



(12) 发明专利

(10) 授权公告号 CN 115631319 B

(45) 授权公告日 2023. 06. 23

(21) 申请号 202211361582.7

G06F 17/16 (2006.01)

(22) 申请日 2022.11.02

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 114937083 A, 2022.08.23

申请公布号 CN 115631319 A

CN 110689562 A, 2020.01.14

(43) 申请公布日 2023.01.20

CN 112258600 A, 2021.01.22

(73) 专利权人 北京科技大学

审查员 刘津

地址 100083 北京市海淀区学院路30号

专利权人 北京科技大学顺德创新学院

(72) 发明人 曾慧 郑锐 刘红敏 樊彬

张利欣

(74) 专利代理机构 北京市广友专利事务所有限

责任公司 11237

专利代理师 张仲波

(51) Int. Cl.

G06T 17/20 (2006.01)

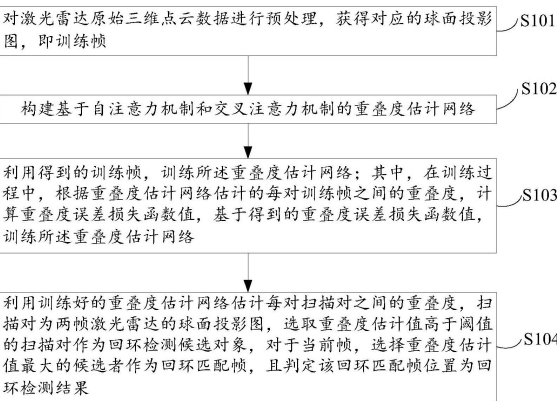
权利要求书3页 说明书9页 附图5页

(54) 发明名称

一种基于交叉注意力网络的回环检测方法

(57) 摘要

本发明提供一种基于交叉注意力网络的回环检测方法,属于计算机视觉技术领域。所述方法包括:对激光雷达原始三维点云数据进行预处理,获得对应的球面投影图,即训练帧;构建基于自注意力机制和交叉注意力机制的重叠度估计网络;利用得到的训练帧,训练所述重叠度估计网络;利用训练好的重叠度估计网络估计每对扫描对之间的重叠度,扫描对为两帧激光雷达的球面投影图,选取重叠度估计值高于阈值的扫描对作为回环检测候选对象,对于当前帧,选择重叠度估计值最大的候选者作为回环匹配帧,且判定该回环匹配帧位置为回环检测结果。采用本发明,能够提高回环检测或地点识别中的准确率和召回率。



1. 一种基于交叉注意力网络的回环检测方法,其特征在于,包括:

对激光雷达原始三维点云数据进行预处理,获得对应的球面投影图,即训练帧;

构建基于自注意力机制和交叉注意力机制的重叠度估计网络;

利用得到的训练帧,训练所述重叠度估计网络;其中,在训练过程中,根据重叠度估计网络估计的每对训练帧之间的重叠度,计算重叠度误差损失函数值,基于得到的重叠度误差损失函数值,训练所述重叠度估计网络;

利用训练好的重叠度估计网络估计每对扫描对之间的重叠度,扫描对为两帧激光雷达的球面投影图,选取重叠度估计值高于阈值的扫描对作为回环检测候选对象,对于当前帧,选择重叠度估计值最大的候选者作为回环匹配帧,且判定该回环匹配帧位置为回环检测结果;

其中,所述球面投影图包括:二维球面投影深度图、二维球面投影强度图以及法向量图;

所述对激光雷达原始三维点云数据进行预处理,获得对应的球面投影图包括:

通过球面投影对激光雷达原始三维点云数据进行预处理,获得二维球面投影深度图和二维球面投影强度图;

由二维球面投影深度图计算得到法向量图;

其中,所述重叠度估计网络包括:孪生网络编码器、交叉注意力网络模块和重叠度预测模块;其中,

所述孪生网络编码器包括:2个子孪生网络编码器,每个子孪生网络编码器都由10层全卷积层堆叠而成,两个子孪生网络编码器的卷积层共享权值;

所述孪生网络编码器,用于将两帧点云数据经过球面投影变换后得到的深度图、强度图和法向量图作为输入,经过10层共享权值的卷积层,获得对应的两个特征块,对输出的两个特征块分别进行向量映射,并加入位置编码,得到特征向量 f_1 和 f_2 ;其中, f_1 和 f_2 作为交叉注意力网络模块的输入;

位置编码生成公式为:

$$PE_{(pos, 2i)} = \sin(pos/10000^{2i/d})$$

$$PE_{(pos, 2i+1)} = \cos(pos/10000^{2i/d})$$

其中, pos 表示某个分块的位置, d 表示一个分块的维度大小, $2i$ 表示偶数, $2i+1$ 表示奇数, 且满足 $2i \leq d, 2i+1 \leq d, i$ 取自然数, $PE_{(pos, 2i)}$ 表示在第 pos 个分块中, 索引为 $2i$ 的向量元素所对应的位置编码值, $PE_{(pos, 2i+1)}$ 表示在第 pos 个分块中, 索引为 $2i+1$ 的向量元素所对应的位置编码值;

其中,所述交叉注意力网络模块包括:若干个交叉注意力块;

每个交叉注意力块包括:2个ECA模块和2个CFA模块;其中,ECA表示自注意力特征增强,CFA表示交叉注意力特征增强;

每个ECA模块通过多头自注意力来关注有用的上下文信息;ECA模块表示为:

$$X_{EC} = X + \text{MultiHead}(X + P_x, X + P_x, X)$$

其中, $\mathbf{X} \in \mathbb{R}^{d \times N_x}$ 是未带位置编码的输入, $\mathbb{R}^{d \times N_x}$ 表示 $d \times N_x$ 维的实数空间, N_x 表示分块的个数, d 表示一个分块的维度; $\mathbf{P}_x \in \mathbb{R}^{d \times N_x}$ 是空间位置编码; $\mathbf{X}_{EC} \in \mathbb{R}^{d \times N_x}$ 是ECA的输出;

MultiHead()表示多头自注意力

每个CFA模块同时接收两条分支上的特征向量,并通过多头交叉注意力来融合接收到的这两个特征向量,实现特征的增强;

在交叉注意力块中,分别把本分支的ECA模块输出的query和另一分支的ECA模块输出的key以及value送入本分支所在的CFA模块,CFA模块表示为:

$$\mathbf{X}_{CF} = \tilde{\mathbf{X}}_{CF} + \text{FFN}(\tilde{\mathbf{X}}_{CF})$$

$$\tilde{\mathbf{X}}_{CF} = \mathbf{X}_q + \text{MultiHead}(\mathbf{X}_q + \mathbf{P}_q, \mathbf{X}_{kv} + \mathbf{P}_{kv}, \mathbf{X}_{kv})$$

其中,query、key、value分别表示查询矩阵、键矩阵、值矩阵, $\mathbf{X}_q \in \mathbb{R}^{d \times N_q}$ 作为一个分支的输入, $\mathbf{X}_{kv} \in \mathbb{R}^{d \times N_{kv}}$ 作为另一分支的输入, $\mathbb{R}^{d \times N_q}$ 、 $\mathbb{R}^{d \times N_{kv}}$ 分别表示 $d \times N_q$ 维、 $d \times N_{kv}$ 维的实数空间, N_q 和 N_{kv} 分别表示各自分支输入特征的分块数目, $\mathbf{P}_q \in \mathbb{R}^{d \times N_q}$ 和 $\mathbf{P}_{kv} \in \mathbb{R}^{d \times N_{kv}}$ 分别是两个输入对应的位置编码, $\tilde{\mathbf{X}}_{CF}$ 表示通过多头注意力层的输出, $\mathbf{X}_{CF}$ 是CFA模块的最终输出;FFN()为全连接的前馈网络,由两个线性变换组成:

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2$$

其中,x表示FFN()的输入,max()表示求其中最大元素, W_1 和 W_2 表示权重矩阵, b_1 和 b_2 表示偏置向量,同一下标代表相同层;

其中,所述重叠度预测模块,用于估计扫描对之间的重叠度;

所述重叠度预测模块包括:CFA模块和带有dropout层的MLP层组成;其中,MLP表示多层感知器;

其中,重叠度误差损失函数表示为:

$$L(I_1, I_2, Y_0) = \text{sigmoid}(s(|\hat{Y}_0 - Y_0| + a) - b)$$

其中, $L(I_1, I_2, Y_0)$ 表示重叠度误差损失函数值, Y_0 表示重叠度的真值, $\hat{Y}_0$ 表示重叠度的估计值, I_1 和 I_2 表示输入的一对训练帧,sigmoid()为激活函数,a和b都表示偏移量,s表示尺度因子;

其中,对于一对激光雷达点云数据 $\mathcal{P}_1$ 、 $\mathcal{P}_2$,分别进行球面投影处理,得到 $\mathcal{V}_1$ 、 $\mathcal{V}_2$;并将第一帧 $\mathcal{P}_1$ 的所有点重投影到第二帧 $\mathcal{V}_2$ 的坐标系上,形成重投影映射 $\mathcal{V}'_1$,计算 $\mathcal{V}'_1$ 和 $\mathcal{V}_2$ 中各对应像素的绝对差值 $\|\mathcal{V}'_1(u, v) - \mathcal{V}_2(u, v)\|$,根据得到的绝对差值

$\|\mathcal{V}'_1(u, v) - \mathcal{V}_2(u, v)\|$,计算得到 $\mathcal{P}_1$ 、 $\mathcal{P}_2$ 两帧之间的重叠度:

$$O_{1,2} = \frac{\sum_{(u,v)} \mathbb{I}\{\|\mathcal{V}'_1(u, v) - \mathcal{V}_2(u, v)\| \leq \epsilon\}}{\min(\text{valid}(\mathcal{V}'_1), \text{valid}(\mathcal{V}_2))}$$

其中,(u,v)表示球面投影后的二维图像坐标, $O_{1,2}$ 表示 $\mathcal{P}_1$ 、 $\mathcal{P}_2$ 两帧之间的重叠度的真值; $\in$ 表示阈值;若 $\|\mathcal{V}'_1(u, v) - \mathcal{V}_2(u, v)\|$ 满足小于等于阈值 $\in$ 的条件,则 $\mathbb{I}\{\|\mathcal{V}'_1(u, v) - \mathcal{V}_2(u, v)\| \leq \epsilon\}$ 取值1,反之为0;valid()代表有效的像素个数。

2.根据权利要求1所述的基于交叉注意力网络的回环检测方法,其特征在于,所述对于当前帧,选择重叠度估计值最大的候选者作为回环匹配帧,且判定该回环匹配帧位置为回

环检测结果包括：

对于当前帧，搜寻在当前帧前N帧之前的历史帧，并且选择重叠度估计值最大的候选者作为回环匹配帧，且判定该回环匹配帧位置为回环检测结果。

一种基于交叉注意力网络的回环检测方法

技术领域

[0001] 本发明涉及计算机视觉技术领域,特别是指一种基于交叉注意力网络的回环检测方法。

背景技术

[0002] 同时定位与建图 (SLAM) 是无人驾驶或机器人自主导航中的一个重要组成部分。经典的SLAM框架通常由五部分组成,分别是传感器信息读取,前端里程计,回环检测,后端优化以及建图。其中,回环检测是判断机器人或无人车是否到达过先前的位置,用于更新当前帧与历史帧之间的位姿约束关系,以减小累积误差,从而得到信息一致的地图。基于激光雷达和基于视觉的回环检测方法较为常见。但同一地点通常会因光照、天气或视角的变化而变化,这在基于视觉的回环检测中是很难解决的,导致回环检测准确率低。

发明内容

[0003] 本发明实施例提供了一种基于交叉注意力网络的回环检测方法,能够提高回环检测或地点识别中的准确率和召回率。

[0004] 本发明实施例提供的基于交叉注意力网络的回环检测方法,包括:

[0005] 对激光雷达原始三维点云数据进行预处理,获得对应的球面投影图,即训练帧;

[0006] 构建基于自注意力机制和交叉注意力机制的重叠度估计网络;

[0007] 利用得到的训练帧,训练所述重叠度估计网络;其中,在训练过程中,根据重叠度估计网络估计的每对训练帧之间的重叠度,计算重叠度误差损失函数值,基于得到的重叠度误差损失函数值,训练所述重叠度估计网络;

[0008] 利用训练好的重叠度估计网络估计每对扫描对之间的重叠度,扫描对为两帧激光雷达的球面投影图,选取重叠度估计值高于阈值的扫描对作为回环检测候选对象,对于当前帧,选择重叠度估计值最大的候选者作为回环匹配帧,且判定该回环匹配帧位置为回环检测结果。

[0009] 进一步地,所述球面投影图包括:二维球面投影深度图、二维球面投影强度图以及法向量图;

[0010] 所述对激光雷达原始三维点云数据进行预处理,获得对应的球面投影图包括:

[0011] 通过球面投影对激光雷达原始三维点云数据进行预处理,获得二维球面投影深度图和二维球面投影强度图;

[0012] 由二维球面投影深度图计算得到法向量图。

[0013] 进一步地,所述重叠度估计网络包括:孪生网络编码器、交叉注意力网络模块和重叠度预测模块;其中,

[0014] 所述孪生网络编码器包括:2个子孪生网络编码器,每个子孪生网络编码器都由10层全卷积层堆叠而成,两个子孪生网络编码器的卷积层共享权值;

[0015] 所述孪生网络编码器,用于将两帧点云数据经过球面投影变换后得到的深度图、

强度图和法向量图作为输入,经过10层共享权值的卷积层,获得对应的两个特征块,对输出的两个特征块分别进行向量映射,并加入位置编码,得到特征向量 f_1 和 f_2 ;其中, f_1 和 f_2 作为交叉注意力网络模块的输入;

[0016] 位置编码生成公式为:

$$[0017] \quad PE_{(pos, 2i)} = \sin(pos/10000^{2i/d})$$

$$[0018] \quad PE_{(pos, 2i+1)} = \cos(pos/10000^{2i/d})$$

[0019] 其中, pos 表示某个分块的位置, d 表示一个分块的维度大小, $2i$ 表示偶数, $2i+1$ 表示奇数, 且满足 $2i \leq d, 2i+1 \leq d, i$ 取自然数, $PE_{(pos, 2i)}$ 表示在第 pos 个分块中, 索引为 $2i$ 的向量元素所对应的位置编码值, $PE_{(pos, 2i+1)}$ 表示在第 pos 个分块中, 索引为 $2i+1$ 的向量元素所对应的位置编码值。

[0020] 进一步地, 所述交叉注意力网络模块包括: 若干个交叉注意力块;

[0021] 每个交叉注意力块包括: 2个ECA模块和2个CFA模块; 其中, ECA表示自注意力特征增强, CFA表示交叉注意力特征增强;

[0022] 每个ECA模块通过多头自注意力来关注有用的上下文信息; ECA模块表示为:

$$[0023] \quad X_{EC} = X + \text{MultiHead}(X + P_x, X + P_x, X)$$

[0024] 其中, $\mathbf{X} \in \mathbb{R}^{d \times N_x}$ 是未带位置编码的输入, $\mathbb{R}^{d \times N_x}$ 表示 $d \times N_x$ 维的实数空间, N_x 表示分块的个数, d 表示一个分块的维度; $\mathbf{P}_x \in \mathbb{R}^{d \times N_x}$ 是空间位置编码; $\mathbf{X}_{EC} \in \mathbb{R}^{d \times N_x}$ 是ECA的输出; $\text{MultiHead}()$ 表示多头自注意力;

[0025] 每个CFA模块同时接收两条分支上的特征向量, 并通过多头交叉注意力来融合接收到的这两个特征向量, 实现特征的增强;

[0026] 在交叉注意力块中, 分别把本分支的ECA模块输出的query和另一分支的ECA模块输出的key以及value送入本分支所在的CFA模块, CFA模块表示为:

$$[0027] \quad \mathbf{X}_{CF} = \tilde{\mathbf{X}}_{CF} + \text{FFN}(\tilde{\mathbf{X}}_{CF})$$

$$[0028] \quad \tilde{\mathbf{X}}_{CF} = \mathbf{X}_q + \text{MultiHead}(\mathbf{X}_q + \mathbf{P}_q, \mathbf{X}_{kv} + \mathbf{P}_{kv}, \mathbf{X}_{kv})$$

[0029] 其中, query、key、value分别表示查询矩阵、键矩阵、值矩阵, $\mathbf{X}_q \in \mathbb{R}^{d \times N_q}$ 作为一个分支的输入, $\mathbf{X}_{kv} \in \mathbb{R}^{d \times N_{kv}}$ 作为另一分支的输入, $\mathbb{R}^{d \times N_q}$ 、 $\mathbb{R}^{d \times N_{kv}}$ 分别表示 $d \times N_q$ 维、 $d \times N_{kv}$ 维的实数空间, N_q 和 N_{kv} 分别表示各自分支输入特征的分块数目, $\mathbf{P}_q \in \mathbb{R}^{d \times N_q}$ 和 $\mathbf{P}_{kv} \in \mathbb{R}^{d \times N_{kv}}$ 分别是两个输入对应的位置编码, $\tilde{\mathbf{X}}_{CF}$ 表示通过多头注意力层的输出, $\mathbf{X}_{CF}$ 是CFA模块的最终输出; $\text{FFN}()$ 为全连接的前馈网络, 由两个线性变换组成:

$$[0030] \quad \text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2$$

[0031] 其中, x 表示 $\text{FFN}()$ 的输入, $\max()$ 表示求其中最大元素, W_1 和 W_2 表示权重矩阵, b_1 和 b_2 表示偏置向量, 同一下标代表相同层。

[0032] 进一步地, 所述重叠度预测模块, 用于估计扫描对之间的重叠度;

[0033] 所述重叠度预测模块包括: CFA模块和带有dropout层的MLP层组成; 其中, MLP表示多层感知器。

[0034] 进一步地,重叠度误差损失函数表示为:

$$[0035] \quad L(I_1, I_2, Y_0) = \text{sigmoid}(s(|\hat{Y}_0 - Y_0| + a) - b)$$

[0036] 其中, $L(I_1, I_2, Y_0)$ 表示重叠度误差损失函数值, Y_0 表示重叠度的真值, $\hat{Y}_0$ 表示重叠度的估计值, I_1 和 I_2 表示输入的一对训练帧, $\text{sigmoid}()$ 为激活函数, a 和 b 都表示偏移量, s 表示尺度因子。

[0037] 进一步地,对于一对激光雷达点云数据 $\mathcal{P}_1$ 、 $\mathcal{P}_2$,分别进行球面投影处理,得到 $\mathcal{V}_1$ 、 $\mathcal{V}_2$;并将第一帧 $\mathcal{P}_1$ 的所有点重投影到第二帧 $\mathcal{V}_2$ 的坐标系上,形成重投影映射 $\mathcal{V}_1'$,计算 $\mathcal{V}_1'$ 和 $\mathcal{V}_2$ 中各对应像素的绝对差值 $\|\mathcal{V}_1'(u, v) - \mathcal{V}_2(u, v)\|$,根据得到的绝对差值 $\|\mathcal{V}_1'(u, v) - \mathcal{V}_2(u, v)\|$,计算得到 $\mathcal{P}_1$ 、 $\mathcal{P}_2$ 两帧之间的重叠度:

$$[0038] \quad O_{1,2} = \frac{\sum_{(u,v)} \mathbb{I}\{\|\mathcal{V}_1'(u, v) - \mathcal{V}_2(u, v)\| \leq \epsilon\}}{\min(\text{valid}(\mathcal{V}_1'), \text{valid}(\mathcal{V}_2))}$$

[0039] 其中, (u, v) 表示球面投影后的二维图像坐标, $O_{1,2}$ 表示 $\mathcal{P}_1$ 、 $\mathcal{P}_2$ 两帧之间的重叠度的真值; ϵ 表示阈值;若 $\|\mathcal{V}_1'(u, v) - \mathcal{V}_2(u, v)\|$ 满足小于等于阈值 ϵ 的条件,则 $\mathbb{I}\{\|\mathcal{V}_1'(u, v) - \mathcal{V}_2(u, v)\| \leq \epsilon\}$ 取值1,反之为0; $\text{valid}()$ 代表有效的像素个数。

[0040] 进一步地,所述对于当前帧,选择重叠度估计值最大的候选者作为回环匹配帧,且判定该回环匹配帧位置为回环检测结果包括:

[0041] 对于当前帧,搜寻在当前帧前N帧之前的历史帧,并且选择重叠度估计值最大的候选者作为回环匹配帧,且判定该回环匹配帧位置为回环检测结果。

[0042] 本发明实施例提供的技术方案带来的有益效果至少包括:

[0043] 本发明实施例中,对激光雷达原始三维点云数据进行预处理,获得对应的球面投影图,即训练帧;构建基于自注意力机制和交叉注意力机制的重叠度估计网络;利用得到的训练帧,训练所述重叠度估计网络;其中,在训练过程中,根据重叠度估计网络估计的每对训练帧之间的重叠度,计算重叠度误差损失函数值,基于得到的重叠度误差损失函数值,训练所述重叠度估计网络;利用训练好的重叠度估计网络估计每对扫描对之间的重叠度,扫描对为两帧激光雷达的球面投影图,选取重叠度估计值高于阈值的扫描对作为回环检测候选对象,对于当前帧,选择重叠度估计值最大的候选者作为回环匹配帧,且判定该回环匹配帧位置为回环检测结果,这样,能够提高回环检测或地点识别中的准确率和召回率。

附图说明

[0044] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例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。

[0045] 图1为本发明实施例提供的基于交叉注意力网络的回环检测方法的流程示意图;

[0046] 图2为本发明实施例提供的原始三维点云转化为球面投影图的原理示意图;

[0047] 图3为本发明实施例提供的重叠度估计网络的结构示意图；

[0048] 图4(a)为本发明实施例提供的ECA模块的结构示意图；

[0049] 图4(b)为本发明实施例提供的CFA模块的结构示意图；

[0050] 图5(a)为本发明实施例提供的在KITTI序列00中重叠度绝对误差分布图图；

[0051] 图5(b)为本发明实施例提供的在KITTI序列00中重叠度准确率-召回率曲线的结构示意图；

[0052] 图6为本发明实施例提供的不同方法在KITTI序列00中的候选者数量-召回率曲线示意图。

具体实施方式

[0053] 为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明实施方式作进一步地详细描述。

[0054] 如图1所示，本发明实施例提供了一种基于交叉注意力网络的回环检测方法，包括：

[0055] S101，对激光雷达原始三维点云数据进行预处理，获得对应的球面投影图，即训练帧；其中，所述球面投影图包括：二维球面投影深度图、二维球面投影强度图以及法向量图。

[0056] 本实施例中，所述对激光雷达原始三维点云数据进行预处理，获得对应的球面投影图包括：

[0057] A1，通过球面投影对激光雷达原始三维点云数据进行预处理，获得二维球面投影深度图和二维球面投影强度图；

[0058] 如图2所示，球面投影关系可以表示为 $\mathcal{V}: \mathbb{R}^3 \mapsto \mathbb{R}^2$ ，其中， $\mathbb{R}^3$ 表示三维实数空间， $\mathbb{R}^2$ 表示二维实数空间。投影平面上的每一个点的像素值都代表了激光雷达传感器与对应的空间点之间的距离，即深度信息。空间上被扫描到的三维点 $p_i = (x, y, z)$ 都能通过以下公式转换到图像坐标 (u, v) ：

$$[0059] \quad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2} [1 - \arctan(y, x) \pi^{-1}] w \\ [1 - (\arcsin(zr^{-1}) + f_{up}) f^{-1}] h \end{pmatrix}$$

[0060] 其中， $r = \|p\|_2$ 为三维点的深度， $f = f_{up} + f_{down}$ 表示激光雷达的垂直视场角， f_{up} 表示上垂直视场角， f_{down} 表示下垂直视场角。而 w, h 分别是投影图的宽度和高度。

[0061] A2，由二维球面投影深度图计算得到法向量图。

[0062] S102，构建基于自注意力机制和交叉注意力机制的重叠度估计网络 (CrossT-Net)；

[0063] 如图3所示，构建的重叠度估计网络是一种端到端的神经网络，所述重叠度估计网络包括：孪生网络编码器、交叉注意力 (Cross Transformer) 网络模块和重叠度预测模块。

[0064] 本实施例中，孪生神经网络编码器由共享权值的全卷积层构成，具体的：所述孪生神经网络编码器包括：2个子孪生网络编码器，每个子孪生网络编码器都由10层全卷积层堆叠而成，两个子孪生网络编码器的卷积层 (CNN) 共享权值，如图3所示；子孪生网络编码器的详细结构如表1所示：

[0065] 表1孪生网络编码器结构

	Operator	Stride	Filters	Size	Output Shape
[0066] 孪生网络编码器	Conv2D	(2,2)	8	(3,9)	31×446×8
	Conv2D	(2,1)	16	(3,9)	15×438×16
	Conv2D	(2,1)	32	(3,9)	7×430×32
	Conv2D	(2,1)	64	(3,9)	3×422×64
	Conv2D	(1,1)	64	(2,9)	2×414×64
	Conv2D	(1,1)	128	(2,9)	1×406×128
	Conv2D	(1,1)	128	(1,9)	1×398×128
	Conv2D	(1,1)	128	(1,9)	1×390×128
	Conv2D	(1,1)	128	(1,3)	1×388×128
	Conv2D	(1,1)	128	(1,3)	1×386×128

[0067] 本实施例中,所述孪生网络编码器,用于将两帧点云数据经过球面投影变换后得到的深度图(占用1个通道)、强度图(占用1个通道)和法向量图(占用3个通道)作为输入,其输入尺寸为 $5 \times 64 \times 900$,对应着总通道数、高度、宽度,两帧的输入分别是input1和input2,经过10层共享权值的卷积层,获得对应的两个特征块,特征块的尺寸皆为 $1 \times 386 \times 128$,对输出的两个特征块分别进行向量映射(embedding),并加入位置编码,得到两个分支的386个128维的特征向量,即386个分块(patch)。

[0068] 本实施例中,用正弦函数来生成空间位置编码,位置编码按如下公式生成:

$$[0069] \quad PE_{(pos, 2i)} = \sin(pos/10000^{2i/d})$$

$$[0070] \quad PE_{(pos, 2i+1)} = \cos(pos/10000^{2i/d})$$

[0071] 其中,pos表示某个分块的位置,d表示一个分块的维度大小($d=128$),i取自然数($i=0,1,2,3,\dots$), $2i$ 表示偶数, $2i+1$ 表示奇数,且满足 $2i \leq d, 2i+1 \leq d$, $PE_{(pos, 2i)}$ 表示在第pos个分块中,索引为 $2i$ 的向量元素所对应的位置编码值, $PE_{(pos, 2i+1)}$ 表示在第pos个分块中,索引为 $2i+1$ 的向量元素所对应的位置编码值。

[0072] 本实施例中,如图3所示,所述交叉注意力网络模块包括:若干个交叉注意力块;每个交叉注意力块包括:2个自注意力特征增强(ego-context augment, ECA)模块和2个交叉注意力特征增强(cross-feature augment, CFA)模块。交叉注意力网络模块,用于实现自注意力和交叉注意力特征增强,从而自适应地关注回环检测相关的关键信息。

[0073] 本实施例中,每个ECA模块通过多头自注意力来关注有用的上下文信息;将两条分支上 $d=128$ 的特征向量 f_1 和 f_2 作为2个ECA模块的输入,先通过一个ECA模块,如图4(a)所示, ECA模块计算可总结为下述公式:

$$[0074] \quad X_{EC} = X + \text{MultiHead}(X + P_x, X + P_x, X)$$

[0075] 其中, $\mathbf{X} \in \mathbb{R}^{d \times N_x}$ 是未带位置编码的输入, $\mathbb{R}^{d \times N_x}$ 表示 $d \times N_x$ 维的实数空间,,d表示一个分块的维度, N_x 表示分块的个数,设置 $N_x = 386$; $\mathbf{P}_x \in \mathbb{R}^{d \times N_x}$ 是空间位置编码;

$\mathbf{X}_{EC} \in \mathbb{R}^{d \times N_x}$ 是ECA的输出;MultiHead()表示多头自注意力;

[0076] 多头自注意力的计算公式定义如下:

$$[0077] \quad \text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\mathbf{H}_1, \dots, \mathbf{H}_{n_h}) \mathbf{W}^O$$

$$[0078] \quad \mathbf{H}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V)$$

$$[0079] \quad \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V}$$

[0080] 其中, $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别是查询(query), 键(key), 值(value) 矩阵, 本发明结构所设置注意力头个数为 $n_h=8$, 其中, 为了防止 $\mathbf{Q}\mathbf{K}^T$ 的内积过大, 因此除以 d_k 的平方根, 默认值 $d_k=d/n_h=16$, 而 $\mathbf{W}_i^Q \in \mathbb{R}^{d_m \times d_k}$, $\mathbf{W}_i^K \in \mathbb{R}^{d_m \times d_k}$, $\mathbf{W}_i^V \in \mathbb{R}^{d_m \times d_v}$, $\mathbf{W}^O \in \mathbb{R}^{n_h d_v \times d_m}$ 皆为参数矩阵, $\mathbb{R}^{n_h d_v \times d_m}$ 表示 $n_h d_v \times d_m$ 维的实数空间, d_k 代表 $\mathbf{Q}, \mathbf{K}$ 矩阵的列数, d_v 代表 $\mathbf{V}$ 矩阵的列数, d_m 代表多头注意力最终输出的矩阵列数, 其中, $d_m=d=128$, $d_k=d_v=16$ 。另外, $\text{Concat}()$ 表示通道维度上的拼接, $\mathbf{H}_i$ 表示第 i 个注意力头矩阵, $\text{Attention}()$ 表示注意力函数, $\text{softmax}()$ 为激活函数, 并进行归一化。

[0081] 每个CFA模块同时接收两条分支上的特征向量, 并通过多头交叉注意力来融合接收到的这两个特征向量, 实现特征的增强;

[0082] 在交叉注意力块中, 分别把本分支的ECA模块输出的查询矩阵query和另一分支的ECA模块输出的键矩阵key以及值矩阵value送入本分支所在的CFA模块, 输出依旧是两分支的386个128维特征向量, 如图4(b)所示, CFA模块计算可总结为下述公式:

$$[0083] \quad \mathbf{X}_{CF} = \tilde{\mathbf{X}}_{CF} + \text{FFN}(\tilde{\mathbf{X}}_{CF})$$

$$[0084] \quad \tilde{\mathbf{X}}_{CF} = \mathbf{X}_q + \text{MultiHead}(\mathbf{X}_q + \mathbf{P}_q, \mathbf{X}_{kv} + \mathbf{P}_{kv}, \mathbf{X}_{kv})$$

[0085] 其中, $\mathbf{X}_q \in \mathbb{R}^{d \times N_q}$ 作为一个分支的输入, $\mathbf{X}_{kv} \in \mathbb{R}^{d \times N_{kv}}$ 作为另一分支的输入, $\mathbf{P}_q \in \mathbb{R}^{d \times N_q}$ 和 $\mathbf{P}_{kv} \in \mathbb{R}^{d \times N_{kv}}$ 分别是两个输入对应的位置编码, N_q 和 N_{kv} 分别代表各自分支输入特征的分块(patch)数目, 其中 $N_q=N_{kv}=N_x=386$, $\tilde{\mathbf{X}}_{CF}$ 表示通过多头注意力层的输出, $\mathbf{X}_{CF}$ 是CFA模块的最终输出; $\text{FFN}()$ 用于增强模型的拟合能力, $\text{FFN}()$ 为一个全连接的前馈网络, 由两个线性变换组成:

$$[0086] \quad \text{FFN}(x) = \max(0, x\mathbf{W}_1 + b_1) \mathbf{W}_2 + b_2$$

[0087] 其中, x 表示 $\text{FFN}()$ 的输入, $\max()$ 表示求其中最大元素, $\mathbf{W}_1$ 和 $\mathbf{W}_2$ 表示权重矩阵, b_1 和 b_2 表示偏置向量, 同一下标代表相同层。

[0088] 本实施例中, 将上述的2个ECA模块和2个CFA模块作为一个交叉注意力块, 将交叉注意力块重复 N 次 (本实施例中, $N=2$), 最后输出为2个 386×128 的向量。

[0089] 本实施例中, 所述重叠度预测模块为解码器, 用于估计扫描对之间的重叠度(Overlap), 即输出overlap的估计值; 所述重叠度预测模块包括: CFA模块和带有dropout层的多层感知器(MLP)层组成。

[0090] 本实施例中, 交叉注意力网络模块输出的两个特征向量作为重叠度预测模块的CFA的输入, 其key和value来自同一个特征向量, 而query来自另一个特征向量。在CFA之后, 这些特征向量被融合成一个新的特征向量, 并把得到的新的特征向量展平为一维, 送入带有dropout层的多层感知器得到单输出, 通过softmax激活函数将估计的重叠度值范围控制

在0到1之间。

[0091] 本实施例中,用全卷积层构成的孪生网络编码器结合基于自注意力与交叉注意力的特征增强模块(即:交叉注意力网络模块),来实现SLAM回环检测中两帧激光雷达点云数据的重叠度估计。

[0092] S103,利用得到的训练帧,训练所述重叠度估计网络;其中,在训练过程中,根据重叠度估计网络估计的每对训练帧之间的重叠度,计算重叠度误差损失函数值,基于得到的重叠度误差损失函数值,训练所述重叠度估计网络;

[0093] 本实施例中,在训练过程中,根据重叠度估计网络输出的每对训练帧之间的重叠度估计值,计算重叠度误差损失函数值;基于得到重叠度误差损失函数值,采取端到端的训练方式并通过反向传播来训练整个重叠度估计网络。

[0094] 本实施例中,重叠度误差损失函数表示为:

$$[0095] \quad L(I_1, I_2, Y_0) = \text{sigmoid}(s(|\hat{Y}_0 - Y_0| + a) - b)$$

[0096] 其中, $L(I_1, I_2, Y_0)$ 表示重叠度误差损失函数值, Y_0 表示重叠度的真值, $\hat{Y}_0$ 表示重叠度的估计值, I_1 和 I_2 表示输入的一对训练帧,而 $\text{sigmoid}(x) = (1 + \exp(-x))^{-1}$ 为激活函数,使得输出值归一化, a 和 b 都表示偏移量, s 表示尺度因子。

[0097] 本实施例中,选择为 $a=0.3, b=13, s=24$ 。

[0098] 本实施例中,所述的重叠度的真值计算方法为:

[0099] 对于一对激光雷达点云数据 $\mathcal{P}_1$ 、 $\mathcal{P}_2$,分别进行球面投影处理,得到 $\mathcal{V}_1$ 、 $\mathcal{V}_2$;并将第一帧 $\mathcal{P}_1$ 的所有点重投影到第二帧 $\mathcal{V}_2$ 的坐标系上,形成重投影映射 $\mathcal{V}_1'$,计算 $\mathcal{V}_1'$ 和 $\mathcal{V}_2$ 中各对应像素(深度值)的绝对差值 $\|\mathcal{V}_1'(u, v) - \mathcal{V}_2(u, v)\|$,其中, (u, v) 表示球面投影后的二维图像坐标,根据得到的绝对差值 $\|\mathcal{V}_1'(u, v) - \mathcal{V}_2(u, v)\|$,计算得到 $\mathcal{P}_1$ 、 $\mathcal{P}_2$ 两帧之间的重叠度:

$$[0100] \quad O_{1, 2} = \frac{\sum_{(u,v)} \mathbb{I}\{\|\mathcal{V}_1'(u, v) - \mathcal{V}_2(u, v)\| \leq \epsilon\}}{\min(\text{valid}(\mathcal{V}_1'), \text{valid}(\mathcal{V}_2))}$$

[0101] 其中, $O_{1, 2}$ 表示 $\mathcal{P}_1$ 、 $\mathcal{P}_2$ 两帧之间的重叠度的真值; ϵ 表示阈值;若 $\|\mathcal{V}_1'(u, v) - \mathcal{V}_2(u, v)\|$ 满足小于等于阈值 ϵ 的条件,则 $\mathbb{I}\{\|\mathcal{V}_1'(u, v) - \mathcal{V}_2(u, v)\| \leq \epsilon\}$ 取值1,反之为0; $\text{valid}()$ 代表有效的像素个数。

[0102] S104,利用训练好的重叠度估计网络估计每对扫描对之间的重叠度,扫描对为两帧激光雷达的球面投影图,选取重叠度估计值高于阈值的扫描对作为回环检测候选对象,对于当前帧,选择重叠度估计值最大的候选者作为回环匹配帧,且判定该回环匹配帧位置为回环检测结果。

[0103] 本实施例中,对于当前帧,搜寻在当前帧前N帧之前的历史帧,并且选择重叠度估计值最大的候选者作为回环匹配帧,且判定该回环匹配帧位置为回环检测结果;若搜寻到无候选者,则判定并非回环。

[0104] 本实施例中,为了验证本发明实施例提供的回环检测方法的有效性,使用KITTI里

程计数据集以及Ford Campus数据集评估测试其性能：

[0105] 本发明实施例提供的回环检测方法(简称:本发明方法)的评估结果与表2中的其他方法进行比较。表明本发明方法在KITTI数据集序列00的F1最大分数(F1 max)和AUC值(准确率-召回率曲线下的面积,Area Under Curve)都优于其他方法,并且本发明方法top-1召回率(Recall@1)也优于大多数其他方法。其中,若加入先验位姿的条件(在实验中设置搜索半径为50m),其表现能有进一步提升。可以注意到,本发明方法的AUC值甚至比其他方法有很大的领先优势。如图5(a)所示,本发明方法在KITTI测试数据集上的重叠度误差分布,88.4%的重叠度估计值误差范围在0.05以内。如图5(b)所示,本发明方法的准确率-召回率曲线与目前最先进的方法OverlapTransformer相比具有不小优势。其他方法更喜欢单独提取一个分支上的特征,相比之下,本发明方法在数据集上表现特别好的原因是本发明方法在早期通过交叉注意加强了扫描对之间的联系,从而更容易专注于关键区域。本发明方法还在未经训练的Ford Campus数据集上测试了本发明的重叠度估计网络。表3显示了本发明方法与其他方法在Ford Campus上的比较,本发明的重叠度估计网络在三个评价指标中均领先。由于本发明没有在Ford Campus数据集上进行训练,因此CrossT-Net在Ford Campus中的表现比在KITTI差是可以接受的。

[0106] 表2KITTI数据集中本发明方法与其他方法对比

[0107]	KITTI	Dataset	Approach	AUC	F1 max	Recall @1
			Histogram	0.826	0.825	0.738
			Scan Context	0.836	0.835	0.820
			LiDAR Iris	0.843	0.848	0.835
			PointNetVLAD	0.856	0.846	0.776
			OverlapNet	0.867	0.865	0.816
			OverlapTransformer	0.907	0.877	0.906
			Ours	0.971	0.915	0.899
			Ours (with search radius)	0.975	0.938	0.903

[0108] 表3Ford Campus数据集中本发明方法与其他方法对比

[0109]	Ford Campus	Dataset	Approach	AUC	F1 max	Recall @1
			Histogram	0.841	0.800	0.812
			Scan Context	0.903	0.842	0.878
			LiDAR Iris	0.907	0.842	0.849
			PointNetVLAD	0.872	0.830	0.862
			OverlapNet	0.854	0.843	0.857
			OverlapTransformer	0.923	0.856	0.914
			Ours	0.934	0.871	0.921

[0110] 如图6所示,本实施例中还验证了KITTI数据集上不同个数的候选者对召回率的影响曲线。从实验结果可以发现,此方法在候选者数量小的条件下具有一定的召回率优势,随着候选者数量增多,各类方法所获得的召回率表现相差不大。

[0111] 为了验证本实施例所述的方法各部分的意义,本实施例中还进行了消融实验。

[0112] 本实施例进行消融实验来探索不同输入和交叉注意力块的数量对结果的影响。不同输入的消融实验结果如表4所示。这项实验证实,如果CrossT-Net只去除法向量图,AUC值和F1分数最大值的性能不会下降太多,但只输入深度图,表现将会下降明显。

[0113] 表4不同输入对结果影响的消融实验

	Dataset	Variant	AUC	F1 max
[0114]	KITTI	Depth	0.915	0.841
		Depth, Intensity	0.965	0.895
		Depth, Intensity, Normal	0.971	0.915

[0115] 本发明还使用不同数量的交叉注意力块来训练CrossT-Net。测试结果如表5所示，当N为0时，使用多次感知机层 (MLP) 代替注意力模块。可以从表5中验证，N取2时，这是最佳选择。同时也表明更多的交叉注意力块并不总是产生更好的结果。

[0116] 表5不同数量的交叉注意力块对结果影响的消融实验

	Number of the blocks	AUC	F1 max
[0117]	$N = 0$ (MLP only)	0.650	0.627
	$N = 1$	0.920	0.882
	$N = 2$	0.971	0.915
	$N = 3$	0.932	0.849
	$N = 4$	0.934	0.856

[0118] 本发明实施例所述的回环检测方法，至少具有以下优点：

[0119] 1) 本发明是一种基于交叉注意力网络模块的回环检测方法，通过深度神经网络仅使用来自激光雷达的球面投影图，不使用任何其他信息来估计两帧间的重叠度，并利用所估计的重叠度 (Overlap) 来预测其是否为回环。

[0120] 2) 增加了结合自注意力机制和交叉注意力机制的交叉注意力网络模块，能够自适应地关注回环检测相关的关键信息，从而提高回环检测或地点识别中的准确率和召回率。

[0121] 3) 本发明可以在没有任何先验姿态的情况下检测环路，并且可以在没有微调的情况下应用于不同的道路环境。

[0122] 4) 本发明在不同的数据集上验证了此方法，并且可以获得比最先进的方法更好的结果。

[0123] 以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

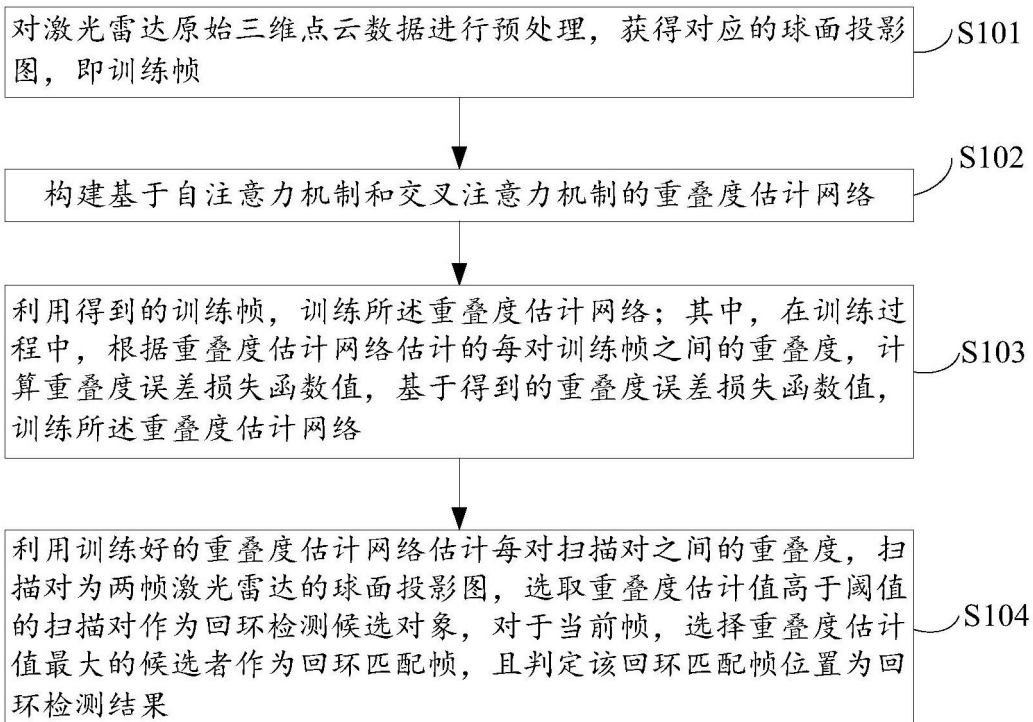


图1

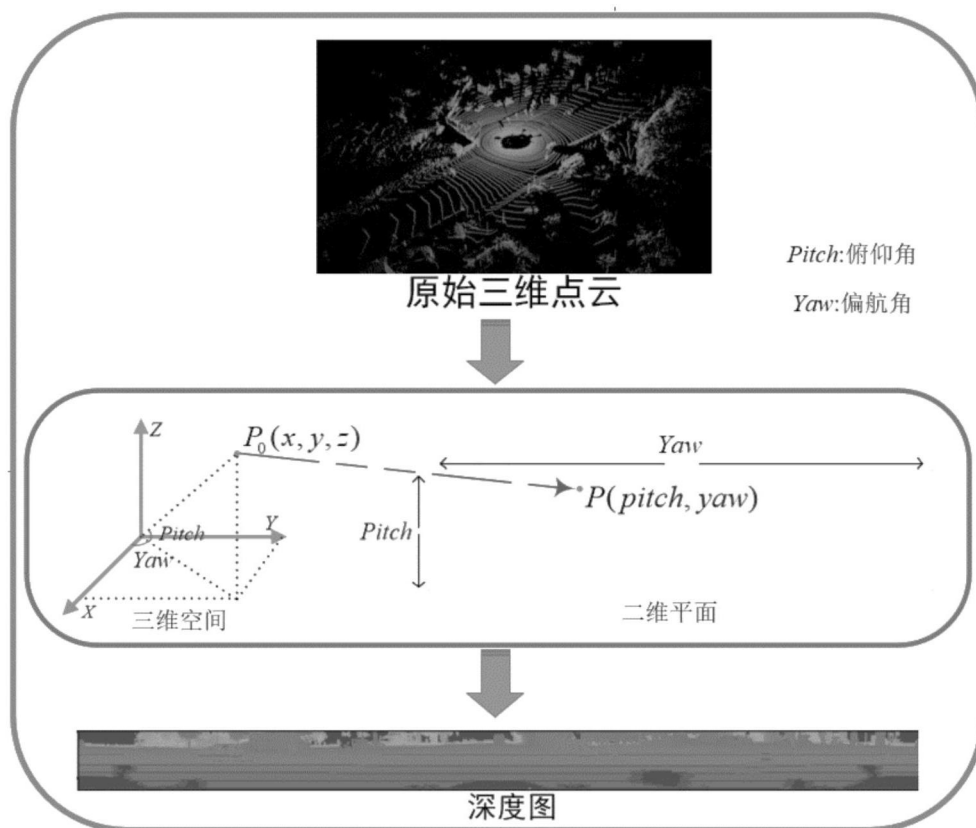


图2

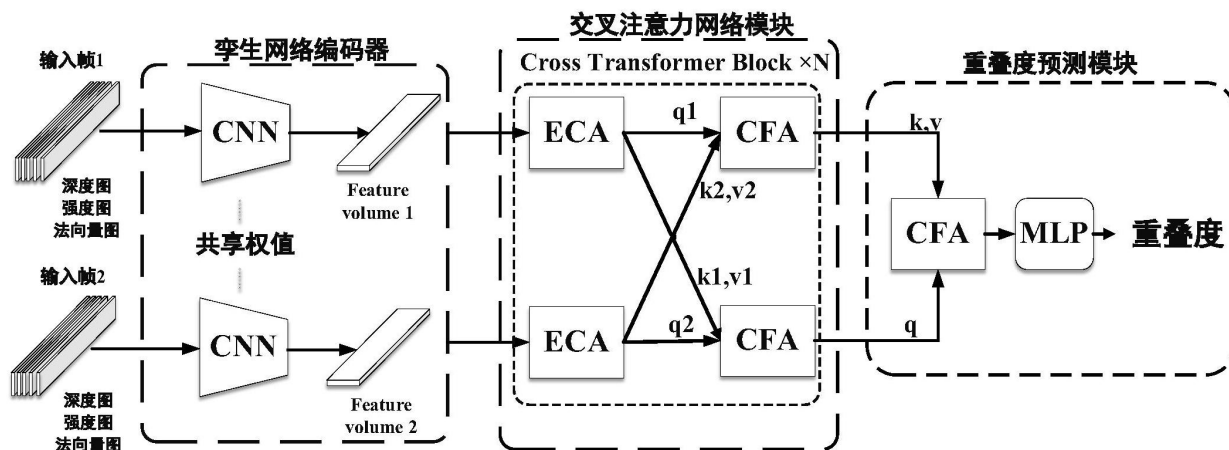


图3

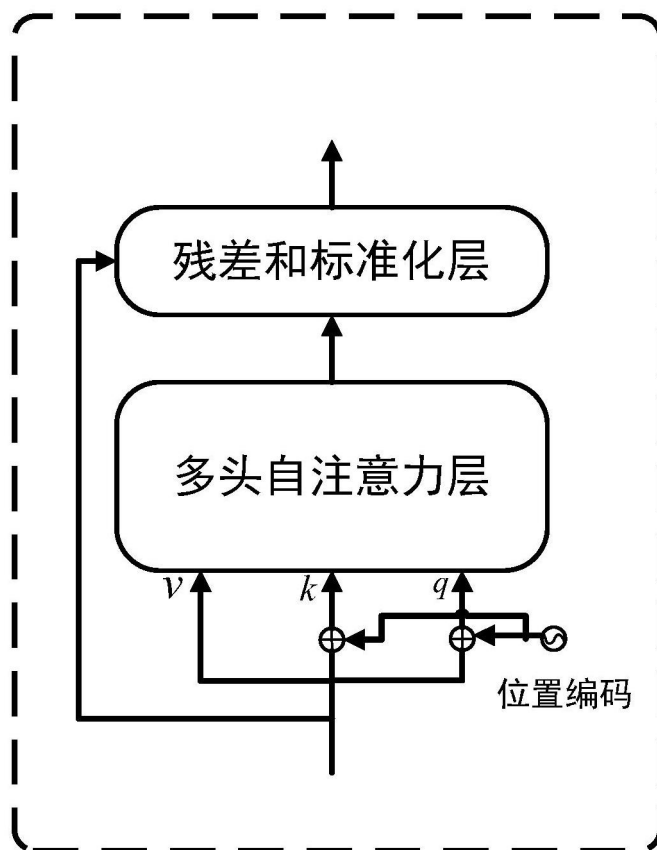


图4(a)

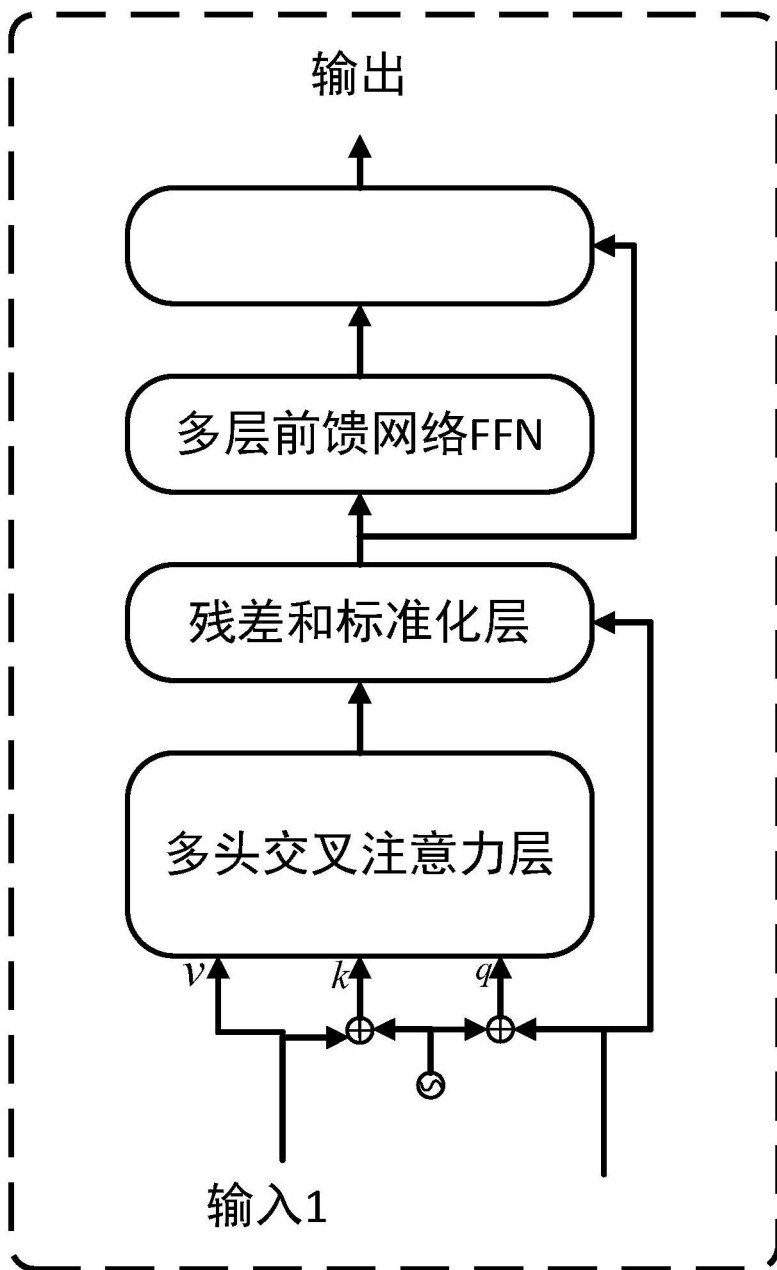


图4(b)

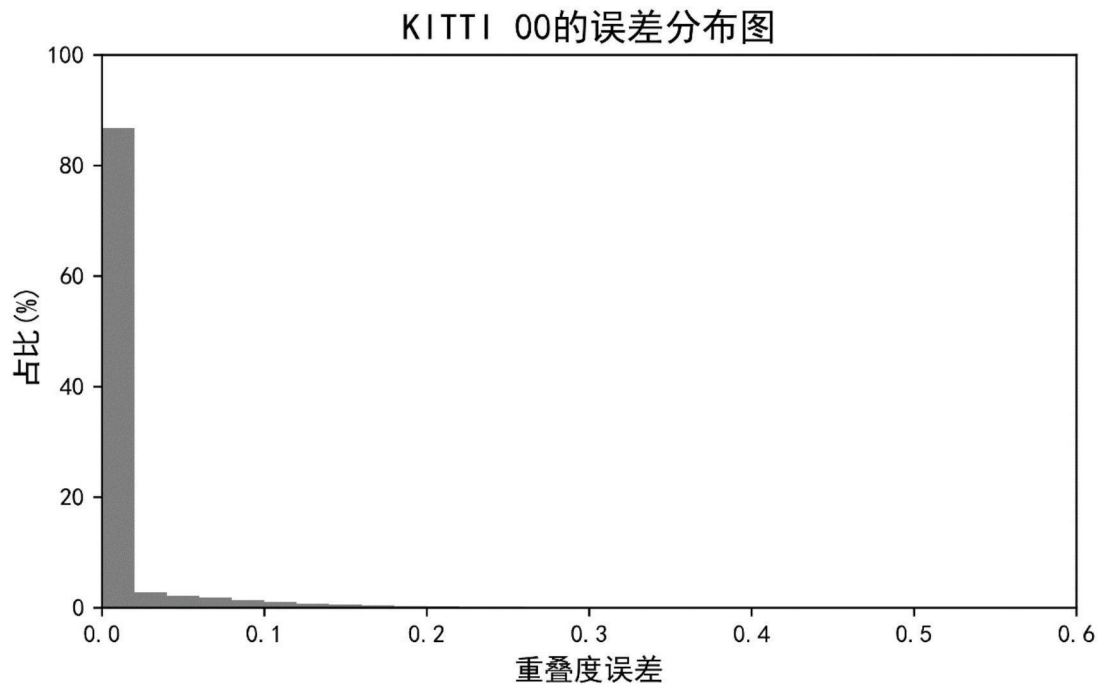


图5(a)

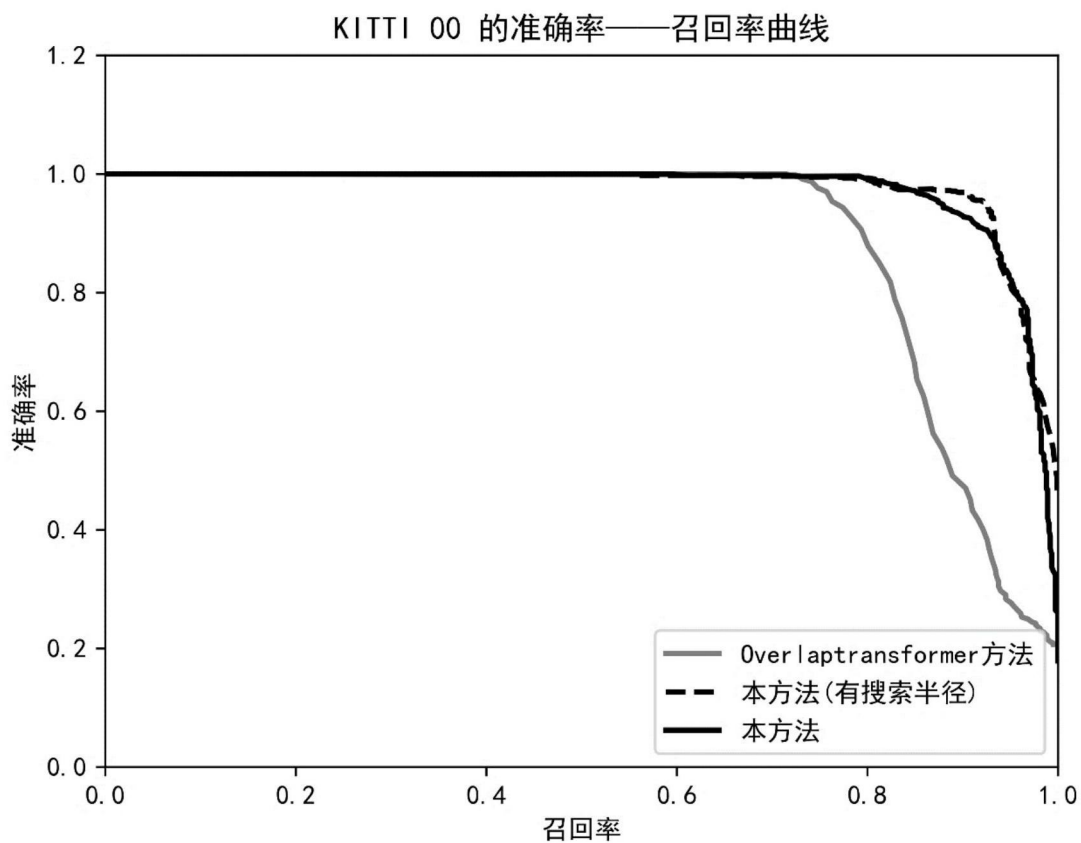


图5(b)

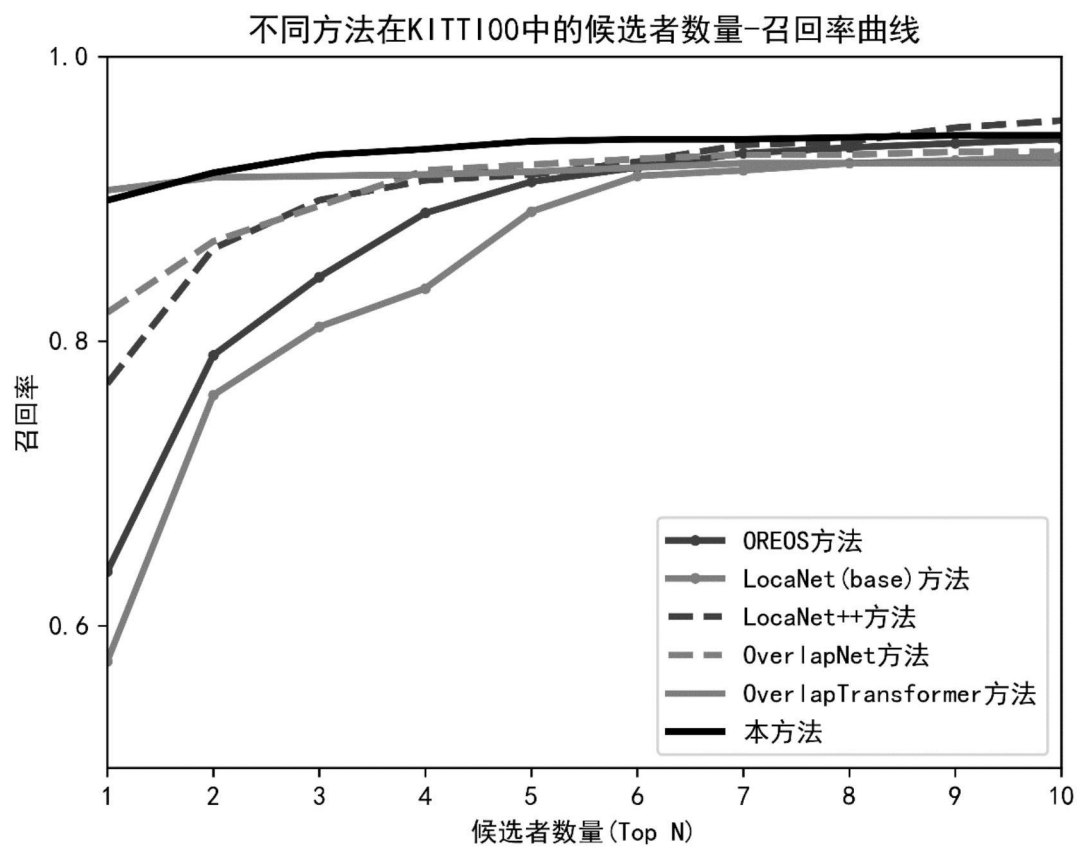


图6