

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020年12月10日(10.12.2020)



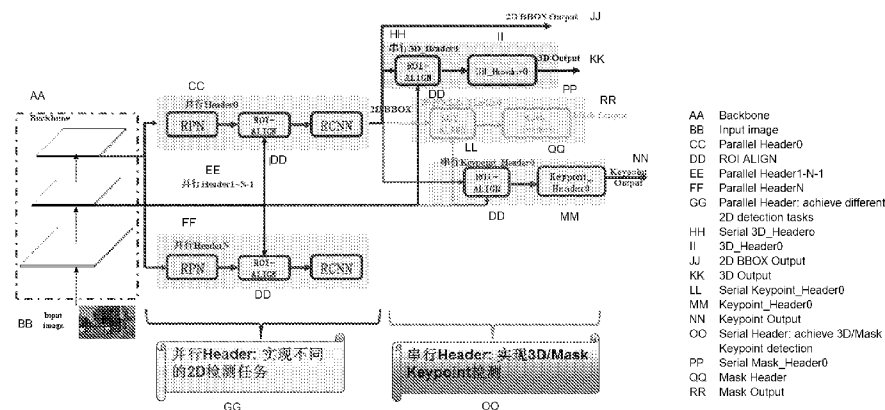
(10) 国际公布号
WO 2020/244653 A1

- (51) 国际专利分类号:
G06K 9/00 (2006.01) *G06N 3/04* (2006.01)
- (21) 国际申请号: PCT/CN2020/094803
- (22) 国际申请日: 2020年6月8日(08.06.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910493331.6 2019年6月6日(06.06.2019) CN
- (71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 江立辉 (JIANG, Lihui); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 屈展 (QU, Zhan); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong

- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。
- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,

(54) Title: OBJECT IDENTIFICATION METHOD AND DEVICE

(54) 发明名称: 物体识别方法及装置



(57) Abstract: Provided is a perceptual network based on multiple headers, relating to the field of artificial intelligence and computer vision. The perceptual network comprises a backbone network and multiple parallel Headers, the multiple parallel Headers being connected to the backbone network. The backbone network is configured to receive an input image, perform convolution processing on the input image, and output feature images corresponding to the image, the feature images having different resolutions; each of the multiple parallel Headers is used for detecting a task object in a task according to the feature images output by the backbone network, and outputting 2D boxes of the area where the task object is located and a confidence level corresponding to each 2D box, wherein each parallel Header completes the detection of different task objects. The task objects are objects that are required to be detected in the task; a higher confidence level represents a greater probability that task objects corresponding to the task will exist in the 2D box corresponding to the confidence level.

WO 2020/244653 A1

IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

- 包括国际检索报告(条约第21条(3))。

(57) 摘要：一种基于多个Header的感知网络，涉及人工智能领域，计算机视觉领域，所述感知网络包括主干网络和多个并行Header，所述多个并行Header和所述主干网络连接；所述主干网络，用于接收输入的图片，并对所述输入的图片进行卷积处理，输出对应所述图片的具有不同分辨率的特征图；所述多个并行Header中的每个Header，用于根据所述主干网络输出的特征图，对一个任务中的任务物体进行检测，输出所述任务物体所在区域的2D框以及每个2D框对应的置信度；其中，所述每个并行Header完成不同的任务物体的检测；其中，所述任务物体为该任务中需要检测的物体；所述置信度越高，表示所述对应该置信度的2D框内存在所述任务所对应的任务物体的概率越大。

物体识别方法及装置

技术领域

本申请涉及人工智能领域，尤其涉及一种物体识别方法及装置。

背景技术

计算机视觉是各个应用领域，如制造业、检验、文档分析、医疗诊断，和军事等领域中各种智能/自主系统中不可分割的一部分，它是一门关于如何运用照相机/摄像机和计算机来获取我们所需的，被拍摄对象的数据与信息的信息的学问。形象地说，就是给计算机安装上眼睛（照相机或摄像机）和大脑（算法）用来代替人眼对目标进行识别、跟踪和测量等，从而使计算机能够感知环境。因为感知可以看作是从感官信号中提取信息，所以计算机视觉也可以看作是研究如何使人工系统从图像或多维数据中“感知”的科学。总的来说，计算机视觉就是用各种成像系统代替视觉器官获取输入信息，再由计算机来代替大脑对这些输入信息完成处理和解释。计算机视觉的最终研究目标就是使计算机能像人那样通过视觉观察和理解世界，具有自主适应环境的能力。

目前视觉感知网络能完成的功能越来越多，包括图片分类、2D 检测、语义分割（Mask）、关键点检测、线性物体检测（比如自动驾驶技术中的车道线或停止线检测）、可行驶区域检测等。另外，视觉感知系统具有成本低、非接触性、体积小、信息量大的特点。随着视觉感知算法的精度不断提高，其成为当今众多人工智能系统的关键技术，得到越来越广泛的应用，如：高级驾驶辅助系统（ADAS, Advanced Driving Assistant System）和自动驾驶系统（ADS, Autonomous Driving System）中对路面上的动态障碍物（人或车）、静态物体（交通灯、交通标志或交通锥状物）的识别，在终端视觉的拍照美颜功能中通过识别人体的 Mask 和关键点实现瘦身效果等。

目前主流的视觉感知网络大多数集中在一种检测任务，如 2D 检测、3D 检测、语义分割、关键点检测等。如果要实现多个功能，则往往需要不同的网络来完成。多个网络同时运行会显著提升硬件的计算量和功耗，降低模型的运行速度，难以实现实时检测。

发明内容

为了减少硬件的计算量和功耗，提升感知网络模型的运算速度，本发明实施例提供一种基于多个头端(Header)的感知网络，所述感知网络包括主干网络和多个并行 Header，所述多个并行 Header 所述和主干网络连接；

所述主干网络，用于接收输入的图片，并对输入的图片进行卷积处理，输出对应所述图片的具有不同分辨率的特征图；

所述一个并行 Header，用于根据所述主干网络输出的特征图，对一个任务中的任务物体进行检测，输出所述任务物体所在区域的 2D 框以及每个 2D 框对应的置信度；其中，所述每个并行 Header 完成不同的任务物体的检测；其中，所述任务物体为该任务中需要检测的物体；所述置信度越高，表示所述对应该置信度的 2D 框内存在所述任务所对应的物体的概率越大。所述一个并行 Header 是上述多个并行 Header 中的任一个，每个并行 header 的

功能是相似的。

可选地,每个并行头端包括候选区域生成网络(RPN)模块、感兴趣区域提取(ROI-ALIGN)模块和区域卷积神经网络(RCNN)模块,所述一个并行头端的 RPN 模块独立于其它并行头端的 RPN 模块;所述一个并行头端的 ROI-ALIGN 模块独立于其它并行头端的 ROI-ALIGN 模块;所述一个并行头端的 RCNN 模块独立于其它并行头端的 RCNN 模块,其中,对于每一个并行头端:

所述 RPN 模块用于:在主干网络提供的一个或者多个特征图上预测所述任务物体所在的区域,并输出匹配所述区域的候选 2D 框;

所述 ROI-ALIGN 模块用于:根据所述 RPN 模块预测得到的区域,从所述主干网络提供的一个特征图中扣取出所述候选 2D 框所在区域的特征;

所述 RCNN 模块用于:通过神经网络对所述候选 2D 框所在区域的特征进行卷积处理,得到所述候选 2D 框属于各个物体类别的置信度;所述各个物体类别为所述一个并行头端对应的任务中的物体类别;通过神经网络对所述候选区域 2D 框的坐标进行调整,使得调整后的 2D 候选框比所述候选 2D 框与实际物体的形状更加匹配,并选择置信度大于预设阈值的调整后的 2D 候选框作为所述区域的 2D 框。

可选地,所述 2D 框为矩形框。

可选地,在本申请实施例的另一方面,所述 RPN 模块用于:基于所属任务对应的物体的模板框(Anchor),在主干网络提供的一个或者多个特征图上对存在该任务物体的区域进行预测以得到候选区域,并输出匹配所述候选区域的候选 2D 框;其中,所述模板框是基于其所属的任务物体的统计特征得到的,所述统计特征包括所述物体的形状和大小。

可选地,在本申请实施例的另一方面,所述感知网络还包括至少一个或多个串行 Header;所述串行 Header 与所述一个并行 Header 连接;

所述串行 Header 用于:利用其连接的并行 Header 提供的所属任务的任务物体的 2D 框,在主干网络上的一个或多个特征图上提取所述 2D 框所在区域的特征,根据所述 2D 框所在区域的特征对所述所属任务的任务物体的 3D 信息、Mask 信息或 Keypoint 信息进行预测。

可选地,所述 RPN 模块在不同的分辨率的特征图上预测不同大小物体所在的区域。

可选地,所述 RPN 模块在低分辨率的特征上完成大物体所在区域的检测,所述 RPN 模块在高分辨率的特征图上完成小物体所在区域的检测。

另一方面,本发明实施例还提供一种物体检测方法,所述方法包括:

接收输入的图片;

对输入的图片进行卷积处理,输出对应所述图片的具有不同分辨率的特征图;

根据所述特征图,针对不同的任务独立检测每个任务中的任务物体,输出所述每个任务物体所在区域的 2D 框以及每个 2D 框对应的置信度;其中,所述任务物体为该任务中需要检测的物体;所述置信度越高,表示所述对应该置信度的 2D 框内存在所述任务所对应的物体的概率越大。

可选地,所述根据所述特征图,针对不同的任务独立检测每个任务中的任务物体,输出所述每个任务物体所在区域的 2D 框以及每个 2D 框对应的置信度,包括:

在一个或者多个特征图上预测所述任务物体所在的区域,并输出匹配所述区域的候选

2D 框;

根据所述任务物体所在的区域, 从一个特征图中扣取出所述候选 2D 框所在区域的特征;

对所述候选 2D 框所在区域的特征进行卷积处理, 得到所述候选 2D 框属于各个物体类别的置信度; 所述各个物体类别为所述一个任务中的物体类别;

通过神经网络对所述候选区域 2D 框的坐标进行调整, 使得调整后的 2D 候选框比所述候选 2D 框与实际物体的形状更加匹配, 并选择置信度大于预设阈值的调整后的 2D 候选框作为所述区域的 2D 框。

可选地, 所述 2D 框为矩形框。

可选地, 在一个或者多个特征图上预测所述任务物体所在的区域, 并输出匹配所述区域的候选 2D 框为:

基于所属任务对应的物体的模板框 (Anchor), 在主干网络提供的一个或者多个特征图上对存在该任务物体的区域进行预测以得到候选区域, 并输出匹配所述候选区域的候选 2D 框; 其中, 所述模板框是基于其所属的任务物体的统计特征得到的, 所述统计特征包括所述物体的形状和大小。

可选地, 所述方法还包括:

基于所属任务的任务物体的 2D 框, 在主干网络上的一个或多个特征图上提取所述 2D 框所在区域的特征, 根据所述 2D 框所在区域的特征对所述所属任务的任务物体的 3D 信息、Mask 信息或 Keypoint 信息进行预测。

可选地, 在低分辨率的特征上完成大物体所在区域的检测, 所述 RPN 模块在高分辨率的特征图上完成小物体所在区域的检测。

另一方面, 本申请实施例提供一种基于部分标注数据训练多任务感知网络的方法, 其特征在于, 所述感知网络包括主干网络和多个并行头端(Header), 所述方法包括:

根据每张图片的标注数据类型, 确定每张图片所属的任务; 其中, 所述每张图片标注一个或者多个数据类型, 所述多个数据类型是所有数据类型的子集, 一个数据类型对应一个任务;

根据每张图片所属的任务, 决定所述每张图片所需训练的 Header;

计算每张图片所述所需训练的 Header 的损失值;

对于每张图片, 通过所述所需训练的 Header 进行梯度回传, 并基于所述损失值调整所述所需训练的 Header 以及主干网络的参数。

可选地, 对所属不同任务的图片进行数据均衡。

本发明实施例还提供一种基于部分标注数据训练多任务感知网络的装置, 所述感知网络包括主干网络和多个并行头端(Header), 所述装置包括:

任务确定模块, 用于根据每张图片的标注数据类型, 确定每张图片所属的任务; 其中, 所述每张图片标注一个或者多个数据类型, 所述多个数据类型是所有数据类型的子集, 一个数据类型对应一个任务;

Header 决定模块, 用于根据每张图片所属的任务, 决定所述每张图片所需训练的 Header;

损失值计算模块，针对每张图片，用于计算 Header 决定模块决定出的 Header 的损失值；

调整模块，针对每张图片，通过计算 Header 决定模块决定出的 Header 进行梯度回传，并基于所述损失值计算模块得到的损失值调整所述所需训练的 Header 以及主干网络的参数。

可选地，所述装置还包括：数据均衡模块，用于对所属不同任务的图片进行数据均衡。

本发明实施例还提供一种感知网络应用系统，该感知网络应用系统包括至少一个处理器，至少一个存储器、至少一个通信接口以及至少一个显示设备。处理器、存储器、显示设备和通信接口通过通信总线连接并完成相互间的通信。

通信接口，用于与其他设备或通信网络通信；

存储器用于存储执行以上方案的应用程序代码，并由处理器来控制执行。所述处理器用于执行所述存储器中存储的应用程序代码。

存储器 2002 存储的代码可执行以上提供的一种基于 Multi-Header 的物体的感知方法也可以是上述实施例提供的训练该感知网络的方法。

显示设备用于显示待识别图像、该图像中感兴趣物体的 2D、3D、Mask、关键点等信息。

本申请实施例提供的感知网络，各个感知任务共用相同的主干网络，成倍节省计算量，提升感知网络模型的运算速度；网络结构易于扩展，只需要增加一个或者若干个 Header 就可以扩展 2D 的检测类型。每个并行 Header 具有独立的 RPN 和 RCNN 模块，仅需要检测其所属的任务的物体，这样在训练过程中，可以避免对未标注的其他任务的物体的误伤。

本申请的这些方面或其他方面在以下实施例的描述中会更加简明易懂。

附图说明

为了更清楚地说明本申请发明实施例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简单地介绍，显而易见地，下面描述中的附图仅仅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。

图 1 为本申请实施例提供的系统架构的结构示意图；

图 2 为本申请实施例提供的 CNN 特征提取模型的示意图；

图 3 是本申请实施例提供的一种芯片硬件结构示意图；

图 4 为本申请实施例提供的一种基于多个并行 Header 的感知网络应用系统框架示意图；

图 5 为本申请实施例提供的基于多个并行 Header 的感知网络结构示意图；

图 6 为本申请实施例提供的 ADAS/AD 的基于多个并行 Header 感知系统结构示意图；

图 7 为本申请实施例提供的一种基础特征生成的流程示意图；

图 8 为本申请实施例提供的另一种 RPN 层的结构示意图；

图 9 为本申请实施例提供的另一种 RPN 层的物体的对应的 Anchor 示意图；

图 10 为本申请实施例提供的另一种 ROI-ALIGN 的过程示意图；

- 图 11 为本申请实施例提供的另一种 RCNN 的实现及结构示意图；
 图 12 为本申请实施例提供的另一种串行 Header 的实现及结构示意图；
 图 13 为本申请实施例提供的另一种串行 Header 的实现及结构示意图；
 图 14 为本申请实施例提供的一种串行 Header 的实现及结构示意图示意图；
 图 15 为本申请实施例提供的一种部分标注数据的训练方法示意图；
 图 16 为本申请实施例提供的另一种部分标注数据的训练方法示意图；
 图 17 为本申请实施例提供的另一种部分标注数据的训练方法示意图；
 图 18 为本申请实施例提供的另一种部分标注数据的训练方法示意图；
 图 19 为本申请实施例提供的一种基于多个并行 Header 的感知网络的应用示意图；
 图 20 为本申请实施例提供的一种基于多个并行 Header 的感知网络的应用示意图；
 图 21 为本申请实施例提供的一种感知方法的流程示意图；
 图 22 为本申请实施例提供的一种 2D 检测流程示意图；
 图 23 为本申请实施例提供的一种终端设备的 3D 检测流程示意图；
 图 24 为本申请实施例提供的一种 Mask 预测流程示意图；
 图 25 为本申请实施例提供的一种关键点坐标预测流程示意图；
 图 26 为本申请实施例提供的一种感知网络的训练流程示意图；
 图 27 为本申请实施例提供的一种基于多个并行 Header 的感知网络实现结构示意图；
 图 28 为本申请实施例提供的一种基于多个并行 Header 的感知网络实现结构示意图；
 图 29 为本申请实施例提供的一种基于部分标注数据训练多任务感知网络的装置图；
 图 30 为本申请实施例提供的一种物体检测方法流程示意图；
 图 31 为本申请实施例提供的一种基于部分标注数据训练多任务感知网络的流程图。

具体实施方式

首先对本申请实施例中用到的缩略语，列表如下：

英文缩略语	完整的英文表述/英文标准用语	中文表述/中文术语
ADAS	Advanced Driving Assistant System	高级辅助驾驶系统
ADS	Autonomous Driving System	自动驾驶系统
CNN	Convolutional Neural Networks	卷积神经网络
FC	Fully Convolutional	全卷积
2D	2 Dementional	2 维
RPN	Region Proposal Network	候选区域生成网络
RCNN	Region Convolutional	区域卷积神经网络

	Neural Networks	
ROI	Region of Interest	感兴趣区域
NMS	Non Maximum Suppression	非最大抑制
Deconv	De-Convolutional Layer	反卷积层
GFlops	Giga Float Operations	千兆浮点运算

表 1

需要说明的是，本发明实施例部分附图为了更符合业界的术语描述，使用了英文描述，同时实施例中也给出了相应的中文的定义。下面结合附图对本申请的实施例进行描述。

本申请实施例主要应用在驾驶辅助、自动驾驶、手机终端等需要完成多种感知任务的领域。本发明的应用系统框架如图 4 所示，视频经过抽帧得到单张图片，该图片送入到本发明的 Mulit-Header 感知网络，得到该图片中感兴趣物体的 2D、3D、Mask（掩膜）、关键点等信息。这些检测结果输出到后处理模块进行处理，比如在自动驾驶系统中送入规划控制单元进行决策、在手机终端中送入美颜算法进行处理得到美颜后的图片。下面分别对 ADAS/ADS 视觉感知系统和手机美颜两种应用场景做简单的介绍

应用场景 1: ADAS/ADS 视觉感知系统

如图 19 所示，在 ADAS 和 ADS 中，需要实时进行多类型的 2D 目标检测，包括：动态障碍物（行人（Pedestrian）、骑行者（Cyclist）、三轮车（Tricycle）、轿车（Car）、卡车（Truck）、公交车（Bus））、静态障碍物（交通锥标（TrafficCone）、交通棍标（TrafficStick）、消防栓（FireHydrant）、摩托车（Motocycle）、自行车（Bicycle））、交通标志（TrafficSign、导向标志（GuideSign）、广告牌（Billboard）、红色交通灯（TrafficLight_Red）/黄色交通灯（TrafficLight_Yellow）/绿色交通灯（TrafficLight_Green）/黑色交通灯（TrafficLight_Black）、路标（RoadSign））。另外，为了准确获取动态障碍物的在 3 维空间所占的区域，还需要对动态障碍物进行 3D 估计，输出 3D 框。为了与激光雷达的数据进行融合，需要获取动态障碍物的 Mask，从而把打到动态障碍物上的激光点云筛选出来；为了进行精确的泊车位，需要同时检测出泊车位的 4 个关键点；为了进行构图定位，需要检测出静态目标的关键点。使用本申请实施例提供的技术方案，可以在一个感知网络中完成上述所有的功能。

应用场景 2: 手机美颜功能

如图 20 所示，在手机中，通过本申请实施例提供的感知网络检测出人体的 Mask 和关键点，可以对人体相应的部位进行放大缩小，比如进行收腰和美臀操作，从而输出美颜的图片。

应用场景 3: 图像分类场景:

物体识别装置在获取待分类图像后，采用本申请的物体识别方法获取待分类图像中的物体的类别，然后可根据待分类图像中物体的类别对待分类图像进行分类。对于摄影师来说，每天会拍很多照片，有动物的，有人物，有植物的。采用本申请的方法可以快速地将照片按照照片中的内容进行分类，可分成包含动物的照片、包含人物的照片和包含植物的照片。

对于图像数量比较庞大的情况，人工分类的方式效率比较低，并且人在长时间处理同一件事情时很容易产生疲劳感，此时分类的结果会有很大的误差；而采用本申请的方法可以快速地将图像进行分类，并且不会有误差。

应用场景 4 商品分类：

物体识别装置获取商品的图像后，然后采用本申请的物体识别方法获取商品的图像中商品的类别，然后根据商品的类别对商品进行分类。对于大型商场或超市中种类繁多的商品，采用本申请的物体识别方法可以快速完成商品的分类，降低了时间开销和人工成本。

本申请实施例提供的方法和装置还可以用于扩充训练数据库，如图 1 所示执行设备 120 的 I/O 接口 112 可以将经执行设备处理过的图像（如包含物体的图像块或图像）和用户输入的物体类别一起作为训练数据发送给数据库 130，以使得数据库 130 维护的训练数据更加丰富，从而为训练设备 130 的训练工作提供更丰富的训练数据。

下面从模型训练侧和模型应用侧对本申请提供的方法进行描述：

本申请实施例提供的训练 CNN 特征提取模型的方法，涉及计算机视觉的处理，具体可以应用于数据训练、机器学习、深度学习等数据处理方法，对训练数据（如本申请中的物体的图像或图像块和物体的类别）进行符号化和形式化的智能信息建模、抽取、预处理、训练等，最终得到训练好的 CNN 特征提取模型的；并且，本申请实施例将输入数据（如本申请中的物体的图像）输入到所述训练好的 CNN 特征提取模型中，得到输出数据（如本申请中得到该图片中感兴趣物体的 2D、3D、Mask、关键点等信息）。

由于本申请实施例涉及大量神经网络的应用，为了便于理解，下面先对本申请实施例涉及的相关术语及神经网络等相关概念进行介绍。

(1) 物体识别，利用图像处理 and 机器学习、计算机图形学等相关方法，确定图像物体的类别。

(2) 神经网络

神经网络可以是由神经元组成的，神经元可以是指以 x_s 和截距 1 为输入的运算单元，该运算单元的输出可以为：

$$h_{w,b}(x) = f(W^T x) = f\left(\sum_{s=1}^n W_s x_s + b\right) \quad (1-1)$$

其中， $s=1, 2, \dots, n$ ， n 为大于 1 的自然数， W_s 为 x_s 的权重， b 为神经单元的偏置。 f 为神经单元的激活函数（activation functions），用于将非线性特性引入神经网络中，来将神经元中的输入信号转换为输出信号。该激活函数的输出信号可以作为下一层卷积层的输入。激活函数可以是 sigmoid 函数。神经网络是将许多个上述单一的神经单元联结在一起形成的网络，即一个神经单元的输出可以是另一个神经单元的输入。每个神经单元的输入可以与前一层的局部接受域相连，来提取局部接受域的特征，局部接受域可以由若干个神经单元组成的区域。

(3) 深度神经网络

深度神经网络（Deep Neural Network, DNN），也称多层神经网络，可以理解为具有很多层隐含层的神经网络，这里的“很多”并没有特别的度量标准。从 DNN 按不同层的位置划分，DNN 内部的神经网络可以分为三类：输入层，隐含层，输出层。一般来说第一层

是输入层，最后一层是输出层，中间的层数都是隐含层。层与层之间是全连接的，也就是说，第 i 层的任意一个神经元一定与第 $i+1$ 层的任意一个神经元相连。虽然 DNN 看起来很复杂，但是就每一层的工作来说，其实并不复杂，简单来说就是如下线性关系表达式：

$\vec{y} = \alpha(W\vec{x} + \vec{b})$ ，其中， \vec{x} 是输入向量， \vec{y} 是输出向量， \vec{b} 是偏移向量， W 是权重矩阵（也称系数）， $\alpha()$ 是激活函数。每一层仅仅是对输入向量 \vec{x} 经过如此简单的操作得到输出向量 \vec{y} 。由于 DNN 层数多，则系数 W 和偏移向量 \vec{b} 的数量也就很多了。这些参数在 DNN 中的定义如下所述：以系数 W 为例：假设在一个三层的 DNN 中，第二层的第 4 个神经元到第三层的第 2 个神经元的线性系数定义为 W_{24}^3 。上标 3 代表系数 W 所在的层数，而下标对应的是输出的第三层索引 2 和输入的第二层索引 4。总结就是：第 $L-1$ 层的第 k 个神经元到第 L 层的第 j 个神经元的系数定义为 W_{jk}^L 。需要注意的是，输入层是没有 W 参数的。在深度

神经网络中，更多的隐含层让网络更能够刻画现实世界中的复杂情形。理论上而言，参数越多的模型复杂度越高，“容量”也就越大，也就意味着它能完成更复杂的学习任务。训练深度神经网络的也就是学习权重矩阵的过程，其最终目的是得到训练好的深度神经网络的所有层的权重矩阵（由很多层的向量 W 形成的权重矩阵）。

（4）卷积神经网络

卷积神经网络（CNN, Convolutional Neuron Network）是一种带有卷积结构的深度神经网络。卷积神经网络包含了一个由卷积层和子采样层构成的特征抽取器。该特征抽取器可以看作是滤波器，卷积过程可以看作是使用一个可训练的滤波器与一个输入的图像或者卷积特征平面(feature map)做卷积。卷积层是指卷积神经网络中对输入信号进行卷积处理的神经元层。在卷积神经网络的卷积层中，一个神经元可以只与部分邻层神经元连接。一个卷积层中，通常包含若干个特征平面，每个特征平面可以由一些矩形排列的神经单元组成。同一特征平面的神经单元共享权重，这里共享的权重就是卷积核。共享权重可以理解为提取图像信息的方式与位置无关。这其中隐含的原理是：图像的某一部分的统计信息与其他部分是一样的。即意味着在某一部分学习的图像信息也能用在另一部分上。所以对于图像上的所有位置，都能使用同样的学习得到的图像信息。在同一卷积层中，可以使用多个卷积核来提取不同的图像信息，一般地，卷积核数量越多，卷积操作反映的图像信息越丰富。

卷积核可以以随机大小的矩阵的形式初始化，在卷积神经网络的训练过程中卷积核可以通过学习得到合理的权重。另外，共享权重带来的直接好处是减少卷积神经网络各层之间的连接，同时又降低了过拟合的风险。

（5）循环神经网络（RNN, Recurrent Neural Networks)是用来处理序列数据的。在传统的神经网络模型中，是从输入层到隐含层再到输出层，层与层之间是全连接的，而对于每一层层内之间的各个节点是无连接的。这种普通的神经网络虽然解决了很多难题，但是却仍然对很多问题却无能为力。例如，你要预测句子的下一个单词是什么，一般需要用到前面的单词，因为一个句子中前后单词并不是独立的。RNN 之所以称为循环神经网络，即一个序列当前的输出与前面的输出也有关。具体的表现形式为网络会对前面的信息进行记忆并应用于当前输出的计算中，即隐含层本层之间的节点不再无连接而是有连接的，并且隐含层的输入不仅包括输入层的输出还包括上一时刻隐含层的输出。理论上，RNN 能够对任何长度的序列数据进行处理。对于 RNN 的训练和对传统的 CNN 或 DNN 的训练一样。同

样使用误差反向传播算法，不过有一点区别：即，如果将 RNN 进行网络展开，那么其中的参数，如 W ，是共享的；而如上举例上述的传统神经网络却不是这样。并且在使用梯度下降算法中，每一步的输出不仅依赖当前步的网络，还依赖前面若干步网络的状态。该学习算法称为基于时间的反向传播算法 Back propagation Through Time (BPTT)。

既然已经有了卷积神经网络，为什么还要循环神经网络？原因很简单，在卷积神经网络中，有一个前提假设是：元素之间是相互独立的，输入与输出也是独立的，比如猫和狗。但现实世界中，很多元素都是相互连接的，比如股票随时间的变化，再比如一个人说了：我喜欢旅游，其中最喜欢的地方是云南，以后有机会一定要去。这里填空，人类应该都知道是填“云南”。因为人类会根据上下文的内容进行推断，但如何让机器做到这一步？RNN 就应运而生了。RNN 旨在让机器像人一样拥有记忆的能力。因此，RNN 的输出就需要依赖当前的输入信息和历史的记忆信息。

(6) 损失函数

在训练深度神经网络的过程中，因为希望深度神经网络的输出尽可能的接近真正想要预测的值，所以可以通过比较当前网络的预测值和真正想要的目标值，再根据两者之间的差异情况来更新每一层神经网络的权重向量（当然，在第一次更新之前通常会有初始化的过程，即为深度神经网络中的各层预先配置参数），比如，如果网络的预测值高了，就调整权重向量让它预测低一些，不断的调整，直到深度神经网络能够预测出真正想要的目标值或与真正想要的目标值非常接近的值。因此，就需要预先定义“如何比较预测值和目标值之间的差异”，这便是损失函数（loss function）或目标函数（objective function），它们是用于衡量预测值和目标值的差异的重要方程。其中，以损失函数举例，损失函数的输出值（loss）越高表示差异越大，那么深度神经网络的训练就变成了尽可能缩小这个 loss 的过程。

(7) 反向传播算法

卷积神经网络可以采用误差反向传播（back propagation, BP）算法在训练过程中修正初始的超分辨率模型中参数的大小，使得超分辨率模型的重建误差损失越来越小。具体地，前向传递输入信号直至输出会产生误差损失，通过反向传播误差损失信息来更新初始的超分辨率模型中参数，从而使误差损失收敛。反向传播算法是以误差损失为主导的反向传播运动，旨在得到最优的超分辨率模型的参数，例如权重矩阵。

下面介绍本申请实施例提供系统架构。

参见图 1，本申请实施例提供了一种系统架构 110。如所述系统架构 110 所示，数据采集设备 170 用于采集训练数据，本申请实施例中训练数据包括：物体的图像或者图像块及物体的类别；并将训练数据存入数据库 130，训练设备 130 基于数据库 130 中维护的训练数据训练得到 CNN 特征提取模型 101（解释说明：这里的 101 就是前面介绍的经训练阶段训练得到的模型，可以是用于特征提取的感知网络等）。下面将以实施例一更详细地描述训练设备 130 如何基于训练数据得到 CNN 特征提取模型 101，该 CNN 特征提取模型 101 能够用于实现本申请实施例提供的感知网络，即，将待识别图像或图像块通过相关预处理后输入该 CNN 特征提取模型 101，即可得到待识别图像或图像块感兴趣物体的 2D、3D、Mask、关键点等信息。本申请实施例中的 CNN 特征提取模型 101 具体可以为 CNN 卷积神

神经网络。需要说明的是，在实际的应用中，所述数据库 130 中维护的训练数据不一定都来自于数据采集设备 170 的采集，也有可能是从其他设备接收得到的。另外需要说明的是，训练设备 130 也不一定完全基于数据库 130 维护的训练数据进行 CNN 特征提取模型 101 的训练，也有可能从云端或其他地方获取训练数据进行模型训练，上述描述不应该作为对本申请实施例的限定。

根据训练设备 130 训练得到的 CNN 特征提取模型 101 可以应用于不同的系统或设备中，如应用于图 1 所示的执行设备 120，所述执行设备 120 可以是终端，如手机终端，平板电脑，笔记本电脑，AR/VR，车载终端等，还可以是服务器或者云端等。在附图 1 中，执行设备 120 配置有 I/O 接口 112，用于与外部设备进行数据交互，用户可以通过客户设备 150 向 I/O 接口 112 输入数据，所述输入数据在本申请实施例中可以包括：待识别图像或者图像块或者图片。

在执行设备 120 对输入数据进行预处理，或者在执行设备 120 的计算模块 111 执行计算等相关的处理（比如进行本申请中感知网络的功能实现）过程中，执行设备 120 可以调用数据存储系统 160 中的数据、代码等以用于相应的处理，也可以将相应处理得到的数据、指令等存入数据存储系统 160 中。

最后，I/O 接口 112 将处理结果，如上述得到的图像或图像块或者图片中感兴趣物体的 2D、3D、Mask、关键点等信息返回给客户设备 150，从而提供给用户。

可选地，客户设备 150，可以是自动驾驶系统中的规划控制单元、手机终端中的美颜算法模块。

值得说明的是，训练设备 130 可以针对不同的目标或称不同的任务，基于不同的训练数据生成相应的目标模型/规则 101，该相应的目标模型/规则 101 即可以用于实现上述目标或完成上述任务，从而为用户提供所需的结果。

在附图 1 中所示情况下，用户可以手动给定输入数据，该手动给定可以通过 I/O 接口 112 提供的界面进行操作。另一种情况下，客户设备 150 可以自动地向 I/O 接口 112 发送输入数据，如果要求客户设备 150 自动发送输入数据需要获得用户的授权，则用户可以在客户设备 150 中设置相应权限。用户可以在客户设备 150 查看执行设备 120 输出的结果，具体的呈现形式可以是显示、声音、动作等具体方式。客户设备 150 也可以作为数据采集端，采集如图所示输入 I/O 接口 112 的输入数据及输出 I/O 接口 112 的输出结果作为新的样本数据，并存入数据库 130。当然，也可以不经过客户设备 150 进行采集，而是由 I/O 接口 112 直接将如图所示输入 I/O 接口 112 的输入数据及输出 I/O 接口 112 的输出结果，作为新的样本数据存入数据库 130。

值得注意的是，附图 1 仅是本发明实施例提供的一种系统架构的示意图，图中所示设备、器件、模块等之间的位置关系不构成任何限制，例如，在附图 1 中，数据存储系统 160 相对执行设备 120 是外部存储器，在其它情况下，也可以将数据存储系统 160 置于执行设备 120 中。

如图 1 所示，根据训练设备 130 训练得到 CNN 特征提取模型 101，该 CNN 特征提取模型 101 在本申请实施例中可以是 CNN 卷积神经网络也可以是下面实施例即将介绍的基于多个 Header 的感知网络。

如前文的基础概念介绍所述，卷积神经网络是一种带有卷积结构的深度神经网络，是一种深度学习（deep learning）架构，深度学习架构是指通过机器学习的算法，在不同的抽象层级上进行多个层次的学习。作为一种深度学习架构，CNN 是一种前馈（feed-forward）人工神经网络，该前馈人工神经网络中的各个神经元可以对输入其中的图像作出响应。

如图 2 所示，卷积神经网络（CNN）210 可以包括输入层 220，卷积层/池化层 230（其中池化层为可选的），以及神经网络层 230。

卷积层/池化层 230:

卷积层:

如图 2 所示卷积层/池化层 230 可以包括如示例 221-226 层，举例来说：在一种实现中，221 层为卷积层，222 层为池化层，223 层为卷积层，224 层为池化层，225 为卷积层，226 为池化层；在另一种实现方式中，221、222 为卷积层，223 为池化层，224、225 为卷积层，226 为池化层。即卷积层的输出可以作为随后的池化层的输入，也可以作为另一个卷积层的输入以继续进行卷积操作。

下面将以卷积层 221 为例，介绍一层卷积层的内部工作原理。

卷积层 221 可以包括很多个卷积算子，卷积算子也称为核，其在图像处理中的作用相当于一个从输入图像矩阵中提取特定信息的过滤器，卷积算子本质上可以是一个权重矩阵，这个权重矩阵通常被预先定义，在对图像进行卷积操作的过程中，权重矩阵通常在输入图像上沿着水平方向一个像素接着一个像素（或两个像素接着两个像素……这取决于步长 stride 的取值）的进行处理，从而完成从图像中提取特定特征的工作。该权重矩阵的大小应该与图像的大小相关，需要注意的是，权重矩阵的纵深维度（depth dimension）和输入图像的纵深维度是相同的，在进行卷积运算的过程中，权重矩阵会延伸到输入图像的整个深度。因此，和一个单一的权重矩阵进行卷积会产生一个单一纵深维度的卷积化输出，但是大多数情况下不使用单一权重矩阵，而是应用多个尺寸（行×列）相同的权重矩阵，即多个同型矩阵。每个权重矩阵的输出被堆叠起来形成卷积图像的纵深维度，这里的维度可以理解为由上面所述的“多个”来决定。不同的权重矩阵可以用来提取图像中不同的特征，例如一个权重矩阵用来提取图像边缘信息，另一个权重矩阵用来提取图像的特定颜色，又一个权重矩阵用来对图像中不需要的噪点进行模糊化等。该多个权重矩阵尺寸（行×列）相同，经过该多个尺寸相同的权重矩阵提取后的特征图的尺寸也相同，再将提取到的多个尺寸相同的特征图合并形成卷积运算的输出。

这些权重矩阵中的权重值在实际应用中需要经过大量的训练得到，通过训练得到的权重值形成的各个权重矩阵可以用来从输入图像中提取信息，从而使得卷积神经网络 210 进行正确的预测。

当卷积神经网络 210 有多个卷积层的时候，初始的卷积层（例如 221）往往提取较多的一般特征，该一般特征也可以称之为低级别的特征；随着卷积神经网络 210 深度的加深，越往后的卷积层（例如 226）提取到的特征越来越复杂，比如高级别的语义之类的特征，语义越高的特征越适用于待解决的问题。

池化层:

由于常常需要减少训练参数的数量，因此卷积层之后常常需要周期性的引入池化层，

在如图 2 中 230 所示例的 221-226 各层，可以是一层卷积层后面跟一层池化层，也可以是多层卷积层后面接一层或多层池化层。在图像处理过程中，池化层的唯一目的就是减少图像的空间大小。池化层可以包括平均池化算子和/或最大池化算子，以用于对输入图像进行采样得到较小尺寸的图像。平均池化算子可以在特定范围内对图像中的像素值进行计算产生平均值作为平均池化的结果。最大池化算子可以在特定范围内取该范围内值最大的像素作为最大池化的结果。另外，就像卷积层中用权重矩阵的大小应该与图像尺寸相关一样，池化层中的运算符也应该与图像的大小相关。通过池化层处理后输出的图像尺寸可以小于输入池化层的图像的尺寸，池化层输出的图像中每个像素点表示输入池化层的图像的对应子区域的平均值或最大值。

神经网络层 230:

在经过卷积层/池化层 230 的处理后，卷积神经网络 210 还不足以输出所需要的输出信息。因为如前所述，卷积层/池化层 230 只会提取特征，并减少输入图像带来的参数。然而为了生成最终的输出信息（所需要的类信息或其他相关信息），卷积神经网络 210 需要利用神经网络层 230 来生成一个或者一组所需要的类的数量的输出。因此，在神经网络层 230 中可以包括多层隐含层（如图 2 所示的 231、232 至 23n）以及输出层 240，该多层隐含层中所包含的参数可以根据具体的任务类型的相关训练数据进行预先训练得到，例如该任务类型可以包括图像识别，图像分类，图像超分辨率重建等等……

在神经网络层 230 中的多层隐含层之后，也就是整个卷积神经网络 210 的最后层为输出层 240，该输出层 240 具有类似分类交叉熵的损失函数，具体用于计算预测误差，一旦整个卷积神经网络 210 的前向传播（如图 2 由 220 至 240 方向的传播为前向传播）完成，反向传播（如图 2 由 240 至 220 方向的传播为反向传播）就会开始更新前面提到的各层的权重值以及偏差，以减少卷积神经网络 210 的损失，及卷积神经网络 210 通过输出层输出的结果和理想结果之间的误差。

需要说明的是，如图 2 所示的卷积神经网络 210 仅作为一种卷积神经网络的示例，在具体的应用中，卷积神经网络还可以以其他网络模型的形式存在。

下面介绍本申请实施例提供的一种芯片硬件结构。

图 3 为本发明实施例提供的一种芯片硬件结构，该芯片包括神经网络处理器 30。该芯片可以被设置在如图 1 所示的执行设备 120 中，用以完成计算模块 111 的计算工作。该芯片也可以被设置在如图 1 所示的训练设备 130 中，用以完成训练设备 130 的训练工作并输出目标模型/规则 101。如图 2 所示的卷积神经网络中各层的算法均可在如图 3 所示的芯片中得以实现。

神经网络处理器 NPU 30，NPU 作为协处理器挂载到主 CPU（Host CPU）上，由 Host CPU 分配任务。NPU 的核心部分为运算电路 303，控制器 304 控制运算电路 303 提取存储器（权重存储器或输入存储器）中的数据并进行运算。

在一些实现中，运算电路 303 内部包括多个处理单元（Process Engine, PE）。在一些实现中，运算电路 303 是二维脉动阵列。运算电路 303 还可以是一维脉动阵列或者能够执行例如乘法和加法这样的数学运算的其它电子线路。在一些实现中，运算电路 303 是通用的矩阵处理器。

举例来说,假设有输入矩阵 A,权重矩阵 B,输出矩阵 C。运算电路从权重存储器 302 中取矩阵 B 相应的数据,并缓存在运算电路中每一个 PE 上。运算电路从输入存储器 301 中取矩阵 A 数据与矩阵 B 进行矩阵运算,得到的矩阵的部分结果或最终结果,保存在累加器 (accumulator) 308 中。

向量计算单元 307 可以对运算电路的输出做进一步处理,如向量乘,向量加,指数运算,对数运算,大小比较等等。例如,向量计算单元 307 可以用于神经网络中非卷积/非 FC 层的网络计算,如池化 (Pooling),批归一化 (Batch Normalization),局部响应归一化 (Local Response Normalization) 等。

在一些实现中,向量计算单元能 307 将经处理的输出的向量存储到统一缓存器 306。例如,向量计算单元 307 可以将非线性函数应用到运算电路 303 的输出,例如累加值的向量,用以生成激活值。在一些实现中,向量计算单元 307 生成归一化的值、合并值,或二者均有。在一些实现中,处理过的输出的向量能够用作到运算电路 303 的激活输入,例如用于在神经网络中的后续层中的使用。

本申请实施例提供的感知网络的运算可以由 303 或 307 执行。

统一存储器 306 用于存放输入数据以及输出数据。

权重数据直接通过存储单元访问控制器 305 (Direct Memory Access Controller, DMAC) 将外部存储器中的输入数据搬运到输入存储器 301 和/或统一存储器 306、将外部存储器中的权重数据存入权重存储器 302,以及将统一存储器 306 中的数据存入外部存储器。

总线接口单元 (Bus Interface Unit, BIU) 310,用于通过总线实现主 CPU、DMAC 和取指存储器 309 之间进行交互。

与控制器 304 连接的取指存储器 (instruction fetch buffer) 309,用于存储控制器 304 使用的指令;

控制器 304,用于调用指存储器 309 中缓存的指令,实现控制该运算加速器的工作过程。

可选地,本申请中此处的输入数据为图片,输出数据为图片中感兴趣物体的 2D、3D、Mask、关键点等信息。

一般地,统一存储器 306,输入存储器 301,权重存储器 302 以及取指存储器 309 均为片上 (On-Chip) 存储器,外部存储器为该 NPU 外部的存储器,该外部存储器可以为双倍数据率同步动态随机存储器 (Double Data Rate Synchronous Dynamic Random Access Memory, 简称 DDR SDRAM)、高带宽存储器 (High Bandwidth Memory, HBM) 或其他可读可写的存储器。

图 1 和图 2 中的程序算法是由主 CPU 和 NPU 共同配合完成的。

其中,图 2 所示的卷积神经网络中各层的运算可以由运算电路 303 或向量计算单元 307 执行。

参见图 5,图 5 为本申请实施例提供的一种多头感知网络的结构示意图。如图 5 所示,该感知网络包括:

其主要由主干网络 (Backbone) 401、多个并行 Header0~N 两部分组成。

主干网络 401,用于接收输入的图片,并对输入的图片进行卷积处理,输出对应所述

图片的具有不同分辨率的特征图；也就是说输出对应所述图片的不同大小的特征图；

也就是说，Backbone 完成基础特征的提取，为后续的检测提供相应的特征。

任一个并行头端，用于根据所述主干网络输出的特征图，对一个任务中的任务物体进行检测，输出所述任务物体所在区域的 2D 框以及每个 2D 框对应的置信度；其中，所述每个并行 Header 完成不同的任务物体的检测；其中，所述任务物体为该任务中需要检测的物体；所述置信度越高，表示所述对应该置信度的 2D 框内存在所述任务所对应的物体的概率越大。

也就是说并行 Header 完成不同的 2D 检测任务，比如并行 Header0 完成车的检测，输出 Car/Truck/Bus 的 2D 框和置信度；并行 Header1 完成人的检测，输出 Pedestrian/Cyclist/Tricycle 的 2D 框和置信度；并行 Header2 完成交通灯的检测，输出 Red_Trafficlight/ Green_Trafficlight/Yellow_TrafficLight/Black_TrafficLight 的 2D 框和置信度。

可选地，如图 5 中的所示，所示感知网络还可以包括多个串行 Header，所述感知网络还包括至少一个或多个串行头端；所述串行头端与一个并行头端连接；这里需要强调的是，虽然图 5 为了更好得展示画出了多个串行 Header，但实际上，串行 Header 并不是必须的，对于只需要检测 2D 框的场景，就不需要包括串行 Header。

所述串行 Header 用于：利用其连接的并行 Header 提供的所属任务的任务物体的 2D 框，在主干网络上的一个或多个特征图上提取所述 2D 框所在区域的特征，根据所述 2D 框所在区域的特征对所述所属任务的任务物体的 3D 信息、Mask 信息或 Keypoint 信息进行预测。

串行 Header 可选地串接在并行 Header 的后面，在检测出该任务的 2D 框的基础上，完成 2D 框内部物体的 3D/Mask/Keypoint 检测。

比如，串行 3D_Header0 完成车辆的朝向、质心和长宽高的估计，从而输出车辆的 3D 框；串行 Mask_Header0 预测车辆的精细掩膜，从而把车辆分割开来；串行 Keypoint_Header0 完成车辆的关键点的估计。

串行 Header 并不是必须的，某些任务不需要进行 3D/Mask/Keypoint 检测，则不需要串接串行 Header，比如交通灯的检测，只需要检测 2D 框，就不用串接串行 Header。另外，某些任务可以根据任务的具体需求，选择串接一个或者多个串行 Header，比如停车场（Parkingslot）的检测，除了需要得到 2D 框外，还需要车位的关键点，因此在这个任务中只需要串接一个串行 Keypoint_Header 即可，不需要 3D 和 Mask 的 Header。

下面对各个模块进行详细描述。

Backbone: 主干网络对输入的图片进行一系列的卷积处理，得到在不同的尺度下的特征图（feature map）。这些特征图将为后续的检测模块提供基础特征。主干网络可以采用多种形式，比如 VGG（Visual Geometry Group，视觉几何组）、Resnet（Residual Neural Network，残差神经网络）、Inception-net（GoogLeNet 的核心结构）等。

并行 Header: 并行 Header 主要根据 Backbone 提供的基础特征，完成一个任务的 2D 框的检测，输出这个任务的物体的 2D 框以及对应的置信度。

可选地，每个任务的并行 Header 包括 RPN、ROI-ALIGN 和 RCNN 三个模块。

RPN 模块: 用于在主干网络提供的一个或者多个特征图上预测所述任务物体所在的区域，并输出匹配所述区域的候选 2D 框；

或者可以这样理解，RPN 全称为候选区域生成网络（Region Proposal Network），其在 Backbone 的一个或者多个 feature map 上预测出可能存在该任务物体的区域，并且给出这些区域的框，这些区域称为候选区域（Proposal）。

比如，当并行 Header0 负责检测车时，其 RPN 层就预测出可能存在车的候选框；当并行 Header1 负责检测人时，其 RPN 层就预测出可能存在人的候选框。当然，这些 Proposal 是不准确的，一方面其不一定含有该任务的物体，另一方面这些框也是不紧致的

ROI-ALIGN 模块：用于根据所述 RPN 模块预测得到的区域，从所述主干网络提供的一个特征图中扣取出所述候选 2D 框所在区域的特征；

也就是说，ROI-ALIGN 模块主要根据 RPN 模块提供的 Proposal，在 Backbone 的某个 feature map 上把每个 Proposal 所在的区域的特征扣取出来，并且 resize 到固定的大小，得到每个 Proposal 的特征。可以理解的是，ROI-ALIGN 模块可以使用但不局限于 ROI-POOLING（感兴趣区域池化）/ROI-ALIGN（感兴趣区域提取）/PS-ROIPOOLING（位置敏感的感兴趣区域池化）/PS-ROIALIGN（位置敏感的感兴趣区域提取）等特征抽取方法。

RCNN 模块：用于通过神经网络对所述候选 2D 框所在区域的特征进行卷积处理，得到所述候选 2D 框属于各个物体类别的置信度；所述各个物体类别为所述一个并行头端对应的任务中的物体类别；通过神经网络对所述候选区域 2D 框的坐标进行调整，使得调整后的 2D 候选框比所述候选 2D 框与实际物体的形状更加匹配，并选择置信度大于预设阈值的调整后的 2D 候选框作为所述区域的 2D 框。

也就是说，RCNN 模块主要是对 ROI-ALIGN 模块提出的每个 Proposal 的特征进行细化处理，得到每个 Proposal 的属于各个类别置信度（比如对于车这个任务，会给出 Background/Car/Truck/Bus 4 个分数），同时对 Proposal 的 2D 框的坐标进行调整，输出更加紧致的 2D 框。这些 2D 框经过 NMS(Non Maximum Suppression, 非极大值抑制) 合并后，作为最后的 2D 框输出。

如上所述，在一些实际应用场景中，该感知网络还可以包括串行 Header。串行 Header 主要是串接在并行 Header 后面，在检测出 2D 框的基础上，进一步进行 3D/Mask/Keypoint 检测。因此串行 Header 有 3 个类型：

串行 3D Header: 串行 3D Header 根据前端的并行 Header 提供的 2D 框（此时的 2D 框是准确的紧致 2D 框），通过 ROI-ALIGN 模块在 Backbone 的某个 feature map 上把这些 2D 框所在的区域的特征扣取出来，然后通过一个小网络(图 5 中的 3D_Header)回归出这个 2D 框内部的物体的质心点坐标、朝向角度和长宽高，从而得到完整的 3D 信息。

串行 Mask Header: 串行 Mask Header 根据前端的并行 Header 提供的 2D 框(此时的 2D 框是准确的紧致 2D 框)，通过 ROI-ALIGN 模块在 Backbone 的某个 feature map 上把这些 2D 框所在的区域的特征扣取出来，然后通过一个小网络(图 5 中的 Mask_Header)回归出这个 2D 框内部的物体的掩膜，从而把这个物体分割出来。

串行 Keypoint Header: 串行 Keypoint Header 根据前端的并行 Header 提供的 2D 框（此时的 2D 框是准确的紧致 2D 框），通过 ROI-ALIGN 模块在 Backbone 的某个 feature map 上把这些 2D 框所在的区域的特征扣取出来，然后通过一个小网络(图 5 中的 Keypoint_Header)回归出这个 2D 框内部的物体的关键点坐标。

可选地,如图 29 所示,本发明实施例还提供一种基于部分标注数据训练多任务感知网络的装置,所述感知网络包括主干网络和多个并行头端(Header),感知网络的结构在前述实施例中已经详细描述,在此不再赘述,所述装置包括:

任务确定模块 2900,用于根据每张图片的标注数据类型,确定每张图片所属的任务;其中,所述每张图片标注一个或者多个数据类型,所述多个数据类型是所有数据类型的子集,一个数据类型对应一个任务;

Header 决定模块 2901,用于根据任务确定模块 2900 确定的每张图片所属的任务,决定所述每张图片所需训练的 Header;

损失值计算模块 2902,针对每张图片,用于计算 Header 决定模块 2901 决定出的 Header 的损失值;

调整模块 2903,针对每张图片,通过计算 Header 决定模块 2901 决定出的 Header 进行梯度回传,并基于所述损失值计算模块 2902 得到的损失值调整所述所需训练的 Header 以及主干网络的参数。

可选地,在一个实施例中,如图 29 的虚线框所示,所述装置还可以包括:

数据均衡模块 2904,用于对所属不同任务的图片进行数据均衡。

如图 6 所示,下面以 ADAS/AD 的视觉感知系统为例,对本发明实施例进行详细的介绍。在 ADAS/AD 的视觉感知系统中,需要实时进行多类型的 2D 目标检测,包括:动态障碍物(Pedestrian、Cyclist、Tricycle、Car、Truck、Bus),静态障碍物(TrafficCone、TrafficStick、FireHydrant、Motorcycle、Bicycle),交通标志(TrafficSign、GuideSign、Billboard)。另外,为了准确获取车辆在 3 维空间所占的区域,还需要对动态障碍物进行 3D 估计,输出 3D 框。为了与激光雷达的数据进行融合,需要获取动态障碍物的 Mask,从而把打到动态障碍物上的激光点云筛选出来;为了进行精确的泊车位,需要同时检测出泊车位的 4 个关键点。使用本实施例提供的技术方案,可以在一个网络中完成上述所有的功能,下面对本实施例进行详细描述。

1、各个 Header 任务的划分以及网络的总体框图

根据需要检测的物体的相似性以及训练样本的丰富和稀缺程度,本实施例中把需要检测的 20 类物体划分为 8 个大类别,如表 2 所示。

Header	检测 2D 物体类别	3D	Mask	Keypoint
0	Car/Truck/Bus	Yes	Yes	No
1	Pedestrian/Cyclist/Tricycle	No	Yes	No
2	Parkingslot	No	No	Yes
3	TrafficLight_Red/Yellow/Green/Black	No	No	No
4	TrafficSign/GuideSign/BillBoard	No	No	No
5	TrafficCone/TrafficStick/FireHydrant	No	No	No
6	Motorcycle/Bicycle	No	No	No

7	RoadSign	No	No	No
---	----------	----	----	----

表 2 各个 Header 需要检测的物体类别以及扩展功能

根据业务的需求，Header0 除了需要完成车的 2D 检测外，还需要进一步完成 3D 和 Mask 检测；Header1 除了需要完成人的 2D 检测外，还需要进一步完成 Mask 的检测；Header2 除了完成泊车位 2D 框的检测外，还需要完成车位关键点的检测。

需要说明的是，表 2 的任务划分只是本实施例中的一个举例，在其它实施例中可以进行不同的任务划分，并不局限于表 2 中的任务划分。

根据表 2 的任务划分，本实施例的感知网络总体结构如图 6 所示。

感知网络其主要包括 3 部分：Backbone、并行 Header 和串行 Header。需要说明的是，如前实施例所述，串行 Header 并不是必须的，原因在上述实施例已经描述，在此不再赘述。其中，8 个并行 Header 同时完成了表 1 中的 8 个大类的 2D 检测，在 Header0~2 后面串接了若干个串行 Header，进一步完成 3D/Mask/Keypoint 的检测。从图 6 可以看出，本发明可以根据业务的需求，对 Header 进行灵活的增加和删减，从而实现不同的功能配置。

2、基础特征生成

基础特征生成流程由图 6 的 Backbone 实施，其对输入的图像进行卷积处理，生成若干不同尺度的卷积特征图，每张特征图是一个 $H*W*C$ 的矩阵，其中 H 是特征图的高度，W 是特征图的宽度、C 是特征图的通道数。

Backbone 可以采用目前多种现有的卷积网络框架，比如 VGG16、Resnet50、Inception-Net 等，下面以 Resnet18 为 Backbone，说明基础特征的生成流程。该流程如图 7 所示。

假设输入的图片的分辨率为 $H*W*3$ （高度 H，宽度 W，通道数为 3，也就是 RGB 三个通道）。输入图片经过 Resnet18 的第一个卷积模块（图中的 Res18-Conv1，该卷积模块由若干卷积层组成，后面的卷积模块类似）进行卷积运算，生成 Featuremap（特征图）C1，这个特征图相对于输入图像进行了 2 次下采样，并且通道数扩充为 64，因此 C1 的分辨率是 $H/4*W/4*64$ ；C1 经过 Resnet18 的第 2 个卷积模块（Res18-Conv2）进行卷积运算，得到 Featuremap C2，这个特征图的分辨率与 C1 一致；C2 继续经过 Resnet18 的第 3 个卷积模块（Res18-Conv3）处理，生成 Featuremap C3，这个特征图相对 C2 进一步下采样，通道数增倍，其分辨率为 $H/8*W/8*128$ ；最后 C3 经过 Res18-Conv4 处理，生成 Featuremap C4，其分辨率为 $H/16*W/16*256$ 。

从图 7 可以看出，Resnet18 对输入图片进行多个层次的卷积处理，得到不同尺度的特征图：C1/C2/C3/C4。底层的特征图的宽度和高度比较大，通道数较少，其主要为图像的底层特征（比如图像边缘、纹理特征），高层的特征图的宽度和高度比较小，通道数较多，其主要为图像的高层特征（比如形状、物体特征）。后续的 2D 检测流程将会基于这些特征图进行进一步的预测。

3、2D 候选区域预测流程

2D 候选区域预测流程由图 6 中每个并行 Header 的 RPN 模块实施，其根据 Backbone 提供的特征图（C1/C2/C3/C4），预测出可能存在该任务物体的区域，并且给出这些区域的候选框（也可以叫候选区域，Proposal）。在本实施例中，并行 Header0 负责检测车，其 RPN 层就预测出可能存在车的候选框；并行 Header1 负责检测人，其 RPN 层就预测出可能

存在人的候选框，以此类推，不再赘述。

RPN 层的基本结构如图 8 所示。在 C4 上通过一个 3*3 的卷积，生成特征图 RPN Hidden。后面每个并行 Header 的 RPN 层将会从 RPN Hidden 中预测 Proposal。具体来说，并行 Header0 的 RPN 层分别通过两个 1*1 的卷积，预测出 RPN Hidden 每个位置处的 Proposal 的坐标以及置信度。这个置信度越高，表示这个 Proposal 存在该任务的物体的概率越大。比如，在并行 Header0 中某个 Proposal 的 score 越大，就表示其存在车的概率越大。每个 RPN 层预测出来的 Proposal 需要经过 Proposal 合并模块，根据 Proposal 之间的重合程度去掉多余的 Proposal(这个过程可以采用但不限制于 NMS 算法)，在剩余的 K 个 Proposal 中挑选出 score 最大的 N(N<K)个 Proposal 作为候选的可能存在物体的区域。从图 8 可以看出，这些 Proposal 是不准确的，一方面其不一定含有该任务的物体，另一方面这些框也是不紧致的。因此，RPN 模块只是一个粗检测的过程，需要后续的 RCNN 模块进行细分。

在 RPN 模块回归 Proposal 的坐标时，并不是直接回归坐标的绝对值，而是回归出相对于 Anchor 的坐标。当这些 Anchor 与实际的物体匹配越高，PRN 能检测出物体的概率越大。本发明中采用多个 Header 的框架，可以对每个 RPN 层的物体的尺度和宽高比设计对应的 Anchor，从而提升每个 PRN 层的查全率。如图 9 所示。

对于并行 Header1，其负责人的检测，而人的主要形态是瘦长型的，因此，可以把 Anchor 设计为瘦长型；对于并行 Header4，其负责交通标志的检测，而交通标志的主要形态是正方形的，因此，可以把 Anchor 设计为正方形。

4、2D 候选区域特征提取流程

2D 候选区域特征提取流程主要由图 6 中的每个并行 Header 中的 ROI-ALIGN 模块实施，其根据 PRN 层提供的 Proposal 的坐标，在 Backbone 提供的一个特征图上把每个 Proposal 所在的特征抽取出来。ROI-ALIGN 的过程如图 10 所示。

本实施例中在 Backbone 的 C4 特征图上进行扣取特征，每个 Proposal 在 C4 上区域如图中的箭头所指的深色区域，在这个区域中采用插值和抽样的方法，扣取出固定分辨率的特征。假设 Proposal 的个数为 N，ROI-ALIGN 抽取出来的特征的宽高都为 14，则 ROI-ALIGN 输出的特征的大小为 $N*14*14*256$ (ROI-ALIGN 抽取出来的特征的通道数与 C4 的通道数相同，均为 256 个通道)。这些特征将会送到后续的 RCNN 模块进行细分。

5、2D 候选区域细分类

2D 候选区域细分类主要由图 6 中每个并行 Header 的 RCNN 模块实施，其根据 ROI-ALIGN 模块提取出来的每个 Proposal 的特征，进一步回归出更加紧致的 2D 框坐标，同时对这个 Proposal 进行分类，输出其属于各个类别的置信度。

RCNN 的可实现形式很多，其中一种实现形式如图 11 所示。下面对其进行分析。

ROI-ALIGN 模块输出的特征大小为 $N*14*14*256$ ，其在 RCNN 模块中首先经过 Resnet18 的第 5 个卷积模块 (Res18-Conv5) 处理，输出的特征大小为 $N*7*7*512$ ，然后通过一个 Global Avg Pool (平均池化层) 进行处理，把输入特征中每个通道内的 7*7 的特征进行平均，得到 $N*512$ 的特征，其中每个 1*512 维的特征向量代表每个 Proposal 的特征。接下来通过 2 个全连接层 FC 分别回归框的精确坐标(输出 $N*4$ 的向量，这 4 个数值分表表示框的中心点 x/y 坐标，框的宽高)，框的类别的置信度 (在 Header0 中，需要给出这个框

是 Background/Car/Truck/Bus 的分数)。最后通过框合并操作,选择分数最大的若干个框,并且通过 NMS 操作去除重复的框,从而得到紧致的框输出。

6、3D 检测流程

3D 检测流程由图 6 中的串行 3D_Header0 完成,其根据“2D 检测”流程提供的 2D 框以及 Backbone 提供的特征图,预测出每个 2D 框内部物体的质心点坐标、朝向角、长、宽、高等 3D 信息。串行 3D_Header 一种可能的实现方式如图 12 所示。

ROI-ALIGN 模块根据并行 Header 提供的准确的 2D 框,在 C4 上提取出每个 2D 框所在区域的特征,假设 2D 框的个数为 M ,那么 ROI-ALIGN 模块输出的特征大小为 $M*14*14*256$,其首先经过 Resnet18 的第 5 个卷积模块 (Res18-Conv5) 处理,输出的特征大小为 $N*7*7*512$,然后通过一个 Global Avg Pool (平均池化层) 进行处理,把输入特征中每个通道的 $7*7$ 的特征进行平均,得到 $M*512$ 的特征,其中每个 $1*512$ 维的特征向量代表每个 2D 框的特征。接下来通过 3 个全连接层 FC 分别回归框中物体的朝向角 (图中的 orientation, $M*1$ 向量)、质心点坐标 (图中的 centroid, $M*2$ 向量,这 2 个数值表示质心的 x/y 坐标) 和长宽高 (图中的 dimension)

7、Mask 检测流程

Mask 检测流程由图 6 中的串行 Mask_Header0 完成,其根据“2D 检测”流程提供的 2D 框以及 Backbone 提供的特征图,预测出每个 2D 框内部物体的精细掩膜。串行 Mask_Header 一种可能的实现方式如图 13 所示。

ROI-ALIGN 模块根据并行 Header 提供的准确的 2D 框,在 C4 上提取出每个 2D 框所在区域的特征,假设 2D 框的个数为 M ,那么 ROI-ALIGN 模块输出的特征大小为 $M*14*14*256$,其首先经过 Resnet18 的第 5 个卷积模块 (Res18-Conv5) 处理,输出的特征大小为 $N*7*7*512$,然后通过经过反卷积层 Deconv 进一步进行卷积,得到 $M*14*14*512$ 的特征,最后通过一个卷积,得到 $M*14*14*1$ 的 Mask 置信度输出。这个输出中每个 $14*14$ 的矩阵代表每个 2D 框中物体的掩膜的置信度,每个 2D 框被均等的划分为 $14*14$ 个区域,这个 $14*14$ 的矩阵就标志每个区域存在物体的可能性。对这个置信度矩阵进行阈值化处理 (比如大于阈值 0.5 的输出 1,否则输出 0) 就可以得到物体的掩膜。

8、Keypoint 检测流程

Keypoint 检测流程由图 6 中的串行 Keypoint_Header2 完成,其根据“2D 检测”流程提供的 2D 框以及 Backbone 提供的特征图,预测出每个 2D 框内部物体的关键点坐标。串行 Keypoint_Header 一种可能的实现方式如图 14 所示。

ROI-ALIGN 模块根据并行 Header 提供的准确的 2D 框,在 C4 上提取出每个 2D 框所在区域的特征,假设 2D 框的个数为 M ,那么 ROI-ALIGN 模块输出的特征大小为 $M*14*14*256$,其首先经过 Resnet18 的第 5 个卷积模块 (Res18-Conv5) 处理,输出的特征大小为 $N*7*7*512$,然后通过一个 Global Avg Pool 进行处理,把输入特征中每个通道的 $7*7$ 的特征进行平均,得到 $M*512$ 的特征,其中每个 $1*512$ 维的特征向量代表每个 2D 框的特征。接下来通过 1 个全连接层 FC 分别回归框中物体的关键点坐标 (图中的 Keypoint, $M*8$ 向量,这 8 个数值表示泊车位的 4 个角点的 x/y 坐标)。

基于表 2 的任务划分,本申请实施例还对该感知网络的训练过程做详细描述。

A. 训练数据的准备

根据表 2 的任务划分, 我们需要为每个任务提供标注数据。比如需要为 Header0 的训练提供车的标注数据, 在数据集上标注出 Car/Truck/Bus 的 2D 框以及类标签; 需要为 Header1 的训练提供人的标注数据, 在数据集上标注出 Pedestrian/Cyclist/Tricycle 的 2D 框以及类标签; 需要为 Header3 提供交通灯的标注数据, 在数据集上标注出 TrafficLight_Red/Yellow/Green/Black 的 2D 框以及类标签, 以此类推。

每一种数据只需要标注特定类型的物体即可, 这样可以进行针对性的采集, 而不用在每一张图片中把所有感兴趣的物体标注出来, 从而降低数据采集和标注的成本。另外, 采用这种方式准备数据具有很灵活的扩展性, 在增加检测物体检测类型的情况下, 只需要增加一个或者多个 Header, 并且提供新增物体的标注数据类型即可, 不需要在原有数据上把新增的物体标注出来。

另外, 为了训练 Header0 中的 3D 检测功能, 需要提供独立的 3D 标注数据, 在数据集上标注出每一辆车的 3D 信息(质心点坐标、朝向角、长、宽、高); 为了训练 Header0 中的 Mask 检测功能, 需要提供独立的 Mask 标注数据, 在数据集上标注出每一辆车的掩膜; 特别的, Header2 中的 Parkingslot 检测需要检测关键点, 这个任务要求数据集同时把停车位的 2D 框和关键点标注出来(实际上, 只需要把关键点标注出来即可, 车位的 2D 框可以通过关键点的坐标自动生成)

一般情况下只需要为每个任务提供独立的训练数据即可, 但是也可以提供混合标注的数据, 比如可以在数据集上同时标注 Car/Truck/Bus/Pedestrian/Cyclist/Tricycle 的 2D 框以及类标签, 这样就可以利用这个数据来同时训练 Header0 和 Header1 的并行 Header; 也可以在数据集上同时标注出 Car/Truck/Bus 的 2D/3D/Mask 数据, 这样这份数据就可以同时训练并行 Header0、串行 3D_Header0 和串行 Mask_Header0。

我们可以给每张图片指定一个标签, 这个标签决定这张图片可以用来训练网络中的哪些 Header, 这在后续的训练流程中会有详细的描述。

为了保证各个 Header 得到均等的训练机会, 需要对数据进行均衡, 具体就是把数量少的数据进行扩展, 扩展的方式包括但不限于复制扩展。对均衡后的数据进行随机打乱, 然后送入到网络中进行训练, 如图 15 所示。

B. 基于部分标注数据训练全功能网络

基于部分标注数据训练全功能网络的过程中根据每张输入图片所属的任务的类型, 计算相应的 Header 的 Loss, 并且通过这个 Loss 进行梯度回传, 并计算相应 Header 以及 Backbone 上的参数的梯度, 然后根据梯度对相应的 Header 以及 Backbone 进行调整。不在当前输入图片的标注任务中的 Header 则不进行调整。

如果一张图片只是标注一个任务的 2D 数据, 那么这张图片送入网络训练的时候, 只训练对应的一个并行 Header。如图 16 所示。

当前图片只标注了交通灯的 2D 框, 那么在训练的时候, 只通过并行 Header3 得到此输入图片的交通灯的预测结果, 并且和真值进行比较, 得到这个 Header 的损失代价 2D_Loss3。由于只产生一个代价损失, 总体的代价损失值 $Final\ Loss = 2D_Loss3$ 。也就是说, 交通灯的输入图片只是流过 Backbone 以及并行 Header3, 其余的 Header 并不参与到训练中, 如图

16 中的无“X”号的粗箭头所示。在得到 Final Loss 后，沿着图 16 中的无“X”号粗箭头的反方向计算并行 Header3 以及 Backbone 中的梯度，然后利用这个梯度更新 Header3 以及 Backbone 的参数，实现对网络的调整，使得网络更好得预测交通灯。

如果一张图片标注了多个任务的 2D 数据，那么这张图片送入网络训练的时候，则会训练对应的多个并行 Header。如图 17 所示。

当前图片同时标注了人和车的 2D 框，那么在训练的时候，就会通过并行 Header0 和并行 Header1 得到此输入图片的人车的预测结果，并且和真值进行比较，得到这 2 个 Header 的损失代价 2D_Loss0/2D_Loss1。由于产生多个代价损失，总体的代价损失值是各个损失的均值，即 $\text{Final Loss} = (2\text{D_Loss0} + 2\text{D_Loss1})/2$ 。也就是说，标注人车的输入图片只是流过 Backbone 以及并行 Header0/1，其余的 Header 并不参与到训练中，如图 17 中的无“X”号粗箭头所示。在得到 Final Loss 后，沿着图 17 中的无“X”号粗箭头的反方向计算并行 Header0/1 以及 Backbone 中的梯度，然后利用这个梯度更新 Header0/1 以及 Backbone 的参数，实现对网络的调整，使得网络更好得预测人车。

串行 Header 的训练需要独立的数据集来进行训练，需要下面以车的 3D 训练为例进行说明。如图 18 所示。

此时的输入图片标注了车的 2D 和 3D 的真值。在训练的时候，数据的流向如图中的无“X”号粗箭头所示，有“X”号粗箭头表示数据流不到的 Header。当这张图片送入网络后，会同时计算出来 2D 和 3D 的损失函数，得到最终的 $\text{Final Loss} = (2\text{D_Loss0} + 3\text{D_Loss0})/2$ 。然后沿着无“X”号粗箭头的反方向计算串行 3D Header0、并行 Header0 以及 Backbone 中的梯度，然后利用这个梯度更新其参数，实现对网络的调整，使得网络更好得预测车的 2D 和 3D。

每张图片送入网络中训练时只会调整对应的 Header 以及 Backbone，使得对应的任务表现更好，在这个过程中，其他任务性能会变差，但是后续轮到该任务的图片时，变差的 Header 又可以调整过来。由于所有任务的训练数据预先进行均衡，各个任务都得到均等的训练机会，因此不会出现某个任务过度训练的情况。通过这种训练方法，使得 Backbone 学习到了各个任务的共有的特征，而每个 Header 学习到其任务特定的特征。

目前感知需要实现的功能越来越多，使用多个网络来实现单点功能会导致总体计算量过大。本发明实施例提出一种基于 Multi-Header 的高性能可扩展的感知网络，各个感知任务共用相同的主干网络，成倍节省计算量和网络的参数量。表 3 给出了 Single-Header 网络实现单个功能的计算量和参数量统计。

Single-Header-Model@720p	GFlops	Parameters(M)
Vehicle(Car/Truck/Tram)	235.5	17.76
Vehicle+Mask+3D	235.6	32.49
Person(Pedestrian/Cyclist/Tricycle)	235.5	17.76
Person+Mask	235.6	23.0
Motocycle/Bicycle	235.5	17.76
TrafficLight(Red/Green/Yellow/Black)	235.6	17.76
TrafficSign(Trafficsign/Guideside/Billboard)	235.5	17.75

TrafficCone/TrafficStick/FireHydrant	235.5	17.75
Parkingslot (with keypoint)	235.6	18.98
全功能网络(多个 single-Header 网络)	1648.9	145.49

表 3 Single-Header 网络的计算量和参数量统计

从表中可以看出, 如果采用 8 个网络实现本实施例中的所有功能, 需要的总计算量为 1648.9 GFlops, 网络参数总量为 145.49M。这个计算量和网络参数量非常巨大, 将会给硬件带来很大的压力。

表 4 给出了采用一个 Multi-Header 的网络实现本实施例所有功能的计算量和参数量。

Multi-Header-Model@720p	GFlops	Parameters(M)
全功能网络(单个 Multi-Header 网络)	236.6	42.16

表 4 Multi-Header 网络的计算量和参数量统计

从表中可以看出, Multi-Header 网络的计算量和参数量仅为 Single-Header 的 1/7 和 1/3, 极大降低了计算消耗。

另外, Multi-Header 网络可以实现与 Single-Header 一样的检测性能, 表 5 给出了 Multi-Header 与 Single-Header 在部分类别上的性能比较。

类别	Single-Header	Multi-Header
Car	91.7	91.6
Tram	81.8	80.1
Pedestrian	73.6	75.2
Cyclist	81.8	83.3
TrafficLight	98.3	97.5
TrafficSign	95.1	94.5
Parkingslot (point precision/recall)	94.01/80.61	95.17/78.89
3D (mean_orien_err/mecentroid_dist_err)	2.95/6.78	2.88/6.34

表 5 Single-Header 与 Multi-Header 网络的检测性能对比

从表中可以看出, 两者的性能相当, 因此, Multi-Header 网络在节省计算量和显存的情况下不会导致性能下降。

本发明实施例提出一种基于 Multi-Header 的高性能可扩展感知网络, 在同一个网络上同时实现不同的感知任务(2D/3D/关键点/语义分割等), 该网络中各个感知任务共用相同的主干网络, 节省计算量, 网络结构易于扩展, 只需要通过增加一个 Header 来增加一个功能。另外, 本发明实施例还提出一种基于部分标注数据训练多任务感知网络的方法, 各个任务使用独立数据集, 不需要在同一张图片上进行全任务标注, 不同任务的训练数据方便进行均衡, 不同任务的数据不会产生相互的抑制。

如图 30 所示, 本发明实施例还提供一种物体检测方法, 该方法包括:

S3001, 接收输入的图片;

S3002, 对输入的图片进行卷积处理, 输出对应所述图片的具有不同分辨率的特征图;

S3003, 根据所述特征图, 针对不同的任务独立检测每个任务中的任务物体, 输出所述每个任务物体所在区域的 2D 框以及每个 2D 框对应的置信度; 其中, 所述任务物体为该任务中需要检测的物体; 所述置信度越高, 表示所述对应该置信度的 2D 框内存在所述任务所对应的物体的概率越大。

可选地, 在一个实施例中, S3002 可以包括如下 4 个步骤:

1、在一个或者多个特征图上预测所述任务物体所在的区域, 并输出匹配所述区域的候选 2D 框;

可选地, 可以基于所属任务对应的物体的模板框 (Anchor), 在主干网络提供的一个或者多个特征图上对存在该任务物体的区域进行预测以得到候选区域, 并输出匹配所述候选区域的候选 2D 框; 其中, 所述模板框是基于其所属的任务物体的统计特征得到的, 所述统计特征包括所述物体的形状和大小。

2、根据所述任务物体所在的区域, 从一个特征图中扣取出所述候选 2D 框所在区域的特征;

3、对所述候选 2D 框所在区域的特征进行卷积处理, 得到所述候选 2D 框属于各个物体类别的置信度; 所述各个物体类别为所述一个任务中的物体类别;

4、通过神经网络对所述候选区域 2D 框的坐标进行调整, 使得调整后的 2D 候选框比所述候选 2D 框与实际物体的形状更加匹配, 并选择置信度大于预设阈值的调整后的 2D 候选框作为所述区域的 2D 框。

可选地, 上述 2D 框可以为矩形框。

可选地, 所述方法还包括:

S3004, 基于所属任务的任务物体的 2D 框, 在主干网络上的一个或多个特征图上提取所述 2D 框所在区域的特征, 根据所述 2D 框所在区域的特征对所述所属任务的任务物体的 3D 信息、Mask 信息或 Keypoint 信息进行预测。

可选地, 可以在低分辨率的特征上完成大物体所在区域的检测, 所述 RPN 模块在高分辨率的特征图上完成小物体所在区域的检测。

如图 31 所示, 本发明实施例还提供一种基于部分标注数据训练多任务感知网络的方法, 所述方法包括:

S3101, 根据每张图片的标注数据类型, 确定每张图片所属的任务; 其中, 所述每张图片标注一个或者多个数据类型, 所述多个数据类型是所有数据类型的子集, 一个数据类型对应一个任务;

S3102, 根据每张图片所属的任务, 决定所述每张图片所需训练的 Header;

S3103, 计算每张图片所述所需训练的 Header 的损失值;

S3104, 对于每张图片, 通过所述所需训练的 Header 进行梯度回传, 并基于所述损失值调整所述所需训练的 Header 以及主干网络的参数。

可选地, 如图 31 的虚线框所示, 在步骤 S3102 之前, 所述方法还包括:

S31020, 对所属不同任务的图片进行数据均衡。

本发明实施例还提供一种基于 multi-Header 的物体感知方法, 本发明实施例提供的感知方法的流程包括“推理”流程和“训练”流程两部分。下面分别介绍。

一、感知流程:

本发明实施例提供的感知方法的流程如图 21 所示。

步骤 S210 中, 图片输入网络;

步骤 S220 中, 进入“基础特征生成”流程。

在该流程中, 图片通过图 5 的 Backbone 进行基础特征提取, 得到不同尺度的特征图。在基础特征生成后, 将会进入图 21 中虚线框所在的核心流程。在核心流程中, 每个任务具有独立的“2D 检测”流程以及可选的“3D 检测”流程、“Mask 检测”流程和“Keypoint 检测”流程。下面对核心流程进行描述。

1、2D 检测流程

“2D 检测”流程根据“基础特征生成”流程产生的特征图, 预测出每个任务的 2D 框以及置信度。具体来说, “2D 检测”流程又可以细分为“2D 候选区域预测”流程、“2D 候选区域特征提取”流程和“2D 候选区域细分类”流程, 如图 22 所示。

“2D 候选区域预测”流程由图 5 中的 RPN 模块实施, 其在“基础特征生成”流程提供的一个或者多个 feature map 上预测出可能存在该任务物体的区域, 并且给出这些区域的框 (Proposal)。

“2D 候选区域特征提取”流程由图 5 中的 ROI-ALIGN 模块实施, 其根据“2D 候选区域预测”流程提供的 Proposal, 在“基础特征生成”流程提供的一个 feature map 上把每个 Proposal 所在的区域的特征扣取出来, 并且 resize 到固定的大小, 得到每个 Proposal 的特征。

“2D 候选区域细分类”流程由图 5 中的 RCNN 模块实施, 其采用神经网络对每个 Proposal 的特征进行进一步预测, 输出每个 Proposal 的属于各个类别置信度, 同时对 Proposal 的 2D 框的坐标进行调整, 输出更加紧致的 2D 框。

2、3D 检测流程

“3D 检测”流程根据“2D 检测”流程提供的 2D 框以及“基础特征生成”流程产生的特征图, 预测出每个 2D 框内部物体的质心点坐标、朝向角和长宽高等 3D 信息。具体来说, “3D 检测”由两个子流程组成, 如图 23 所示。

子流程分析如下:

“2D 候选区域特征提取”流程由图 5 中的 ROI-ALIGN 模块实施, 其根据 2D 框的坐标, 在“基础特征生成”流程提供的一个 feature map 上把每个 2D 框所在的区域的特征扣取出来, 并且 resize 到固定的大小, 得到每个 2D 框的特征。

“3D 质心/朝向/长宽高预测”流程由图 5 中的 3D_Header 实施, 其主要根据每个 2D 框的特征, 回归出 2D 框内部物体的质心点坐标、朝向角度和长宽高等 3D 信息。

3、Mask 检测流程

“Mask 检测”流程根据“2D 检测”流程提供的 2D 框以及“基础特征生成”流程产生的特征图, 预测出每个 2D 框内部物体的细致的掩膜。具体来说, “Mask 检测”由两个子流程组成, 如图 24 所示。

子流程分析如下:

“2D 候选区域特征提取”流程由图 5 中的 ROI-ALIGN 模块实施，其根据 2D 框的坐标，在“基础特征生成”流程提供的一个 feature map 上把每个 2D 框所在的区域的特征扣取出来，并且 resize 到固定的大小，得到每个 2D 框的特征。

“掩膜预测”流程由图 5 中的 Mask_Header 实施，其主要根据每个 2D 框的特征，回归出 2D 框内部物体所在的掩膜。

4、Keypoint 检测流程

“Keypoint 预测”流程根据“2D 检测”流程提供的 2D 框以及“基础特征生成”流程产生的特征图，预测出每个 2D 框内部物体的掩膜。具体来说，“Keypoint 预测”由两个子流程组成，如图 25 所示。

子流程分析如下：

“2D 候选区域特征提取”流程由图 5 中的 ROI-ALIGN 模块实施，其根据 2D 框的坐标，在“基础特征生成”流程提供的一个 feature map 上把每个 2D 框所在的区域的特征扣取出来，并且 resize 到固定的大小，得到每个 2D 框的特征。

“关键点坐标预测”流程由图 5 中的 Keypoint_Header 实施，其主要根据每个 2D 框的特征，回归出 2D 框内部物体的关键点的坐标。

二、训练流程

本发明实施例的训练流程如图 26 所示。

其中红框部分为核心的训练流程。下面对核心训练流程进行介绍。

1、任务间数据均衡流程

各个任务的数据量是极度不均衡的，比如含有人的图片数量会比交通标志的大得多，为了让各个任务的 Header 得到均等的训练机会，必须对任务间的数据进行均衡。具体就是把数量少的数据进行扩展，扩展的方式包括但不限于复制扩展。

2、根据图片所属任务计算 Loss 流程

每张图片根据其标注的数据类型可以分属于一个或者多个任务，比如一张图片中仅标注交通标志，那么这张图片就只属于交通标志这个任务；如果一张图片同时标注了人和车，那么这种图片就同时属于人和车 2 个任务。在计算 Loss 的时候，只计算当前图片所属任务所对应的 Header 的 Loss，其余任务的 Loss 并不计算。比如，当前输入的训练图片属于人和车的任务，则此时仅计算人和车的 Header 的 Loss，其余的（比如交通灯、交通标志）的 Loss 不计算。

3、根据图片所属任务回传梯度流程

在计算 Loss 后，需要进行梯度回传。此时，仅通过当前任务的 Header 进行梯度回传，不在当前任务的 Header 不参与梯度回传。这样，就可以对当前图片调整当前的 Header，使得当前的 Header 更好的学习当前的任务。由于各个任务的数据已经经过均衡化处理，各个 Header 能够得到均等的训练机会，所以在这样反复调整过程中，各个 Header 学习到任务相关的特征，而 Backbone 则学习到任务间共用的特征。

本申请实施例综合考虑了现有方法的不足,提出一种基于 Multi-Header 的高性能可扩展感知网络,在同一个网络上同时实现不同的感知任务(2D/3D/关键点/语义分割等),该网络中各个感知任务共用相同的主干网络,显著节省计算量。另外,其网络结构易于扩展,只需要通过增加一个或者若干个 Header 来进行功能扩展。

此外,本申请实施例还提出一种基于部分标注数据训练多任务感知网络的方法,各个任务使用独立数据集,不需要在同一张图片上进行全任务标注,不同任务的训练数据方便进行均衡,不同任务的数据不会产生相互的抑制。

如图 5 所示的感知网络可以以图 27 中的结构来实现,图 27 是感知网络的一个应用系统示意图,图中所示,该感知网络 2000 包括至少一个处理器 2001,至少一个存储器 2002、至少一个通信接口 2003 以及至少一个显示设备 2004。处理器 2001、存储器 2002、显示设备 2004 和通信接口 2003 通过通信总线连接并完成相互间的通信。

通信接口 2003,用于与其他设备或通信网络通信,如以太网,无线接入网(radio access network ,RAN),无线局域网(wireless local area networks, WLAN)等。

存储器 2002 可以是只读存储器(read-only memory, ROM)或可存储静态信息和指令的其他类型的静态存储设备,随机存取存储器(random access memory, RAM)或者可存储信息和指令的其他类型的动态存储设备,也可以是电可擦可编程只读存储器(electrically erasable programmable read-only memory, EEPROM)、只读光盘(compact disc read-only memory, CD-ROM)或其他光盘存储、光碟存储(包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等)、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质,但不限于此。存储器可以是独立存在,通过总线与处理器相连接。存储器也可以和处理器集成在一起。

其中,所述存储器 2002 用于存储执行以上方案的应用程序代码,并由处理器 2001 来控制执行。所述处理器 2001 用于执行所述存储器 2002 中存储的应用程序代码。

存储器 2002 存储的代码可执行以上提供的一种基于 Multi-Header 的物体的感知方法。

显示设备 2004 用于显示待识别图像、该图像中感兴趣物体的 2D、3D、Mask、关键点等信息。

处理器 2001 还可以采用或者一个或多个集成电路,用于执行相关程序,以实现本申请实施例的基于 Multi-Header 的物体的感知方法或模型训练方法。

处理器 2001 还可以是一种集成电路芯片,具有信号的处理能力。在实现过程中,本申请的推荐方法的各个步骤可以通过处理器 2001 中的硬件的集成逻辑电路或者软件形式的指令完成。在实现过程中,本申请实施例的训练方法的各个步骤可以通过处理器 2001 中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器 2001 还可以是通用处理器、数字信号处理器(digital signal processing, DSP)、ASIC、现成可编程门阵列(field programmable gate array, FPGA)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。可以实现或者执行本申请实施例中的公开的各方法、步骤及模块框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬

件及软件模块组合执行完成。软件模块可以位于随机存储器，闪存、只读存储器，可编程只读存储器或者电可擦写可编程存储器、寄存器等领域成熟的存储介质中。该存储介质位于存储器 2002，处理器 2001 读取存储器 2002 中的信息，结合其硬件完成本申请实施例的物体感知方法或模型训练方法。

通信接口 2003 使用例如但不限于收发器一类的收发装置，来实现推荐装置或训练装置与其他设备或通信网络之间的通信。例如，可以通过通信接口 2003 获取待识别图片或者训练数据。

总线可包括在装置各个部件（例如，存储器 2002、处理器 2001、通信接口 2003、显示设备 2004）之间传送信息的通路。在一种可能的实施例中，处理器 2001 具体执行以下步骤：接收输入的图片；对输入的图片进行卷积处理，输出对应所述图片的具有不同分辨率的特征图；根据所述主干网络提供的特征图，针对不同的任务独立检测每个任务所对应的物体，输出所述每个任务所对应物体的候选区域的 2D 框以及每个 2D 框对应的置信度。

在一种可能的实施例中，在执行根据所述主干网络提供的特征图，针对不同的任务独立检测每个任务所对应的物体，输出所述每个任务所对应物体的候选区域的 2D 框以及每个 2D 框对应的置信度的步骤时，处理器 2001 具体执行如下步骤：在一个或者多个特征图上对存在该任务物体的区域进行预测以得到候选区域，并输出匹配所述候选区域的候选 2D 框；根据所述 RPN 模块得到的候选区域，从一个特征图中扣取出所述候选区域所在区域的特征；对所述候选区域的特征进行细化处理，得到所述候选区域对应各个物体类别的置信度；所述各个物体为对应的一个任务中的物体；对所述候选区域的坐标进行调整得到第二候选 2D 框，所述第二 2D 候选框比所述候选 2D 框与实际物体更加匹配，并选择置信度大于预设阈值的 2D 候选框作为候选区域的 2D 框。

在一种可能的实施例中，在执行在一个或者多个特征图上对存在该任务物体的区域进行预测以得到候选区域，并输出匹配所述候选区域的候选 2D 框时，处理器 2001 具体执行如下步骤：

基于所对应任务的物体的模板框（Anchor），在一个或者多个特征图上对存在该任务的物体的区域进行预测以得到候选区域，并输出匹配所述候选区域的候选 2D 框；其中：所述模板框是基于其所属任务物体的统计特征得到的，所述统计特征包括所述物体的形状和大小。

在一种可能的实施例中，处理器 2001 还执行如下步骤：

基于对应任务的物体的 2D 框，在主干网络上的一个或多个特征图上提取所述物体的特征，对所述物体的 3D、Mask 或 Keypoint 进行预测。

在一种可能的实施例中，在低分辨率的特征图上完成大物体的候选区域的检测，在高分辨率的特征图上完成小物体的候选区域的检测。

在一种可能的实施例中，所述 2D 框为矩形框。

可选地，如图 28 所示，感知网络的结构可以由服务器实现，服务器可以以图 28 中的结构来实现，该服务器 2110 包括至少一个处理器 2101，至少一个存储器 2102 以及至少一个通信接口 2103。处理器 2101、存储器 2102 和通信接口 2103 通过通信总线连接并完成相互间的通信。

通信接口 2103, 用于与其他设备或通信网络通信, 如以太网, RAN, WLAN 等。

存储器 2102 可以是 ROM 或可存储静态信息和指令的其他类型的静态存储设备, RAM 或者可存储信息和指令的其他类型的动态存储设备, 也可以是 EEPROM) CD-ROM 或其他光盘存储、光碟存储 (包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等)、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质, 但不限于此。存储器可以是独立存在, 通过总线与处理器相连接。存储器也可以和处理器集成在一起。

其中, 所述存储器 2102 用于存储执行以上方案的应用程序代码, 并由处理器 2101 来控制执行。所述处理器 2101 用于执行所述存储器 2102 中存储的应用程序代码。

存储器 2102 存储的代码可执行以上提供的一种基于 Multi-Header 的物体的感知方法。

处理器 2101 还可以采用或者一个或多个集成电路, 用于执行相关程序, 以实现本申请实施例的基于 Multi-Header 的物体的感知方法或模型训练方法。

处理器 2101 还可以是一种集成电路芯片, 具有信号的处理能力。在实现过程中, 本申请的推荐方法的各个步骤可以通过处理器 2101 中的硬件的集成逻辑电路或者软件形式的指令完成。在实现过程中, 本申请实施例的训练方法的各个步骤可以通过处理器 2101 中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器 2101 还可以是通用处理器、DSP、ASIC、FPGA 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。可以实现或者执行本申请实施例中的公开的各方法、步骤及模块框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成, 或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器, 闪存、只读存储器, 可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器 2102, 处理器 2101 读取存储器 2102 中的信息, 结合其硬件完成本申请实施例的物体感知方法或模型训练方法。

通信接口 2103 使用例如但不限于收发器一类的收发装置, 来实现推荐装置或训练装置与其他设备或通信网络之间的通信。例如, 可以通过通信接口 2103 获取待识别图片或者训练数据。

总线可包括在装置各个部件 (例如, 存储器 2102、处理器 2101、通信接口 2103) 之间传送信息的通路。在一种可能的实施例中, 处理器 2101 具体执行以下步骤: 在一个或者多个特征图上对存在该任务物体的区域进行预测以得到候选区域, 并输出匹配所述候选区域的候选 2D 框; 根据所述 RPN 模块得到的候选区域, 从一个特征图中扣取出所述候选区域所在区域的特征; 对所述候选区域的特征进行细化处理, 得到所述候选区域对应各个物体类别的置信度; 所述各个物体为对应的一个任务中的物体; 对所述候选区域的坐标进行调整得到第二候选 2D 框, 所述第二 2D 候选框比所述候选 2D 框与实际物体更加匹配, 并选择置信度大于预设阈值的 2D 候选框作为候选区域的 2D 框。

本申请提供一种计算机可读介质, 该计算机可读介质存储用于设备执行的程序代码, 该程序代码包括用于执行如图 21、22、23、24 或 25 所示实施例的物体感知方法的相关内容。

本申请提供一种计算机可读介质，该计算机可读介质存储用于设备执行的程序代码，该程序代码包括用于执行如图 26 所示实施例的训练方法的相关内容。

本申请提供一种包含指令的计算机程序产品，当该计算机程序产品在计算机上运行时，使得计算机执行上述如图 21、22、23、24 或 25 所示实施例的感知方法的相关内容。

本申请提供一种包含指令的计算机程序产品，当该计算机程序产品在计算机上运行时，使得计算机执行上述如图 26 所示实施例的训练方法的相关内容。

本申请提供一种芯片，所述芯片包括处理器与数据接口，所述处理器通过所述数据接口读取存储器上存储的指令，执行如图 21、22、23、24、25 或 26 所示实施例的感知方法的相关内容。

本申请提供一种芯片，所述芯片包括处理器与数据接口，所述处理器通过所述数据接口读取存储器上存储的指令，执行如图 26 所示实施例的训练方法的相关内容。

可选地，作为一种实现方式，所述芯片还可以包括存储器，所述存储器中存储有指令，所述处理器用于执行所述存储器上存储的指令，当所述指令被执行时，所述处理器用于执行如图 21、22、23、24 或 25 所示实施例的感知方法的相关内容，或者执行如图 26 所示实施例的训练方法的相关内容。

需要说明的是，对于前述的各方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本发明并不受所描述的动作顺序的限制，因为依据本发明，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本发明所必须的。

在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中未详述的部分，可以参见其他实施例的相关描述。

综上所述，本申请实施例的有益效果总结如下：

(1) 各个感知任务共用相同的主干网络，成倍节省计算量；网络结构易于扩展，只需要增加一个或者若干个 Header 就可以扩展 2D 的检测类型。每个并行 Header 具有独立的 RPN 和 RCNN 模块，仅需要检测其所属的任务的物体，这样在训练过程中，可以避免对未标注的其他任务的物体的误伤。另外，采用独立 RPN 层，可以针对每个任务的物体的尺度和宽高比定制专门的 Anchor，从而提高 Anchor 与物体的重合比例，进而提高 RPN 层对物体的查全率。

(2) 可以以灵活方便的方式，实现 3D、Mask、Keypoint 的检测功能。并且这些功能扩展与 2D 部分共用相同的主干网络，不会显著增加计算量。使用一个网络实现多项功能，易于在芯片上实现

(3) 各个任务使用独立的数据集，不需要在同一张图片上把所有任务标注出来，节省标注成本。任务扩展灵活简单，在增加新的任务的时候，只需要提供新任务的数据即可，不要在原来的数据上把新的物体标注出来。不同任务的训练数据方便进行均衡，使得各个任务得到均等的训练机会，避免数量大的数据淹没数量少的数据。

在本申请所提供的几个实施例中，应该理解到，所揭露的装置，可通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如所述单元的划分，仅仅为一种

逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性或其它的形式。

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单元来实现本实施例方案的目的。

另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

所述集成的单元如果以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器中。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全部或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器中，包括若干指令用以使得一台计算机设备（可为个人计算机、服务器或者网络设备）执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储器包括：U盘、ROM、RAM、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通过程序来指令相关的硬件来完成，该程序可以存储于一计算机可读存储器中，存储器可以包括：闪存盘、ROM、RAM、磁盘或光盘等。

以上对本申请实施例进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本发明的限制。

权利要求

1、一种基于多个头端(Header)的感知网络，其特征在于，所述感知网络包括主干网络和多个并行 Header，所述多个并行 Header 和所述主干网络连接；

所述主干网络，用于接收输入的图片，并对所述输入的图片进行卷积处理，输出对应所述图片的具有不同分辨率的特征图；

所述多个并行 Header 中的每个 Header，用于根据所述主干网络输出的特征图，对一个任务中的任务物体进行检测，输出所述任务物体所在区域的 2D 框以及每个 2D 框对应的置信度；其中，所述每个并行 Header 完成不同的任务物体的检测；其中，所述任务物体为该任务中需要检测的物体；所述置信度越高，表示所述对应该置信度的 2D 框内存在所述任务所对应的任务物体的概率越大。

2、根据权利要求 1 所述的感知网络，其特征在于，所述每个并行头端包括候选区域生成网络(RPN)模块、感兴趣区域提取(ROI-ALIGN)模块和区域卷积神经网络(RCNN)模块，所述每个并行头端的 RPN 模块独立于其它并行头端的 RPN 模块；所述每个并行头端的 ROI-ALIGN 模块独立于其它并行头端的 ROI-ALIGN 模块；所述每个并行头端的 RCNN 模块独立于其它并行头端的 RCNN 模块，其中，对于所述每个并行头端：

所述 RPN 模块用于：在所述主干网络提供的一个或者多个特征图上预测所述任务物体所在的区域，并输出匹配所述区域的候选 2D 框；

所述 ROI-ALIGN 模块用于：根据所述 RPN 模块预测得到的区域，从所述主干网络提供的一个特征图中扣取出所述候选 2D 框所在区域的特征；

所述 RCNN 模块用于：通过神经网络对所述候选 2D 框所在区域的特征进行卷积处理，得到所述候选 2D 框属于各个物体类别的置信度；所述各个物体类别为所述并行头端对应的任务中的物体类别；通过神经网络对所述候选区域 2D 框的坐标进行调整，使得调整后的 2D 候选框比所述候选 2D 框与实际物体的形状更加匹配，并选择置信度大于预设阈值的调整后的 2D 候选框作为所述区域的 2D 框。

3、根据权利要求 1 或 2 所述的感知网络，其特征在于，所述 2D 框为矩形框。

4、根据权利要求 2 所述的感知网络，其特征在于，

所述 RPN 模块用于：基于所属任务对应的物体的模板框(Anchor)，在主干网络提供的一个或者多个特征图上对存在该任务物体的区域进行预测以得到候选区域，并输出匹配所述候选区域的候选 2D 框；其中，所述模板框是基于其所属的任务物体的统计特征得到的，所述统计特征包括所述物体的形状和大小。

5、根据权利要求 1-4 任一项所述的感知网络，其特征在于，所述感知网络还包括一个或多个串行 Header；所述串行 Header 与一个所述并行 Header 连接；

所述串行 Header 用于：利用其连接的所述并行 Header 提供的所属任务的任务物体的 2D 框，在所述主干网络上的一个或多个特征图上提取所述 2D 框所在区域的特征，根据所述 2D 框所在区域的特征对所述所属任务的任务物体的 3D 信息、Mask 信息或 Keypoint 信息进行预测。

6、根据权利要求 1-4 任一项所述的感知网络，其特征在于，所述 RPN 模块用于在不

同的分辨率的特征图上预测不同大小物体所在的区域。

7、根据权利要求 6 所述的感知网络，其特征在于，所述 RPN 模块用于在低分辨率的特征图上完成大物体所在区域的检测，在高分辨率的特征图上完成小物体所在区域的检测。

8、一种物体检测方法，其特征在于，所述方法包括：

接收输入的图片；

对所述输入的图片进行卷积处理，输出对应所述图片的具有不同分辨率的特征图；

根据所述特征图，针对不同的任务独立检测每个任务中的任务物体，输出所述每个任务物体所在区域的 2D 框以及每个 2D 框对应的置信度；其中，所述任务物体为该任务中需要检测的物体；所述置信度越高，表示对应所述该置信度的 2D 框内存在所述任务所对应的任务物体的概率越大。

9、根据权利要求 8 所述的物体检测方法，其特征在于，所述根据所述特征图，针对不同的任务独立检测每个任务中的任务物体，输出所述每个任务物体所在区域的 2D 框以及每个 2D 框对应的置信度，包括：

在一个或者多个特征图上预测所述任务物体所在的区域，并输出匹配所述区域的候选 2D 框；

根据所述任务物体所在的区域，从一个特征图中扣取出所述候选 2D 框所在区域的特征；

对所述候选 2D 框所在区域的特征进行卷积处理，得到所述候选 2D 框属于各个物体类别的置信度；所述各个物体类别为所述一个任务中的物体类别；

通过神经网络对所述候选区域 2D 框的坐标进行调整，使得调整后的 2D 候选框比所述候选 2D 框与实际物体的形状更加匹配，并选择置信度大于预设阈值的调整后的 2D 候选框作为所述区域的 2D 框。

10、根据权利要求 9 所述的物体检测方法，其特征在于，所述 2D 框为矩形框。

11、根据权利要求 9 所述的物体检测方法，其特征在于，在一个或者多个特征图上预测所述任务物体所在的区域，并输出匹配所述区域的候选 2D 框为：

基于所属任务对应的物体的模板框（Anchor），在所述主干网络提供的一个或者多个特征图上对存在该任务物体的区域进行预测以得到候选区域，并输出匹配所述候选区域的候选 2D 框；其中，所述模板框是基于其所属的任务物体的统计特征得到的，所述统计特征包括所述物体的形状和大小。

12、根据权利要求 8-11 任一项所述的物体检测方法，其特征在于，所述方法还包括：

基于所属任务的任务物体的 2D 框，在主干网络上的一个或多个特征图上提取所述 2D 框所在区域的特征，根据所述 2D 框所在区域的特征对所述所属任务的任务物体的 3D 信息、Mask 信息或 Keypoint 信息进行预测。

13、根据权利要求 8-12 任一项所述的物体检测方法，其特征在于，在低分辨率的特征图上完成大物体所在区域的检测，在高分辨率的特征图上完成小物体所在区域的检测。

14、一种基于部分标注数据训练多任务感知网络的方法，其特征在于，所述感知网络包括主干网络和多个并行头端(Header)，所述方法包括：

根据每张图片的标注数据类型，确定每张图片所属的任务；其中，所述每张图片标注

一个或者多个数据类型，所述多个数据类型是所有数据类型的子集，所述所有数据类型中的每一个数据类型对应一个任务；

根据所述每张图片所属的任务，决定所述每张图片所需训练的 Header；

计算所述每张图片所需训练的 Header 的损失值；

对于所述每张图片，通过所述所需训练的 Header 进行梯度回传，并基于所述损失值调整所述所需训练的 Header 以及所述主干网络的参数。

15、如权利要求 14 所述的训练多任务感知网络的方法，其特征在于，计算所述每张图片所需训练的 Header 的损失值之前，所述方法还包括：

对属于不同任务的图片进行数据均衡。

16、一种基于部分标注数据训练多任务感知网络的装置，其特征在于，所述感知网络包括主干网络和多个并行头端(Header)，所述装置包括：

任务确定模块，用于根据每张图片的标注数据类型，确定每张图片所属的任务；其中，所述每张图片标注一个或者多个数据类型，所述多个数据类型是所有数据类型的子集，所述所有数据类型中的每一个数据类型对应一个任务；

Header 决定模块，用于根据所述每张图片所属的任务，决定所述每张图片所需训练的 Header；

损失值计算模块，针对所述每张图片，用于计算所述 Header 决定模块决定出的 Header 的损失值；

调整模块，针对所述每张图片，通过计算所述 Header 决定模块决定出的 Header 进行梯度回传，并基于所述损失值计算模块得到的损失值调整所述所需训练的 Header 以及主干网络的参数。

17、如权利计 16 所述的训练多任务感知网络的方法，其特征在于，所述装置还包括：数据均衡模块，用于对属与不同任务的图片进行数据均衡。

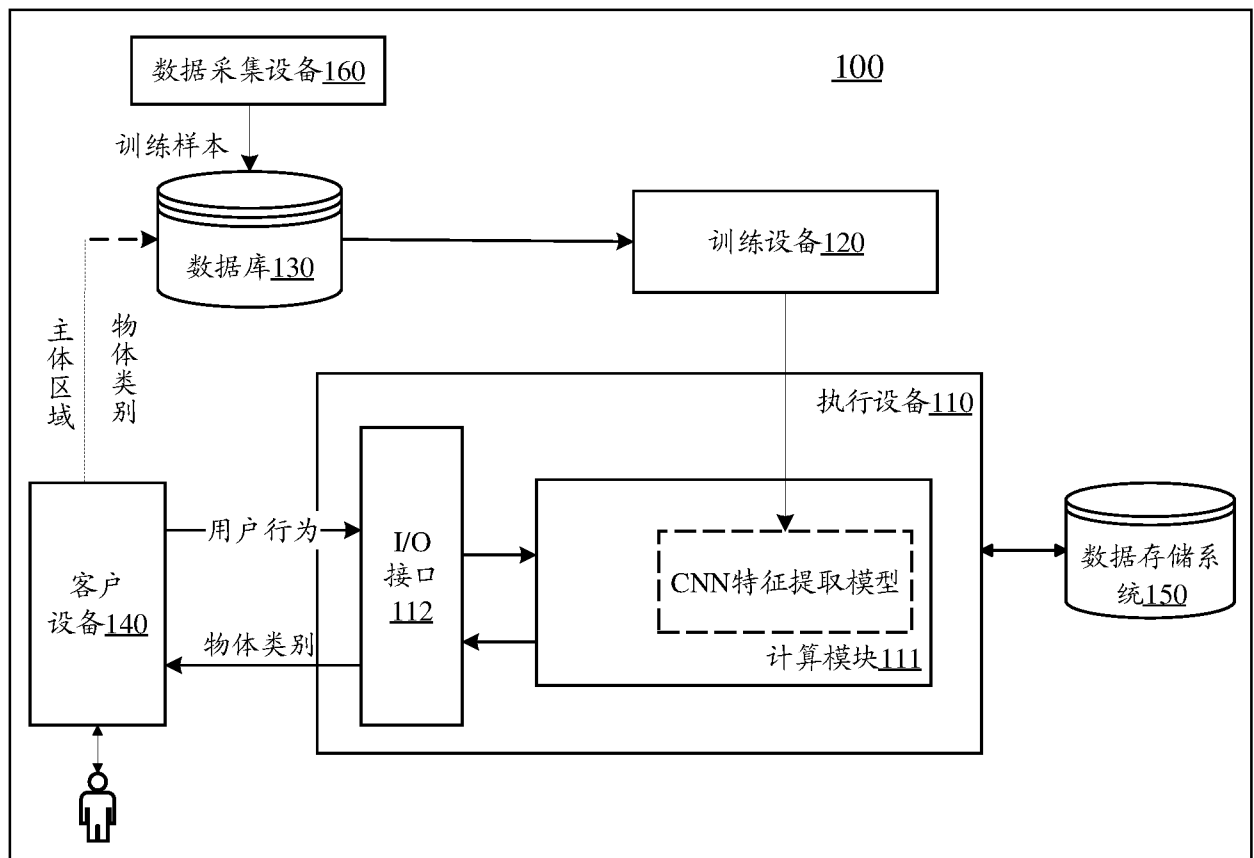


图 1

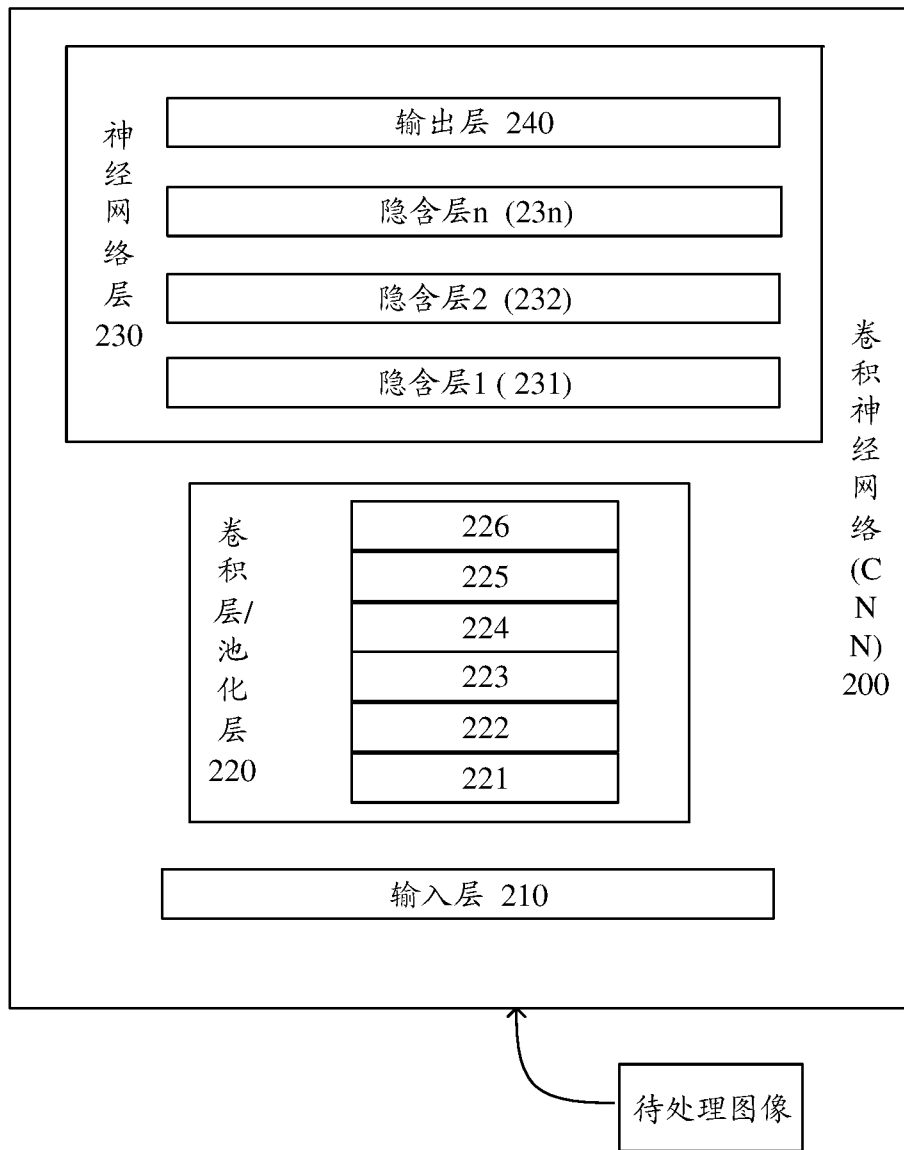


图 2

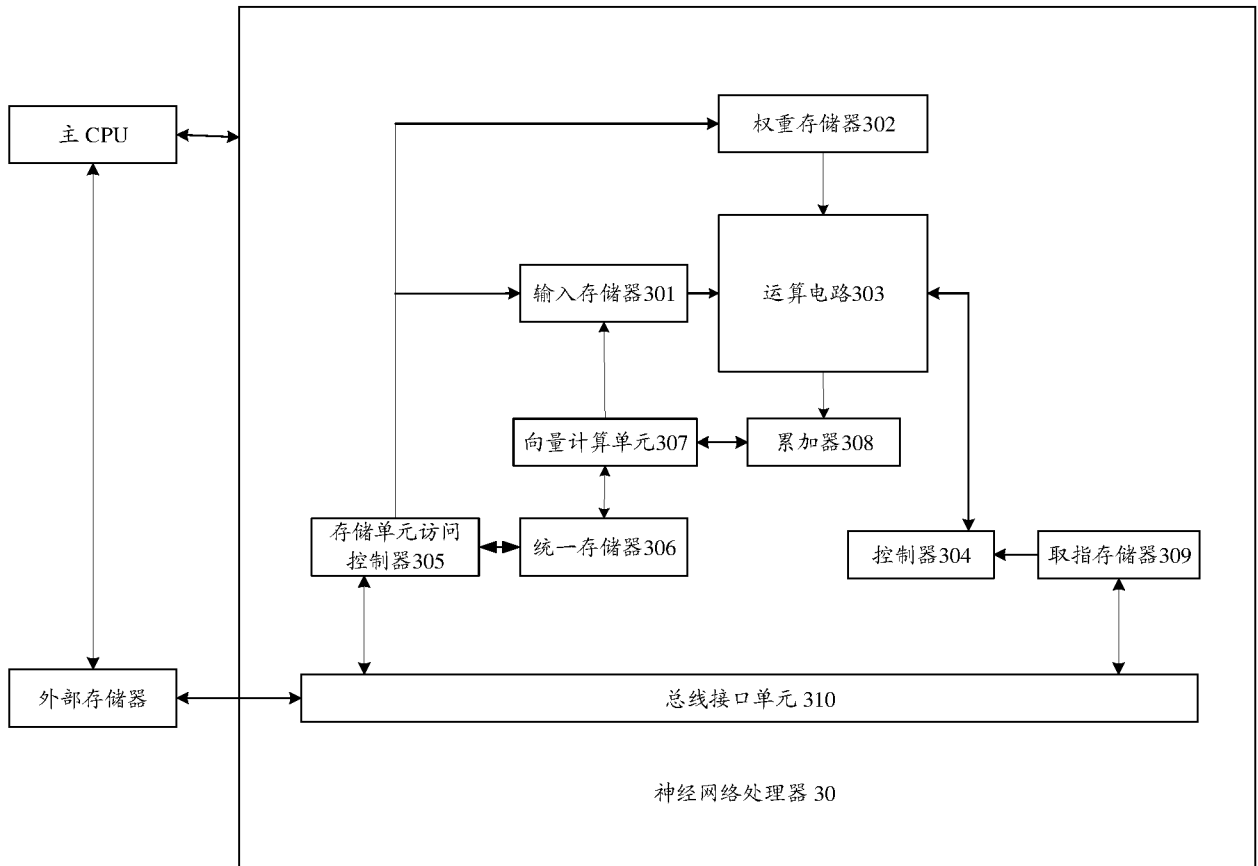


图 3

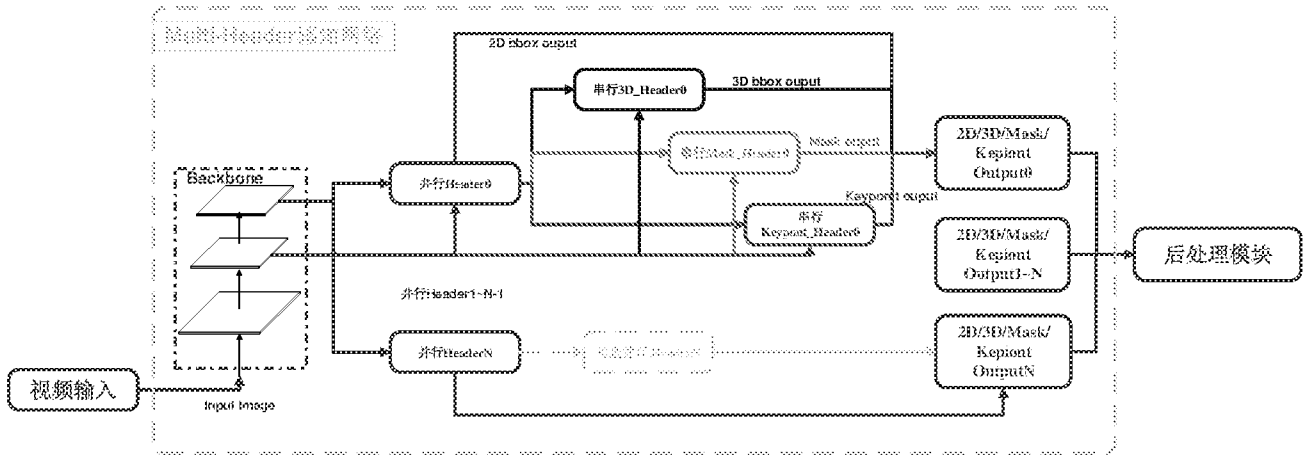


图 4

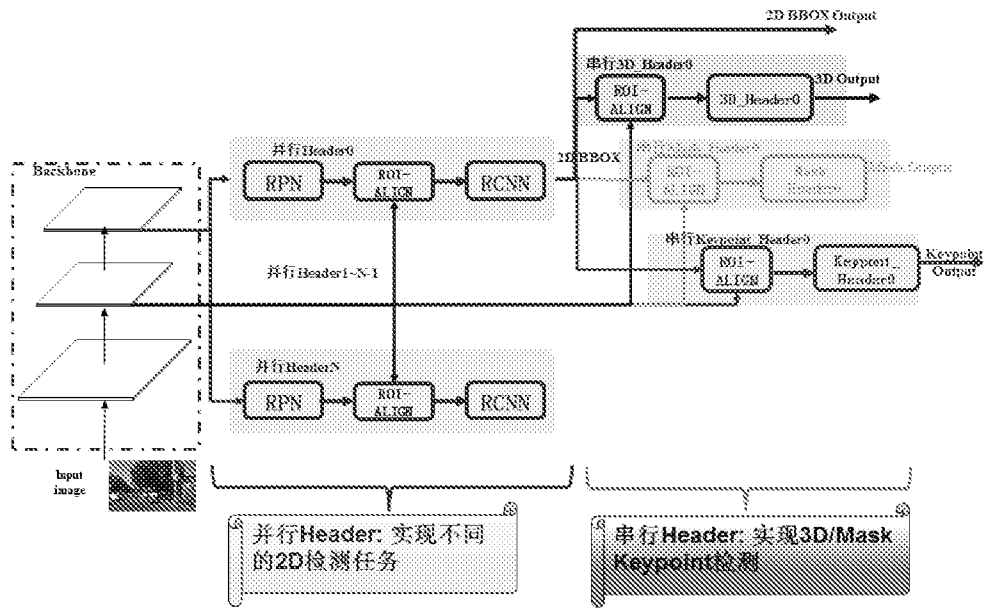


图 5

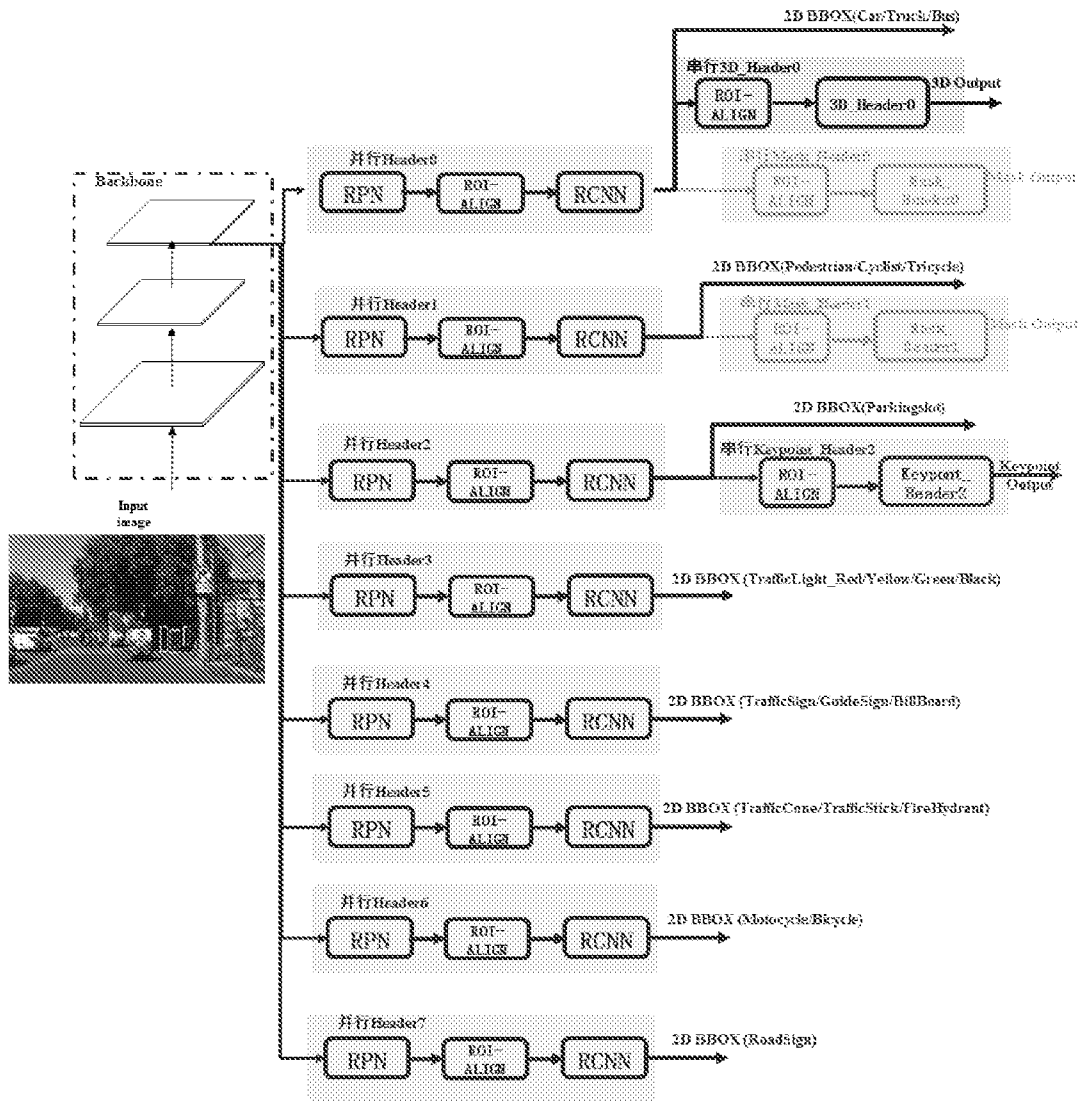


图 6

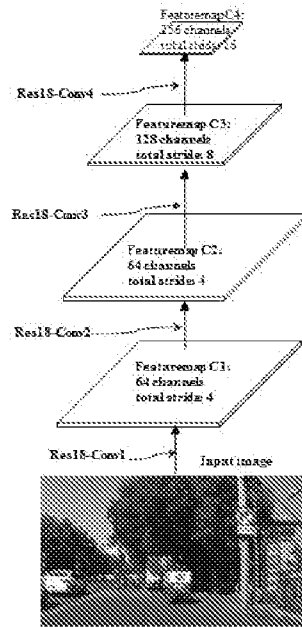


图 7

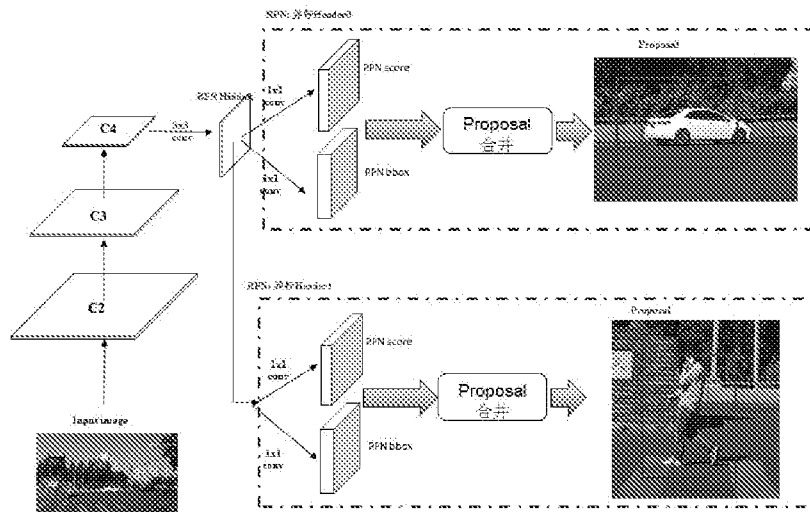


图 8

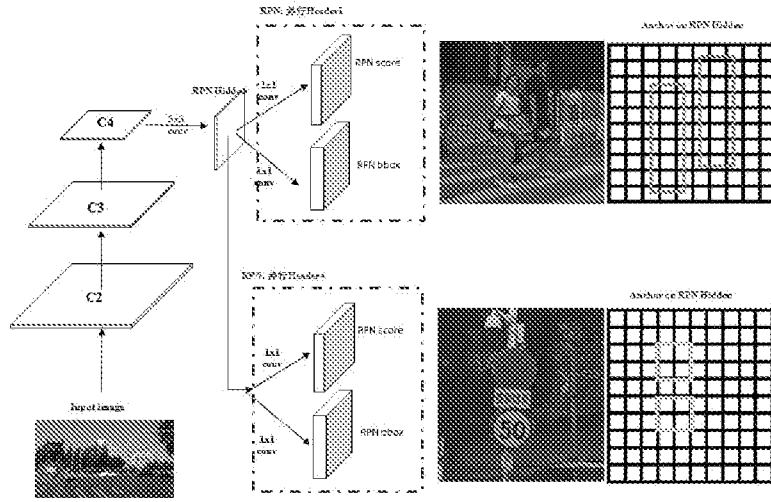


图 9

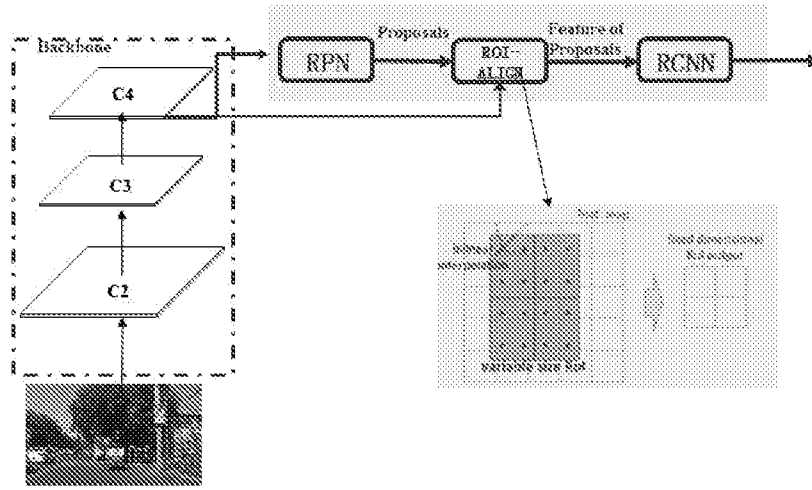


图 10

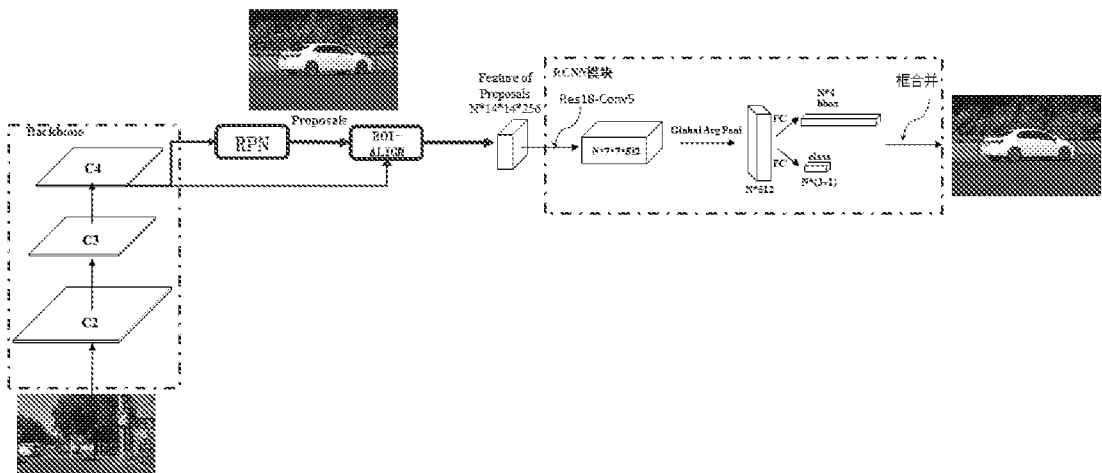


图 11

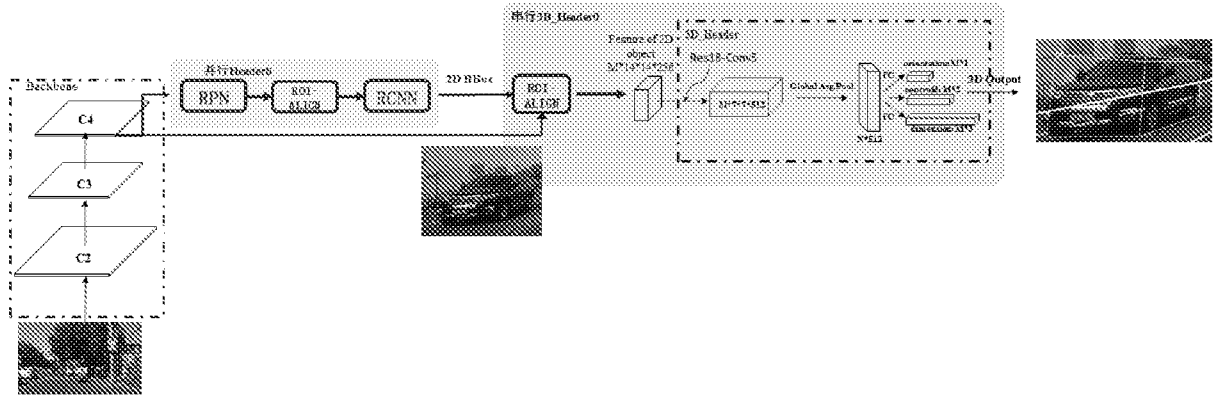


图 12

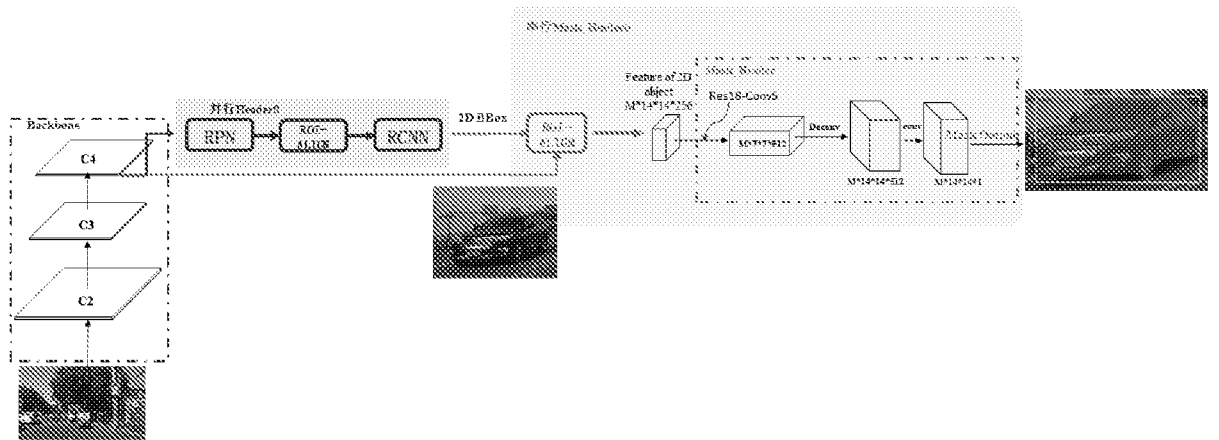


图 13

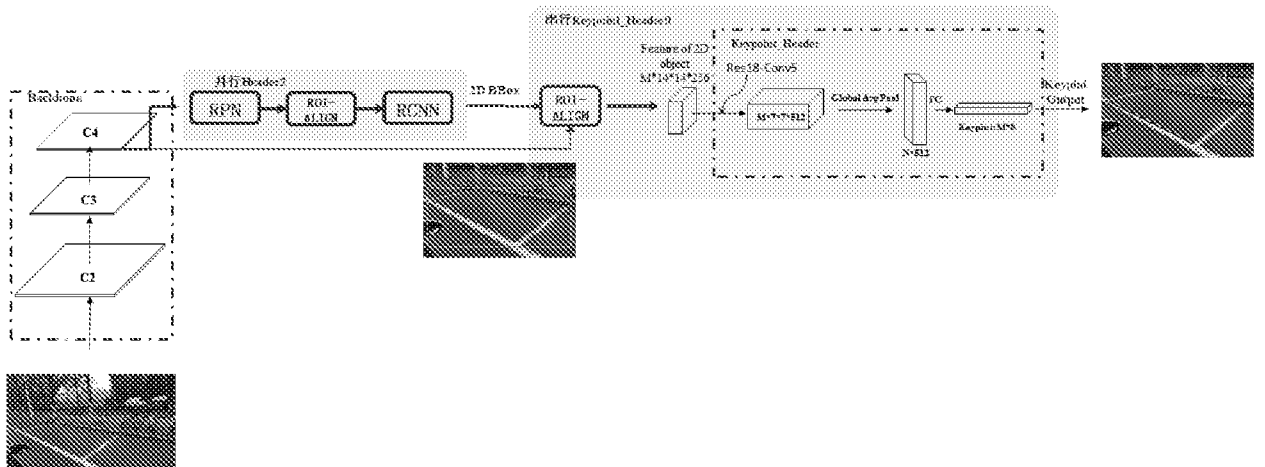


图 14

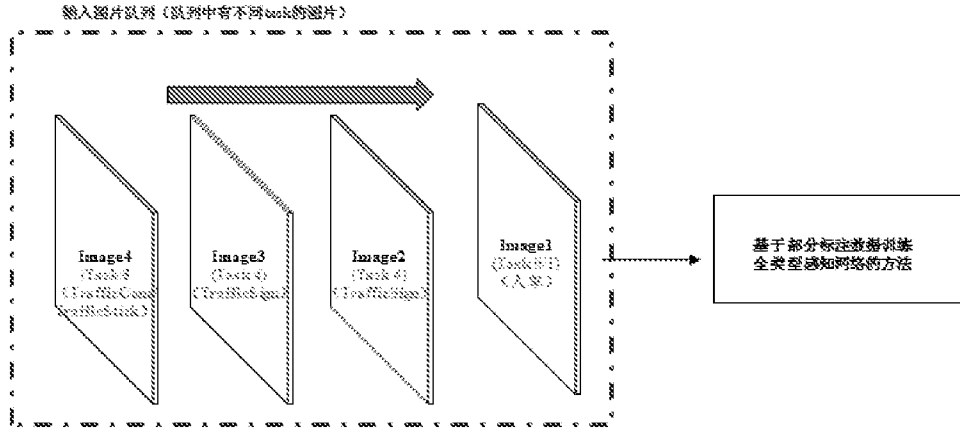


图 15

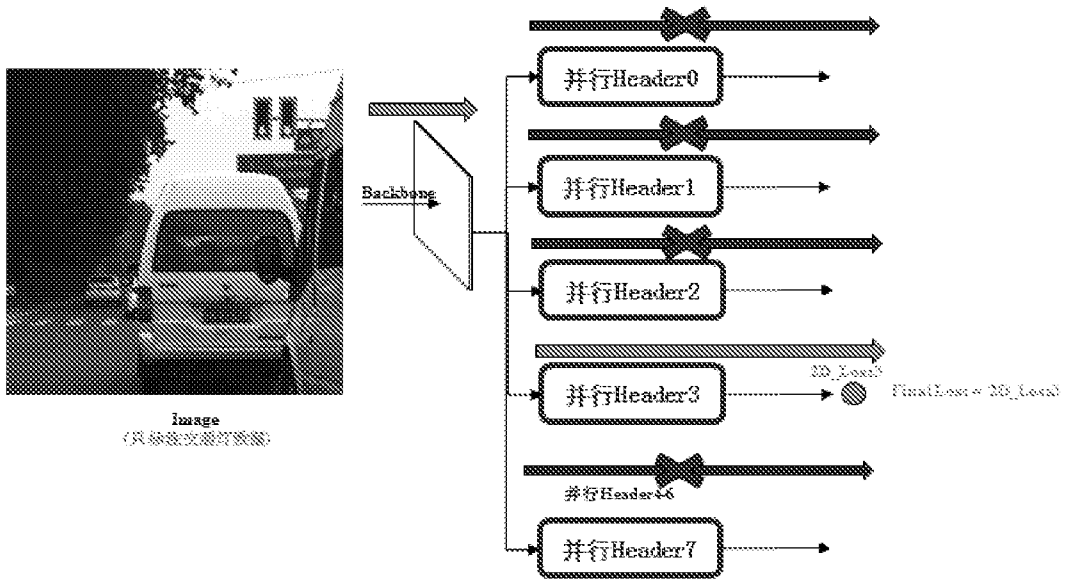


图 16

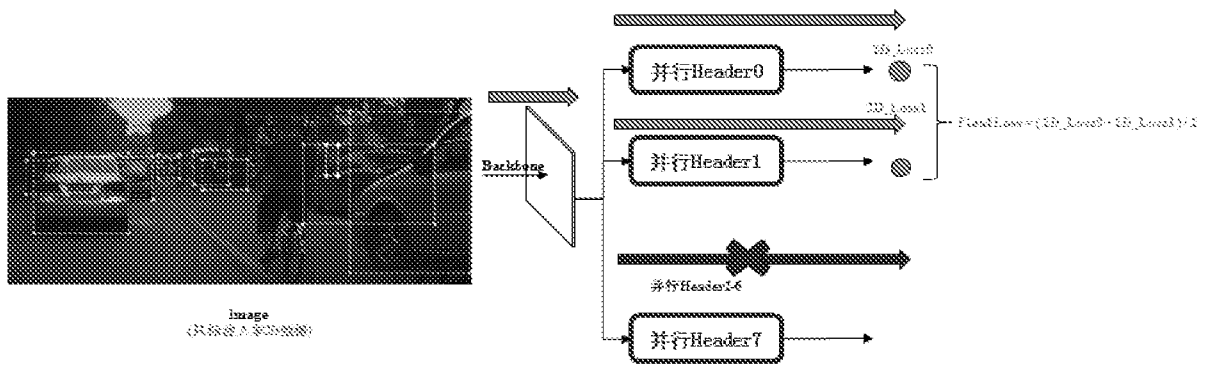


图 17

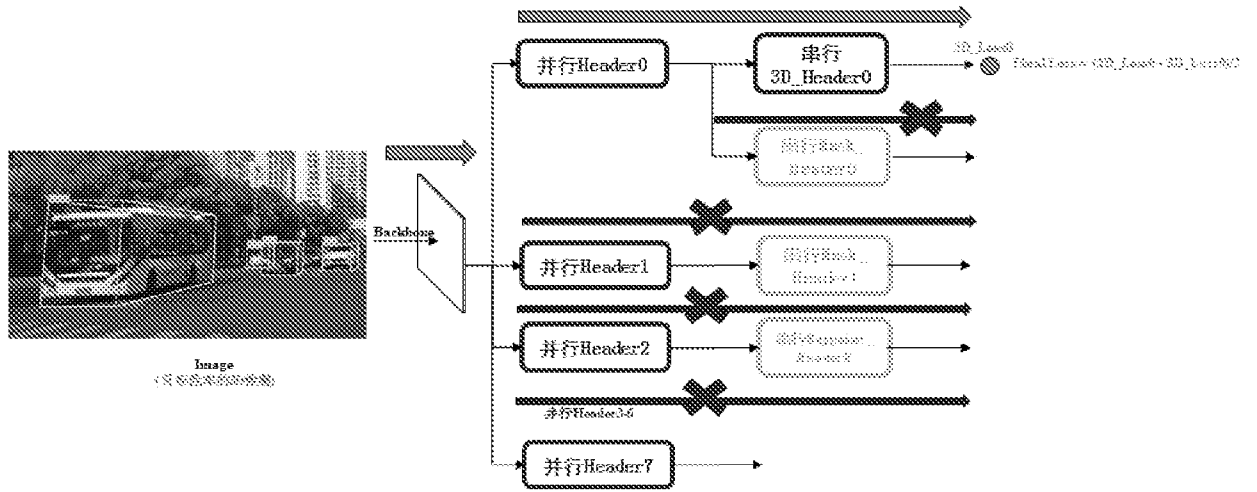


图 18

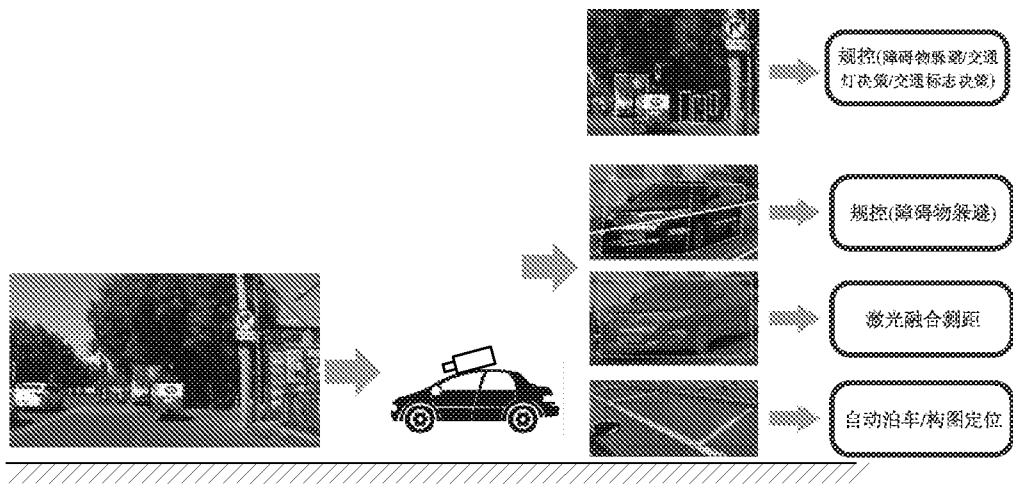


图 19

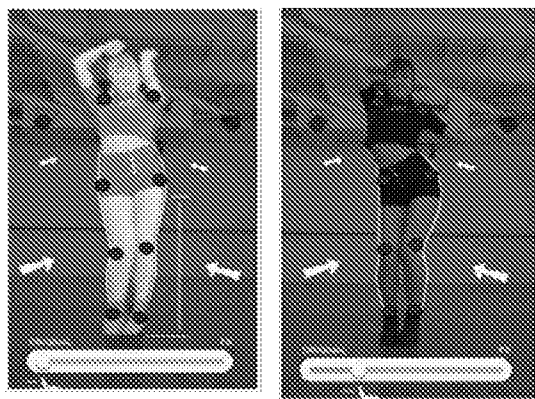


图 20

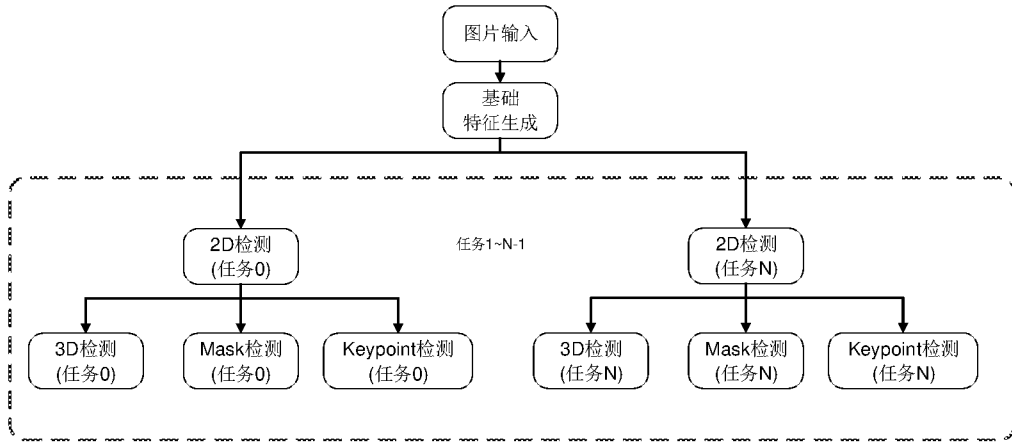


图 21

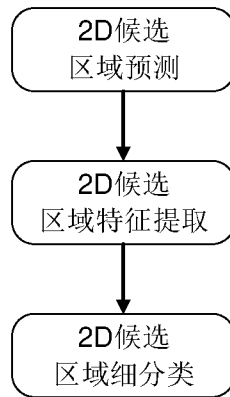


图 22

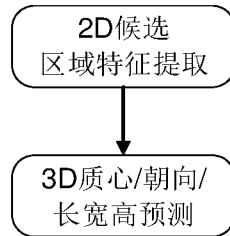


图 23

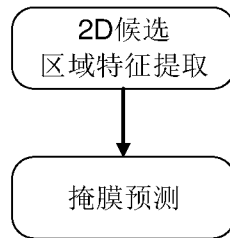


图 24

13/15

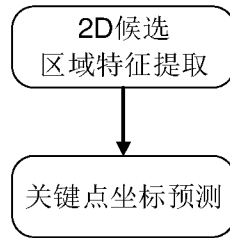


图 25

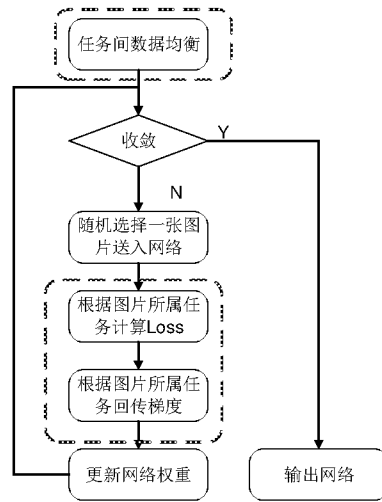


图 26

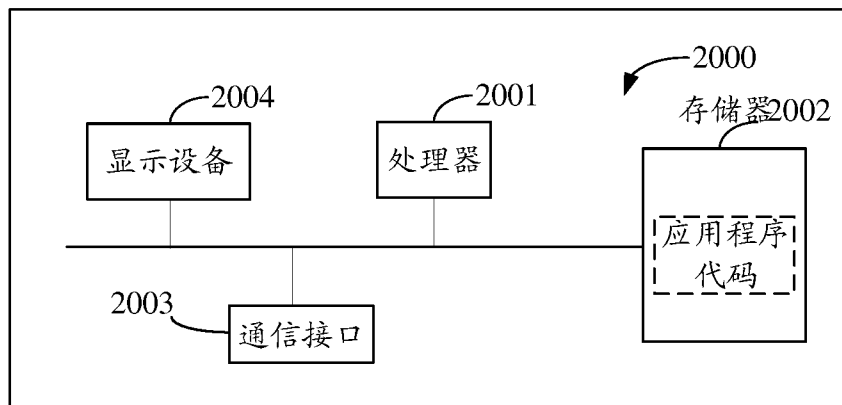


图 27

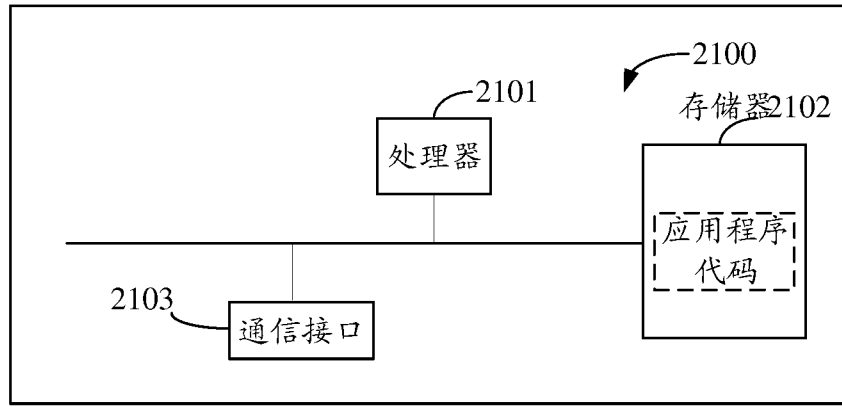


图 28

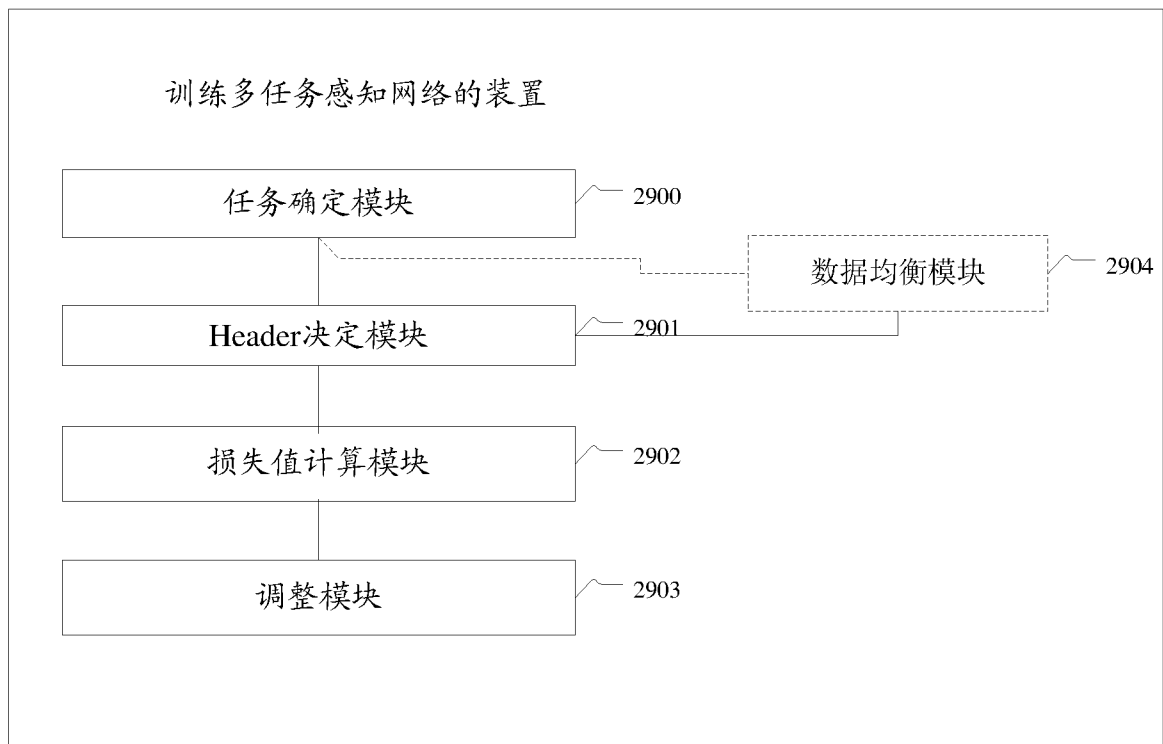


图 29

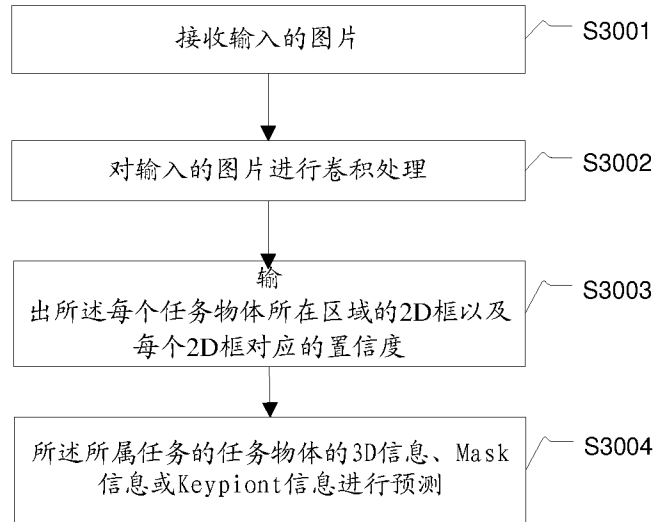


图 30

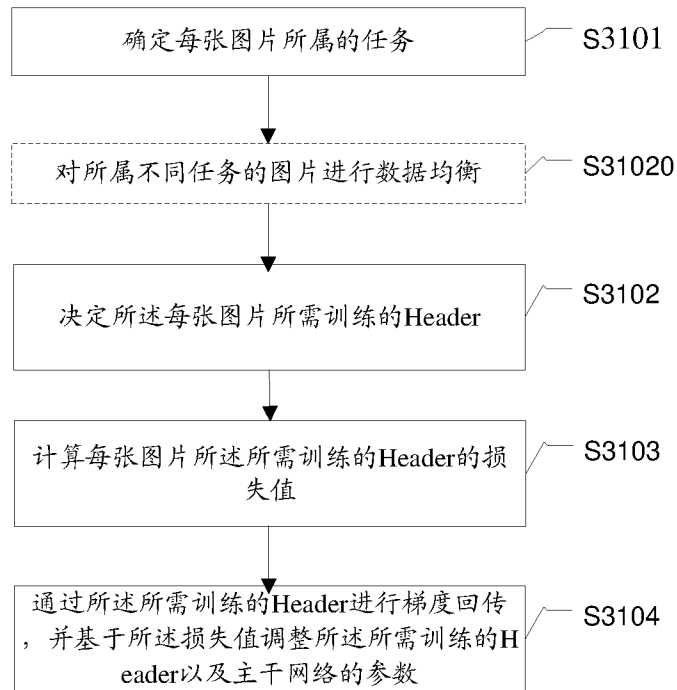


图 31

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/094803

A. CLASSIFICATION OF SUBJECT MATTER		
G06K 9/00(2006.01)i; G06N 3/04(2006.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) G06K; G06N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNPAT, CNKI, WPI, EPODOC, IEEE: 骨干网络, 主干网络, 特征, 提取, 检测, 头部, 端部, 多, 不同, 分辨率, 尺度, 尺寸, backbone, head, Header, detect+, RPN, ROIAlign, ROI w Align, R w CNN, RCNN, region, interest, convolutional, neural, network, resolution, scale, receptive w field, RF, multi, different		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110298262 A (HUAWEI TECHNOLOGIES CO., LTD.) 01 October 2019 (2019-10-01) claims 1-17	1-17
X	CN 109784194 A (SHANGHAI TUSEN FUTURE ARTIFICIAL INTELLIGENCE TECHNOLOGY CO., LTD.) 21 May 2019 (2019-05-21) description paragraphs 0075-0099, 0133-0137, figures 4B, 6-7	1, 3, 6-8, 10, 13-17
Y	CN 109784194 A (SHANGHAI TUSEN FUTURE ARTIFICIAL INTELLIGENCE TECHNOLOGY CO., LTD.) 21 May 2019 (2019-05-21) description paragraphs 0075-0099, 0133-0137, figures 4B, 6-7	2, 4, 5, 9, 11, 12
Y	CN 109712118 A (WUHAN SANJIANG CLP TECHNOLOGY CO., LTD.) 03 May 2019 (2019-05-03) description paragraphs 0037, 0058, figures 2, 5	2, 4, 5, 9, 11, 12
A	CN 108520229 A (BEIJING MEGVII TECHNOLOGY LTD.) 11 September 2018 (2018-09-11) entire document	1-17
A	CN 109815922 A (CASCO SIGNAL LTD.) 28 May 2019 (2019-05-28) entire document	1-17
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 28 August 2020		Date of mailing of the international search report 11 September 2020
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088 China		Authorized officer
Facsimile No. (86-10)62019451		Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/094803

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 10223610 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 05 March 2019 (2019-03-05) entire document	1-17
A	US 2017124409 A1 (NEC LABORATORLES AMERICA, INC.) 04 May 2017 (2017-05-04) entire document	1-17

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/094803

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110298262	A	01 October 2019	None			
CN	109784194	A	21 May 2019	None			
CN	109712118	A	03 May 2019	None			
CN	108520229	A	11 September 2018	None			
CN	109815922	A	28 May 2019	None			
US	10223610	B1	05 March 2019	None			
US	2017124409	A1	04 May 2017	WO	2017079521	A1	11 May 2017
				DE	112016005062	T5	12 July 2018
				JP	2018538612	A	27 December 2018

国际检索报告

国际申请号

PCT/CN2020/094803

<p>A. 主题的分类</p> <p>G06K 9/00(2006.01)i; G06N 3/04(2006.01)i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																							
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06K; G06N</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNPAT, CNKI, WPI, EPDOC, IEEE: 骨干网络, 主干网络, 特征, 提取, 检测, 头部, 端部, 多, 不同, 分辨率, 尺度, 尺寸, backbone, head, Header, detect+, RPN, ROIAlign, ROI w Align, R w CNN, RCNN, region, interest, convolutional, neural, network, resolution, scale, receptive w field, RF, multi, different</p>																							
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 110298262 A (华为技术有限公司) 2019年 10月 1日 (2019 - 10 - 01) 权利要求1-17</td> <td>1-17</td> </tr> <tr> <td>X</td> <td>CN 109784194 A (上海图森未来人工智能科技有限公司) 2019年 5月 21日 (2019 - 05 - 21) 说明书第0075-0099, 0133-0137段, 附图4B、6-7</td> <td>1, 3, 6-8, 10, 13-17</td> </tr> <tr> <td>Y</td> <td>CN 109784194 A (上海图森未来人工智能科技有限公司) 2019年 5月 21日 (2019 - 05 - 21) 说明书第0075-0099, 0133-0137段, 附图4B、6-7</td> <td>2, 4, 5, 9, 11, 12</td> </tr> <tr> <td>Y</td> <td>CN 109712118 A (武汉三江中电科技有限责任公司) 2019年 5月 3日 (2019 - 05 - 03) 说明书第0037, 0058段, 附图2、5</td> <td>2, 4, 5, 9, 11, 12</td> </tr> <tr> <td>A</td> <td>CN 108520229 A (北京旷视科技有限公司) 2018年 9月 11日 (2018 - 09 - 11) 全文</td> <td>1-17</td> </tr> <tr> <td>A</td> <td>CN 109815922 A (卡斯柯信号有限公司) 2019年 5月 28日 (2019 - 05 - 28) 全文</td> <td>1-17</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 110298262 A (华为技术有限公司) 2019年 10月 1日 (2019 - 10 - 01) 权利要求1-17	1-17	X	CN 109784194 A (上海图森未来人工智能科技有限公司) 2019年 5月 21日 (2019 - 05 - 21) 说明书第0075-0099, 0133-0137段, 附图4B、6-7	1, 3, 6-8, 10, 13-17	Y	CN 109784194 A (上海图森未来人工智能科技有限公司) 2019年 5月 21日 (2019 - 05 - 21) 说明书第0075-0099, 0133-0137段, 附图4B、6-7	2, 4, 5, 9, 11, 12	Y	CN 109712118 A (武汉三江中电科技有限责任公司) 2019年 5月 3日 (2019 - 05 - 03) 说明书第0037, 0058段, 附图2、5	2, 4, 5, 9, 11, 12	A	CN 108520229 A (北京旷视科技有限公司) 2018年 9月 11日 (2018 - 09 - 11) 全文	1-17	A	CN 109815922 A (卡斯柯信号有限公司) 2019年 5月 28日 (2019 - 05 - 28) 全文	1-17
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 110298262 A (华为技术有限公司) 2019年 10月 1日 (2019 - 10 - 01) 权利要求1-17	1-17																					
X	CN 109784194 A (上海图森未来人工智能科技有限公司) 2019年 5月 21日 (2019 - 05 - 21) 说明书第0075-0099, 0133-0137段, 附图4B、6-7	1, 3, 6-8, 10, 13-17																					
Y	CN 109784194 A (上海图森未来人工智能科技有限公司) 2019年 5月 21日 (2019 - 05 - 21) 说明书第0075-0099, 0133-0137段, 附图4B、6-7	2, 4, 5, 9, 11, 12																					
Y	CN 109712118 A (武汉三江中电科技有限责任公司) 2019年 5月 3日 (2019 - 05 - 03) 说明书第0037, 0058段, 附图2、5	2, 4, 5, 9, 11, 12																					
A	CN 108520229 A (北京旷视科技有限公司) 2018年 9月 11日 (2018 - 09 - 11) 全文	1-17																					
A	CN 109815922 A (卡斯柯信号有限公司) 2019年 5月 28日 (2019 - 05 - 28) 全文	1-17																					
<p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p> <table border="0"> <tr> <td> <p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件</p> </td> <td> <p>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</p> <p>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</p> <p>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</p> <p>“&” 同族专利的文件</p> </td> </tr> </table>			<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件</p>	<p>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</p> <p>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</p> <p>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</p> <p>“&” 同族专利的文件</p>																			
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件</p>	<p>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</p> <p>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</p> <p>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</p> <p>“&” 同族专利的文件</p>																						
<p>国际检索实际完成的日期</p> <p>2020年 8月 28日</p>		<p>国际检索报告邮寄日期</p> <p>2020年 9月 11日</p>																					
<p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>		<p>授权官员</p> <p>曹妹妹</p> <p>电话号码 86-(10)-53961315</p>																					

C. 相关文件		
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	US 10223610 B1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2019年 3月 5日 (2019 - 03 - 05) 全文	1-17
A	US 2017124409 A1 (NEC LABORATORLES AMERICA, INC.) 2017年 5月 4日 (2017 - 05 - 04) 全文	1-17

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/094803

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	110298262	A	2019年 10月 1日	无			
CN	109784194	A	2019年 5月 21日	无			
CN	109712118	A	2019年 5月 3日	无			
CN	108520229	A	2018年 9月 11日	无			
CN	109815922	A	2019年 5月 28日	无			
US	10223610	B1	2019年 3月 5日	无			
US	2017124409	A1	2017年 5月 4日	WO	2017079521	A1	2017年 5月 11日
				DE	112016005062	T5	2018年 7月 12日
				JP	2018538612	A	2018年 12月 27日