如上面相对于图1提到的那些方案，替代方案可以代替基于FCN的估计算法来实施。此外，本领域技术人员将容易理解，本文给出的特定方程仅仅是说明性的，并且可以使用其他方程代替它们来产生类似的结果。

[0056] FCN-MT模型(单个图像)

[0057] 一些块级优化方法通过将密集图像特征映射到车辆密度来避免个体车辆检测和跟踪。它们在权重矩阵中嵌入道路几何图形，并获得有希望的结果。评估这些方法得出以下见解：1)其验证了将密集特征映射到车辆密度用于车辆计数的有效性；2)考虑摄像机视角对减少计数误差很重要；以及3)其揭示了考虑邻近像素的相关性的必要性。然而，这种方法存在局限性：1)规模不变特征变换(SIFT)特征不能有效区分每个像素；2)它们的性能高度依赖于背景减除；以及3)它们的性能对不同的场景和环境条件敏感。

[0058] 考虑到上述城市交通摄像机的挑战和块级优化方法的局限性，本发明人已经寻求

了一种能够解决以下挑战性和关键性问题的更鲁棒和更一般化的模型:1)提取代表性和区别性密集特征;2)了解具有附加丰富信息的城市交通,如车辆计数、类型和速度;3)克服诸如高遮挡和低图像分辨率等挑战来检测小型车辆;4)并入交通视频的时间信息;5)使模型适应多台摄像机和不同的环境条件;以及6)了解不同的摄像机视角。

[0059] 利用分层特征学习和深度神经网络表现出的最先进的性能,用深度多任务模型来代替具有完全卷积回归网络的线性回归模型。为了解决上述问题,本发明人开发了:1)基于完全卷积神经网络的深度多任务学习框架(FCN-MT),用于联合学习车辆密度和车辆类型;2)结合交通流量时间信息的深度时空网络;以及3)多域对抗训练机制,使模型适应不同的数据域。以下章节详细描述这些技术。

[0060] 为了克服先前模型的局限性,使用深度多任务模型来联合学习车辆密度和车辆类型。如图10所示。为了产生与输入图像具有相同大小的密度图,FCN被设计成通过密集前馈计算和反向传播来执行每次整幅图像(whole-image-at-a-time)的像素级预测。在网络末端增加了密度图的多级监督,以提高密度估计的精确度。使用基于FCN的密度估计任务联合学习车辆类型和边界框。为了检测具有小规模和高遮挡的车辆,估计的密度图被作为车辆检测的先验。该方法将该模型与现有的基于检测的计数方法区别开来,后者首先检测个体车辆,然后对车辆数量计数。在测试阶段中,系统将一幅图像作为输入,并输出车辆密度、类别置信度得分和对于每个像素的预测位置偏移。全局车辆计数以残差学习方式从估计的密度图回归。

[0061] 车辆类型检测被分解成两项任务:车辆边界框回归和车辆类型分类。我们将目标边界框的左上顶点和右下顶点分别定义为$b_l=(x_l,y_l)$和$b_r=(x_r,y_r)$,然后位于输出特征图中(x_p,y_p)处的每个像素p用5维向量$\tilde{T}_p=\{\tilde{s},dx_l=\tilde{x}_p-x_l,dy_l=\tilde{y}_p-y_l,dx_r=\tilde{x}_p-x_r,dy_r=\tilde{y}_p-y_r\}$描述边界框,其中$\tilde{s}$是处于某种车辆类型的置信度得分,并且$(dx_l,dy_l,dx_r,dy_r)$表示从输出像素位置到目标边界框的边界的距离。估计的密度图作为检测的先验。鼓励高密度区域输出小边界框,而鼓励低密度区域输出大边界框。图10示出了示例FCN-MT的总体结构,其包含卷积网络1000、反卷积网络1004、特征组合和选择部件1008以及多任务学习部件1012。

[0062] 基于FCN的车辆计数和检测面临三大挑战:1)车辆规模的变化;2)降低的特征分辨率;以及3)高遮挡和小车辆规模导致对于车辆检测的高遗漏率。为了避免规模变化引起的大误差,基于FCN-MT的示例方法联合地执行全局计数回归和密度估计。用残差学习框架将全局计数回归重组为参照每个帧中的密度之和学习残差函数,而不是从最后的特征图直接回归全局车辆计数。这种设计避免了学习不相关的函数,并简化了网络训练。

[0063] 第二个挑战是降低的特征分辨率,其由针对图像分类最初设计的传统深度卷积神经网络的最大池化(max-pooling)和步幅(striding)的重复组合引起的。这导致特征图具有显著降低的空间分辨率。为了解决这个问题,本方法通过组合浅层的外观特征和深层的语义特征来使用更密集的特征图。然后在组合的特征量(feature volume)之后增加具有1x1核的卷积层,以执行特征选择。所选特征可以更好地区分前景和背景。因此,整个网络能够准确估计车辆密度,而无需前景分割。

[0064] 为了解决高遮挡和小车辆规模的第三个挑战,不直接检测个体车辆,而是用估计

密度图的先验来回归车辆边界框。鼓励具有正密度值的区域具有高置信度得分，而具有零密度值的区域应当具有低置信度得分。也鼓励具有大密度值的区域具有小边界框，而具有小密度值的区域具有大边界框。

[0065] 车辆密度估计：在网络的最后一级处，基于FCN-MT的方法学联合学习车辆密度、车辆计数和车辆类型。在这个示例中，车辆密度由最后的卷积1x 1层从特征图预测。欧几里德距离被用来度量估计密度和地面实况(ground truth)之间的差异。在这个示例中，针对密度图估计的损失函数定义如下：

[0066] $$L_D(\Theta)=\frac{1}{2N}\sum_{i=1}^{N}\sum_{p=1}^{P}\|F(X_i(p);\Theta)-F_i(p)\|_2^2 \quad (1)$$

[0067] 其中，Θ是FCN-MT模型的参数向量，N是训练图像的数量，并且$F_i(p)$是像素p的地面实况密度。

[0068] 全局计数回归：第二任务(全局计数回归)被重组为参照密度之和学习残差函数，它由两部分组成：1)基本计数：整个图像上密度图的积分；以及2)偏移计数：在反卷积网络的卷积3×3层之后，由两个完全连接的层从特征图预测。将这两个部分加在一起得到估计的车辆计数，如以下方程所示：

[0069] $$C(i)=B(D(i);\gamma)+\sum_{p=1}^{P}D(i,p) \quad (2)$$

[0070] 其中，γ是两个完全连接层的可学习参数；B(D(i)；ϒ)是学习到的偏差，而D(i,p)表示图像i中每个像素p的密度。假设优化残差映射比优化原始的、未参考的映射更容易。考虑到对于一些帧的车辆计数可能具有非常大的值，在该示例中，采用Huber损失来度量估计的计数和地面实况计数之间的差异。针对一个帧的计数损失定义如下：

[0071] $$L_\delta(i)=\begin{cases}\frac{1}{2}(C(i)-C_t(i))^2 & \text{对于 } |C(i)-C_t(i)|\le\delta\\ \delta|C(i)-C_t(i)|-\frac{1}{2}\delta^2 & \text{其他}\end{cases} \quad (3)$$

[0072] 其中$C_t(i)$是帧i的地面实况车辆计数，C(i)是帧i的估计的损失，而δ是用来控制训练集中异常值的阈值。

[0073] 车辆类型检测：第三任务的损失由两部分组成：车辆类型分类损失和边界框回归损失。在一个示例中，车辆类型分类损失定义如下：

[0074] $$L_T(\tilde{s},s^*)=\|\tilde{s}-s^*\|^2 \quad (4)$$

[0075] 其中，$\tilde{s}$是预测的置信度得分，而s^*是地面实况标签。在一个示例中，边界框回归损失定义如下：

[0076] $$L_B(\tilde{d},d^*)=\sum_{b\in\{x_l,y_l,x_r,y_r\}}\left\|\tilde{d}_b-d_b^*\right\|^2 \quad (5)$$

[0077] 其中，$\tilde{d}_b=(\tilde{d}_{x_l},\tilde{d}_{y_l},\tilde{d}_{x_r},\tilde{d}_{y_r})$表示预测的位置偏移，而$d_b^*=(d_{x_l}^*,d_{y_l}^*,d_{x_r}^*,d_{y_r}^*)$表示目

标。然后，在一个示例中，网络的总体损失函数被定义为：

[0078] $$L = L_D + \lambda \sum_{i=1}^{N} L_\delta(i) + \alpha L_T + \beta L_B \quad (6)$$

[0079] 其中，λ、α和β分别是全局计数损失、车辆类型分类损失和边界框回归损失的权重。通过同时学习三个相关的任务，可以用更少的参数更好地训练每个任务。本领域技术人员将容易理解，如本文中的其他方程一样，方程(4)-(6)仅仅是示例性的，并且可以用其他方程代替。

[0080] FCN-rLSTM模型和网络体系结构(时序图像)

[0081] 图3示出了在交通密度和/或流量估计场景的环境中的示例FCN-rLSTM网络300，如由交通摄像机(未示出)所获取的城市交叉路口的五个时序视频帧304(1)至304(5)所描绘的。在该示例中，交通摄像机具有大视角、低分辨率和低帧率，并且场景通常具有高遮挡，如较大的车辆遮挡较小车辆的至少一些部分。在FCN-rLSTM网络300中，FCN 308和LSTM 312被组合在残差学习框架中，利用FCN的优势进行密集视觉预测，并利用LSTM的优势进行时间相关性建模。视频帧被输入到FCN 308中，并且输出密度图被馈送到LSTM的堆栈中，以参照每个帧中的密度之和来学习残差函数。全局车辆计数最终通过将学习到的残差与密度求和来生成。

[0082] 由于交通摄像机视频的低空间和时间分辨率以及高遮挡排除了现有的用于车辆计数的检测或基于运动的方法，所以当前示例应用FCN 308将密集(像素级)特征映射到车辆密度，并避免检测或跟踪个体车辆。基于FCN 308的密度估计允许任意的输入分辨率，并输出与输入图像大小相同的车辆密度图。现有的对象计数文献估计对象密度图，并直接对整个图像上的密度求和以获得对象计数。但是当视频具有大视角和超大车辆(例如，大型公共汽车或大型卡车)时，这种方法会有很大的误差。然而，通过以残差学习方式连接FCN 308和LSTM 312，当前的FCN-rLSTM网络300被用于联合估计车辆密度和车辆计数。这种设计利用FCN 308的优势进行像素级预测，并利用LSTM 312的优势学习复杂的时间动态。通过考虑车辆计数的时间相关性，计数精度显著提高。然而，训练组合的FCN 308和LSTM 312网络并不容易。增加了FCN 308和LSTM 312的残差连接，以加速训练过程。得到的端到端可训练网络具有高收敛速度，并进一步提高了计数精度。

[0083] 图4中示出了图3的示例FCN-rLSTM网络300的体系结构400以及示例详细参数。如图4所示，网络体系结构400包括卷积网络404、反卷积网络408、超空洞(hyperatrous)特征组合412和LSTM层416。在这个示例中，大小为3x3的小核被应用于卷积网络404的卷积(conv)层和反卷积网络408的反卷积(deconv)层。此外，在这个示例中，较高层中滤波器通道的数量增加，以补偿最大池化造成的空间信息损失。

[0084] 为了保持特征图分辨率，使用超空洞组合。在超空洞组合中，空洞卷积(atrous convolution)被集成到卷积网络和第二最大池化层之后的特征图中，并且空洞卷积层被一起组合到更深的特征量中。空洞卷积相当于通过在非零滤波器抽头之间插入孔来对上采样滤波。它更密集地计算特征映射，然后简单地将特征响应双线性插值回原始图像大小。与常规卷积相比，空洞卷积在不增加参数数量的情况下有效地扩大了滤波器的视场。

[0085] 经过几个空洞卷积层后，来自第二最大池化层和空洞卷积层的特征被组合。然后，

在组合的特征量之后，使用具有1x1核的卷积层来执行特征重新加权，以鼓励重新加权的特征量更好地区分的前景和背景像素。组合的和重新加权的特征量是包含两个反卷积层的反卷积网络的输入。在FCN308(图3)的顶部，具有1x1核的卷积层充当回归器，以将特征映射到车辆密度。

[0086] 为了从连续帧结合车辆计数的时间相关性，我们组合LSTM 312(图3)和FCN 308，以联合学习车辆密度和计数。RNN保持内部隐藏状态来对序列的动态时间行为建模。LSTM通过给RNN神经元增加三个门来扩展RNN：遗忘门f_t；输入门i_t；和输出门o_t。这些门使LSTM 312(图3)能够学习序列中的长期依赖性，并使其更易于优化。LSTM 312有效地处理了RNN训练期间经常出现的梯度消失/爆炸问题。它还包含单元激活向量c_t和隐藏输出向量h_t。FCN的输出密度图被重新形成到1D向量x_t中，这个向量被馈送到三个LSTM层。在这个示例中，每个LSTM层有100个隐藏单元，并且针对5帧的窗口展开。这些门应用具有权重参数W_{hi}、W_{hf}、W_{ho}、W_{xi}、W_{xf}、和W_{xo}的sigmoid非线性σ_i、σ_f、σ_o和tanh非线性σ_c和σ_h，这些权重参数将不同的输入和门与记忆单元、输出、以及偏置b_i、b_f和b_o相连。常用的更新公式定义如下：

[0087] $$i_t=\sigma_i(x_tW_{xi}+h_{t-1}W_{hi}+w_{ci}\odot c_{t-1}+b_i)$$

[0088] $$f_t=\sigma_f(x_tW_{xf}+h_{t-1}W_{hf}+w_{cf}\odot c_{t-1}+b_f)$$

[0089] $$c_t=f_t\odot c_{t-1}+i_t\odot\sigma_c(x_tW_{xc}+h_{t-1}W_{hc}+b_c)$$

[0090] $$o_t=\sigma_o(x_tW_{xo}+h_{t-1}W_{ho}+w_{co}\odot c_t+b_o)$$

[0091] $$h_t=\sigma_t\odot\sigma_h(c_t) \quad (7)$$

[0092] 为了加速训练，FCN 308(图3)和LSTM 312以图5中(b)所示的残差学习方式连接。每个帧上学习的密度图的和500被用作基本计数504，并且LSTM 312的最后一层的输出隐藏向量508被馈送到一个完全连接的层512中，以学习基本计数和最终估计计数之间的残差。与图5中(a)所示的FCN和LSTM的直连相比，残差连接简化了训练过程并提高了计数精度。

[0093] 空间-时间多任务学习

[0094] FCN-rLSTM网络300(图3)的地面实况监督包括两种类型的信息：像素级密度图和针对每帧的全局车辆计数。地面实况监督的生成取决于对象是如何被标注的。如果每个对象的中心被标记为点d，则对于帧i的地面实况车辆计数就是标记点的总数。针对图像i中每个像素p的地面实况密度$F_i^0(p)$被定义为以覆盖像素p的每个点标注为中心的2D高斯核的和：

[0095] $$F_i^0(p)=\sum_{d\in D_i}N(p;d,\delta) \quad (8)$$

[0096] 其中，D_i是点标注的集合，d是每个标注点，高斯核的δ由透视图决定。如果每个对象由边界框$B=(x_1,y_1,x_2,y_2)$标注，其中(x_1,y_1)是左顶点的坐标，(x_2,y_2)是右底点的坐标，则对于帧i的地面实况车辆计数是帧i中边界框的总数。每个边界框B的中心o是：$o_x=1/2(x_1+x_2)$，$o_y=1/2(y_1+y_2)$。然后，图像i中每个像素p的地面实况密度$F_i^0(p)$被定义为：

[0097] $$F_i^0(p)=\sum_{o\in O_i}N(p;o,\delta) \quad (9)$$

[0098] 其中，参数O_i是帧i中边界框中心的集合。高斯核的δ由边界框的长度决定。

[0099] FCN 308(图3)的任务是估计像素级密度图，而LSTM 312的任务是估计对于每帧的全局车辆计数。这两项任务是通过端到端地训练整个FCN-rLSTM网络来联合完成的。车辆密

度通过FCN 308最后的卷积1x1层从特征图预测。欧几里德距离被用来度量估计密度和地面实况之间的差异。密度图估计的损失函数定义如下：

[0100] $$L_D = \frac{1}{2N}\sum_{i=1}^{N}\sum_{p=1}^{P}\left\|F_i(p;\Theta) - F_i^0(p))\right\|_2^2 \quad (10)$$

[0101] 其中，N是批大小(batch size)，$F_i(p)$是图像i中像素p的估计的车辆密度，Θ是FCN的参数。第二个任务(全局计数回归)从包括两个部分的LSTM层学习：(i)基本计数：整个图像上密度图的积分；和(ii)残差计数：通过LSTM层学习。将两者求和得到估计的车辆计数：

[0102] $$C_i = G(F_i;\Gamma,\Phi) + \sum_{p=1}^{P} F_i(p) \quad (11)$$

[0103] 其中，$G(F_i;\Gamma,\Phi)$是估计的残差计数，F_i是对于帧i的估计的密度图，Γ是LSTM的可学习参数，而是Φ是完全连接层的可学习参数。假设优化残差映射比优化原始映射更容易。全局计数估计的损失是：

[0104] $$L_C = \frac{1}{2N}\sum_{i=1}^{N}(C_i - C_i^0)^2 \quad (12)$$

[0105] 其中，C_i^0是帧i的地面实况车辆计数，C_i是帧i的估计计数。那么网络的总体损失函数定义为：

[0106] $L = L_D + \lambda L_C$ (13)

[0107] 其中，λ是全局计数损失的权重，应当对其进行调整以达到最佳精度。通过同时学习两个相关的任务，每个任务可以用更少的参数得到更好的训练。

[0108] 损失函数通过基于批的Adam优化器进行优化，尽管也可以使用其他优化器。下面的算法1概述了FCN-rLSTM训练过程的示例。由于FCN-rLSTM能够适应不同的输入图像分辨率以及车辆规模和视角的变化，因此对不同场景具有鲁棒性。

[0109]

算法1：FCN-rLSTM训练算法

输入： 图像：$\{I_{11},...,I_{nm}\}$，其中n是序列的数量，并且m是展开帧的数量

标签： 密度图：$\{F^0_{11},...,F^0_{nm}\}$

输出： FCN、LSTM和FC的参数：Θ,Γ,Φ

for $i = 1$ *to max_iteration* **do**
  for $j = 1$ *to unroll_number* **do**
    $F_{ij} = \text{FCN}(I_{ij};\Theta)$
    $L_{Dj} = L_2(F_{ij}, F^0_{ij})$
    $C_{\text{residual}} = \text{FC}(\text{LSTM}(F_{ij};\Gamma);\Phi)$
    $C_{ij} = \sum F_{ij} + C_{\text{residual}}$
    $L_{Cj} = L_2(\sum F^0_{ij}, C_{ij})$
  end
  $L = \sum L_{Dj} + \lambda \sum L_{Cj}$
  $\Theta,\Gamma,\Phi \leftarrow \text{Adam}(L,\Theta,\Gamma,\Phi)$
end

[0110] 使用FCN-rLSTM的示例结果

[0111] 本节讨论实验和定量结果。首先，对上述示例FCN-rLSTM方法进行了评估，并与公共数据集Web-CamT上的当前技术方法进行了比较。2.接下来，在公共数据集TRANCOS上评估了示例FCN-rLSTM方法。最后，为了验证FCN-rLSTM模型的鲁棒性和泛化(generation)能力，在公众人群计数数据集UCSD上对FCN-rLSTM方法进行了评估。以下描述了这些实验中的每一个。

[0112] WebCamT上的定量评估

[0113] WebCamT是对于大规模城市摄像机视频的公共数据集，其具有低分辨率(352x240)、低帧率(1帧/秒)和高遮挡。边界框和车辆计数都适用于60,000帧。数据集被划分为训练集和测试集，分别为45,850帧和14,150帧，涵盖多个摄像机和不同的天气条件。

[0114] 在WebCamT的14,150个测试帧上评估上述示例FCN-rLSTM方法，其包含来自8个摄像机的61个视频。这些视频涵盖不同的场景、拥堵状态、摄像机视角、天气条件和一天中的时间。训练集包含45,850帧，分辨率相同，但来自不同的视频。训练集和测试集都被分成两组：市区摄像机和绿化路摄像机。平均绝对误差(MAE)用于进行评估。对于FCN-rLSTM，车辆计数损失的权重为0.01。学习速率由0.0001初始化，并由训练过程中的一阶和二阶动量进行调整。为了测试建议的超空洞组合、FCN和LSTM以及残差连接的组合的功效，如下表1所示，评估了FCN-rLSTM的不同配置。在表1中："空洞"表示空洞卷积；"超级"表示特征图的超列组合；"直接连接"表示FCN与LSTM的直接组合；"残差连接"表示以残差方式连接FCN和LSTM。

[0115] 表1:FCN-rLSTM的不同配置

[0116]

配置	空洞	超级	直接连接	残差连接
FCN-A	√	X	X	X

FCN-A	X	√	X	X
FCN-HA	√	√	X	X
FCN-dLSTM	√	√	√	X
FCN-rLSTM	√	√	X	√

[0117] 数据增强:为了使模型对各种摄像机和天气条件更加鲁棒,对训练图像应用了几种数据增强技术:1)水平翻转;2)随机裁切;3)随机亮度;和4)随机对比度。注意,也可以应用其他数据增强技术。

[0118] 基线方法:将本示例的FCN-rLSTM方法与三种方法进行了比较:基线1:学会计数(Learning to count)(A.Z.V.Lempitsky在2010年的ACM的Advances in Neural Information Processing Systems(NIPS)中的“Learning to count objects in images”)。这项工作将每个像素的特征映射到针对整个图像具有均匀权重的对象密度中。为了比较,使用VLFeat(A.Vedaldi和B.Fulkerson,“VLFeat:An open and portable library of computer vision algorithms”,2008年)针对每个像素提取密集SIFT特征,并学习视觉单词。基线2:Hydra(D.Onoro-Rubio和R.J.Lopez-Sastre于2016年在Springer的European Conference on Computer Vision,第615-629页中的“Towards perspective-free object counting with deep learning”)。它学习多规模非线性回归模型,该模型使用在多个规模上提取的图像块的金字塔来执行最终密度预测。Hydra 3s模型是在与FCN-rLSTM相同的训练集上训练的。基线3:FCN(S.Zhang、G.Wu、J.P.Costeira和J.M.F.Moura于2017年在IEEE的Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,第5898-5907页中的“Understanding traffic density from large-scale web camera data”)。其基于FCN开发了一个深度多任务模型来联合估计车辆密度和车辆计数。FCN接受了在与FCN-rLSTM相同的训练集上的训练。

[0119] 实验结果:以下表2比较了建议方法和基线方法的误差。从结果可以看出,示例FCN-rLSTM方法优于所有基线方法和所有其他配置。由于测试数据涵盖不同的拥堵状态、摄像机视角、天气条件和一天中的时间,这些结果验证了FCN-rLSTM的有效性和鲁棒性。为了对建议的技术进行消融分析,还评估了表2所示不同配置的性能。

[0120] 表2:WebCamT上的结果比较

[0121]

方法	市区	绿化路
基线1	5.91	5.19
基线2	3.55	3.64
基线3	2.74	2.52
FCN-A	3.07	2.75
FCN-H	2.48	2.30
FCN-HA	2.04	2.04
FCN-dLSTM	1.80	1.82
FCN-rLSTM	1.53	1.63

[0122] 使用超空洞组合,FCN-HA本身已经优于所有基线方法,并获得了比FCN-A和FCN-H更好的精确度,这验证了超空洞组合的有效性。FCN-rLSTM比FCN-HA和FCN-dLSTM(直接连接)的准确率高,这验证了FCN和LSTM的残差连接的功效。从FCN-HA和FCN-rLSTM相对于绿化

路摄像机和市区摄像机的计数结果比较(未示出)的评估中,可以得出结论,FCN-rLSTM更好地估计了车辆计数,并减少了由过大车辆引起的大计数误差。在没有前景分割的情况下,所学习的密度图仍然可以在晴天、雨天和多云的、密集和稀疏场景中区分背景和前景。

[0123] 除了FCN-rLSTM实现的高精确度外,建议方法的收敛性也显著提高。FCN和LSTM的残差连接也比直接连接实现了更快的收敛。

[0124] TRANCOS上的定量评估

[0125] 还在公共数据集TRANCOS上评估示例FCN-rLSTM方法,以验证其功效。TRANCOS是监控摄像机视频中不同交通场景的1,244幅图像的集合。它总共有46,796辆标注的车辆,并为每幅图像提供一个感兴趣区域(ROI)。TRANCOS的图像来自非常不同的场景,并且没有提供透视图。地面实况车辆密度图由每个标注车辆中心的2D高斯核生成。

[0126] 表3:TRANCOS数据集上的结果比较

[0127]

方法	MAE	方法	MAE
基线1	13.76	基线3	5.31
基线2-CCNN	12.49	FCN-HA	4.21
基线2-Hydra	10.99	FCN-rLSTM	4.38

[0128] 建议方法和基线方法的MAE在上面的表3中进行了比较。基线2-CCNN是上面引用的Onoro-Rubio等人的Hydra参考文献中的网络的基本版本,并且基线2-Hydra通过利用图像块金字塔学习多规模回归模型来增强性能,以执行最终的密度预测。在上面引用的Onero-Rubio等人的Hydra参考文件中,所有的基线方法和建议方法都在823幅图像上进行了训练并在分离之后的421帧图像上进行了测试。从结果可以看出,与基线2-Hydra相比,FCN-HA将MAE从10.99显著降至4.21,并且与基线3相比,将MAE从5.31显著降至4.21。由于TRANCOS的训练和测试图像是来自不同摄像机和视频的随机样本,因此它们缺乏一致的时间信息。因此,FCN-rLSTM无法从训练数据中学习时间模式。FCN-rLSTM的性能不如FCN-HA那么好,但它已经超过了所有的基线方法。当将当前示例FCN-rLSTM方法应用于其他数据集时,可以为具有时间相关性的数据集选择FCN-rLSTM配置,并且可以为不具有时间相关性的数据集选择FCN-HA配置。示例FCN-rLSTM方法的估计数显然比基线方法更精确。FCN-rLSTM和FCN-HA对车辆计数的估计精度相当。

[0129] UCSD数据集的定量评价

[0130] 为了验证当前示例FCN-rLSTM方法在不同计数任务中的通用性和鲁棒性,FCN-rLSTM方法被评估,并与行人计数数据集UCSD上的基线进行比较。该数据集包含从一台监控摄像机中选择的2,000帧。帧大小为158x 238,帧速率为10fps。每帧的平均人数约为25人。数据集为每个视频帧提供ROI。601到1,400帧用作训练数据,其余1,200帧用作测试数据。以下表4示出了当前示例FCN-rLSTM方法和现有方法的结果,从中可以看出,FCN-rLSTM优于所有基线方法和FCN-HA配置。这些结果表明,与其他类型的计数任务相比,当前示例FCN-rLSTM方法是鲁棒的。

[0131] 表4:UCSD数据集上的结果比较

[0132]

方法	MAE	MSE
核岭回归	2.16	7.45
岭回归	2.25	7.82

高斯处理回归	2.24	7.97
累积属性回归	2.07	6.86
跨场景DNN	1.6	3.31
基线3	1.67	3.41
FCN-HA	1.65	3.37
FCN-rLSTM	1.54	3.02

[0133] 估计算法的训练

[0134] 训练本公开的估计算法(如图1的估计系统112的估计算法120)由于许多因素而极具挑战性,包括高度可变的照明条件、高度可变的天气条件、遮挡高发率以及被计数的对象类型的高度可变性。使挑战更加复杂的事实在于,需要对训练图像进行标注,以便创建地面实况图像用于训练。一般来说,由于人类准确地标注图像的能力,人类标注是优选的。然而,由于上述挑战性因素的巨大可变性,需要大量的标注图像,因此人工标注将是极其昂贵的,尤其是当本公开的估计系统被大规模部署时,如在城市或区域的交通摄像机的整个网络中,或者甚至在整个网络的一部分中,其中数十个、数百个或更多个摄像机在整个道路网络中提供数十个、数百个或更多个位置的图像。因此,需要成本较低的方法来创建用于训练本公开的估计算法的图像。本节中描述了两种这样的方法,即合成图像方法和混合图像方法。合成图像方法允许创建训练图1的估计系统112的算法120所需的更精确的密度图。混合图像方法允许无成本创建合成标注图像。

[0135] 合成训练图像生成

[0136] 现在参考图6和图7,图7示出了生成像素级掩模604(图6)的集合的示例方法700,并且图6示出了示例合成训练图像生成系统600。(应当注意,600系列数字指的是图6中的元件,700系列数字指的是图7中的元件。)在块705处,接收训练真实图像608的集合。在该示例中,训练真实图像608是由摄像机捕获的实际图像,该摄像机是当使用方法700进行训练的估计系统(如图1的估计系统112)被部署时将提供实时图像以供估计系统分析的摄像机。

[0137] 在块710处,每个训练真实图像608由人类标注员用一个或更多个矩形(如矩形612A)手动标记(为了方便起见,仅标记了一对),每个矩形表示最小限度地限定图像内相应真实对象(这里指车辆)的边界框。这种手动标注创建标注训练图像612的集合。在块715处,真实的标注训练图像612被用于训练基于ANN的检测器616,该检测器616在运行时间期间将利用像素级掩模604自动标注真实图像620。

[0138] 在块720处,伪对象624(这里指车辆)的集合及其背景掩模628被合成。在块725处,使用伪对象624、背景掩模628和从标注训练图像612提取的被提取的边界框区域632的集合来训练基于ANN的分割器636。一旦检测器616和分割器636已经被训练,自动生成逐像素掩模604的过程可以开始。

[0139] 在块730处,在训练之后,检测器616在真实图像620内发现对象(这里指车辆),以生成相应的自动标注图像640,自动标注图像640被标注为包括边界框640A,边界框640A界定检测器检测到的相应的各个对象。在块735处,分割器636为每个自动标注图像640分割出对于每个边界框640A的背景,以生成相应的边界框掩模644。在块740处,边界框掩模644在对应于自动标注图像的边界框640A的位置处被插入到背景掩模648,以创建像素级掩模604。此时,像素级掩模604可用于生成密度图,用于训练任何合适的估计算法,如图1的估计

算法120。这种密度图可能比那些从标注的矩形生成的密度图更精确。本领域技术人员将容易理解，一旦检测器616和分割器636已经被适当地训练，它们可以被用于为用矩形标注的任意数量的图像生成许多像素级车辆掩模，而几乎没有任何成本，除了为用于执行方法700的计算系统供电的成本。

[0140] 混合训练图像生成

[0141] 图8示出了生成混合训练图像的视频的方法800。在块805处，从拍摄自真实场景的输入视频中去除前景。在场景是道路的车辆交通实现的环境中，该去除产生空道路的视频。(为了方便，从这一点开始，方法800的描述基于车辆交通实现。然而，本领域技术人员将容易理解方法800如何能够适用于其他实现。)要进行去除，首先检测视频中的前景。前景包括汽车、它们的阴影和一些图像伪像。在简单的情形中，可以使用高斯混合模型(GMM)，该模型将每个像素独立地建模为属于背景或前景。其次，使用下面的公式7逐帧生成背景视频。每个新帧的非掩模区域用更新速率η更新背景：

[0142]
$$\begin{aligned} 背景[t] = & 背景[t-1] \\ & -\eta\left(1-掩模[t]\right)\otimes 背部[t-1] \\ & +\eta\left(1-掩模[t]\right)\otimes 帧[t] \end{aligned} \tag{7}$$

[0143] 在块810处，识别真实场景的3D几何形状。由于目标是将车辆再现为如从特定的观看点来观看一样，因此有必要首先学习作为估计系统的实施方式一部分的对于每个摄像机的场景3D几何形状。场景3D几何形状包括摄像机外在参数(位置和旋转)和固有参数。在透视摄像机模型下，固有参数包括焦距、中心偏移、偏斜和非线性失真。对于许多监控摄像机，可以假设存在零中心偏移、无偏斜或非线性失真。为了进一步简化任务，地面被认为是平坦的。在块815处，道路车道被标注以创建交通模型。

[0144] 对于原始视频的每一帧，逐帧生成一个合成视频。合成混合帧包括四个步骤。在块820处，车辆的3D模型被放置在靠近车道中心的帧中。假设车辆不改变车道，汽车之间的距离可以自动选择，并且可以从高(稀疏交通)到低(对应于交通堵塞)变化。根据在块815处创建的交通模型，在场景中布置多个随机3D CAD车辆模型。场景中的照明条件是根据源视频中的天气设置的。在一个示例中，天气被建模为以下类型之一：晴天、多云、潮湿和下雨。这些类型反映了场景获得的照明量、阴影以及在视频中体验到的地面外观。对于潮湿和下雨的天气，地面被制成是反射性的。在晴天时，太阳的角度基于录制视频帧时太阳的真实位置来设定。附加地，周围形状的阴影投射在汽车上。

[0145] 在块825处，车辆和它们投射在道路上的阴影被再现在透明背景上。为了数据增强的目的，如果需要，可以改变再现图像的饱和度和亮度。场景深度图也可以被再现并用于推断场景中车辆之间的相互遮挡。在一个示例中，如果车辆被遮挡超过50％，则从训练图像中将其省略。

[0146] 在块830处，车辆与背景混合。图像饱和度和亮度可以变化，以增加数据的可变性。再现的图像覆盖在帧背景之上，车辆阴影被模糊以隐藏再现引擎引入的噪声。在块835处，为了模仿真实视频中的伪影，再现的车辆可以在将其覆盖在背景上之前可选地被锐化，并且最终图像可以可选地用高压缩的JPEG算法重新编码。一旦混合合成视频完成，它可以用

于训练合适的估计算法，如图1的估计算法120。

[0147] 利用对抗学习的多摄像机域自适应

[0148] 目前，许多城市都被数百台交通摄像机监控。以纽约市(NYC)为例，目前NYC安装了564台网络摄像机。不同的摄像机有不同的场景、背景和视角。即使是同一台摄像机，视角、天气和光照也会随着时间而改变。所有这些因素导致不同的数据域。然而，很难为具有不同数据域的所有这些摄像机标记大量的训练图像。在当前环境中，具有充足标记图像的摄像机被称为“源摄像机”，而没有标记信息的摄像机被称为“目标摄像机”。这里的一个关键问题是如何使在源摄像机上训练的深度模型适应目标摄像机。由于可能有数十个、数百个或更多的摄像机要处理，导致多个源摄像机域和目标域，本发明人已经开发了利用对抗学习方法的多摄像机域自适应(MDA)，以使在多个源摄像机上训练的深度神经网络适应不同的目标摄像机。

[0149] MDA方法学习特征，这些特征对源摄像机上的主要学习任务来说是区别性的，而当在源摄像机和目标摄像机之间迁移时是不加区分的。图11示出了示例MDA系统1100和方法的主要特征。MDA系统包括深度特征提取器1104和深度标签预测器1108。无监督的多摄像机域自适应是通过添加经由梯度反转层1116连接到特征提取器1104的多域分类器1112来实现的，梯度反转层1116在基于反向传播的训练期间将梯度乘以某个负常数。否则，训练将以标准方式进行，并且最小化标签预测损失(对于源示例)和域分类损失(对于所有示例)。梯度反转确保所学习的特征分布在不同的域上是相似的(对于域分类器1112来说尽可能不可区分)，从而导致域不变的特征。

[0150] 通过使用几个标准层和一个新的梯度反转层进行增强，MDA方法可以在几乎任何前馈模型中实现。所得到的增强的体系结构可以使用标准反向传播和随机梯度下降来训练，并且因此可以容易地集成到例如本文公开的FCN-MT和FCN-rLSTM方法中。由于大部分现有的域自适应工作集中于对于深度神经网络的分类和单个源/目标域自适应，因此所提出的MDA方法可能是第一次尝试使来自多个源域的完全卷积网络适应不同的目标域以用于回归任务。

[0151] 本发明人从理论上分析了多摄像机域自适应问题，并基于理论结果开发了对抗学习策略。具体地说，当存在具有标记实例的多个源域和具有未标记实例的一个目标域时，证明了新的泛化界限(generalization bounds)用于域自适应。从技术上讲，界限是通过首先提出来自多个域的两个分布集合之间的广义散度度量来推导的。通过使用集中不等式和Vapnik-Chervonenkis(VC)理论中的工具将目标风险与经验源风险相结合，证明了目标风险的可能近似正确(PAC)界限。与现有界限相比，新界限不需要关于目标域分布的专家知识，也不需要针对多个源域的最佳组合规则。获得的结果也暗示，天真地将更多的源域纳入训练并不总是有益的，发明人在他们的实验中证实了这一点。

[0152] 该界限导致使用对抗神经网络的多摄像机域自适应(MDA)的有效的实现。MDA方法使用神经网络作为丰富的函数逼近器来实例化导出的泛化界限。经过适当的转换后，MDA方法可以被视为泛化界限的计算效率近似，因此目标是优化网络参数以最小化界限。本发明人目前已经开发了两种MDA方法：直接优化最坏情况泛化界限的硬版本(Hard-Max MDA)和作为硬版本平滑近似的软版本(Soft-Max MDA)，导致更高数据效率的模型并优化任务自适应界限。本公开MDA方法的优化是极大极小鞍点问题，其可以被解释为两个参与者相互竞争

以学习不变特征的零和游戏。MDA方法将特征提取、域分类和任务学习组合在一个训练过程中。同时更新的随机优化用于优化每次迭代中的参数。用于实施MDA的示例网络体系结构在图12中示出。

[0153] 以下是使用多域对抗网络从多个摄像机对对象(如车辆)进行计数的的示例机制。如上所述，具有大量训练图像的摄像机被视为源摄像机，而没有标记数据的摄像机被视为目标摄像机。研究了不同源摄像机与目标摄像机之间的关系，并根据与目标摄像机的距离对源摄像机进行排序。最初的k个摄像机被选择来形成k个源域。因此，可在不同数量的源上评估MDA方法。Hard-Max MDA和Soft-Max MDA都是根据下面的示例算法2、基于基本车辆计数FCN来实施的。记录真实计数和估计计数之间的平均绝对误差(MAE)。然后，针对不同的源组合比较MAE，选择MAE最低的组合作为对于目标摄像机的训练集合。在一个示例中，MDA和这个源选择机制可以被应用于使用多个摄像机来对车辆计数。

[0154]

算法2 多源域自适应

for $t = 1$ to ∞ **do**
  Sample $\{S_i^{(t)}\}_{i=1}^k$ and $T^{(t)}$ from $\{\widehat{\mathcal{D}}_{S_i}\}_{i=1}^k$ and $\widehat{\mathcal{D}}_T$, each of size m
  for $i = 1$ to k **do**
    $\widehat{\varepsilon}_i^{(t)} \leftarrow \widehat{\varepsilon}_{S_i^{(t)}}(h) - \min_{h' \in \mathcal{H}\Delta\mathcal{H}} \widehat{\varepsilon}_{T^{(t)}, S_i^{(t)}}(h')$
    Compute $w_i^{(t)} := \exp(\widehat{\varepsilon}_i^{(t)})$
  end for
  # 硬版本
  Select $i^{(t)} := \arg\max_{i \in [k]} \widehat{\varepsilon}_i^{(t)}$
  Backpropagate gradient of $\widehat{\varepsilon}_{i^{(t)}}^{(t)}$
  # 平滑版本
  for $i = 1$ to k **do**
    Normalize $w_i^{(t)} \leftarrow w_i^{(t)} / \sum_{i' \in [k]} w_{i'}^{(t)}$
  end for
  Backpropagate gradient of $\sum_{i \in [k]} w_i^{(t)} \widehat{\varepsilon}_i^{(t)}$
end for

[0155] 示例计算系统

[0156] 应当注意，本文描述的一个或更多个方面和方法实施例可以在根据本说明书的教导编程的一个或更多个机器中和/或使用根据本说明书的教导编程的一个或更多个机器(例如，一个或更多个计算机、一个或更多个通信网络设备、一个或更多个配电网络设备、其任意组合和/或网络等)方便地实施，这对计算机领域的普通技术人员来说是明显的。对于软件领域的普通技术人员来说明显的是，基于本公开的教导，熟练的程序员可以容易地准备适当的软件编码。上面讨论的采用软件和/或软件模块的方面和实施方式也可以包括适当的硬件，用于帮助实现软件和/或软件模块的机器可执行指令。

[0157] 这种软件可以是采用机器可读存储介质的计算机程序产品。机器可读存储介质可以是能够存储和/或编码供机器(例如，计算设备)执行的指令序列并且使得机器执行本文描述的方法和/或实施例中的任何一个的任何介质。机器可读存储介质的示例包括但不限

于磁盘、光盘(例如,CD、CD-R、DVD、DVD-R等)、磁光盘、只读存储器“ROM”设备、随机存取存储器“RAM”设备、磁卡、光卡、固态存储器设备、EPROM、EEPROM及其任意组合。本文使用的机器可读介质旨在包括单个介质以及物理上分离的介质的集合,例如,紧凑光盘的集合或者与计算机存储器组合的一个或更多个硬盘驱动器。如本文所使用的,机器可读存储介质不包括信号传输的暂时形式。

[0158] 这种软件还可以包括作为数据信号在数据载体(如载波)上承载的信息(例如,数据)。例如,机器可执行信息可以作为体现在数据载体中的数据承载信号被包括,在该数据信号中,信号编码用于供机器(例如,计算设备)执行的指令序列或其一部分以及使机器执行本文所描述方法和/或实施例中的任一个的任何相关信息(例如,数据结构和数据)。

[0159] 计算设备的示例包括但不限于膝上型计算机、计算机工作站、终端计算机、服务器计算机、手持设备(例如,平板计算机、智能手机等)、网络设备、网络路由器、网络交换机、网桥、能够执行指定机器要采取的动作的指令序列的任何机器、及其任意组合。在一个示例中,计算设备可以包括和/或被包括在信息亭中。

[0160] 图9示出了计算机系统900示例形式中计算设备的一个实施例的图示性表示,其可以在该计算机系统900内执行指令集,用于执行本公开的任何一个或更多个方面和/或方法。还可以设想,多个计算设备可以用于实现被特别配置的指令集,用于使一个或更多个设备包含和/或执行本公开的任何一个或更多个方面和/或方法。计算机系统900包括处理器904和存储器908,它们经由总线912相互通信并与其他部件通信。总线912可以包括几种类型的总线结构中的任何一种,包括但不限于使用多种总线体系结构中的任何一种的存储器总线、存储器控制器、外围总线、本地总线及其任意组合。

[0161] 存储器908可以包括各种部件(例如,机器可读介质),包括但不限于随机存取存储器部件、只读部件及其任意组合。在一个示例中,基本输入/输出系统916(BIOS)可以存储在存储器908中,该基本输入/输出系统916包括诸如在启动期间帮助在计算机系统900内的元件之间转移信息的基本例程。存储器908还可以包括(例如,存储在一个或更多个机器可读介质上)体现本公开的任何一个或更多个方面和/或方法的指令(例如,软件)920。在另一个示例中,存储器908还可以包括任何数量的软件类型,包括但不限于操作系统、一个或更多个应用程序、其他程序模块、程序数据及其任意组合。

[0162] 计算机系统900还可以包括储存设备924。储存设备(例如,储存设备924)的示例包括但不限于硬盘驱动器、磁盘驱动器、与光学介质结合的光盘驱动器、固态存储器设备及其任意组合。储存设备924可以通过适当的接口(未示出)连接到总线912。示例接口包括但不限于SCSI、高级技术附件A(TA)、串行ATA、通用串行总线(USB)、IEEE 994(FIREWIRE)及其任意组合。在一个示例中,储存设备924(或其一个或更多个部件)可以可移除地与计算机系统900相接(例如,经由外部端口连接器(未示出))。特别地,储存设备924和相关联的机器可读介质928可以为机器可读指令、数据结构、程序模块和/或计算机系统900的其他数据提供非易失性和/或易失性存储。在一个示例中,软件920可以完全或部分驻留在机器可读介质928内。在另一个示例中,软件920可以完全或部分驻留在处理器904内。

[0163] 计算机系统900还可包括输入设备932。在一个示例中,计算机系统900的用户可以经由输入设备932将命令和/或其他信息输入到计算机系统900中。输入设备932的示例包括但不限于字母数字输入设备(例如键盘)、定点设备、操纵杆、游戏手柄、音频输入设备(例如

麦克风、语音响应系统等)、光标控制设备(例如,鼠标)、触摸板、光学扫描仪、视频捕获设备(例如,静态摄像机、视频摄像机)、触摸屏及其任意组合。输入设备932可以经由各种接口(未示出)中的任何一种接口连接到总线912,这些接口包括但不限于串行接口、并行接口、游戏端口、USB接口、FIREWIRE接口、到总线912的直接接口及其任意组合。输入设备932可以包括触摸屏接口,其可以是显示器936的一部分或者与显示器936分离,这将在下面进一步讨论。输入设备932可以用作用户选择设备,用于在如上所述的图形接口中选择一个或更多个图形表示。

[0164] 用户还可以经由储存设备924(例如,可移除盘驱动器、闪存驱动器等)和/或网络接口设备940向计算机系统900输入命令和/或其他信息。诸如网络接口设备940的网络接口设备可以用于将计算机系统900连接到一个或更多个各种网络(如网络944),和连接到这些网络的一个或更多个远程设备948。网络接口设备的示例包括但不限于网络接口卡(例如,移动网络接口卡、LAN卡)、调制解调器、及其任意组合。网络的示例包括但不限于广域网(例如,因特网、企业网络)、局域网(例如,与办公室、建筑物、校园或其他相对较小的地理空间相关联的网络)、电话网络、与电话/语音提供商相关联的数据网络(例如,移动通信提供商数据和/或语音网络)、两个计算设备之间的直接连接、及其任意组合。诸如网络944的网络可以采用有线和/或无线通信模式。通常,可以使用任何网络拓扑结构。信息(例如,数据、软件920等)可以经由网络接口设备940传送至计算机系统900和/或从计算机系统900传送。

[0165] 计算机系统900还可以包括视频显示适配器952,用于将可显示图像传送到显示设备,如显示设备936。显示设备的示例包括但不限于液晶显示器(LCD)、阴极射线管(CRT)、等离子显示器、发光二极管(LED)显示器及其任意组合。显示适配器952和显示设备936可以与处理器904结合使用,以提供本公开各方面的图形表示。除了显示设备,计算机系统900还可以包括一个或更多个其他外围输出设备,包括但不限于音频扬声器、打印机及其任意组合。这种外围输出设备可以经由外围接口956连接到总线912。外围接口的示例包括但不限于串行端口、USB连接、FIREWIRE连接、并行连接及其任意组合。

[0166] 前文是对本发明的说明性实施例的详细描述。注意,在本说明书及其所附权利要求中,除非另有明确说明或指示,诸如在短语“X、Y和Z中的至少一个”和“X、Y和Z中的一个或更多个”中使用的连接语言应该被理解为意味着连接列表中的每一项可以表示为排除该列表中的所有其他项的任何数目或者表示为结合连接列表中的任何或所有其他项的任何数目,每个项也可以表示为任何数目。应用这个一般规则,前述示例中的连接短语(其中连接列表由X、Y和Z组成)应各自涵盖:X中的一个更多个;Y中的一个或更多个;Z中的一个或更多个;X中的一个或更多个和Y中的一个或更多个;Y中的一个或更多个和Z中的一个或更多个;X中的一个或更多个和Z中的一个或更多个;以及X中的一个或更多个、Y中的一个或更多个和Z中的一个或更多个。

[0167] 可以在不偏离本发明的精神和范围的情况下做出各种修改和添加。上述各种实施例中的每一个实施例的特征可以适当地与其他所描述的实施例的特征组合,以便在相关联的新实施例中提供多种特征组合。此外,虽然前文描述多个分开的实施例,但是本文描述的内容仅仅是对本发明的原理的应用的说明。此外,虽然本文的特定方法可被说明和/或描述为以特定顺序执行,但是在本领域技术范围内,排序是高度可变的以实现本公开的各方面。因此,本说明书旨在仅通过示例的方式做出,而不以其他方式限制本发明的范围。

[0168] 已经在上面公开并在附图中示出示例性实施例。本领域技术人员将理解，在不背离本发明的精神和范围的情况下，可以对本文具体公开的实施例进行各种改变、省略和添加。

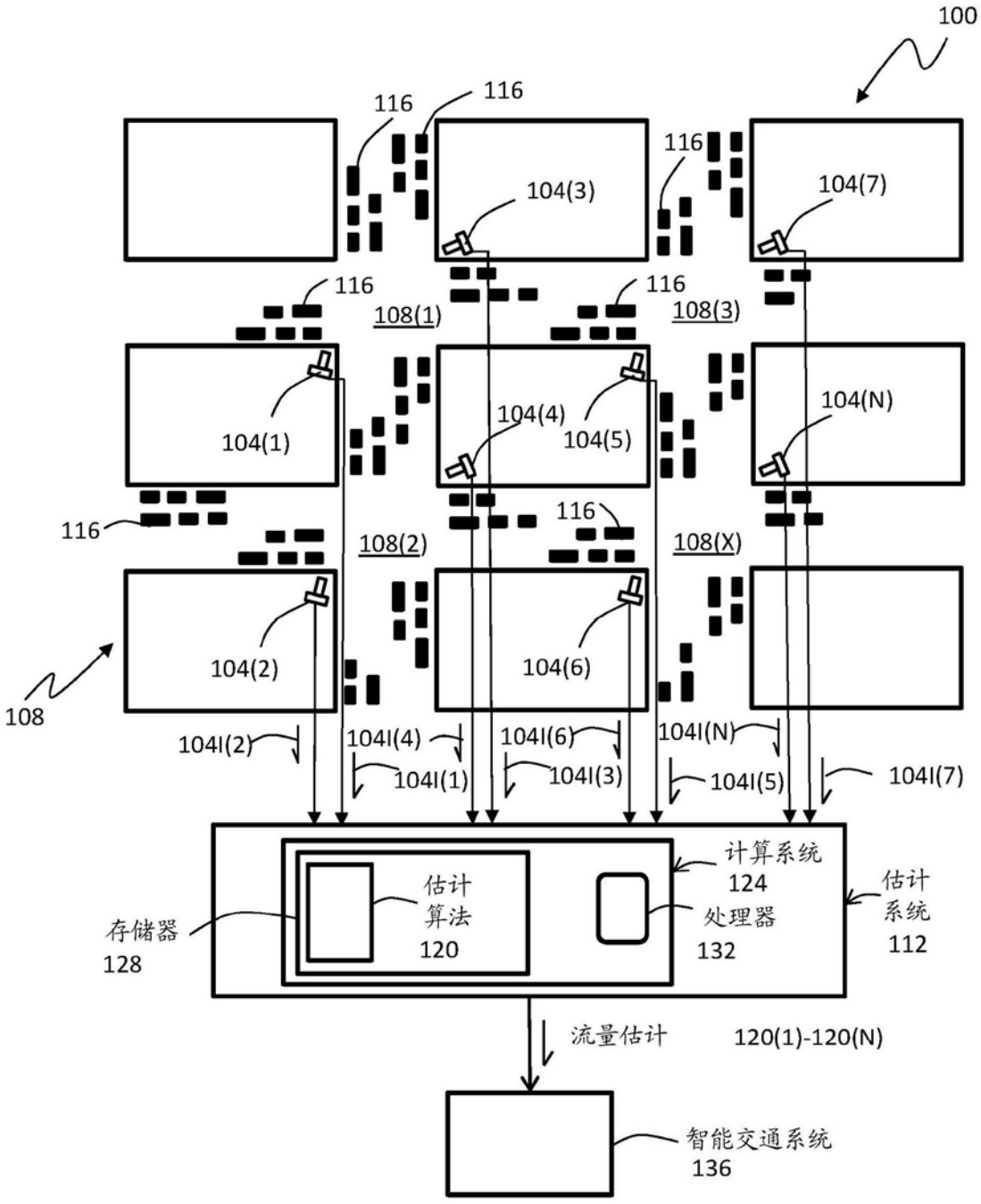

100
116
116
116
104(3)
104(7)
116
116
108(1)
108(3)
104(1)
104(4)
104(5)
104(N)
116
116
108(2)
108(X)
104(2)
104(6)
108
104I(2)
104I(4)
104I(6)
104I(N)
104I(1)
104I(3)
104I(5)
104I(7)
计算系统
124
估计
算法
120
处理器
132
估计
系统
112
存储器
128
流量估计
120(1)-120(N)
智能交通系统
136

图1

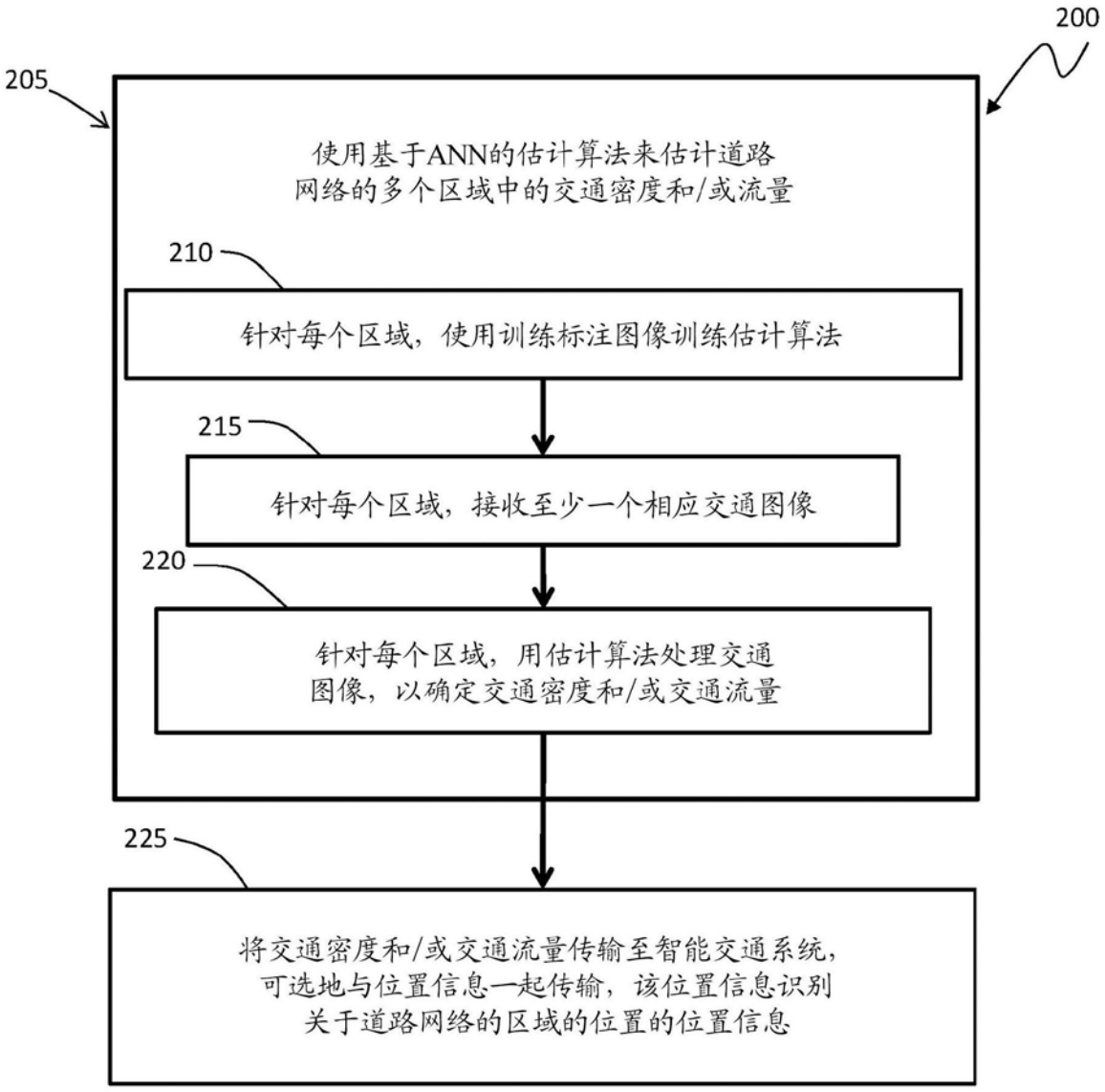

200
205
使用基于ANN的估计算法来估计道路网络的多个区域中的交通密度和/或流量
210
针对每个区域，使用训练标注图像训练估计算法
215
针对每个区域，接收至少一个相应交通图像
220
针对每个区域，用估计算法处理交通图像，以确定交通密度和/或交通流量
225
将交通密度和/或交通流量传输至智能交通系统，可选地与位置信息一起传输，该位置信息识别关于道路网络的区域的位置的位置信息

图2

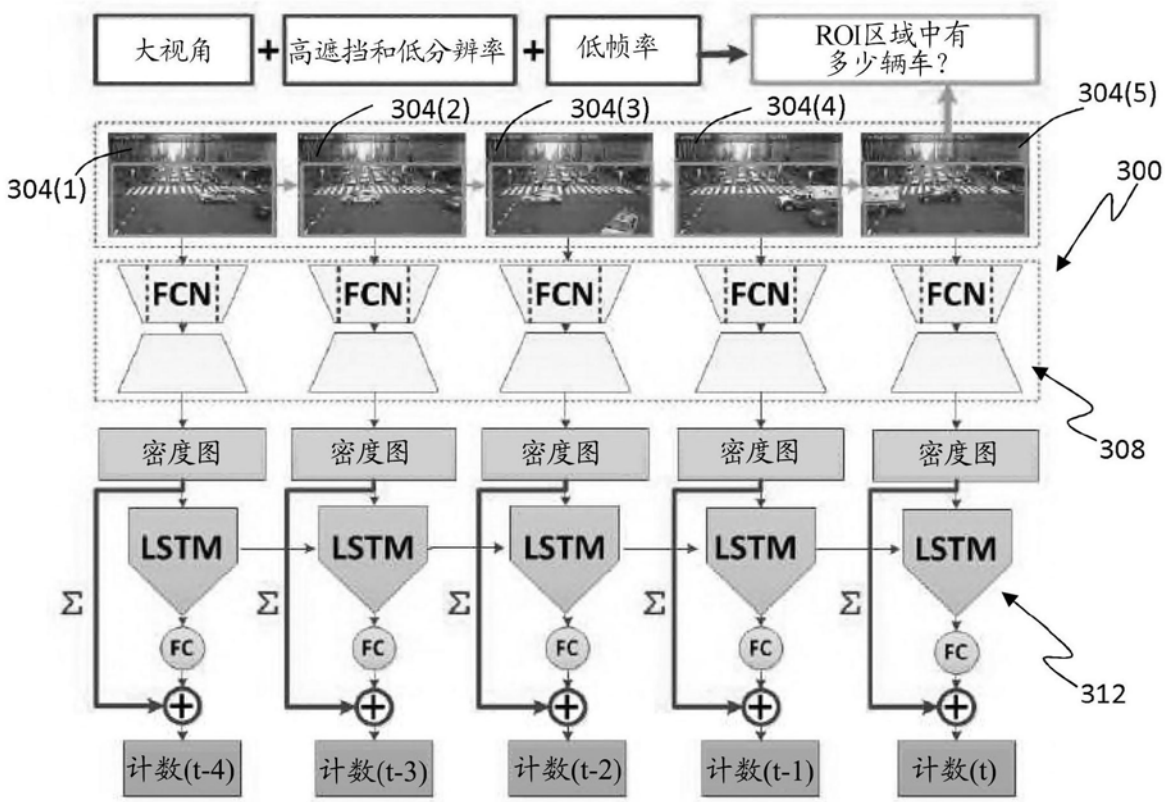

大视角
高遮挡和低分辨率
低帧率
ROI区域中有多少辆车？
304(1)
304(2)
304(3)
304(4)
304(5)
300
FCN
308
密度图
LSTM
FC
312
计数(t-4)
计数(t-3)
计数(t-2)
计数(t-1)
计数(t)

图3

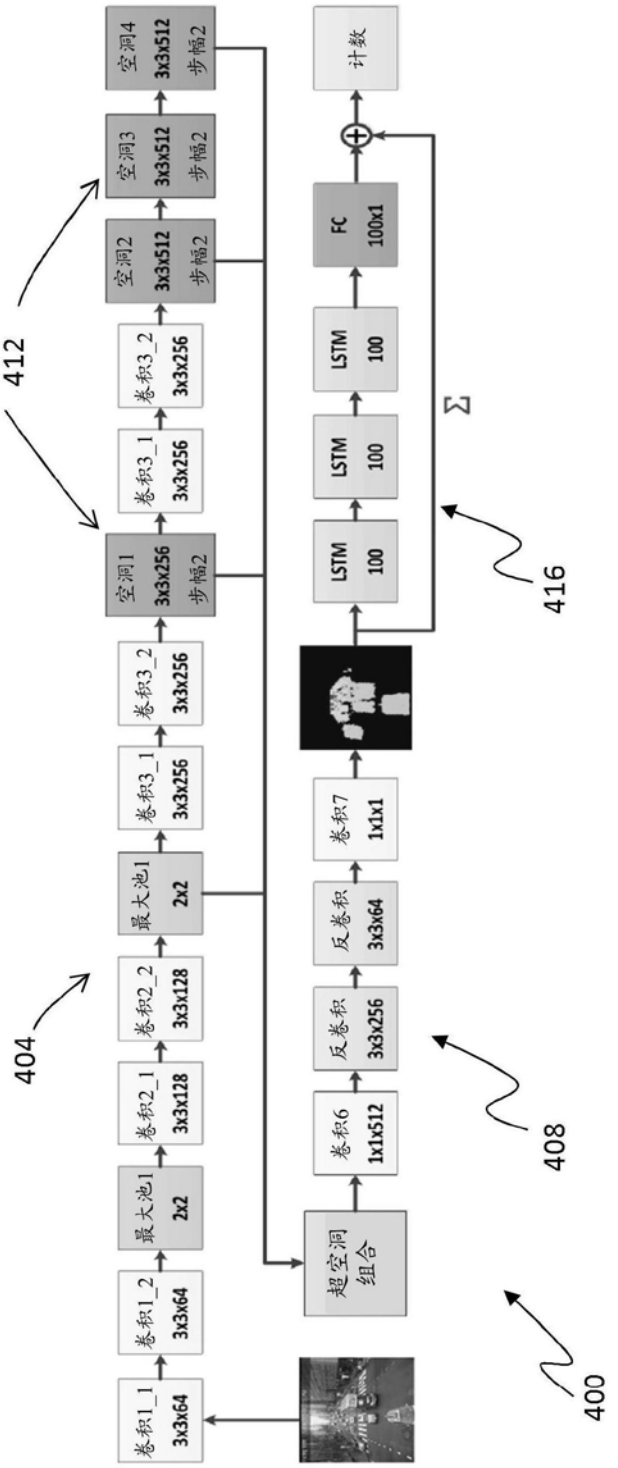

卷积1_1 3x3x64
卷积1_2 3x3x64
最大池1 2x2
卷积2_1 3x3x128
卷积2_2 3x3x128
最大池1 2x2
卷积3_1 3x3x256
卷积3_2 3x3x256
空洞1 3x3x256 步幅2
卷积3_1 3x3x256
卷积3_2 3x3x256
空洞2 3x3x512 步幅2
空洞3 3x3x512 步幅2
空洞4 3x3x512 步幅2
超空洞组合
卷积6 1x1x512
反卷积 3x3x256
反卷积 3x3x64
卷积7 1x1x1
LSTM 100
LSTM 100
LSTM 100
FC 100x1
计数
404
412
408
416
400

图4

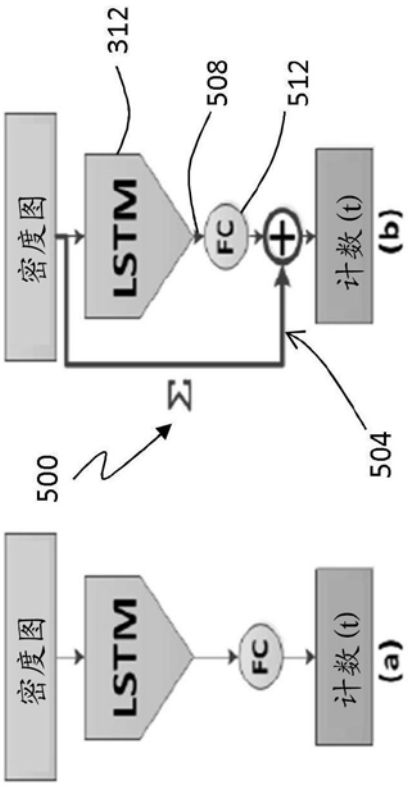


图5

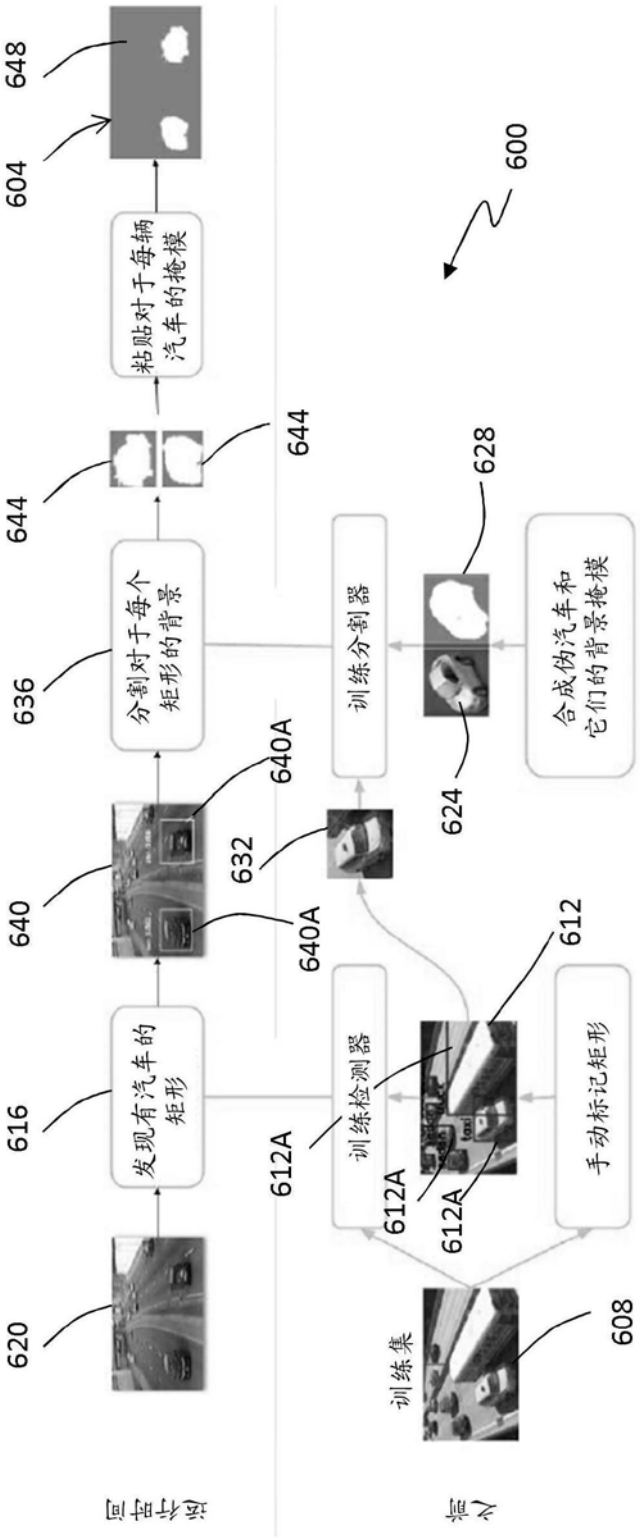

600
训练集
608
手动标记矩形
612
训练检测器
612A
612A
612A
发现有汽车的矩形
616
620
640
640A
640A
632
训练分割器
624
628
合成伪汽车和它们的背景掩模
分割对于每个矩形的背景
636
644
644
644
粘贴对于每辆汽车的掩模
604
648
分割背景
合成

图6

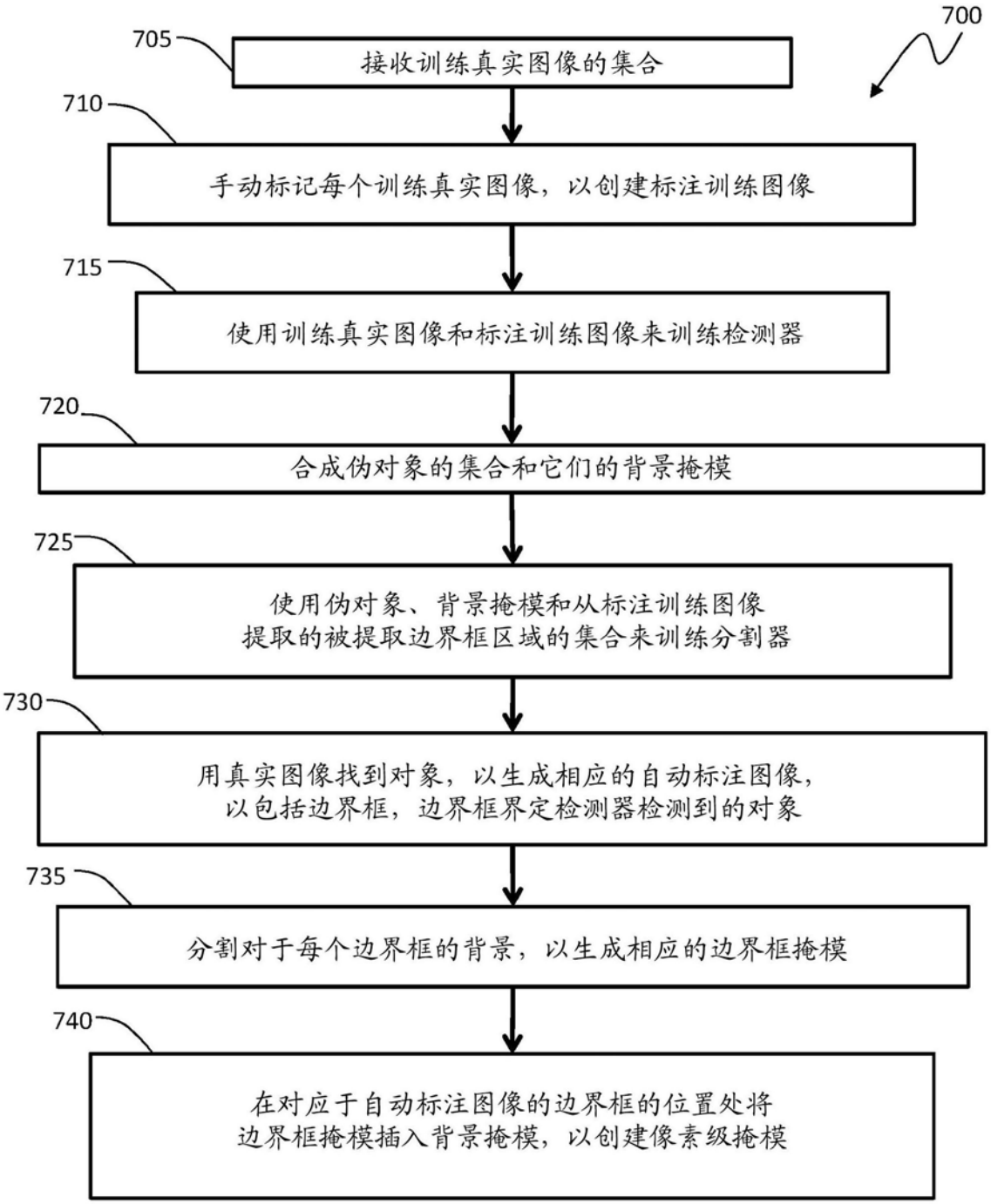

700
705
接收训练真实图像的集合
710
手动标记每个训练真实图像，以创建标注训练图像
715
使用训练真实图像和标注训练图像来训练检测器
720
合成伪对象的集合和它们的背景掩模
725
使用伪对象、背景掩模和从标注训练图像
提取的被提取边界框区域的集合来训练分割器
730
用真实图像找到对象，以生成相应的自动标注图像，
以包括边界框，边界框界定检测器检测到的对象
735
分割对于每个边界框的背景，以生成相应的边界框掩模
740
在对应于自动标注图像的边界框的位置处将
边界框掩模插入背景掩模，以创建像素级掩模

图7

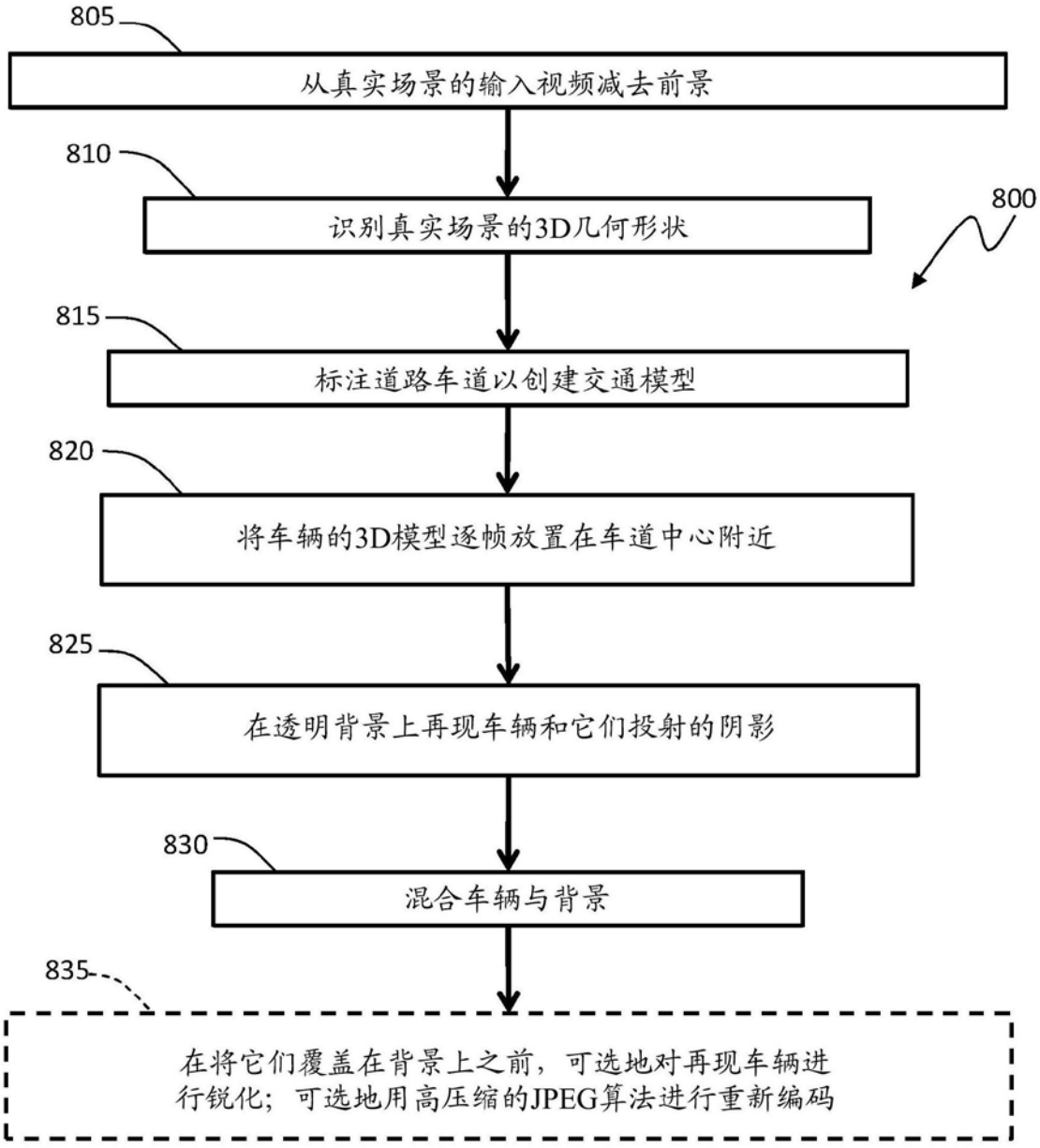

805
从真实场景的输入视频减去前景
810
识别真实场景的3D几何形状
800
815
标注道路车道以创建交通模型
820
将车辆的3D模型逐帧放置在车道中心附近
825
在透明背景上再现车辆和它们投射的阴影
830
混合车辆与背景
835
在将它们覆盖在背景上之前，可选地对再现车辆进行锐化；可选地用高压缩的JPEG算法进行重新编码

图8

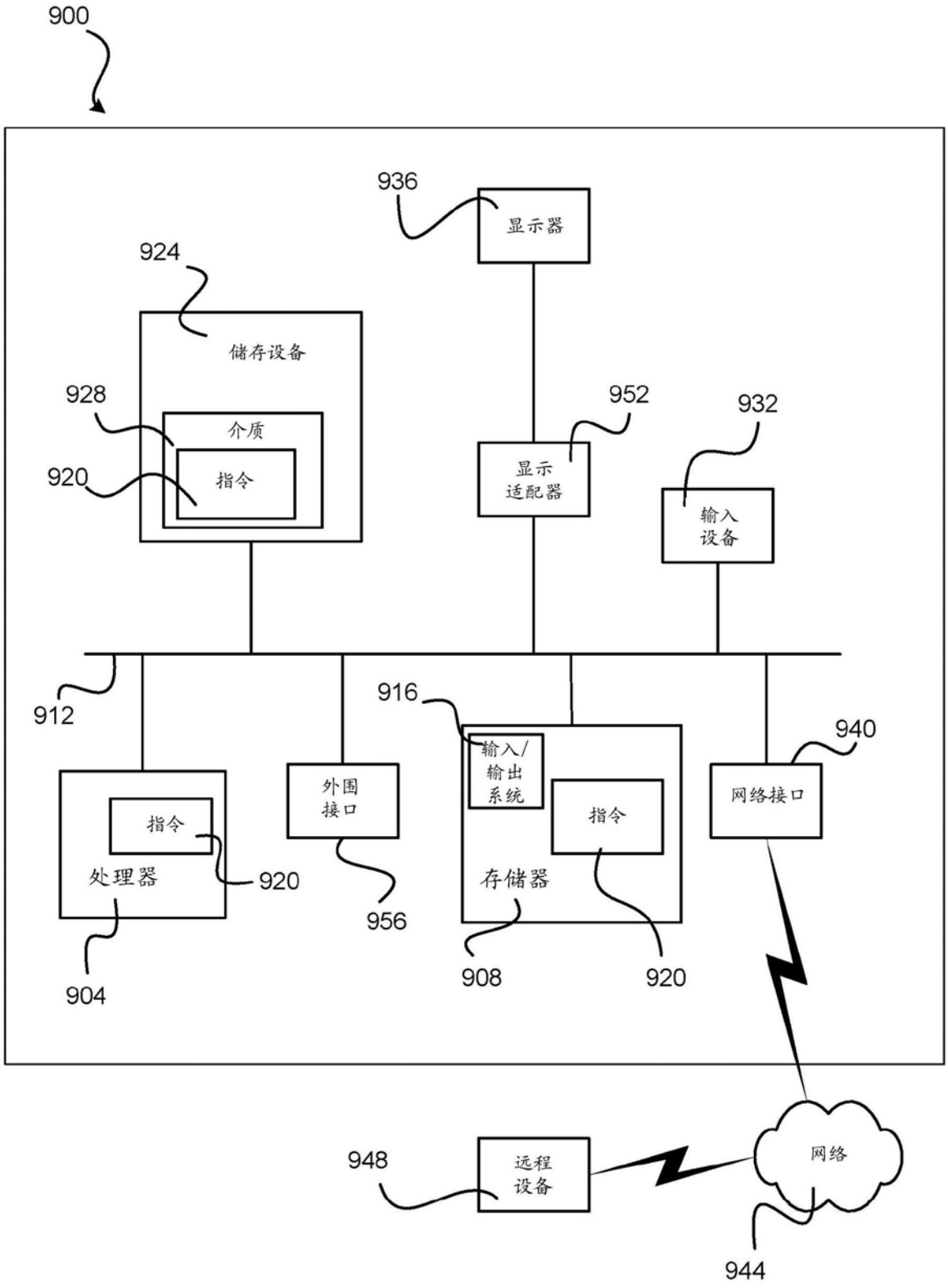

900
936
显示器
924
储存设备
928
介质
920
指令
952
显示
适配器
932
输入
设备
912
916
输入/
输出
系统
指令
存储器
940
网络接口
指令
处理器
920
外围
接口
956
904
908
920
948
远程
设备
网络
944

图9

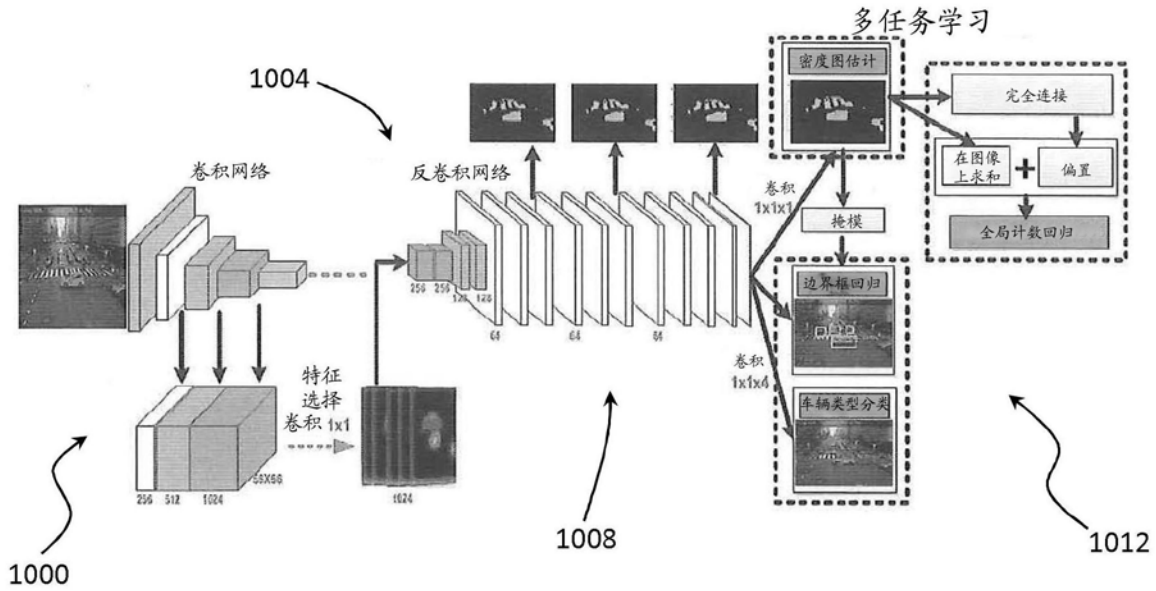

多任务学习
1004
卷积网络
反卷积网络
密度图估计
完全连接
在图像上求和
偏置
全局计数回归
卷积 1x1x1
掩模
边界框回归
卷积 1x1x4
车辆类型分类
特征选择卷积 1x1
1000
1008
1012

图10

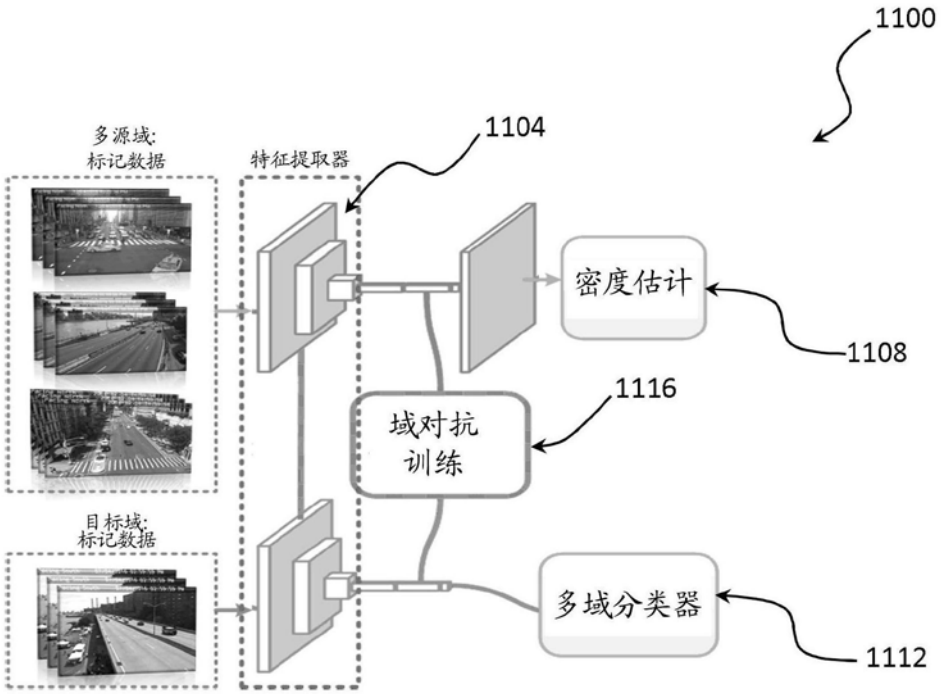

1100
多源域:
标记数据
特征提取器
1104
密度估计
1108
1116
域对抗训练
目标域:
标记数据
多域分类器
1112

图11

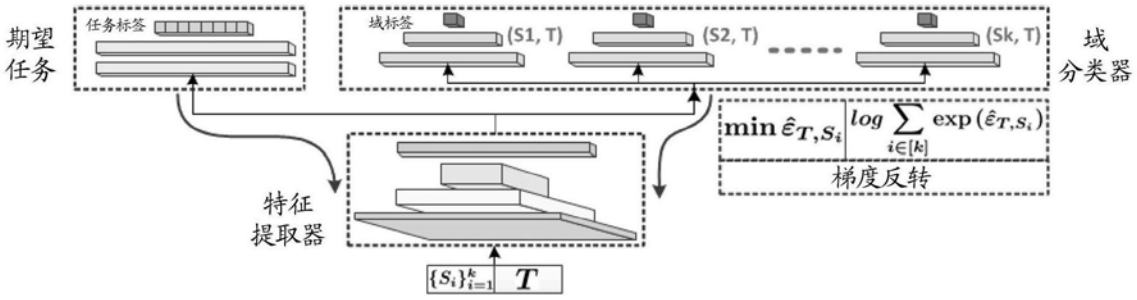

期望
任务
任务标签
域标签
(S1, T)
(S2, T)
(Sk, T)
域
分类器
$\min \hat{\varepsilon}_{T,S_i}$
$log \sum_{i\in[k]} \exp(\hat{\varepsilon}_{T,S_i})$
梯度反转
特征
提取器
$\{S_i\}_{i=1}^{k}$
T

图12