



(12) 发明专利

(10) 授权公告号 CN 115482532 B

(45) 授权公告日 2025.07.04

(21) 申请号 202211008001.1

G06V 10/26 (2022.01)

(22) 申请日 2022.08.22

G06V 10/80 (2022.01)

(65) 同一申请的已公布的文献号

G06V 10/82 (2022.01)

申请公布号 CN 115482532 A

G06T 3/4038 (2024.01)

(43) 申请公布日 2022.12.16

(56) 对比文件

(73) 专利权人 香港理工大学深圳研究院

Fan, J et al..A Multi-Granularity Scene Segmentation Network for Human-Robot Collaboration Environment

地址 518057 广东省深圳市南山区粤海街

道高新技术产业园南区粤兴一道18号

香港理工大学产学研大楼205室

Perception.《2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)》.2022,全文.

(72) 发明人 范峻铭 郑湃 李树飞

审查员 傅志良

(74) 专利代理机构 深圳市君胜知识产权代理事务

所(普通合伙) 44268

专利代理师 谢松 吴志益

(51) Int. Cl.

G06V 20/70 (2022.01)

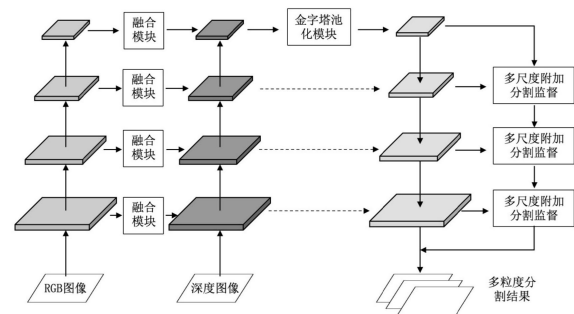
权利要求书3页 说明书12页 附图6页

(54) 发明名称

一种基于深度学习的多颗粒度人机共融环境感知方法

(57) 摘要

本发明公开了一种基于深度学习的多颗粒度人机共融环境感知方法,获取人机共融场景的RGB图像和深度图像;将所述RGB图像和所述深度图像输入编码网络,得到编码图像;将所述编码图像输入金字塔池化模块,得到池化图像;将所述池化图像输入解码网络,得到解码图像;将所述解码图像输入多粒度分割输出模块,得到不同粒度等级的场景分割图像;其中,所述粒度等级包括区域等级、实体等级以及实体的部分等级。不同粒度等级的场景分割图像为协作机器人提供更加完善的环境感知能力,使其能够根据环境和任务的不同自适应地切换不同粒度的环境感知分割结果,从而更好地进行后续协作行为决策和运动规划。



1. 一种基于深度学习的多颗粒度人机共融环境感知方法,其特征在于,该方法包括步骤:

获取人机共融场景的RGB图像和深度图像;其中,所述RGB图像和所述深度图像是对同一人机共融场景拍摄得到的图像;

将所述RGB图像和所述深度图像输入编码网络,得到编码图像;

将所述编码图像输入金字塔池化模块,得到池化图像;

将所述池化图像输入解码网络,得到解码图像;

将所述解码图像输入多粒度分割输出模块,得到不同粒度等级的场景分割图像;

其中,所述粒度等级包括区域等级、实体等级以及实体的部分等级;所述编码网络包括:第一下采样模块、第二下采样模块、第三下采样模块、第四下采样模块、第一融合模块、第二融合模块、第三融合模块以及第四融合模块;所述金字塔池化模块包括:池化层、第四卷积层以及上采样层;

所述解码网络包括:第一上采样模块、第二上采样模块以及第三上采样模块;所述第一融合模块与所述第三上采样模块跳跃连接;所述第二融合模块与所述第二上采样模块跳跃连接;所述第三融合模块与所述第一上采样模块跳跃连接;所述第三上采样模块包括:第一卷积模块、第一上采样深度卷积模块、不同粒度等级的细化分割模块;所述细化分割模块包括:第一上采样卷积模块和第一卷积层;所述第一融合模块包括:第一池化卷积模块和第二池化卷积模块;所述多粒度分割输出模块包括:不同粒度等级的分割输出模块,各粒度等级的分割输出模块包括:第二上采样卷积模块、第二卷积层、第二上采样深度卷积模块以及第三上采样深度卷积模块;

所述将所述池化图像输入解码网络,得到解码图像,包括:

将所述池化图像输入第一上采样模块,得到不同粒度等级的第一细化分割图像以及第一上采样图像;

将所述第一上采样图像和第一细化分割图像输入所述第二上采样模块,得到不同粒度等级的第二细化分割图像以及第二上采样图像;

将所述第二上采样图像和第二细化分割图像输入所述第三上采样模块,得到不同粒度等级的第三细化分割图像以及第三上采样图像,并将所述第三细化分割图像以及所述第三上采样图像作为解码图像。

2. 根据权利要求1所述的基于深度学习的多颗粒度人机共融环境感知方法,其特征在于,所述将所述RGB图像和所述深度图像输入编码网络,得到编码图像,包括:

将所述深度图像和所述RGB图像分别输入所述第一下采样模块,得到第一下采样深度图像和第一下采样RGB图像;

将所述第一下采样深度图像和所述第一下采样RGB图像输入所述第一融合模块,得到第一融合图像;

将所述第一下采样深度图像和所述第一融合图像分别输入所述第二下采样模块,得到第二下采样深度图像和第二下采样RGB图像;

将所述第二下采样深度图像和所述第二下采样RGB图像输入所述第二融合模块,得到第二融合图像;

将所述第二下采样深度图像和所述第二融合图像分别输入所述第三下采样模块,得到

第三下采样深度图像和第三下采样RGB图像；

将所述第三下采样深度图像和所述第三下采样RGB图像输入所述第三融合模块,得到第三融合图像；

将所述第三下采样深度图像和所述第三融合图像分别输入所述第四下采样模块,得到第四下采样深度图像和第四下采样RGB图像；

将所述第四下采样深度图像和所述第四下采样RGB图像输入所述第四融合模块,得到第四融合图像,并将所述第四融合图像作为编码图像。

3. 根据权利要求2所述的基于深度学习的多颗粒度人机共融环境感知方法,其特征在于,所述将所述第二上采样图像和第二细化分割图像输入所述第三上采样模块,得到不同粒度等级的第三细化分割图像以及第三上采样图像,包括:

将所述第二上采样图像输入第一卷积模块,得到第一卷积图像；

将所述第一卷积图像输入所述第一上采样深度卷积模块,得到第一深度卷积图像；

将所述第一深度卷积图像和所述第一融合图像拼接,得到第三上采样图像；

将所述第一卷积图像和粒度等级的第二细化分割图像输入对应粒度等级的细化分割模块,得到对应粒度等级的第三细化分割图像。

4. 根据权利要求3所述的基于深度学习的多颗粒度人机共融环境感知方法,其特征在于,所述将所述第一卷积图像和粒度等级的第二细化分割图像输入对应粒度等级的细化分割模块,得到对应粒度等级的第三细化分割图像,包括:

将粒度等级的第二细化分割图像输入所述第一上采样卷积模块,得到第一上采样卷积图像；

将所述第一卷积图像和所述第一上采样卷积图像拼接后输入所述第一卷积层,得到对应粒度等级的第三细化分割图像。

5. 根据权利要求2所述的基于深度学习的多颗粒度人机共融环境感知方法,其特征在于,所述将所述第一下采样深度图像和所述第一下采样RGB图像,输入所述第一融合模块,得到第一融合图像,包括:

将所述第一下采样深度图像输入所述第一池化卷积模块,得到第一池化卷积图像；

将所述第一下采样RGB图像输入所述第二池化卷积模块,得到第二池化卷积图像；

将所述第一池化卷积图像和所述第二池化卷积图像拼接,得到第一融合图像。

6. 根据权利要求2所述的基于深度学习的多颗粒度人机共融环境感知方法,其特征在于,所述将所述解码图像输入多粒度分割输出模块,得到不同粒度等级的场景分割图像,包括:

将所述第三上采样图像输入所述第二卷积层,得到第二卷积图像；

将粒度等级的第三细化分割图像输入对应粒度等级的分割输出模块中的第二上采样卷积模块,得到第二上采样卷积图像；

将所述第二卷积图像和所述第二上采样卷积图像拼接后输入所述第二上采样深度卷积模块,得到第二深度卷积图像；

将所述第二深度卷积图像输入所述第三上采样深度卷积模块,得到对应粒度等级的场景分割图像。

7. 根据权利要求1所述的基于深度学习的多颗粒度人机共融环境感知方法,其特征在

于,所述将所述编码图像输入金字塔池化模块,得到池化图像,包括:

将所述编码图像的不同区域输入所述池化层,得到不同尺寸的区域池化图像;

将不同尺寸的区域池化图像分别输入所述第四卷积层,得到不同尺寸的区域卷积图像;

将不同尺寸的区域卷积图像分别输入所述上采样层,得到对应的区域上采样图像;各区域上采样图像的尺寸均与所述编码图像的尺寸相同;

将所述编码图像和所有所述区域上采样图像拼接,得到所述池化图像。

8.根据权利要求1所述的基于深度学习的多颗粒度人机共融环境感知方法,其特征在于,基于加权交叉熵损失函数对所述解码网络进行训练;

所述加权交叉熵损失函数为:

$$\mathcal{L}_{wce} = - \sum_{i \in \mathcal{C}} w_i \cdot y_i \log(p_i)$$

其中, \mathcal{C} 表示实体的类别, \mathcal{L}_{wce} 表示加权交叉熵损失函数, w_i 表示第*i*个实体的类别的权重, y_i 表示第*i*个实体的类别的标签图像, p_i 表示第*i*个实体的类别的不同粒度等级的场景分割图像。

9.一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至8中任一项所述方法的步骤。

10.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至8中任一项所述的方法的步骤。

一种基于深度学习的多颗粒度人机共融环境感知方法

技术领域

[0001] 本发明涉及人机协同智能制造装配领域,尤其涉及一种基于深度学习的多颗粒度人机共融环境感知方法。

背景技术

[0002] 工业机器人是智能制造系统不可或缺的重要组成部分,然而传统的工业机器人部署成本高企,需要围绕其进行整个生产线的改造和适应,导致生产线的柔性不足;另外,中小企业受限于资金和规模,也无法大规模部署应用传统工业机器人。在此背景下,协作机器人以及人机共融生产制造模式则逐渐开始受到越来越多的关注,由人类承担对柔性、触觉、灵活性等要求比较高的工作环节,协作机器人则利用其快速精准的优势来负责重复性和程序化的工作环节,既能满足产线柔性化和灵活性的以及大规模个性化的需求,又能提高生产效率,减轻工作人员作业劳动负担。

[0003] 这种近距离的人机共融协作模式,则需要机器人能够对其所处的人机共融环境进行实时准确的感知,从而根据协作人员的动作和环境变化适应性地调整其工作姿态和作业任务。早期的环境感知技术主要仅依靠深度相机或触觉传感来判断机器人和周围障碍物的距离,而近年来随着人工智能技术的发展,更多地开始采用深度神经网络对场景进行语义分割,即根据视觉图像的语义信息将图像划分为不同的类别区域。

[0004] 然而现有场景语义分割方法主要仍然停留在单一颗粒度语义分割模式(例如,无论场景大小,人体总被当作一个整体进行分割),而没有考虑到协作机器人在作业过程中可能有不同的感知距离和不同精细度的任务。例如,在中等距离情况下,机器人或许只需要将工人身体作为一个整体来识别即可避免碰撞;然而在近距离人机协作装配的情况下,机器人则需要更精细地分割判别人体的手部、手臂、身体等不同部位,从而进行更加精细的运动控制和协作动作。而这种具体场景和任务的多变性则要求协作机器人的场景感知算法应该具有多种不同粒度的语义理解能力。

发明内容

[0005] 本发明提供了一种基于深度学习的多颗粒度人机共融环境感知算法模型,通过同时进行区域、实体、部分三个由粗到细的粒度等级的场景分割,为协作机器人提供更加完善的环境感知能力,使其能够根据环境和任务的不同自适应地切换不同粒度的环境感知分割结果,从而更好地进行后续协作行为决策和运动规划。

[0006] 为了实现上述目的,本发明通过以下技术方案实现:

[0007] 一种基于深度学习的多颗粒度人机共融环境感知方法,其中,该方法包括步骤:

[0008] 获取人机共融场景的RGB图像和深度图像;其中,所述RGB图像和所述深度图像是对同一人机共融场景拍摄得到的图像;

[0009] 将所述RGB图像和所述深度图像输入编码网络,得到编码图像;

[0010] 将所述编码图像输入金字塔池化模块,得到池化图像;

- [0011] 将所述池化图像输入解码网络,得到解码图像;
- [0012] 将所述解码图像输入多粒度分割输出模块,得到不同粒度等级的场景分割图像;
- [0013] 其中,所述粒度等级包括区域等级、实体等级以及实体的部分等级;
- [0014] 所述解码网络包括:第一上采样模块、第二上采样模块以及第三上采样模块;所述将所述池化图像输入解码网络,得到解码图像,包括:
- [0015] 将所述池化图像输入第一上采样模块,得到不同粒度等级的第一细化分割图像以及第一上采样图像;
- [0016] 将所述第一上采样图像和第一细化分割图像输入所述第二上采样模块,得到不同粒度等级的第二细化分割图像以及第二上采样图像;
- [0017] 将所述第二上采样图像和第二细化分割图像输入所述第三上采样模块,得到不同粒度等级的第三细化分割图像以及第三上采样图像,并将所述第三细化分割图像以及所述第三上采样图像作为解码图像。
- [0018] 所述的基于深度学习的多颗粒度人机共融环境感知方法,其中,所述编码网络包括:第一下采样模块、第二下采样模块、第三下采样模块、第四下采样模块、第一融合模块、第二融合模块、第三融合模块以及第四融合模块;
- [0019] 所述将所述RGB图像和所述深度图像输入编码网络,得到编码图像,包括:
- [0020] 将所述深度图像和所述RGB图像分别输入所述第一下采样模块,得到第一下采样深度图像和第一下采样RGB图像;
- [0021] 将所述第一下采样深度图像和所述第一下采样RGB图像输入所述第一融合模块,得到第一融合图像;
- [0022] 将所述第一下采样深度图像和所述第一融合图像分别输入所述第二下采样模块,得到第二下采样深度图像和第二下采样RGB图像;
- [0023] 将所述第二下采样深度图像和所述第二下采样RGB图像输入所述第二融合模块,得到第二融合图像;
- [0024] 将所述第二下采样深度图像和所述第二融合图像分别输入所述第三下采样模块,得到第三下采样深度图像和第三下采样RGB图像;
- [0025] 将所述第三下采样深度图像和所述第三下采样RGB图像输入所述第三融合模块,得到第三融合图像;
- [0026] 将所述第三下采样深度图像和所述第三融合图像分别输入所述第四下采样模块,得到第四下采样深度图像和第四下采样RGB图像;
- [0027] 将所述第四下采样深度图像和所述第四下采样RGB图像输入所述第四融合模块,得到第四融合图像,并将所述第四融合图像作为编码图像。
- [0028] 所述的基于深度学习的多颗粒度人机共融环境感知方法,其中,所述第一融合模块与所述第三上采样模块跳跃连接;所述第二融合模块与所述第二上采样模块跳跃连接;所述第三融合模块与所述第一上采样模块跳跃连接;
- [0029] 所述第三上采样模块包括:第一卷积模块、第一上采样深度卷积模块、不同粒度等级的细化分割模块;
- [0030] 所述将所述第二上采样图像和第二细化分割图像输入所述第三上采样模块,得到不同粒度等级的第三细化分割图像以及第三上采样图像,包括:

- [0031] 将所述第二上采样图像输入第一卷积模块,得到第一卷积图像;
- [0032] 将所述第一卷积图像输入所述第一上采样深度卷积模块,得到第一深度卷积图像;
- [0033] 将所述第一深度卷积图像和所述第一融合图像拼接,得到第三上采样图像;
- [0034] 将所述第一卷积图像和粒度等级的第二细化分割图像输入对应粒度等级的细化分割模块,得到对应粒度等级的第三细化分割图像。
- [0035] 所述的基于深度学习的多颗粒度人机共融环境感知方法,其中,所述细化分割模块包括:第一上采样卷积模块和第一卷积层;所述将所述第一卷积图像和粒度等级的第二细化分割图像输入对应粒度等级的细化分割模块,得到对应粒度等级的第三细化分割图像,包括:
- [0036] 将粒度等级的第二细化分割图像输入所述第一上采样卷积模块,得到第一上采样卷积图像;
- [0037] 将所述第一卷积图像和所述第一上采样卷积图像拼接后输入所述第一卷积层,得到对应粒度等级的第三细化分割图像。
- [0038] 所述的基于深度学习的多颗粒度人机共融环境感知方法,其中,所述第一融合模块包括:第一池化卷积模块和第二池化卷积模块;
- [0039] 所述将所述第一下采样深度图像和所述第一下采样RGB图像,输入所述第一融合模块,得到第一融合图像,包括:
- [0040] 将所述第一下采样深度图像输入所述第一池化卷积模块,得到第一池化卷积图像;
- [0041] 将所述第一下采样RGB图像输入所述第二池化卷积模块,得到第二池化卷积图像;
- [0042] 将所述第一池化卷积图像和所述第二池化卷积图像拼接,得到第一融合图像。
- [0043] 所述的基于深度学习的多颗粒度人机共融环境感知方法,其中,所述多粒度分割输出模块包括:不同粒度等级的分割输出模块,各粒度等级的分割输出模块包括:第二上采样卷积模块、第二卷积层、第二上采样深度卷积模块以及第三上采样深度卷积模块;
- [0044] 所述将所述解码图像输入多粒度分割输出模块,得到不同粒度等级的场景分割图像,包括:
- [0045] 将所述第三上采样图像输入所述第二卷积层,得到第二卷积图像;
- [0046] 将粒度等级的第三细化分割图像输入对应粒度等级的分割输出模块中的第二上采样卷积模块,得到第二上采样卷积图像;
- [0047] 将所述第二卷积图像和所述第一上采样卷积图像拼接后输入所述第二上采样深度卷积模块,得到第二深度卷积图像;
- [0048] 将所述第二深度卷积图像输入所述第三上采样深度卷积模块,得到对应粒度等级的场景分割图像。
- [0049] 所述的基于深度学习的多颗粒度人机共融环境感知方法,其中,所述金字塔池化模块包括:池化层、第四卷积层以及上采样层;所述将所述编码图像输入金字塔池化模块,得到池化图像,包括:
- [0050] 将所述编码图像的不同区域输入所述池化层,得到不同尺寸的区域池化图像;
- [0051] 将不同尺寸的区域池化图像分别输入所述第四卷积层,得到不同尺寸的区域卷积

图像；

[0052] 将不同尺寸的区域卷积图像分别输入所述上采样层,得到对应的区域上采样图像;各区域上采样图像的尺寸均与所述编码图像的尺寸相同;

[0053] 将所述编码图像和所有所述区域上采样图像拼接,得到所述池化图像。

[0054] 所述的基于深度学习的多颗粒度人机共融环境感知方法,其中,基于加权交叉熵损失函数对所述解码网络进行训练;

[0055] 所述加权交叉熵损失函数为:

$$[0056] \quad \mathcal{L}_{wce} = - \sum_{i \in \mathcal{C}} w_i \cdot y_i \log(p_i)$$

[0057] 其中, \mathcal{C} 表示实体的类别, \mathcal{L}_{wce} 表示加权交叉熵损失函数, w_i 表示第*i*个实体的类别的权重, y_i 表示第*i*个实体的类别的标签图像, p_i 表示第*i*个实体的类别的不同粒度等级的场景分割图像。

[0058] 一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其中,所述处理器执行所述计算机程序时实现如上任一项所述方法的步骤。

[0059] 一种计算机可读存储介质,其上存储有计算机程序,其中,所述计算机程序被处理器执行时实现如上任一项所述的方法的步骤。

[0060] 有益效果:

[0061] 获取RGB图像和深度图像后,将RGB图像和深度图像输入编码网络,提取并融合RGB图像和深度图像中的特征,得到编码图像。得到编码图像后,将编码图像输入金字塔池化模块进行全局-局部特征融合提取,得到池化图像。得到池化图像后,将池化图像输入解码图像,得到解码图像。得到解码图像后,输入多粒度分割输出模块,得到不同粒度等级的场景分割图像,为协作机器人提供更加完善的环境感知能力,使其能够根据环境和任务的不同自适应地切换不同粒度的环境感知分割结果,从而更好地进行后续协作行为决策和运动规划。

附图说明

[0062] 图1为本发明的针对人机共融装配环境的多粒度场景分割准则的示意图;

[0063] 图2为本发明的编码-解码模型的第一结构示意图;

[0064] 图3为本发明的编码-解码模型的第二结构示意图;

[0065] 图4为本发明中下采样模块的示意图;

[0066] 图5为本发明中融合模块的示意图;

[0067] 图6为本发明中上采样模块的示意图;

[0068] 图7为本发明中多粒度分割输出模块的示意图;

[0069] 图8为本发明中金字塔池化模块的示意图;

[0070] 图9为本发明中RGB图像、深度图像、不同粒度等级的场景分割图像以及不同粒度等级的标签图像的对比示意图;

[0071] 图10为本发明的基于深度学习的多颗粒度人机共融环境感知方法的整体流程图。

具体实施方式

[0072] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。此外,下面所描述的本发明各个实施方式中所涉及到的技术特征只要彼此之间未构成冲突就可以相互组合。

[0073] 请同时参阅图1-图10,本发明提供了一种基于深度学习的多颗粒度人机共融环境感知方法的一些实施例。

[0074] 如图2和图3所示,本发明提供了一种编码-解码模型,编码-解码模型包括:编码网络、金字塔池化模块、解码网络以及多粒度分割输出模块。

[0075] 编码网络用于对图像进行下采样,金字塔池化模块用于全局-局部特征融合,解码网络用于对图像进行上采样,多粒度分割输出模块用于输出不同粒度等级的场景分割图像。粒度等级是指图像中分割图像的尺寸等级,本申请的编码-解码模型应用于人机共融场景,人机共融场景是指协作机器人以及人机共融生产制造的场景,也就是说,在人机共融场景中有机人、工作人员、生产制造工具、生产制造原料及产品、划分的区域等,人机共融场景中各目标的尺寸有大有小,在不同精细度的任务中,机器人需要进行不同精细度的图像分割。因此,将人机共融场景中各目标按照粒度等级进行划分,具体粒度等级的级数可以根据需要设定,如图1所示,本发明中将粒度等级分成三个等级,即区域等级(Area-level)、实体等级(Entity-level)以及实体的部分等级(Part-level),区域等级是指图像中尺寸较大且连接形成区域的等级,例如,区域等级还可以具体分为人机协作区域、仓储区域、办公区域、自由区域、不可达区域等。实体等级是指图像中尺寸较小的目标的等级,例如,实体等级还可以具体分为人体、机器人、无人小车、工具、电池、电脑、桌子、椅子等。实体的部分等级是指图像中实体的各部位的等级,例如,人体的部分等级可以分为手臂、手掌、头、脚、躯干等,机器人的部分等级可以分为底座、主体夹具等,无人小车的部分等级可以分为底座、轮子、载物台,工具的部分等级可以分为把手、前端等,电池的部分等级可以分为外壳、连接线、短电池组、长电池组等,电脑的部分等级可以分为机箱、键盘、鼠标、显示器等。

[0076] 编码网络包括:第一下采样模块、第二下采样模块、第三下采样模块、第四下采样模块、第一融合模块、第二融合模块、第三融合模块以及第四融合模块。编码网络包括四个下采样模块和4个融合模块,下采样模块用于对图像进行下采样,融合模块用于深度图像和RGB图像进行融合。下采样模块包括:1个深度卷积层和2个卷积层。融合模块包括:2个池化卷积模块,池化卷积模块包括:1个池化层和2个卷积层。

[0077] 金字塔池化模块包括:1个池化层、1个卷积层以及1个上采样层。

[0078] 解码网络包括:第一上采样模块、第二上采样模块以及第三上采样模块。解码网络包括3个上采样模块,通过上采样模块对图像进行上采样。上采样模块包括:1个卷积模块、1个上采样深度卷积模块以及3个细化分割模块;卷积模块包括:5个卷积层,上采样深度卷积模块包括:1个上采样层和1个深度卷积层,细化分割模块包括:1个上采样卷积模块和1个卷积层,上采样卷积模块包括:1个上采样层和2个卷积层。

[0079] 多粒度分割输出模块包括:不同粒度等级的分割输出模块,由于粒度等级有3个,分割输出模块也有3个。分割输出模块包括:1个上采样卷积模块、1个卷积层以及2个上采样深度卷积模块,上采样卷积模块包括:1个上采样层和2个卷积层,上采样深度卷积模块包

括:1个上采样层和1个深度卷积层。

[0080] 如图1-图3以及图10所示,本发明的一种基于深度学习的多颗粒度人机共融环境感知方法,包括如下步骤:

[0081] 步骤S100、获取人机共融场景的RGB图像和深度图像;其中,所述RGB图像和所述深度图像是对同一人机共融场景拍摄得到的图像。

[0082] 步骤S200、将所述RGB图像和所述深度图像输入编码网络,得到编码图像。

[0083] 步骤S300、将所述编码图像输入金字塔池化模块,得到池化图像。

[0084] 步骤S400、将所述池化图像输入解码网络,得到解码图像。

[0085] 步骤S500、将所述解码图像输入多粒度分割输出模块,得到不同粒度等级的场景分割图像。

[0086] 对人机共融场景进行拍摄,得到RGB图像和深度图像,RGB图像和深度图像可以采用RGB-D相机拍摄得到,也可以分别由RGB相机和深度相机分别拍摄得到。RGB图像和深度图像是对同一人机共融场景拍摄得到的图像,也就是说,RGB图像和深度图像中目标是相同的目标,且目标的位置位于相同位置。

[0087] 获取RGB图像和深度图像后,将RGB图像和深度图像输入编码网络,提取并融合RGB图像和深度图像中的特征,得到编码图像。得到编码图像后,将编码图像输入金字塔池化模块进行全局-局部特征融合提取,得到池化图像。得到池化图像后,将池化图像输入解码网络,得到解码图像。得到解码图像后,输入多粒度分割输出模块,得到不同粒度等级的场景分割图像,为协作机器人提供更加完善的环境感知能力,使其能够根据环境和任务的不同自适应地切换不同粒度的环境感知分割结果,从而更好地进行后续协作行为决策和运动规划。

[0088] 在一种实现方式中,解码网络主要由三个连续的上采样模块组成,逐渐解码和放大特征图。所述解码网络包括:第一上采样模块、第二上采样模块以及第三上采样模块;步骤S400包括:

[0089] 步骤S410、将所述池化图像输入第一上采样模块,得到不同粒度等级的第一细化分割图像以及第一上采样图像。

[0090] 步骤S420、将所述第一上采样图像和第一细化分割图像输入所述第二上采样模块,得到不同粒度等级的第二细化分割图像以及第二上采样图像。

[0091] 步骤S430、将所述第二上采样图像和第二细化分割图像输入所述第三上采样模块,得到不同粒度等级的第三细化分割图像以及第三上采样图像,并将所述第三细化分割图像以及所述第三上采样图像作为解码图像。

[0092] 具体地,得到池化图像后,将池化图像输入第一上采样模块,得到第一上采样图像和不同粒度等级的第一细化分割图像,例如,可以得到3个粒度等级的第一细化分割图像,具体有区域等级的第一细化分割图像,实体等级的第一细化分割图像以及部分等级的第一细化分割图像。

[0093] 然后将第一上采样图像和第一细化分割图像输入第二上采样模块,得到不同粒度等级的第二细化分割图像以及第二上采样图像。可以理解的是,采用区域等级的第一细化分割图像作为第二上采样模块的输入,则得到的是区域等级的第二细化分割图像。采用实体等级的第一细化分割图像作为第二上采样模块的输入,则得到的是实体等级的第二细化

分割图像。

[0094] 再将第二上采样图像和第二细化分割图像输入第三上采样模块,得到不同粒度等级的第三细化分割图像以及第三上采样图像,则可以将第三细化分割图像以及第三上采样图像作为解码图像。

[0095] 在一种实现方式中,所述编码网络包括:第一下采样模块、第二下采样模块、第三下采样模块、第四下采样模块、第一融合模块、第二融合模块、第三融合模块以及第四融合模块;步骤S200包括:

[0096] 步骤S210、将所述深度图像和所述RGB图像分别输入所述第一下采样模块,得到第一下采样深度图像和第一下采样RGB图像。

[0097] 步骤S220、将所述第一下采样深度图像和所述第一下采样RGB图像输入所述第一融合模块,得到第一融合图像。

[0098] 步骤S230、将所述第一下采样深度图像和所述第一融合图像分别输入所述第二下采样模块,得到第二下采样深度图像和第二下采样RGB图像。

[0099] 步骤S240、将所述第二下采样深度图像和所述第二下采样RGB图像输入所述第二融合模块,得到第二融合图像。

[0100] 步骤S250、将所述第二下采样深度图像和所述第二融合图像分别输入所述第三下采样模块,得到第三下采样深度图像和第三下采样RGB图像。

[0101] 步骤S260、将所述第三下采样深度图像和所述第三下采样RGB图像输入所述第三融合模块,得到第三融合图像。

[0102] 步骤S270、将所述第三下采样深度图像和所述第三融合图像分别输入所述第四下采样模块,得到第四下采样深度图像和第四下采样RGB图像。

[0103] 步骤S280、将所述第四下采样深度图像和所述第四下采样RGB图像输入所述第四融合模块,得到第四融合图像,并将所述第四融合图像作为编码图像。

[0104] 具体地,第一下采样模块、第二下采样模块、第三下采样模块、第四下采样模块采用相同结构的下采样模块,如图4所示,该下采样模块包括:依次连接的 7×7 深度卷积层、 1×1 卷积层以及 1×1 卷积层。

[0105] 编码器网络的目的是提取RGB和深度特征,并在不同阶段聚合它们,以便更好地利用RGB和深度图中的互补信息。在编码和解码过程中均采用逐步的方式进行,在编码过程中,采用4个下采样模块。采用逐步融合的方式,将RGB图像和深度图像融合,具体采用4个融合模块进行融合。在编码器网络的主干网络的每个阶段,通过融合模块将深度特征融合到RGB分支中。

[0106] 先采用第一下采样模块分别对深度图像和RGB图像进行下采样,且分别得到第一下采样深度图像和第一下采样RGB图像。然后采用第一融合模块将第一下采样深度图像和第一下采样RGB图像,得到第一融合图像。接着进行下一步的下采样,经过4次下采样后得到第四融合图像,并将第四融合图像作为编码图像。

[0107] 为了加快训练,融合模块与和上采样模块采用跳跃连接,具体地,所述第一融合模块与所述第三上采样模块跳跃连接;所述第二融合模块与所述第二上采样模块跳跃连接;所述第三融合模块与所述第一上采样模块跳跃连接。所述第一上采样模块包括:第一卷积模块、第一上采样深度卷积模块、不同粒度等级的细化分割模块。在一种实现方式中,步骤

S410包括:

[0108] 步骤S411、将所述池化图像输入所述第一上采样模块的第一卷积模块,得到第一卷积图像。

[0109] 步骤S412、将所述第一卷积图像输入所述第一上采样模块的第一上采样深度卷积模块,得到第一深度卷积图像。

[0110] 步骤S413、将所述第一深度卷积图像和所述第三融合图像拼接,得到第一上采样图像。

[0111] 步骤S414、将所述第一卷积图像输入对应粒度等级的细化分割模块,得到对应粒度等级的第一细化分割图像。

[0112] 需要说明的是,由于池化图像的尺寸较小,第一上采样模块中不同粒度等级的细化分割模块没有细化分割图像的输入。仅是将第一卷积图像作为对应粒度等级的细化分割模块的输入,得到对应粒度等级的第一细化分割图像。在一种实现方式中,所述细化分割模块包括:第一上采样卷积模块和第一卷积层;步骤S414包括:

[0113] 步骤S4141、将第一卷积图像输入粒度等级的细化分割模块的第一上采样卷积模块,得到第一上采样卷积图像。

[0114] 步骤S4142、将所述第一卷积图像和所述第一上采样卷积图像拼接后输入所述第一卷积层,得到对应粒度等级的第一细化分割图像。

[0115] 在一种实现方式中,所述第二上采样模块包括:第一卷积模块、第一上采样深度卷积模块、不同粒度等级的细化分割模块;步骤S420包括:

[0116] 步骤S421、将所述第一上采样图像输入所述第二上采样模块的第一卷积模块,得到第一卷积图像。

[0117] 步骤S422、将所述第一卷积图像输入所述第二上采样模块的第一上采样深度卷积模块,得到第一深度卷积图像。

[0118] 步骤S423、将所述第一深度卷积图像和所述第二融合图像拼接,得到第二上采样图像。

[0119] 步骤S424、将所述第一卷积图像和粒度等级的第二细化分割图像输入对应粒度等级的细化分割模块,得到对应粒度等级的第二细化分割图像。

[0120] 在一种实现方式中,所述细化分割模块包括:第一上采样卷积模块和第一卷积层;步骤S424包括:

[0121] 步骤S4241、将粒度等级的第二细化分割图像输入对应粒度等级的细化分割模块的第一上采样卷积模块,得到第一上采样卷积图像。

[0122] 步骤S4242、将所述第一卷积图像和所述第一上采样卷积图像拼接后输入所述第一卷积层,得到对应粒度等级的第一细化分割图像。

[0123] 在一种实现方式中,所述第三上采样模块包括:第一卷积模块、第一上采样深度卷积模块、不同粒度等级的细化分割模块;步骤S430包括:

[0124] 步骤S431、将所述第二上采样图像输入所述第三上采样模块的第一卷积模块,得到第一卷积图像。

[0125] 步骤S432、将所述第一卷积图像输入所述第三上采样模块的第一上采样深度卷积模块,得到第一深度卷积图像。

[0126] 步骤S433、将所述第一深度卷积图像和所述第一融合图像拼接,得到第三上采样图像。

[0127] 步骤S434、将所述第一卷积图像和粒度等级的第二细化分割图像输入对应粒度等级的细化分割模块,得到对应粒度等级的第三细化分割图像。

[0128] 在一种实现方式中,所述细化分割模块包括:第一上采样卷积模块和第一卷积层;步骤S434包括:

[0129] 步骤S4341、将粒度等级的第二细化分割图像输入所述第一上采样卷积模块,得到第一上采样卷积图像。

[0130] 步骤S4342、将所述第一卷积图像和所述第一上采样卷积图像拼接后输入所述第一卷积层,得到对应粒度等级的第三细化分割图像。

[0131] 具体地,如图6所示,第一卷积模块包括:依次连接的 3×3 卷积层、 3×1 卷积层、 1×3 卷积层、 3×1 卷积层、 1×3 卷积层。第一上采样深度卷积模块包括:依次连接的上采样层和 3×3 深度卷积层。第一上采样卷积模块包括:1个上采样层、2个 3×3 卷积层。第一卷积层采用 1×1 卷积层。

[0132] 在一种实现方式中,所述第一融合模块包括:第一池化卷积模块和第二池化卷积模块;步骤S220包括:

[0133] 步骤S221、将所述第一下采样深度图像输入所述第一融合模块的第一池化卷积模块,得到第一池化卷积图像。

[0134] 步骤S222、将所述第一下采样RGB图像输入所述第一融合模块的第二池化卷积模块,得到第二池化卷积图像。

[0135] 步骤S223、将所述第一池化卷积图像和所述第二池化卷积图像拼接,得到第一融合图像。

[0136] 在一种实现方式中,所述第二融合模块包括:第一池化卷积模块和第二池化卷积模块;步骤S240包括:

[0137] 步骤S241、将所述第二下采样深度图像输入所述第二融合模块的第一池化卷积模块,得到第一池化卷积图像。

[0138] 步骤S242、将所述第二下采样RGB图像输入所述第二融合模块的第二池化卷积模块,得到第二池化卷积图像。

[0139] 步骤S243、将所述第一池化卷积图像和所述第二池化卷积图像拼接,得到第二融合图像。

[0140] 在一种实现方式中,所述第三融合模块包括:第一池化卷积模块和第二池化卷积模块;步骤S260包括:

[0141] 步骤S261、将所述第三下采样深度图像输入所述第三融合模块的第一池化卷积模块,得到第一池化卷积图像。

[0142] 步骤S262、将所述第三下采样RGB图像输入所述第三融合模块的第二池化卷积模块,得到第二池化卷积图像。

[0143] 步骤S263、将所述第一池化卷积图像和所述第二池化卷积图像拼接,得到第三融合图像。

[0144] 在一种实现方式中,所述第四融合模块包括:第一池化卷积模块和第二池化卷积

模块;步骤S280包括:

[0145] 步骤S281、将所述第四下采样深度图像输入所述第四融合模块的第一池化卷积模块,得到第一池化卷积图像。

[0146] 步骤S282、将所述第四下采样RGB图像输入所述第四融合模块的第二池化卷积模块,得到第二池化卷积图像。

[0147] 步骤S283、将所述第一池化卷积图像和所述第二池化卷积图像拼接,得到第四融合图像。

[0148] 具体地,如图5所示,第一池化卷积模块包括:依次连接的池化层、 1×1 卷积层以及 1×1 卷积层。第二池化卷积模块包括:依次连接的池化层、 1×1 卷积层以及 1×1 卷积层。

[0149] 在一种实现方式中,所述多粒度分割输出模块包括:不同粒度等级的分割输出模块,各粒度等级的分割输出模块包括:第二上采样卷积模块、第二卷积层、第二上采样深度卷积模块以及第三上采样深度卷积模块。步骤S500具体包括:

[0150] 步骤S510、将所述第三上采样图像输入所述第二卷积层,得到第二卷积图像。

[0151] 步骤S520、将粒度等级的第三细化分割图像输入对应粒度等级的分割输出模块中的第二上采样卷积模块,得到第二上采样卷积图像。

[0152] 步骤S530、将所述第二卷积图像和所述第一上采样卷积图像拼接后输入所述第二上采样深度卷积模块,得到第二深度卷积图像。

[0153] 步骤S540、将所述第二深度卷积图像输入所述第三上采样深度卷积模块,得到对应粒度等级的场景分割图像。

[0154] 具体地,如图7所示,第二上采样卷积模块包括:1个上采样层、2个 3×3 卷积层。第二卷积层采用 3×3 卷积层。第二上采样深度卷积模块包括:依次连接的上采样层和 3×3 深度卷积层。第三上采样深度卷积模块包括:依次连接的上采样层和 3×3 深度卷积层。多粒度分割输出模块将特征图恢复到输入大小,从而可以生成像素级分割结果。

[0155] 在编码网络的特征提取得到编码图像之后,使用金字塔池化模块来处理具有不同池化尺度的特征,能够聚合全局和局部上下文信息。如图8所示,所述金字塔池化模块包括:池化层、第四卷积层以及上采样层。步骤S300包括:

[0156] 步骤S310、将所述编码图像的不同区域输入所述池化层,得到不同尺寸的区域池化图像。

[0157] 步骤S320、将不同尺寸的区域池化图像分别输入所述第四卷积层,得到不同尺寸的区域卷积图像。

[0158] 步骤S330、将不同尺寸的区域卷积图像分别输入所述上采样层,得到对应的区域上采样图像;各区域上采样图像的尺寸均与所述编码图像的尺寸相同。

[0159] 步骤S340、将所述编码图像和所有所述区域上采样图像拼接,得到所述池化图像。

[0160] 对于分割监督,采用交叉熵损失和附加权重项,基于其在地面实况分割图中呈现的像素数量,则基于加权交叉熵损失函数对所述解码网络进行训练;所述加权交叉熵损失函数为:

$$[0161] \quad \mathcal{L}_{wce} = - \sum_{i \in \mathcal{C}} w_i \cdot y_i \log(p_i)$$

[0162] 其中, \mathcal{C} 表示实体的类别, \mathcal{L}_{wce} 表示加权交叉熵损失函数, w_i 表示第*i*个实体的类别的权重, y_i 表示第*i*个实体的类别的标签图像, p_i 表示第*i*个实体的类别的不同粒度等级的场景分割图像。

[0163] 如图9所示, RGB-B的竖栏下两个图像分别RGB图像和深度图像, Area的竖栏下两个图像分别为区域等级的标签图像(即真值, GT)和区域等级的场景分割图像(即预测图像, Pred), Entity的竖栏下两个图像分别为实体等级的标签图像和实体等级的场景分割图像, Part的竖栏下两个图像分别为部分等级的标签图像和部分等级的场景分割图像。

[0164] 该发明的有益之处是:

[0165] 1) 其将多个粒度等级的场景感知划分集成到了一个统一的编码-解码模型里面, 高效地复用了深度卷积神经网络所提取的多尺度特征来同时进行多粒度场景语义分割, 提高了计算效率, 减少了模型运行时间消耗, 能够更好地满足实时性要求较高的人机协作应用。

[0166] 2) 相对于现有的场景语义分割深度学习模型的编码部分, 该发明所提出的模型采用了更加先进的图像-深度双编码网络来提取场景视觉信息, 并提出了一种基于通道注意力机制的图像-深度特征融合模块来对所提取的多尺度视觉信息进行融合, 能够更好地获取更加鲁棒和表征能力更强的特征。

[0167] 3) 相对于现有的场景语义分割深度学习模型的解码部分, 该发明所提出的模型解码部分提出了多尺度附加分割任务监督、由下到上的逐层精细化连接、多分支分割输出端等针对多粒度场景分割所提出的网络组成模块, 能够更好地对多尺度解码过程进行优化, 并且提供所需要的最终多粒度分割输出。

[0168] 本发明编码-解码模型采用如下步骤训练得到:

[0169] 获取人机共融场景的RGB图像和深度图像;

[0170] 将所述RGB图像和所述深度图像输入编码-解码模型, 得到不同粒度等级的场景分割图像;

[0171] 根据不同粒度等级的场景分割图像和标签图像, 对所述编码-解码模型的参数进行更新;

[0172] 当满足预设训练条件时, 得到训练好的编码-解码模型。

[0173] 具体地, 在训练过程中, 基于加权交叉熵损失函数, 进行反向传播计算编码-解码模型的网络参数的梯度并根据梯度进行参数迭代更新。

[0174] 基于上述任意一实施例的基于深度学习的多颗粒度人机共融环境感知方法, 本发明还提供了一种计算机设备的实施例, 本发明的计算机设备, 包括存储器和处理器, 所述存储器存储有计算机程序, 所述处理器执行所述计算机程序时实现如下步骤:

[0175] 获取人机共融场景的RGB图像和深度图像; 其中, 所述RGB图像和所述深度图像是对同一人机共融场景拍摄得到的图像;

[0176] 将所述RGB图像和所述深度图像输入编码网络, 得到编码图像;

[0177] 将所述编码图像输入金字塔池化模块, 得到池化图像;

[0178] 将所述池化图像输入解码网络, 得到解码图像;

[0179] 将所述解码图像输入多粒度分割输出模块, 得到不同粒度等级的场景分割图像;

[0180] 处理器执行计算机程序时还实现如下步骤:

[0181] 将所述池化图像输入第一上采样模块,得到不同粒度等级的第一细化分割图像以及第一上采样图像;

[0182] 将所述第一上采样图像和第一细化分割图像输入所述第二上采样模块,得到不同粒度等级的第二细化分割图像以及第二上采样图像;

[0183] 将所述第二上采样图像和第二细化分割图像输入所述第三上采样模块,得到不同粒度等级的第三细化分割图像以及第三上采样图像,并将所述第三细化分割图像以及所述第三上采样图像作为解码图像。

[0184] 基于上述任意一实施例的基于深度学习的多颗粒度人机共融环境感知方法,本发明还提供了一种计算机可读存储介质的实施例,本发明的计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现如下步骤:

[0185] 获取人机共融场景的RGB图像和深度图像;其中,所述RGB图像和所述深度图像是对同一人机共融场景拍摄得到的图像;

[0186] 将所述RGB图像和所述深度图像输入编码网络,得到编码图像;

[0187] 将所述编码图像输入金字塔池化模块,得到池化图像;

[0188] 将所述池化图像输入解码网络,得到解码图像;

[0189] 将所述解码图像输入多粒度分割输出模块,得到不同粒度等级的场景分割图像;

[0190] 计算机程序被处理器执行时还实现如下步骤:

[0191] 将所述池化图像输入第一上采样模块,得到不同粒度等级的第一细化分割图像以及第一上采样图像;

[0192] 将所述第一上采样图像和第一细化分割图像输入所述第二上采样模块,得到不同粒度等级的第二细化分割图像以及第二上采样图像;

[0193] 将所述第二上采样图像和第二细化分割图像输入所述第三上采样模块,得到不同粒度等级的第三细化分割图像以及第三上采样图像,并将所述第三细化分割图像以及所述第三上采样图像作为解码图像。

[0194] 以上所述,仅为本发明较佳的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,根据本发明的技术方案及其发明构思加以等同替换或改变,都应涵盖在本发明的保护范围之内。

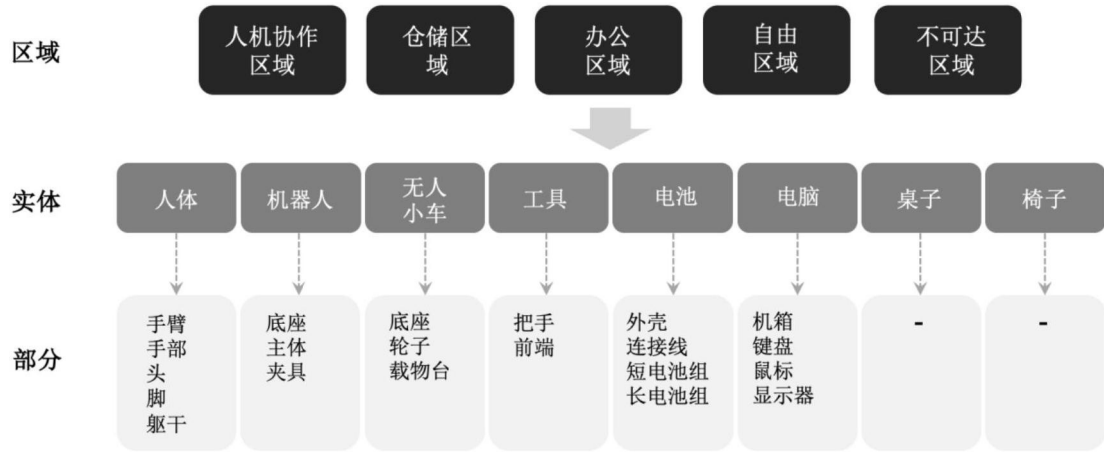


图1

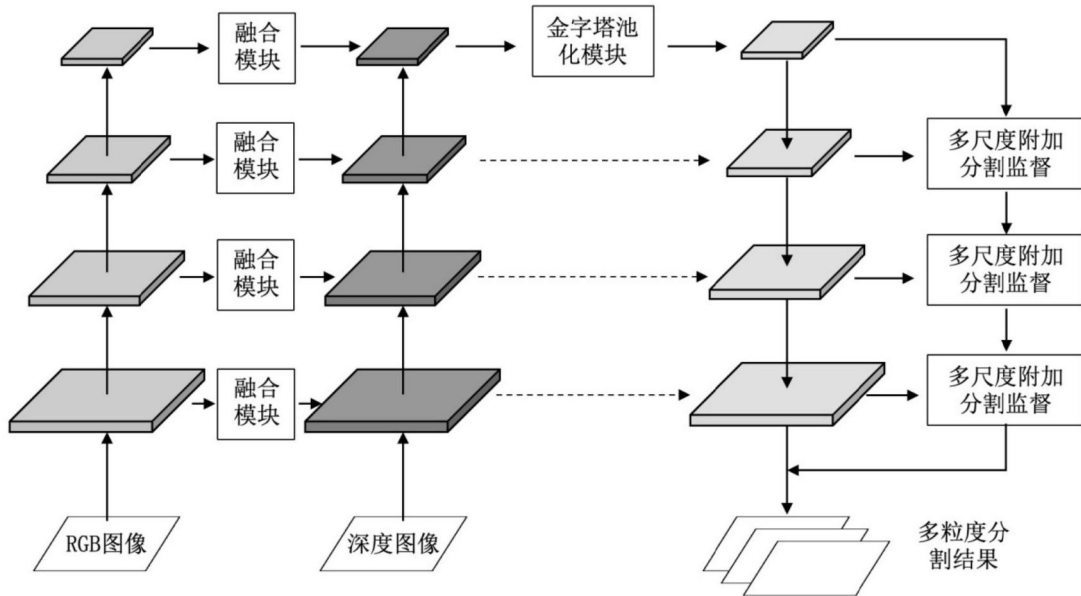


图2

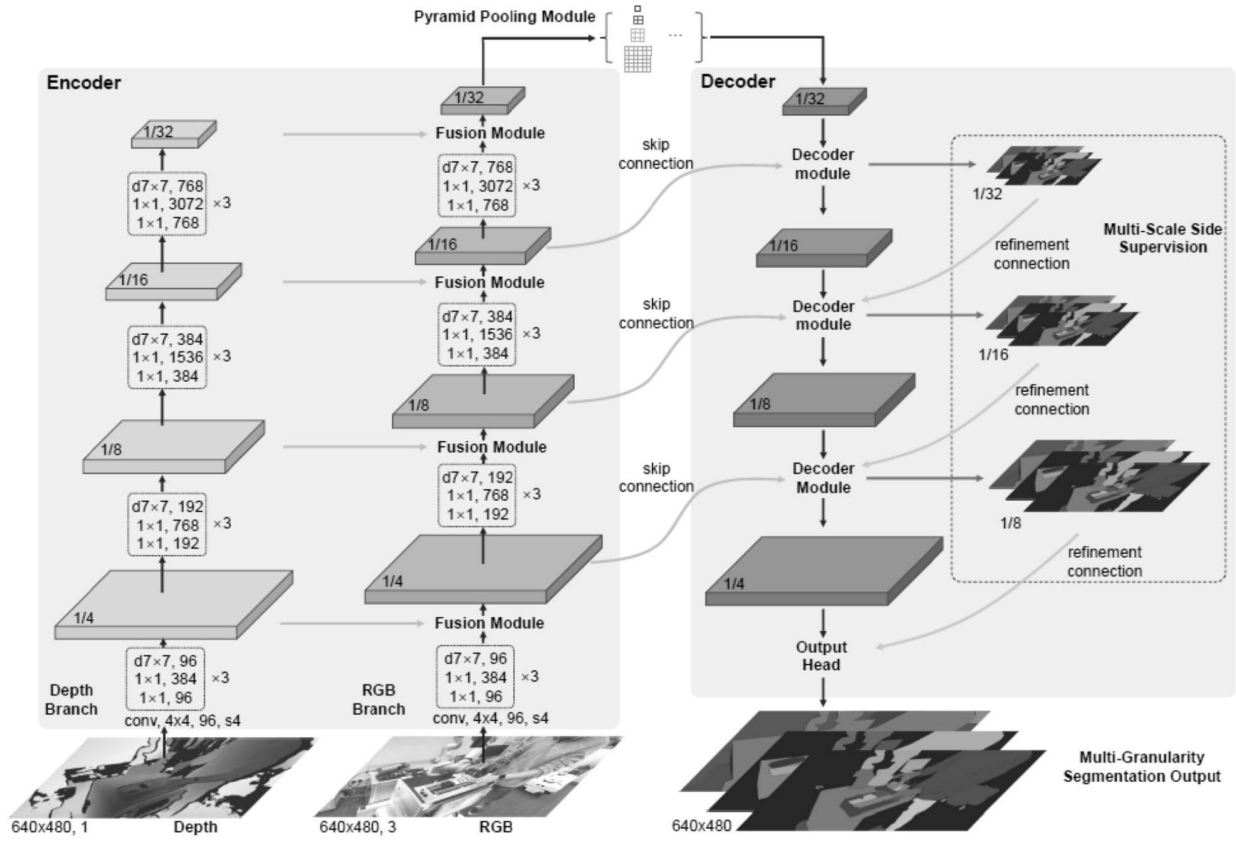


图3

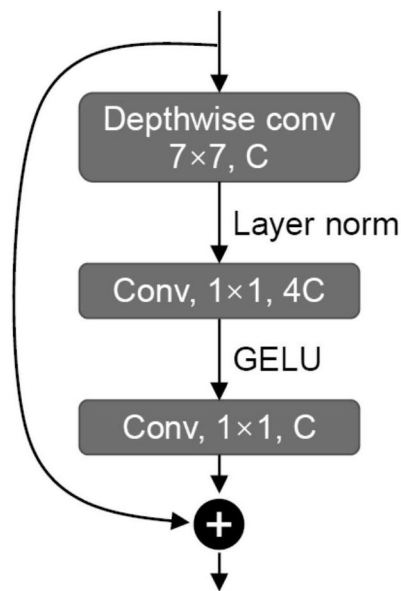


图4

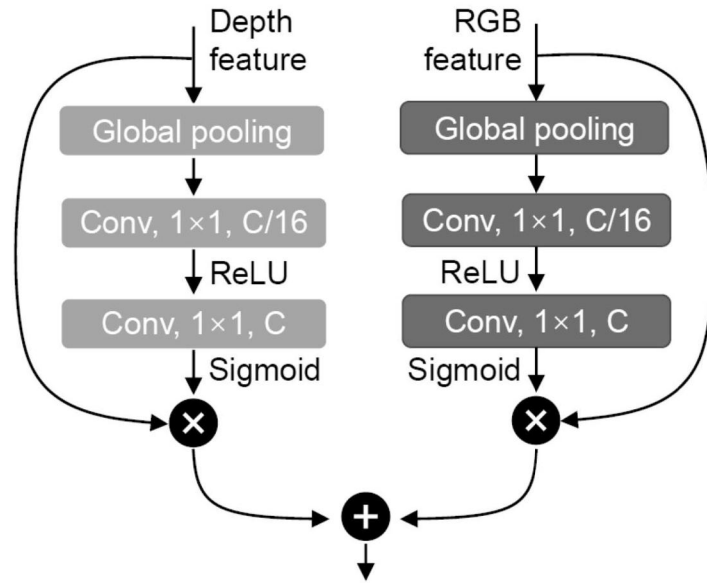


图5

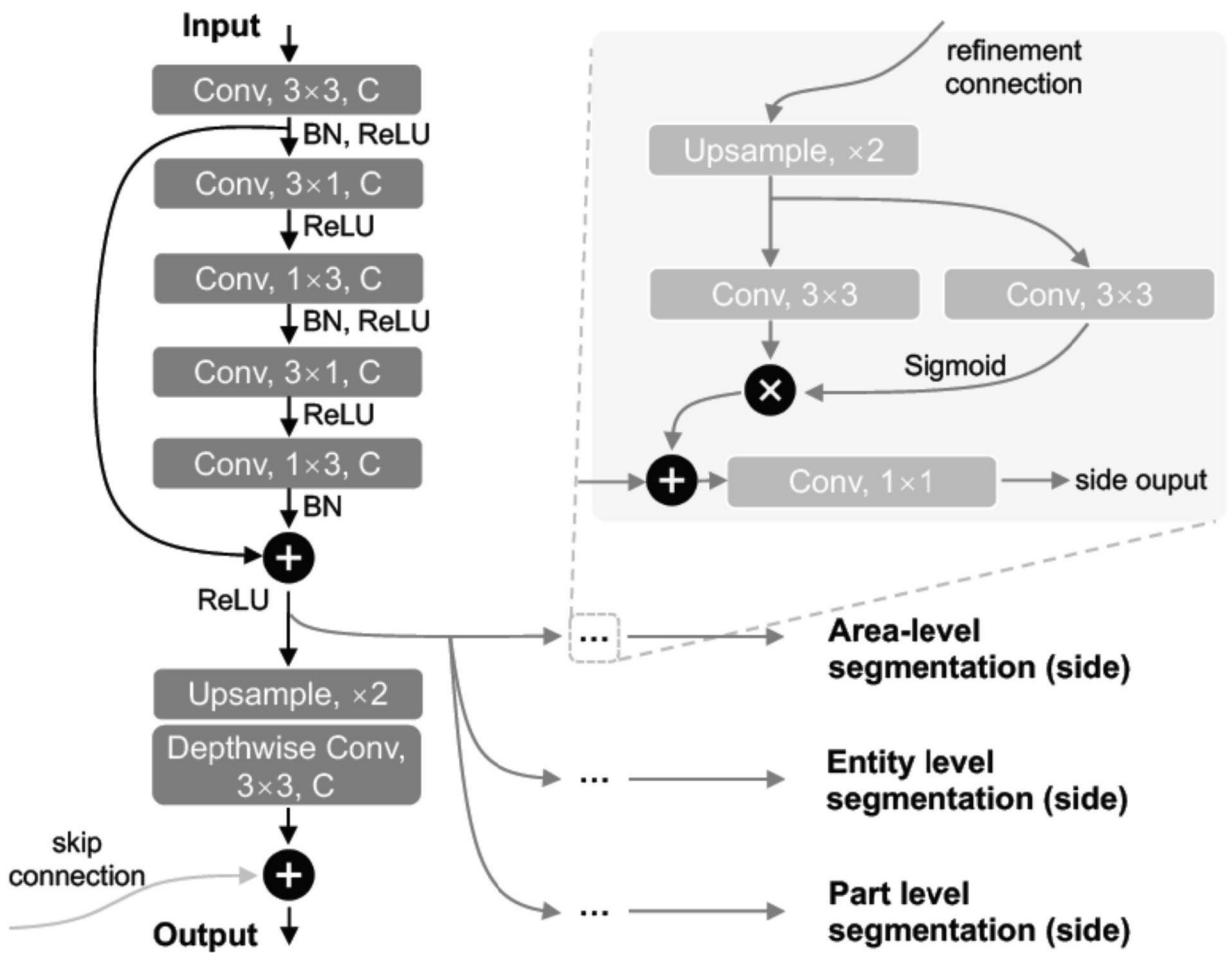


图6

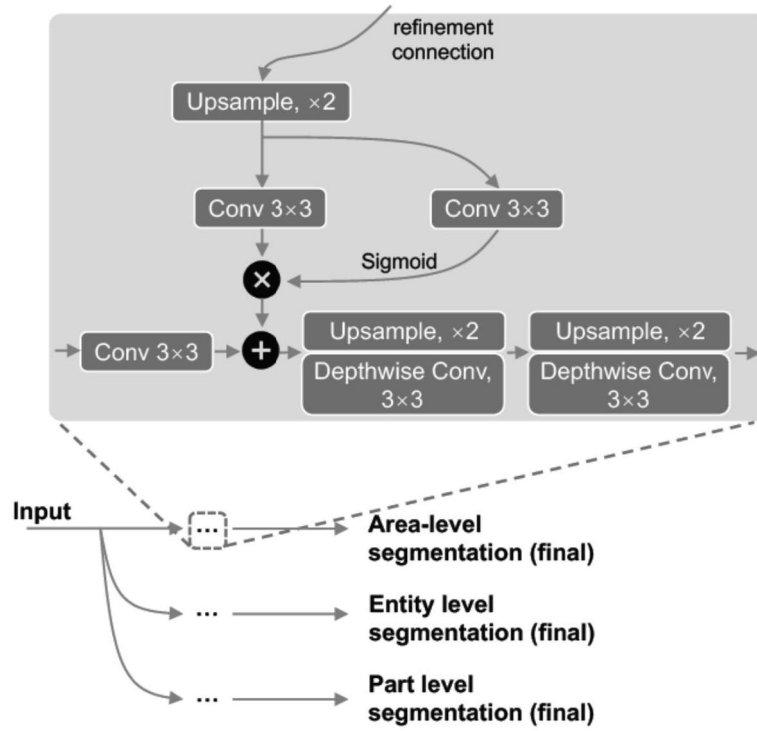


图7

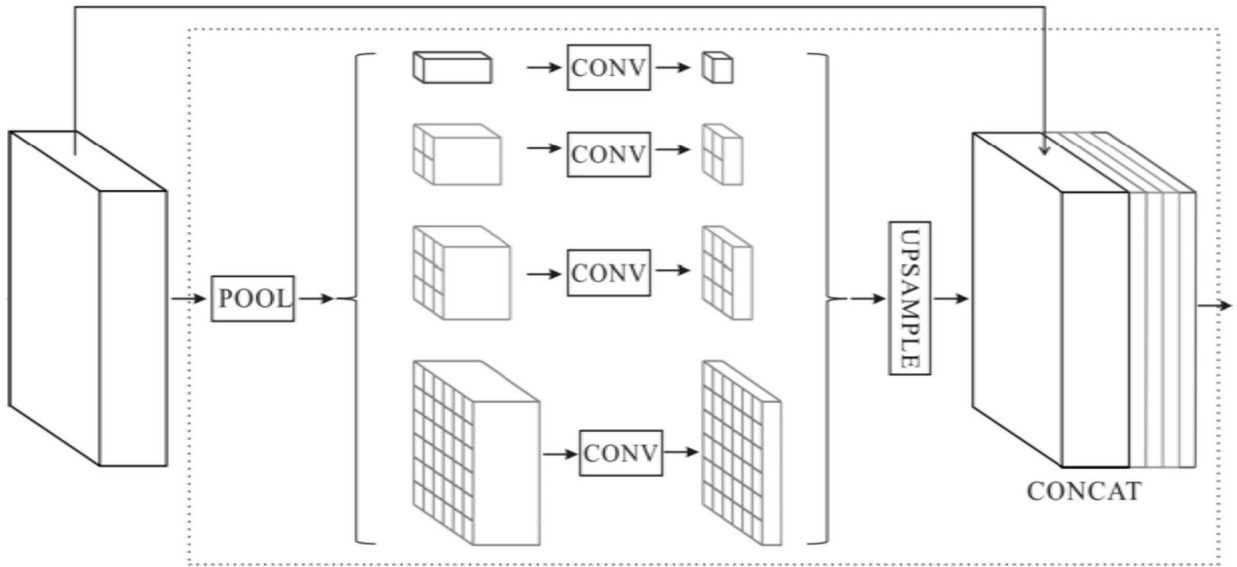


图8

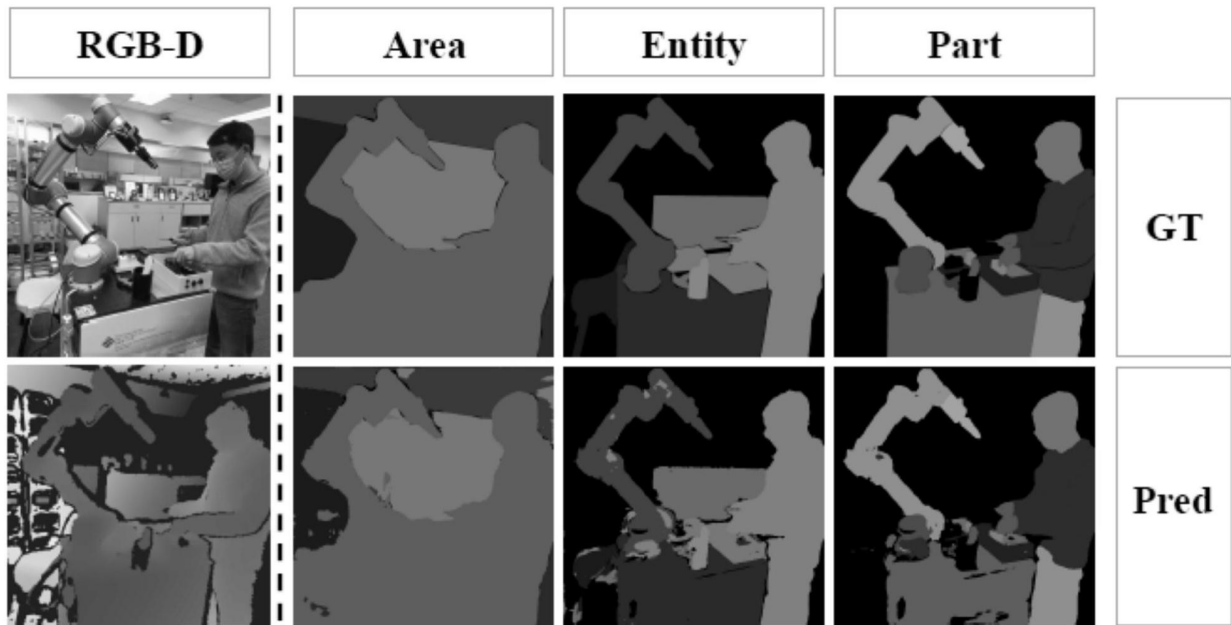


图9

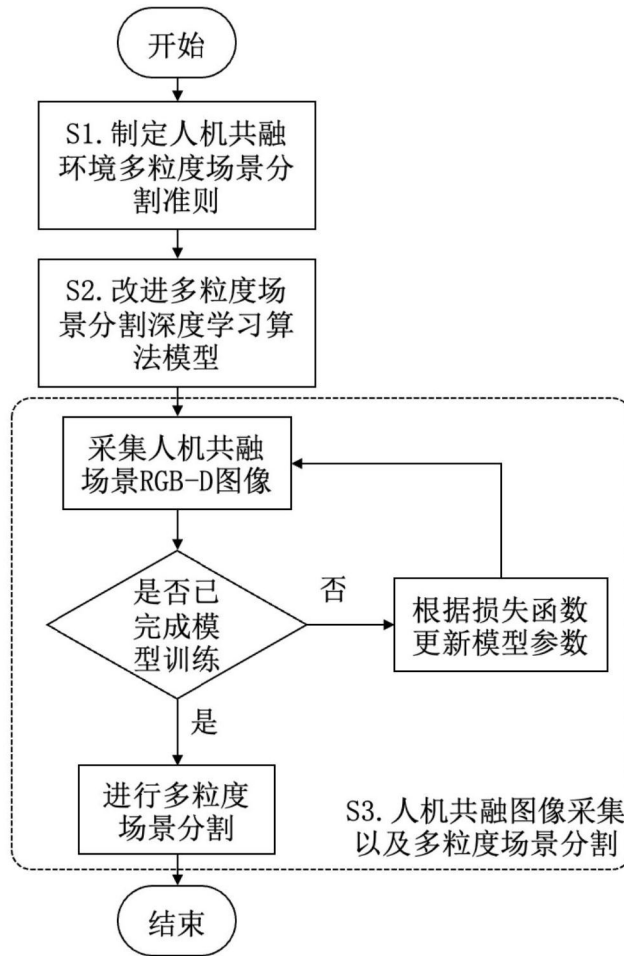


图10