

(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 11 月 19 日 (19.11.2020)



(10) 国际公布号
WO 2020/227933 A1

- (51) 国际专利分类号:
G06T 7/00 (2017.01)
- (21) 国际申请号: PCT/CN2019/086883
- (22) 国际申请日: 2019 年 5 月 14 日 (14.05.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 深圳大学(SHENZHEN UNIVERSITY) [CN/CN]; 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。
- (72) 发明人: 邹文斌(ZOU, Wenbin); 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。 卓圣楷(ZHUO, Shengkai); 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。 庄兆永(ZHUANG, Zhaoyong); 中国广东

省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。 吴迪(WU, Di); 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。 李霞(LI, Xia); 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。 徐晨(XU, Chen); 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。

(74) 代理人: 深圳市恒申知识产权事务所(普通合伙)(HENSEN INTELLECTUAL PROPERTY FIRM); 中国广东省深圳市福田区南园路 68 号上步大厦 10H, Guangdong 518000 (CN)。

(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,

(54) Title: SIX-DEGREE-OF-FREEDOM ATTITUDE ESTIMATION METHOD AND APPARATUS, AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 一种六自由度姿态估计方法、装置及计算机可读存储介质

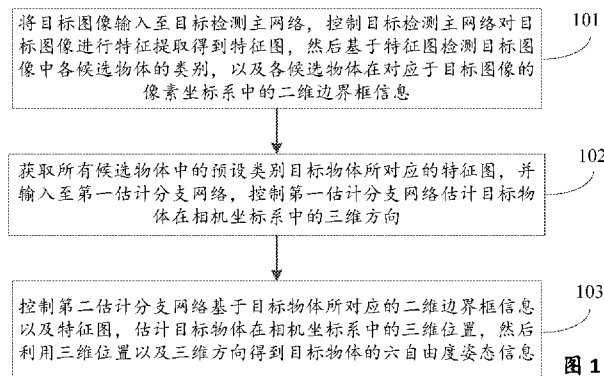


图 1

- 101 Input a target image into a target detection main network, control the target detection main network to perform feature extraction on the target image to obtain a feature map, and then detect, on the basis of the feature map, a category of each candidate object in the target image, and two-dimensional bounding box information of each candidate object in a pixel coordinate system corresponding to the target image
- 102 Acquire a feature map corresponding to a target object of a preset category in all the candidate objects, input the feature map into a first estimation branch network, and control the first estimation branch network to estimate a three-dimensional direction of the target object in a camera coordinate system
- 103 Control a second estimation branch network to estimate, on the basis of the two-dimensional bounding box information and the feature map corresponding to the target object, a three-dimensional position of the target object in the camera coordinate system, and then obtain six-degree-of-freedom attitude information of the target object by using the three-dimensional position and the three-dimensional direction

(57) Abstract: Disclosed are a six-degree-of-freedom attitude estimation method and apparatus, and a computer-readable storage medium. The method comprises: controlling a target detection main network to perform feature extraction on an input image, and then detecting and outputting a category and two-dimensional bounding box information of each candidate object in the image; inputting, into a first estimation branch network, a feature map of a target object of a preset category in all the candidate objects, so as to estimate a three-dimensional direction of the target object in a camera coordinate system; and controlling a second estimation branch network to

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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, 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))。

estimate, on the basis of the two-dimensional bounding box information and the feature map of the target object, a three-dimensional position of the target object in the camera coordinate system, and then obtaining six-degree-of-freedom attitude information of the target object by using the three-dimensional position and the three-dimensional direction. By means of implementing the present invention, the three-dimensional direction and the three-dimensional position of the target object are respectively estimated by different network branches, such that end-to-end six-degree-of-freedom attitude estimation of an object in a surrounding environment of the target object is realized, thereby effectively improving operation speed and operation accuracy.

(57) 摘要: 根据本发明实施例公开的一种六自由度姿态估计方法、装置及计算机可读存储介质, 控制目标检测主网络对输入图像进行特征提取, 然后检测并输出图像中各候选物体的类别以及二维边界框信息; 将所有候选物体中的预设类别目标物体的特征图输入至第一估计分支网络, 估计出目标物体在相机坐标系中的三维方向; 控制第二估计分支网络基于目标物体的二维边界框信息以及特征图估计目标物体在相机坐标系中的三维位置, 然后利用三维位置与三维方向得到目标物体的六自由度姿态信息。通过本发明的实施, 由不同网络分支分别估计目标物体的三维方向和三维位置, 实现端到端的目标物体周围环境中对象的六自由度姿态估计, 有效提升了运算速度以及运算的准确性。

一种六自由度姿态估计方法、装置及计算机可读存储介质

技术领域

本发明涉及空间定位技术领域，尤其涉及一种六自由度姿态估计方法、装置及计算机可读存储介质。

背景技术

随着人工智能技术的快速发展，自动化技术例如车辆自动驾驶、智能机器人控制等越来越得到业界的重视，其中，目标控制对象周围环境的感知是自动化控制操作的基础。

以车辆自动驾驶为例，车辆周围环境感知是自动驾驶系统中最核心的技术，车辆周围环境感知包括行人道检测、车行道检测、车道线检测、车辆检测、行人检测等图像（周围环境）中目标检测与语义分割技术。车辆多自由度姿态估计是传统目标检测与语义分割在三维空间的扩展，其主要任务是精准定位和识别出车辆驾驶视频序列或单帧图像中的所有车辆对象，且同时对所检测到的车辆进行三维空间中的多自由度姿态估计。目前在进行车辆的多自由度姿态估计时，通常所采用的是结合深度学习与几何约束方法的多阶段车辆六自由度姿态估计网络，该方法分为两步实现汽车的六自由度姿态估计，首先通过深度神经网络对输入的单目 RGB 图像中的车辆进行检测，同时对所检测的车辆进行长宽高和三自由度方向估计，然后再利用几何约束关系计算出车辆在实际驾驶场景三维空间中三自由度位置。

尽管上述基于深度学习的多自由度姿态估计方法可以实现目标控制对象周围环境的感知，并在相关场景下取得了不错的成绩，但上述模型仍存在训练与测试过程比较繁琐、不能实现端到端的训练和测试、姿态估计速度慢等缺陷，这制约了自动化技术在控制准确性要求和实时性要求高的场景下的应用，从而在实际应用中具有较大的局限性。

技术问题

本发明实施例的主要目的在于提供一种六自由度姿态估计方法、装置及计算机可读存储介质，至少能够解决相关技术中采用结合深度学习与几何约束的方法进行目标控制对象周围环境的感知时，模型的训练与测试过程比较繁琐、不能实现端到端的训练和测试、周围环境中对象的姿态估计速度慢的问题。

技术方案

为实现上述目的，本发明实施例第一方面提供了一种六自由度姿态估计方

法，应用于包括目标检测主网络、第一估计分支网络以及第二估计分支网络的整体卷积神经网络，该方法包括：

将目标图像输入至所述目标检测主网络，控制所述目标检测主网络对所述目标图像进行特征提取得到特征图，然后基于所述特征图检测所述目标图像中各候选物体的类别，以及所述各候选物体在对应于所述目标图像的像素坐标系中的二维边界框信息；

获取所有候选物体中的预设类别目标物体所对应的特征图，并输入至所述第一估计分支网络，控制所述第一估计分支网络估计所述目标物体在相机坐标系中的三维方向；

控制所述第二估计分支网络基于所述目标物体所对应的二维边界框信息以及特征图，估计所述目标物体在所述相机坐标系中的三维位置，然后利用所述三维位置以及所述三维方向得到所述目标物体的六自由度姿态信息。

为实现上述目的，本发明实施例第二方面提供了一种六自由度姿态估计装置，应用于包括目标检测主网络、第一估计分支网络以及第二估计分支网络的整体卷积神经网络，该装置包括：

检测模块，用于将目标图像输入至所述目标检测主网络，控制所述目标检测主网络对所述目标图像进行特征提取得到特征图，然后基于所述特征图检测所述目标图像中各候选物体的类别，以及所述各候选物体在对应于所述目标图像的像素坐标系中的二维边界框信息；

第一估计模块，用于获取所有候选物体中的预设类别目标物体所对应的特征图，并输入至所述第一估计分支网络，控制所述第一估计分支网络估计所述目标物体在相机坐标系中的三维方向；

第二估计模块，用于控制所述第二估计分支网络基于所述目标物体所对应的二维边界框信息以及特征图，估计所述目标物体在所述相机坐标系中的三维位置，然后利用所述三维位置以及所述三维方向得到所述目标物体的六自由度姿态信息。

为实现上述目的，本发明实施例第三方面提供了一种电子装置，该电子装置包括：处理器、存储器和通信总线；

所述通信总线用于实现所述处理器和存储器之间的连接通信；

所述处理器用于执行所述存储器中存储的一个或者多个程序，以实现上述任意一种六自由度姿态估计方法的步骤。

为实现上述目的,本发明实施例第四方面提供了一种计算机可读存储介质,该计算机可读存储介质存储有一个或者多个程序,所述一个或者多个程序可被一个或者多个处理器执行,以实现上述任意一种六自由度姿态估计方法的步骤。

有益效果

根据本发明实施例提供的六自由度姿态估计方法、装置及计算机可读存储介质,控制目标检测主网络对输入的目标图像进行特征提取,然后检测并输出目标图像中各候选物体的类别,以及各候选物体的二维边界框信息;获取所有候选物体中的预设类别目标物体所对应的特征图,并输入至第一估计分支网络,控制第一估计分支网络估计目标物体在相机坐标系中的三维方向;控制第二估计分支网络基于目标物体的二维边界框信息以及特征图估计目标物体在相机坐标系中的三维位置,然后利用三维位置与三维方向得到目标物体的六自由度姿态信息。通过本发明的实施,由不同网络分支分别估计目标物体的三维方向和三维位置,实现端到端的目标物体周围环境中对象的六自由度姿态估计,有效提升了运算速度以及运算的准确性。

本发明其他特征和相应的效果在说明书的后面部分进行阐述说明,且应当理解,至少部分效果从本发明说明书中的记载变的显而易见。

附图说明

为了更清楚地说明本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为本发明第一实施例提供的六自由度姿态估计方法的基本流程示意图;

图2为本发明第一实施例提供的整体网络框架示意图;

图3为本发明第一实施例提供的目标检测方法的流程示意图;

图4为本发明第一实施例提供的多尺度特征提取示意图;

图5为本发明第一实施例提供的候选区域提取示意图;

图6为本发明第一实施例提供的候选区域特征图池化示意图;

图7为本发明第二实施例提供的六自由度姿态估计装置的结构示意图;

图8为本发明第三实施例提供的电子装置的结构示意图。

本发明的实施方式

为使得本发明的发明目的、特征、优点能够更加的明显和易懂，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而非全部实施例。基于本发明中的实施例，本领域技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

第一实施例：

为了解决相关技术中采用结合深度学习与几何约束的方法进行目标控制对象周围环境的感知时，模型的训练与测试过程比较繁琐、不能实现端到端的训练和测试、周围环境中对象的姿态估计速度慢的技术问题，本实施例提出了一种六自由度姿态估计方法，应用于包括目标检测主网络、第一估计分支网络以及第二估计分支网络的整体卷积神经网络，如图 1 所示为本实施例提供的六自由度姿态估计方法的基本流程示意图，本实施例提出的六自由度姿态估计方法包括以下的步骤：

步骤 101、将目标图像输入至目标检测主网络，控制目标检测主网络对目标图像进行特征提取得到特征图，然后基于特征图检测目标图像中各候选物体的类别，以及各候选物体在对应于目标图像的像素坐标系中的二维边界框信息。

具体的，本实施例的目标检测主网络对输入的图像进行特征提取，然后检测并输出图像中物体的类别和物体的二维边界框。应当说明的是，本实施例中的目标图像可以为通过单目摄像头采集的单目 RGB 图像，另外，候选物体也即感兴趣的物体，其类型可以根据具体应用场景进行选择，例如在车辆自动驾驶的应用场景下，候选物体可以包括行人、车辆等。

如图 2 为本实施例提供的整体网络框架示意图，图 2 中 A 所标识的方框指示本实施例提供的一种目标检测主网络，可选的，该目标检测主网络包括多尺度特征提取网络、候选区域提取网络、候选区域特征图池化层以及物体分类与边界框回归全连接层。基于该目标检测主网络的网络架构，本实施例提供了一种目标检测方法，如图 3 所示为本实施例提供的目标检测方法的流程示意图，具体包括以下步骤：

步骤 301、利用多尺度特征提取网络对目标图像进行多尺度特征提取得到不同尺度的特征图；

步骤 302、利用候选区域提取网络从不同尺度的特征图中，提取出对应于预设候选区域的特征图；

步骤 303、利用候选区域特征图池化层将所有候选区域特征图进行池化操作，将所有候选区域特征图进行尺寸统一；

步骤 304、将尺寸统一的候选区域特征图输入至物体分类与边界框回归全连接层进行候选区域分类检测与边界框回归，得到各候选区域的候选物体的类别，以及各候选物体在对应于目标图像的像素坐标系中的二维边界框信息。

具体的，本实施例中的目标检测主网络由多尺度特征提取网络、候选区域提取网络、候选区域特征图池化层、物体分类与边界框回归全连接层这四个模块组成。其中，以车辆自动化驾驶为例，车辆驾驶过程中周围车辆在相机坐标系中的移动范围较大，导致相机坐标系中不同位置的车辆在像素坐标系中成像的尺寸大小差别较大。本实施例采用多尺度特征提取网络提取输入图像特征，利用深度卷积神经网络中固有的多尺度、多层级的金字塔结构提取单一尺寸输入图像中目标物体的不同尺度特征，使得检测系统具有一定的尺度不变性，可以有效检测出图像中不同尺寸的对象。

进一步地，在本实施例一种可选的实施例中，多尺度特征提取网络为基于 ResNet-101 的多尺度特征提取网络，基于 ResNet-101 的多尺度特征提取网络包括自下向上的深层语义特征提取路径和自上向下的深层语义特征融合路径；具体的请参阅图 4，在利用基于 ResNet-101 的多尺度特征提取网络对目标图像进行多尺度特征提取时，将目标图像输入至自下向上的深层语义特征提取路径所提取出的各层语义特征经过 1×1 卷积核卷积后，通过横向连接的方式与自上向下的深层语义特征融合路径中相同层的语义特征相加融合，得到不同尺度的特征图。通过横向连接方式利用了底层语义的位置细节信息，从而使得融合特征更加精细。

另外，在本实施例中，采用候选区域特征提取网络对多尺度特征图中进行候选区域（也即感兴趣区域）选取。如图 5 所示，候选区域特征提取网络是一个全卷积神经网络，对于任意尺度的图像特征图，采用一个大小为 $n \times n$ 的窗口在特征图上滑动，对于每次滑动都有以该窗口中点为锚点的 3 种不同尺寸和 3 种不同纵横比的锚点框生成，将图像特征图中各个锚点框区域内特征图映射成一个 256 维的特征向量，然后将该特征向量分别输入分类全连接层和边界框回归全连接层中即可得到该锚点框对应输入图像中的候选区域位置 and 该区域是与不是物体的概率（也即置信度）。由于候选区域提取过程中采用了滑动机制和不同尺寸与纵横比的锚点框，因此候选区域提取网络对输入图像中的目标物体

既有平移不变性又有尺度不变性。

还应当说明的是，对于输入图像中一系列任意大小的候选区域，其对应的特征图大小不一，因此不能将其直接输入对大小有固定要求的全连接层中进行候选区域分类检测与边界框回归。基于此，本实施例利用空间金字塔池化层思想设计候选区域特征图池化层，如图 6 所示，首先对于候选区域提取网络输出的任意大小候选区域，将其对应的特征图均匀的划分为 $W \times H$ 块，然后对每一小块特征子图进行最大池化操作即可得到大小统一为 $W \times H$ 的特征图，进而将候选区域特征图输入至物体分类与边界框回归全连接层中进行映射。本发明中采用的候选区域特征池化空间为 7×7 ，即 $W = H = 7$ 。

应当理解的是，本实施例中的物体分类与边界框回归全连接层包括物体分类全连接层与物体边界框回归器两个子模块，请继续参阅图 2，候选区域特征图池化层的输出特征图经过两层 1024 维的全连接层映射后，采用 softmax 函数对候选区域进行例如行人、自行车、汽车、摩托车等候选物体分类，同时也对该候选物体在图像中的二维边界框位置进行估计。

步骤 102、获取所有候选物体中的预设类别目标物体所对应的特征图，并输入至第一估计分支网络，控制第一估计分支网络估计目标物体在相机坐标系中的三维方向。

具体的，本实施例中当目标检测主网络的全连接层对候选区域预测的物体为预设类别的目标物体（例如汽车）时，则将其原始的池化候选区域特征图输入至第一估计分支网络，对该目标物体模型在相机坐标系（实际驾驶环境）中的三维方向进行估计。

请再次参阅图 2，图 2 中 B 所标识的方框指示本实施例提供的一种第一估计分支网络，可选的，该第一估计分支网络为：分类与三维方向估计分支网络。对应的，控制第一估计分支网络估计目标物体在相机坐标系中的三维方向包括：控制分类与三维方向估计分支网络估计目标物体的子类别，以及目标物体在相机坐标系中的三维方向。具体的，可以对目标物体对应区域的特征图经过两层 100 维的全连接层映射后再采用 softmax 函数对该“目标物体候选区域”进行子类别检测，同时对该目标物体模型在相机坐标系（实际驾驶环境）中三维方向进行估计。

步骤 103、控制第二估计分支网络基于目标物体所对应的二维边界框信息以及特征图，估计目标物体在相机坐标系中的三维位置，然后利用三维位置以

及三维方向得到目标物体的六自由度姿态信息。

具体的，在第一估计分支网络估计出目标物体在相机坐标系（实际驾驶环境）中的三维方向之后，由第二估计分支网络融合第一估计分支网络提供的信息并计算出各目标物体在相机坐标系（实际驾驶环境）中的三维位置，从而实现端到端的目标物体的六自由度姿态估计。应当说明的是，在本实施例的一种实现中，第二估计分支网络基于目标物体所对应的二维边界框信息以及特征图，估计目标物体在相机坐标系中的三维位置时，可以是先将二维边界框信息转换为相机坐标系下的边界框信息，以及通过第一估计网络分支将区域特征图转化为特定维度的向量，然后将转换后的信息输入至第二估计分支网络，将转换后的边界框信息与区域特征信息通过一种级联的方式进行融合，输出三维位置，进而与第一估计分支网络输出的三维方向组成目标物体的六自由度姿态信息。由于这一过程是端到端实现的，可大幅提升运算速度且避免的多阶段处理的误差传递，因此保证了目标物体姿态估计的速率和准确率，进而保证了系统对周围环境感知的及时性和准确性，这将大幅提升自动化控制的决策、控制等性能。还应当理解的是，本实施例所得到的六自由度姿态信息可用于可视化目标来得到可视化结果，可以更直观的将目标向用户进行表现。

请再次参阅图 2，图 2 中 C 所标识的方框指示本实施例提供的一种第二估计分支网络，对应于第一估计分支网络为分类与三维方向估计分支网络而还输出目标物体的子类别的情况，利用三维位置以及三维方向得到目标物体的六自由度姿态信息包括：利用三维位置、三维方向以及目标物体的子类别，得到各子类别的目标物体的六自由度姿态信息。具体的，将目标检测主网络全连接层的目标边界框位置特征输入两层 100 维的全连接层，利用目标物体在图像中的二维边界框信息映射后，同时融合来自分类与三维方向估计分支网络的目标物体子类别、目标物体三维方向等信息以提高计算准确率，计算出目标物体在相机坐标系中的三维位置。

进一步可选的，基于本实施例图 2 所提供的网络框架，为了最小化误差，本实施例的整体卷积神经网络的损失函数为： $\text{Loss} = \text{loss}_{\text{det}} + \text{loss}_{\text{inst}}$ ；其中，

$$\text{loss}_{\text{det}} = \text{loss}_{\text{cls}} + \text{loss}_{\text{box}}$$

$$\text{loss}_{\text{cls}}(p_i, p_i^*) = -\log p_i = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M p_{ij} \log p_{ij}^*$$

$$\begin{aligned}
\text{loss}_{\text{box}}(t_i, t_i^*) &= \sum_{i \in \{x_1, y_1, x_2, y_2\}} \text{smooth}_{L_1}(t_i - t_i^*) \\
\text{smooth}_{L_1}(x) &= \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \\
\text{loss}_{\text{inst}} &= \lambda_{\text{obj_cls}} \text{loss}_{\text{obj_cls}} + \lambda_{\text{rot}} \text{loss}_{\text{rot}} + \lambda_{\text{trans}} \text{loss}_{\text{trans}} \\
\text{loss}_{\text{obj_cls}}(c, \hat{c}) &= - \sum_i [c^i \log \hat{c}^i + (1 - c^i) \log(1 - \hat{c}^i)] \\
\text{loss}_{\text{rot}}(q, \hat{q}) &= \left\| \hat{q} - \frac{q}{\|q\|} \right\|_1 \\
\text{loss}_{\text{trans}}(t, \hat{t}) &= \begin{cases} \frac{(t - \hat{t})^2}{2\delta}, & \text{if } |t - \hat{t}| < \delta \\ |t - \hat{t}| - \frac{1}{2}\delta, & \text{otherwise} \end{cases}
\end{aligned}$$

其中， loss_{det} 为目标检测主网络的全连接层的损失函数； $\text{loss}_{\text{inst}}$ 为第一估计分支网络以及第二估计分支网络的全连接层的损失函数， $\text{loss}_{\text{obj_cls}}$ 为第一估计分支网络中的分类估计损失函数， loss_{rot} 为第一估计分支网络中的三维方向估计损失函数， q 为目标物体在相机坐标系中三维方向的估计四元数， $\hat{q}$ 为目标物体在相机坐标系中三维方向的真实四元数， $\text{loss}_{\text{trans}}$ 为第二估计分支网络的三维位置估计损失函数， t 为目标物体在相机坐标系中三维位置的估计坐标， $\hat{t}$ 为目标物体在相机坐标系中三维位置的真实坐标， $\lambda_{\text{obj_cls}}$ 、 λ_{rot} 、 λ_{trans} 分别为对应于各损失函数的权重超参数。

根据本发明实施例提供的六自由度姿态估计方法，控制目标检测主网络对输入的目标图像进行特征提取，然后检测并输出目标图像中各候选物体的类别，以及各候选物体的二维边界框信息；获取所有候选物体中的预设类别目标物体所对应的特征图，并输入至第一估计分支网络，控制第一估计分支网络估计目标物体在相机坐标系中的三维方向；控制第二估计分支网络基于目标物体的二维边界框信息以及特征图估计目标物体在相机坐标系中的三维位置，然后利用三维位置与三维方向得到目标物体的六自由度姿态信息。通过本发明的实施，由不同网络分支分别估计目标物体的三维方向和三维位置，实现端到端的目标物体周围环境中对象的六自由度姿态估计，有效提升了运算速度以及运算的准确性。

第二实施例：

为了解决相关技术中采用结合深度学习与几何约束的方法进行目标控制对象周围环境的感知时，模型的训练与测试过程比较繁琐、不能实现端到端的训练和测试、周围环境中对象的姿态估计速度慢的技术问题，本实施例示出了一种六自由度姿态估计装置，应用于包括目标检测主网络、第一估计分支网络以及第二估计分支网络的整体卷积神经网络，具体请参见图 7，本实施例的六自由度姿态估计装置包括：

检测模块 701，用于将目标图像输入至目标检测主网络，控制目标检测主网络对目标图像进行特征提取得到特征图，然后基于特征图检测目标图像中各候选物体的类别，以及各候选物体在对应于目标图像的像素坐标系中的二维边界框信息；

第一估计模块 702，用于获取所有候选物体中的预设类别目标物体所对应的特征图，并输入至第一估计分支网络，控制第一估计分支网络估计目标物体在相机坐标系中的三维方向；

第二估计模块 703，用于控制第二估计分支网络基于目标物体所对应的二维边界框信息以及特征图，估计目标物体在相机坐标系中的三维位置，然后利用三维位置以及三维方向得到目标物体的六自由度姿态信息。

具体的，本实施例中目标检测主网络对输入的图像进行特征提取，然后检测并输出图像中物体的类别和物体的二维边界框。然后当目标检测主网络的全连接层对候选区域预测的物体为预设类别的目标物体（例如汽车）时，将目标物体对应的特征图输入至第一估计分支网络，对该目标物体模型在相机坐标系（实际驾驶环境）中的三维方向进行估计。在第一估计分支网络估计出目标物体在相机坐标系（实际驾驶环境）中的三维方向之后，由第二估计分支网络融合第一估计分支网络提供的信息并计算出各目标物体在相机坐标系（实际驾驶环境）中的三维位置，从而实现端到端的目标物体的六自由度姿态估计。由于这一过程是端到端实现的，可大幅提升运算速度且避免的多阶段处理的误差传递，因此保证了目标物体姿态估计的速率和准确率，进而保证了系统对周围环境感知的及时性和准确性，这将大幅提升自动化控制的决策、控制等性能。

在本实施例的一些实施方式中，目标检测主网络包括多尺度特征提取网络、候选区域提取网络、候选区域特征图池化层以及物体分类与边界框回归全连接层；对应的，检测模块 701 具体用于将目标图像输入至目标检测主网络，利用

多尺度特征提取网络对目标图像进行多尺度特征提取得到不同尺度的特征图；利用候选区域提取网络从不同尺度的特征图中，提取出对应于预设候选区域的特征图；利用候选区域特征图池化层将所有候选区域特征图进行池化操作，将所有候选区域特征图进行尺寸统一；将尺寸统一的候选区域特征图输入至物体分类与边界框回归全连接层进行候选区域分类检测与边界框回归，得到各候选区域的候选物体的类别，以及各候选物体在对应于目标图像的像素坐标系中的二维边界框信息。

进一步地，在本实施例的一些实施方式中，多尺度特征提取网络为基于 ResNet-101 的多尺度特征提取网络，基于 ResNet-101 的多尺度特征提取网络包括自下向上的深层语义特征提取路径和自上向下的深层语义特征融合路径；对应的，检测模块 701 在利用多尺度特征提取网络对目标图像进行多尺度特征提取得到不同尺度的特征图时，具体用于将目标图像输入至自下向上的深层语义特征提取路径所提取出的各层语义特征经过 1×1 卷积核卷积后，通过横向连接的方式与自上向下的深层语义特征融合路径中相同层的语义特征相加融合，得到不同尺度的特征图。

在本实施例的一些实施方式中，第一估计分支网络为：分类与三维方向估计分支网络；对应的，第一估计模块 702 具体用于获取所有候选物体中的预设类别目标物体所对应的特征图，并输入至分类与三维方向估计分支网络，控制分类与三维方向估计分支网络估计目标物体的子类别，以及目标物体在相机坐标系中的三维方向。第二估计模块 703 具体用于控制第二估计分支网络基于目标物体所对应的二维边界框信息以及特征图，估计目标物体在相机坐标系中的三维位置，然后利用三维位置、三维方向以及目标物体的子类别，得到各子类别的目标物体的六自由度姿态信息。

应当说明的是，前述实施例中的六自由度姿态估计方法均可基于本实施例提供的六自由度姿态估计装置实现，所属领域的普通技术人员可以清楚的了解，为描述的方便和简洁，本实施例中所描述的六自由度姿态估计装置的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

采用本实施例提供的六自由度姿态估计装置，控制目标检测主网络对输入的目标图像进行特征提取，然后检测并输出目标图像中各候选物体的类别，以及各候选物体的二维边界框信息；获取所有候选物体中的预设类别目标物体所对应的特征图，并输入至第一估计分支网络，控制第一估计分支网络估计目标

物体在相机坐标系中的三维方向；控制第二估计分支网络基于目标物体的二维边界框信息以及特征图估计目标物体在相机坐标系中的三维位置，然后利用三维位置与三维方向得到目标物体的六自由度姿态信息。通过本发明的实施，由不同网络分支分别估计目标物体的三维方向和三维位置，实现端到端的目标物体周围环境中对象的六自由度姿态估计，有效提升了运算速度以及运算的准确性。

第三实施例：

本实施例提供了一种电子装置，参见图 8 所示，其包括处理器 801、存储器 802 及通信总线 803，其中：通信总线 803 用于实现处理器 801 和存储器 802 之间的连接通信；处理器 801 用于执行存储器 802 中存储的一个或者多个计算机程序，以实现上述实施例一中的六自由度姿态估计方法中的至少一个步骤。

本实施例还提供了一种计算机可读存储介质，该计算机可读存储介质包括在用于存储信息(诸如计算机可读指令、数据结构、计算机程序模块或其他数据)的任何方法或技术中实施的易失性或非易失性、可移除或不可移除的介质。计算机可读存储介质包括但不限于 RAM (Random Access Memory, 随机存取存储器), ROM (Read-Only Memory, 只读存储器), EEPROM (Electrically Erasable Programmable read only memory, 带电可擦可编程只读存储器)、闪存或其他存储器技术、CD-ROM (Compact Disc Read-Only Memory, 光盘只读存储器), 数字多功能盘(DVD)或其他光盘存储、磁盒、磁带、磁盘存储或其他磁存储装置、或者可以用于存储期望的信息并且可以被计算机访问的任何其他的介质。

本实施例中的计算机可读存储介质可用于存储一个或者多个计算机程序，其存储的一个或者多个计算机程序可被处理器执行，以实现上述实施例一中的方法的至少一个步骤。

本实施例还提供了一种计算机程序，该计算机程序可以分布在计算机可读介质上，由可计算装置来执行，以实现上述实施例一中的方法的至少一个步骤；并且在某些情况下，可以采用不同于上述实施例所描述的顺序执行所示出或描述的至少一个步骤。

本实施例还提供了一种计算机程序产品，包括计算机可读装置，该计算机可读装置上存储有如上所示的计算机程序。本实施例中该计算机可读装置可包括如上所示的计算机可读存储介质。

可见，本领域的技术人员应该明白，上文中所公开方法中的全部或某些步骤、系统、装置中的功能模块/单元可以被实施为软件（可以用计算装置可执行的计算机程序代码来实现）、固件、硬件及其适当的组合。在硬件实施方式中，在以上描述中提及的功能模块/单元之间的划分不一定对应于物理组件的划分；例如，一个物理组件可以具有多个功能，或者一个功能或步骤可以由若干物理组件合作执行。某些物理组件或所有物理组件可以被实施为由处理器，如中央处理器、数字信号处理器或微处理器执行的软件，或者被实施为硬件，或者被实施为集成电路，如专用集成电路。

此外，本领域普通技术人员公知的是，通信介质通常包含计算机可读指令、数据结构、计算机程序模块或者诸如载波或其他传输机制之类的调制数据信号中的其他数据，并且可包括任何信息递送介质。所以，本发明不限制于任何特定的硬件和软件结合。

以上内容是结合具体的实施方式对本发明实施例所作的进一步详细说明，不能认定本发明的具体实施只局限于这些说明。对于本发明所属技术领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干简单推演或替换，都应当视为属于本发明的保护范围。

权利要求书

1. 一种六自由度姿态估计方法，应用于包括目标检测主网络、第一估计分支网络以及第二估计分支网络的整体卷积神经网络，其特征在于，包括：

将目标图像输入至所述目标检测主网络，控制所述目标检测主网络对所述目标图像进行特征提取得到特征图，然后基于所述特征图检测所述目标图像中各候选物体的类别，以及所述各候选物体在对应于所述目标图像的像素坐标系中的二维边界框信息；

获取所有候选物体中的预设类别目标物体所对应的特征图，并输入至所述第一估计分支网络，控制所述第一估计分支网络估计所述目标物体在相机坐标系中的三维方向；

控制所述第二估计分支网络基于所述目标物体所对应的二维边界框信息以及特征图，估计所述目标物体在所述相机坐标系中的三维位置，然后利用所述三维位置以及所述三维方向得到所述目标物体的六自由度姿态信息。

2. 如权利要求 1 所述的六自由度姿态估计方法，其特征在于，所述目标检测主网络包括多尺度特征提取网络、候选区域提取网络、候选区域特征图池化层以及物体分类与边界框回归全连接层；

所述控制所述目标检测主网络对所述目标图像进行特征提取得到特征图，然后基于所述特征图检测所述目标图像中各候选物体的类别，以及所述各候选物体在对应于所述目标图像的像素坐标系中的二维边界框信息包括：

利用所述多尺度特征提取网络对所述目标图像进行多尺度特征提取得到不同尺度的特征图；

利用所述候选区域提取网络从所述不同尺度的特征图中，提取出对应于预设候选区域的特征图；

利用所述候选区域特征图池化层将所有候选区域特征图进行池化操作，将所述所有候选区域特征图进行尺寸统一；

将尺寸统一的候选区域特征图输入至所述物体分类与边界框回归全

连接层进行候选区域分类检测与边界框回归,得到各候选区域的候选物体的类别,以及所述各候选物体在对应于所述目标图像的像素坐标系中的二维边界框信息。

3.如权利要求2所述的六自由度姿态估计方法,其特征在于,所述多尺度特征提取网络为基于 ResNet-101 的多尺度特征提取网络,所述基于 ResNet-101 的多尺度特征提取网络包括自下向上的深层语义特征提取路径和自上向下的深层语义特征融合路径;

所述利用所述多尺度特征提取网络对所述目标图像进行多尺度特征提取得到不同尺度的特征图包括:

将所述目标图像输入至所述自下向上的深层语义特征提取路径所提取出的各层语义特征经过 1×1 卷积核卷积后,通过横向连接的方式与所述自上向下的深层语义特征融合路径中相同层的语义特征相加融合,得到不同尺度的特征图。

4.如权利要求1所述的六自由度姿态估计方法,其特征在于,所述第一估计分支网络为:分类与三维方向估计分支网络;

所述控制所述第一估计分支网络估计所述目标物体在相机坐标系中的三维方向包括:

控制所述分类与三维方向估计分支网络估计所述目标物体的子类别,以及所述目标物体在相机坐标系中的三维方向;

所述利用所述三维位置以及所述所述三维方向得到所述目标物体的六自由度姿态信息包括:

利用所述三维位置、所述三维方向以及所述目标物体的子类别,得到各子类别的所述目标物体的六自由度姿态信息。

5.如权利要求4所述的六自由度姿态估计方法,其特征在于,所述整体卷积神经网络的损失函数为: $\text{Loss} = \text{loss}_{\text{det}} + \text{loss}_{\text{inst}}$; 其中,

$$\text{loss}_{\text{det}} = \text{loss}_{\text{cls}} + \text{loss}_{\text{box}}$$

$$\text{loss}_{\text{cls}}(p_i, p_i^*) = -\log p_i = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M p_{ij} \log p_{ij}^*$$

$$\begin{aligned} \text{loss}_{\text{box}}(t_i, t_i^*) &= \sum_{i \in \{x_1, y_1, x_2, y_2\}} \text{smooth}_{L_1}(t_i - t_i^*) \\ \text{smooth}_{L_1}(x) &= \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \\ \text{loss}_{\text{inst}} &= \lambda_{\text{obj_cls}} \text{loss}_{\text{obj_cls}} + \lambda_{\text{rot}} \text{loss}_{\text{rot}} + \lambda_{\text{trans}} \text{loss}_{\text{trans}} \\ \text{loss}_{\text{obj_cls}}(c, \hat{c}) &= - \sum_i [c^i \log \hat{c}^i + (1 - c^i) \log(1 - \hat{c}^i)] \\ \text{loss}_{\text{rot}}(q, \hat{q}) &= \left\| \hat{q} - \frac{q}{\|q\|} \right\|_1 \\ \text{loss}_{\text{trans}}(t, \hat{t}) &= \begin{cases} \frac{(t - \hat{t})^2}{2\delta}, & \text{if } |t - \hat{t}| < \delta \\ |t - \hat{t}| - \frac{1}{2}\delta, & \text{otherwise} \end{cases} \end{aligned}$$

其中， loss_{det} 为所述目标检测主网络的全连接层的损失函数； $\text{loss}_{\text{inst}}$ 为所述第一估计分支网络以及第二估计分支网络的全连接层的损失函数， $\text{loss}_{\text{obj_cls}}$ 为所述第一估计分支网络中的分类估计损失函数， loss_{rot} 为所述第一估计分支网络中的三维方向估计损失函数， q 为所述目标物体在所述相机坐标系中三维方向的估计四元数， $\hat{q}$ 为所述目标物体在所述相机坐标系中三维方向的真实四元数， $\text{loss}_{\text{trans}}$ 为所述第二估计分支网络的三维位置估计损失函数， t 为所述目标物体在所述相机坐标系中三维位置的估计坐标， $\hat{t}$ 为所述目标物体在所述相机坐标系中三维位置的真实坐标， $\lambda_{\text{obj_cls}}$ 、 λ_{rot} 、 λ_{trans} 分别为对应于各损失函数的权重超参数。

6. 一种六自由度姿态估计装置，应用于包括目标检测主网络、第一估计分支网络以及第二估计分支网络的整体卷积神经网络，其特征在于，包括：

检测模块，用于将目标图像输入至所述目标检测主网络，控制所述目标检测主网络对所述目标图像进行特征提取得到特征图，然后基于所述特征图检测所述目标图像中各候选物体的类别，以及所述各候选物体在对应于所述目标图像的像素坐标系中的二维边界框信息；

第一估计模块,用于获取所有候选物体中的预设类别目标物体所对应的特征图,并输入至所述第一估计分支网络,控制所述第一估计分支网络估计所述目标物体在相机坐标系中的三维方向;

第二估计模块,用于控制所述第二估计分支网络基于所述目标物体所对应的二维边界框信息以及特征图,估计所述目标物体在所述相机坐标系中的三维位置,然后利用所述三维位置以及所述三维方向得到所述目标物体的六自由度姿态信息。

7.如权利要求6所述的六自由度姿态估计装置,其特征在于,所述目标检测主网络包括多尺度特征提取网络、候选区域提取网络、候选区域特征图池化层以及物体分类与边界框回归全连接层;

所述检测模块具体用于将目标图像输入至所述目标检测主网络,利用所述多尺度特征提取网络对所述目标图像进行多尺度特征提取得到不同尺度的特征图;利用所述候选区域提取网络从所述不同尺度的特征图中,提取出对应于预设候选区域的特征图;利用所述候选区域特征图池化层将所有候选区域特征图进行池化操作,将所述所有候选区域特征图进行尺寸统一;将尺寸统一的候选区域特征图输入至所述物体分类与边界框回归全连接层进行候选区域分类检测与边界框回归,得到各候选区域的候选物体的类别,以及所述各候选物体在对应于所述目标图像的像素坐标系中的二维边界框信息。

8.如权利要求6所述的六自由度姿态估计装置,其特征在于,所述第一估计分支网络为:分类与三维方向估计分支网络;

第一估计模块具体用于获取所有候选物体中的预设类别目标物体所对应的特征图,并输入至所述分类与三维方向估计分支网络,控制所述分类与三维方向估计分支网络估计所述目标物体的子类别,以及所述目标物体在相机坐标系中的三维方向;

第二估计模块具体用于控制所述第二估计分支网络基于所述目标物体所对应的二维边界框信息以及特征图,估计所述目标物体在所述相机坐标系中的三维位置,然后利用所述三维位置、所述三维方向以及所述目标物体的子类别,得到各子类别的所述目标物体的六自由度姿态信息。

9. 一种电子装置，其特征在于，包括：处理器、存储器和通信总线；
所述通信总线用于实现所述处理器和存储器之间的连接通信；
所述处理器用于执行所述存储器中存储的一个或者多个程序，以实现如权利要求 1 至 5 中任意一项所述的六自由度姿态估计方法的步骤。
10. 一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有一个或者多个程序，所述一个或者多个程序可被一个或者多个处理器执行，以实现如权利要求 1 至 5 中任意一项所述的六自由度姿态估计方法的步骤。

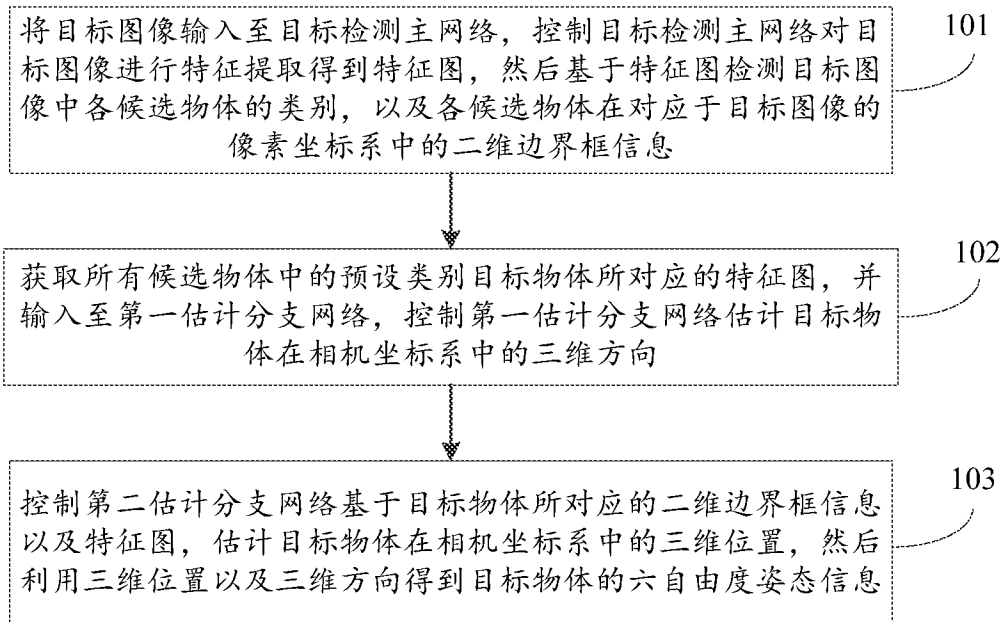


图 1

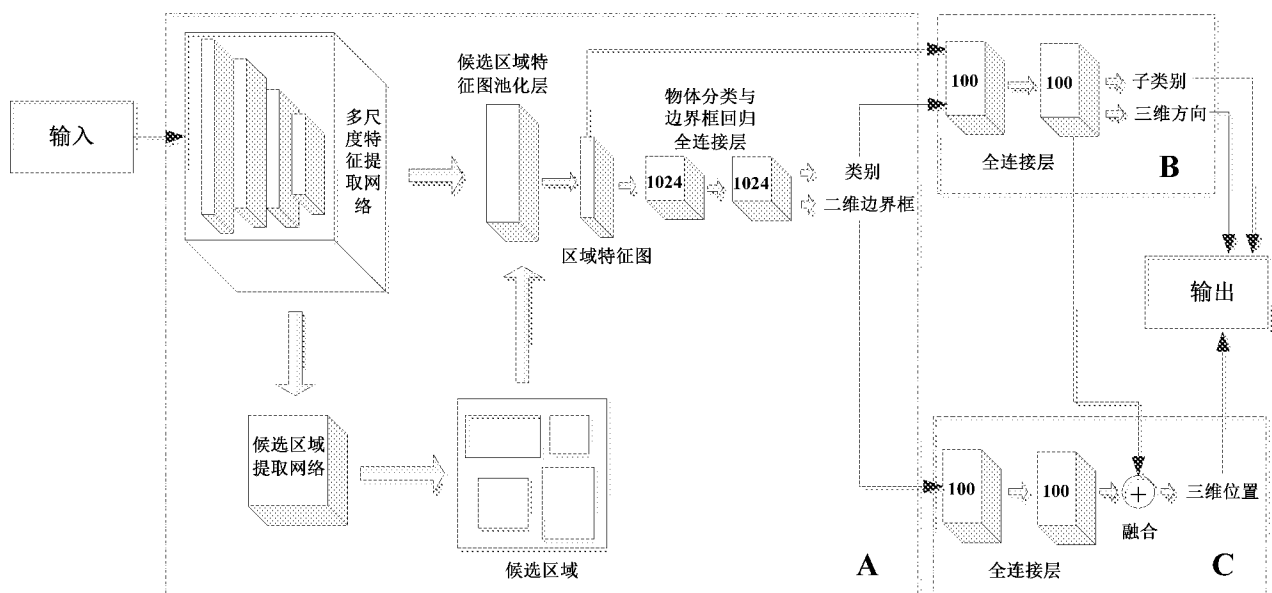


图 2

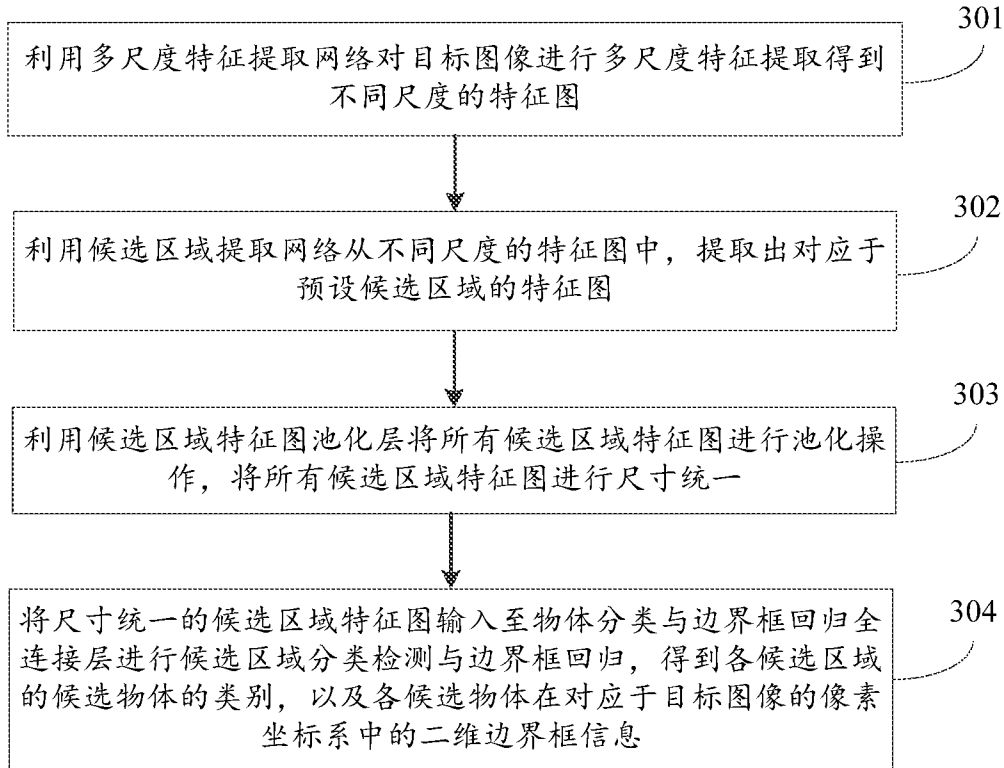


图 3

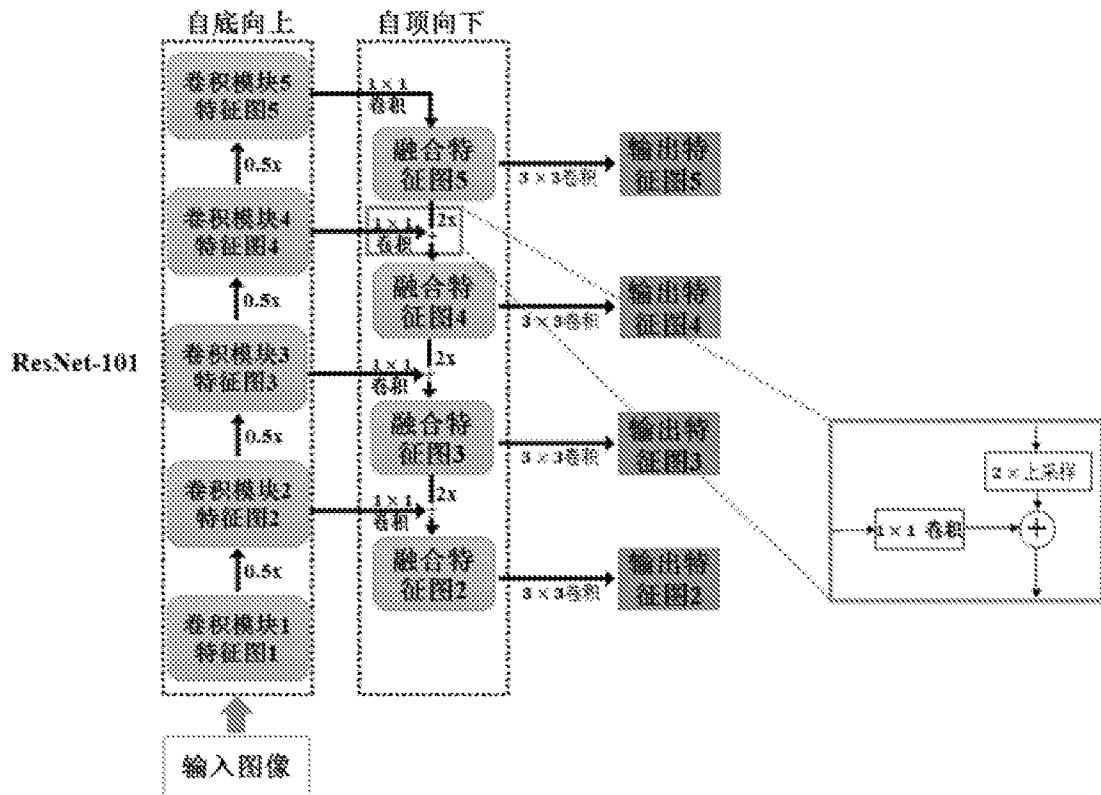


图 4

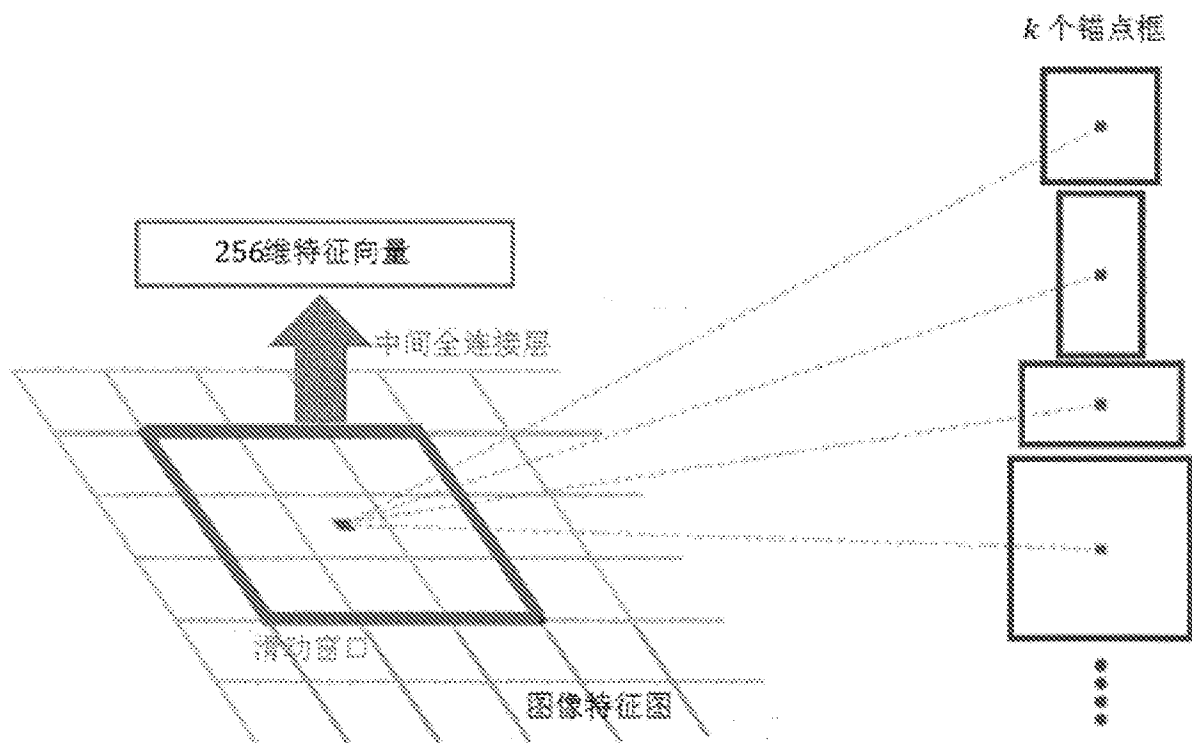


图 5

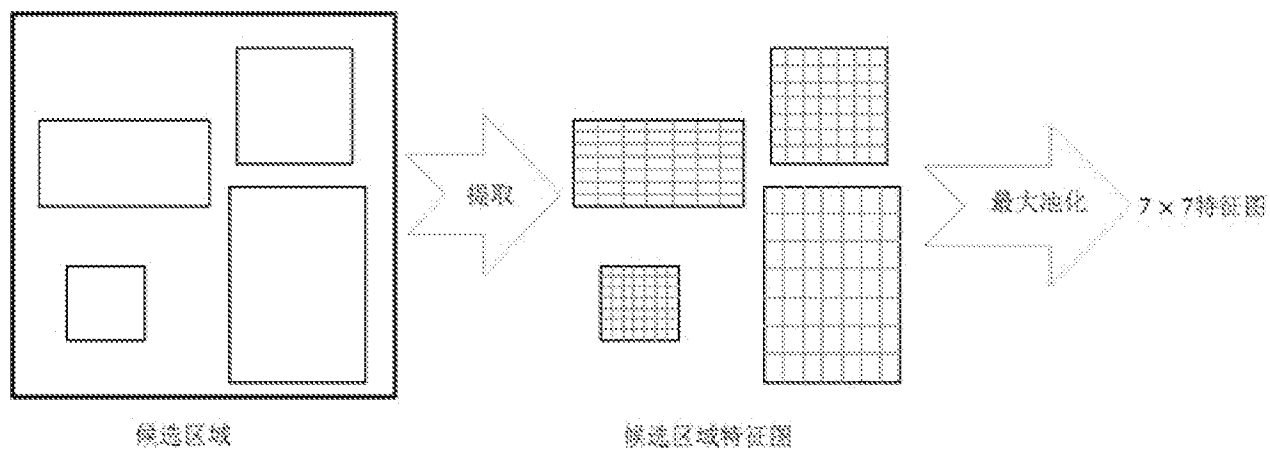


图 6

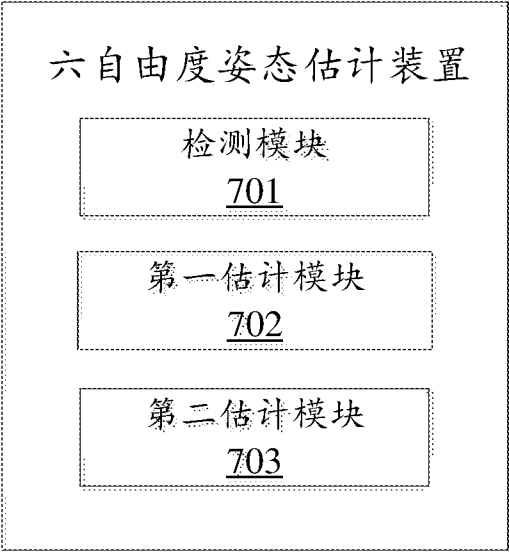


图 7

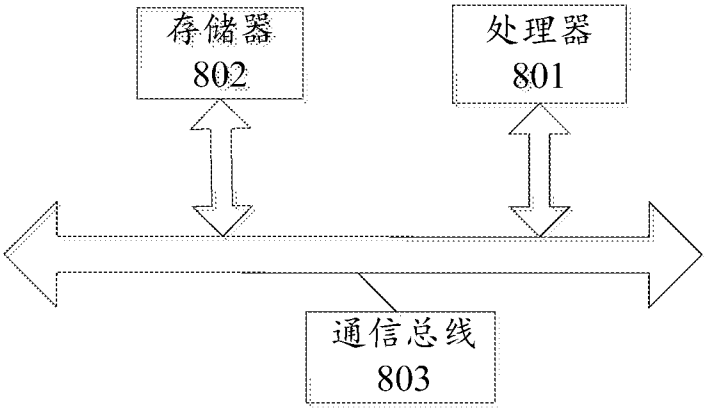


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/086883

A. CLASSIFICATION OF SUBJECT MATTER

G06T 7/00(2017.01)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)

G06T

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)

CNABS, CNTXT, VEN, USTXT, WOTXT, EPTXT, CNKI: 自由度, 姿态, 图像, 神经网络, 三维, 位置, 方向, 特征, degree of freedom, posture, image, neural network, 3D, position, direction, feature

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 105809689 A (HARBIN ENGINEERING UNIVERSITY) 27 July 2016 (2016-07-27) description, paragraphs 4-36, and figure 1	1-10
A	CN 104463108 A (SHANDONG UNIVERSITY) 25 March 2015 (2015-03-25) description, paragraphs 25-38, and figure 1	1-10
A	US 8311954 B2 (NEC LABORATORIES AMERICA INC) 13 November 2012 (2012-11-13) entire document	1-10



Further documents are listed in the continuation of Box C.



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

15 January 2020

Date of mailing of the international search report

13 February 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

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/086883

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	105809689	A	27 July 2016	CN	105809689	B	26 October 2018
CN	104463108	A	25 March 2015	CN	104463108	B	31 July 2018
US	8311954	B2	13 November 2012	US	2010049675	A1	25 February 2010

国际检索报告

国际申请号

PCT/CN2019/086883

A. 主题的分类

G06T 7/00 (2017.01) i

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

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G06T

包含在检索领域中的除最低限度文献以外的检索文献

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))

CNABS, CNTXT, VEN, USTXT, WOTXT, EPTXT, CNKI: 自由度, 姿态, 图像, 神经网络, 三维, 位置, 方向, 特征, degree of freedom, posture, image, neural network, 3D, position, direction, feature

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	CN 105809689 A (哈尔滨工程大学) 2016年 7月 27日 (2016 - 07 - 27) 说明书第4-36段, 图1	1-10
A	CN 104463108 A (山东大学) 2015年 3月 25日 (2015 - 03 - 25) 说明书第25-38段, 图1	1-10
A	US 8311954 B2 (NEC LABORATORIES AMERICA INC) 2012年 11月 13日 (2012 - 11 - 13) 全文	1-10

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“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)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“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

“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

“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

“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

“&” 同族专利的文件

国际检索实际完成的日期

2020年 1月 15日

国际检索报告邮寄日期

2020年 2月 13日

ISA/CN的名称和邮寄地址

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

传真号 (86-10)62019451

受权官员

古志春

电话号码 86-(20)-28950386

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/086883

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	105809689	A	2016年 7月 27日	CN	105809689	B	2018年 10月 26日		
CN	104463108	A	2015年 3月 25日	CN	104463108	B	2018年 7月 31日		
US	8311954	B2	2012年 11月 13日	US	2010049675	A1	2010年 2月 25日		