结合局部特征匹配和点云配准的姿态估计网络

张 靖[1],于 伶[2]
1. 安徽工程大学 计算机与信息学院,安徽 芜湖 241000
2. 长三角哈特机器人产业技术研究院,安徽 芜湖 241000

摘 要:目的 现有的物体六自由度姿态估计方法侧重于处理已训练过的对象,针对未知物体纹理细节较弱或在有遮挡、光照复杂的非结构化环境中的六自由度姿态估计仍是一个具有挑战性的问题。**方法** 提出了一个物体六自由度姿态估计的网络,首先通过局部特征匹配获取图像对中足够多的匹配点对;其次,通过传感器读取匹配点对的深度信息得到相应点云数据并将其作为后续点云精准配准的初始点云进行配准,最终得到源点云相对于目标点云的旋转矩阵和平移向量,即未知物体在机器人坐标系下的六自由度姿态。**结果** 该网络在基准数据集上估计的 6D 姿态结果较好,$d_{\text{ADD-S}}$ 值为 80.0%;在 Occlusion Linemod 数据集上 $d_{\text{ADD-S}}$ 值也达到了 78.0%,均表现出了非常优异的性能。**结论** 该网络泛化性比较好,不仅能够准确地估计严重遮挡、背景杂波和光照差等条件下物体的六自由度姿态,而且对随机噪声也具有较好的鲁棒性。

关键词:姿态估计; 局部特征匹配;点云配准;未知物体
中图分类号:TP391 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2026.0002.018

Pose Estimation Network Combining Local Feature Matching and Point Cloud Registration

ZHANG Jing[1], YU Ling[2]
1. School of Computer and Information, Anhui Polytechnic University, Wuhu 241000, Anhui, China
2. Yangtze River Delta Hart Robot Industry Technology Research Institute, Wuhu 241000, Anhui, China

Abstract: Objective Existing six-degree-of-freedom (6-DoF) object pose estimation methods mainly focus on dealing with trained objects. Estimating the 6-DoF pose of unknown objects with weak texture details or in unstructured environments with occlusion and complex lighting remains a challenging problem. **Methods** A novel network for 6-DoF object pose estimation was proposed. First, a sufficient number of matching point pairs were obtained from image pairs through local feature matching. Second, the depth information of these matching point pairs was read via sensors to generate the corresponding point cloud data, which was used as the initial point cloud for subsequent high-precision point cloud registration. Finally, through the registration process, the rotation matrix and translation vector of the source point cloud relative to the target point cloud were obtained. The matrix and vector represented the 6-DoF pose of the unknown object in the robot coordinate system. **Results** The proposed network demonstrated strong performance in 6D pose estimation on benchmark datasets, with a $d_{\text{ADD-S}}$ value of 80.0%. It also achieved a $d_{\text{ADD-S}}$ value of 78.0% on the Occlusion Linemod dataset, demonstrating outstanding performance. **Conclusion** This network exhibits good generalization ability. It can accurately estimate the 6-DoF pose of objects under conditions such as severe occlusion, background clutter, and poor lighting. Moreover, it shows good robustness against random noise.

Keywords: pose estimation; local feature matching; point cloud registration; unknown object

收稿日期:2024-02-16 **修回日期:**2024-04-17 **文章编号:**1672-058X(2026)02-0139-07
基金项目:安徽省教育厅科学研究重点项目(KJ2020A0364)资助.
作者简介:张靖(1995—),女,浙江磐安人,硕士研究生,从事机器人技术研究.
通信作者:于伶(1972—),女,哈尔滨道里人,高级工程师,从事智能制造研究. Email:yuling3000@163.com.
引用格式:张靖,于伶.结合局部特征匹配和点云配准的姿态估计网络[J].重庆工商大学学报(自然科学版),2026,43(2):139-145.
ZHANG Jing, YU Ling. Pose estimation network combining local feature matching and point cloud registration[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(2): 139-145.

物体六自由度(Six Degrees Of Freedom,6D)姿态表示物体坐标系和相机坐标系之间的几何转换,可以用旋转矩阵和平移向量表示。6D姿态估计是机器人抓取与操作、增强现实、自动驾驶等应用中的关键步骤[1]。

从6D姿态估计的研究现状来看,一些研究者使用CNN来检测在3D对象模型上预定义的一组关键点[2],接着使用最小二乘拟合算法得到物体的6D位姿,该算法能够有效地使用刚性目标几何约束配合额外的深度信息,促进网络学习和优化。然而,这些关键点是稀疏匹配,这使得在遇到视角变化、遮挡或缺乏纹理的物体时,很难通过该方法得到满意的姿态估计效果;或是通过预测像素级的三维坐标,构建出密集的2D-3D对应地图[3-4],相对于使用CNN识别预设关键点的方法,这类方法在物体6D姿态估计时能够体现出更高的准确度。然而,该方法的效果更加依赖于所采用的训练数据集,因此在估计未知物体的姿态时可能无法实现预期的效果。

还有研究者通过基于模板的机制[5-8]进行物体姿态的估计,这些方法一般通过匹配输入图像和对应的3D模板或是另外生成的一系列模板达到物体姿态估计的效果。文献[8]提出了PointNet网络,该网络能够利用深度学习对3D点云数据及其全局函数进行提取,同时提出了对称函数以应对点无序性问题。该网络能够有效地执行分类和分割任务,但是其提取的点云特征仅包含非常有限的局部信息,因此在物体姿态估计方面的效果不尽如人意。基于模板的姿态估计方法很难保证估计精度,因为这类方法的精度很大程度上受限于遮挡和视点数量的限制。

为了使物体6D姿态估计算法摆脱对3D模板的依赖,研究者们提出了类别级姿态估计范式[9-11],该类方法通常采用同一类别不同实例的训练样本,以训练能够学习物体外观和形状的类别级表示的网络,从而可将所得模型推广应用于新的同一类别物体的估计过程中。然而,这种方法同样需要对同一类别中大量的样本进行训练,此外,当遇到一个新实例的外观或形状与训练过的样本显著不同时,并不能保证此类类别级方法的泛化能力。后来也有许多工作都是遵循了类别级姿态估计范式的原理[12-13],然后可以借助基于像素及每个类别的共享归一化对象坐标建立的对应关系,进一步精准地估计物体姿态。但这种方法的不足之处是,某些类别实例的形状和外观可能差异较大,导致训练过的网络在这些差异大的实例上的泛化能力不足,且在训练过程中依然通过生成精确的CAD模型来进行姿态估计。更严重的是,不同的类别往往需要训练不同的网络,无法在实际应用中快速推广。另外,此类方法在有遮挡或光照复杂的场景下精确度会受到很大影响。

通过以上比较探讨,每种方法都有各自不同的优缺点,基于关键点拟合位姿的方法和构建密集的2D-3D对应地图适于具有丰富的纹理或几何细节的目标;细节较弱时,更适合使用基于模板的方法或者类别级物体的6D位姿估计方法,但无法适用于有遮挡或光照复杂的场景,需要进一步对其泛化性进行研究。随着深度学习和卷积神经网络技术的不断发展,6D位姿估计方法已取得了很大进步,但还有很大的发展空间。针对6D位姿估计存在遮挡、光照复杂或是纹理稀疏的问题,本文提出了结合局部特征匹配和点云配准的姿态估计网络,使位姿估计更简便、更有效率,在实践中可以获得更好的结果。该网络首先通过LoFTR[16]局部特征匹配方法得到足够多的RGB图像对中的关键点匹配情况,再获取各匹配对的深度信息作为ICP算法中的初始点云,再通过点云配准的方法求解物体6D姿态参数。

LoFTR局部特征匹配方法不需要任何特征点的先验获取,可以直接执行端到端的匹配,此外,LoFTR在对象纹理较弱的情况下表现优异,能够直接获得足够稳定、足够多的特征点匹配对。点云配准在物体姿态估计领域应用广泛,因为点云的几何特征是通用的,因此点云配准方法非常适用于未知物体的点云匹配。常见的配准算法有迭代最近点(ICP)算法[14]、正态分布法(NDT)以及奇异值分解法(SVD)。其中,ICP算法配准效果最为显著,同时,本文通过LoFTR局部特征匹配方法解决了ICP配准算法对配准点云初始状态敏感、计算速度慢、求解最优目标函数时易陷入局部最优解等缺点[15],并且不需要任何模板,可以直接通过物体的一些姿态图片即可进行6D位姿估计,网络泛化性高,可以推广应用于未知物体的6D位姿估计。

1 姿态估计网络

针对6D位姿估计存在遮挡、光照复杂和纹理稀疏的问题,本文的姿态估计网络结合了局部特征匹配和点云配准方法,目标是训练一个不需要任何人工注释、快速且高精度的网络,使其能够直接得到未知物体的6D姿态。首先通过LoFTR局部特征匹配方法进行关键点匹配,再通过传感器读取深度信息得到相应的初始配对点云,最后利用ICP点云配准方法得到源点云相对于目标点云的旋转矩阵和平移向量,从而得到物体在摄像机坐标系下的6D姿态。姿态估计网络如图1所示,主要由LoFTR、获取深度信息、ICP点云配准组成。

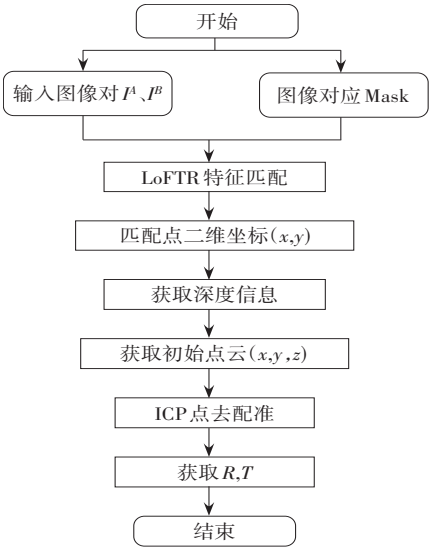

图 1　姿态估计网络流程图

Fig. 1　Flow diagram of pose estimation network

1.1　局部特征匹配

点云配准被广泛应用于物体 6D 姿态估计领域。点云配准的核心任务是确定两个不同视角下点云之间的旋转矩阵和平移向量,以便正确对齐两个点云。ICP 算法被认为是点云配准中最为经典的方法,但仅在优良的初值情况下,才能够获得良好的算法收敛性;且该算法搜索对应点的过程慢、代价大。因此,本文使用 LoFTR 局部特征匹配方法,首先从 RGB 图像对中得出足够多的关键点匹配对,具体流程如图 2 所示。

LoFTR 局部特征匹配方法使用了 Transformer 模块中的自注意层和互注意层来处理卷积网络中分离的密集局部特征。首先,通过从低特征分辨率(仅为 1/8 图像维度)的密集匹配中,选择具有高匹配置信度,并利用基于相关的方法将它们细化到高分辨率亚像素级别,转换成具有反映上下文和位置信息的特征。此外,多次运用自注意力层和互注意力层来学习匹配优先级,使其在低纹理、运动模糊或图像模式重复的区域,能够生成数量充足、质量优良的匹配结果。具体流程如图 3 所示,接下来将具体介绍该模块网络。

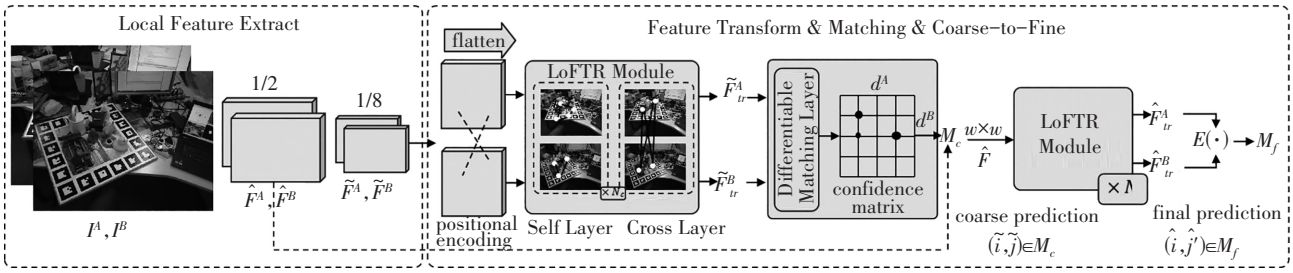

图 2　LoFTR 局部特征匹配方法

Fig. 2　LoFTR local feature matching method

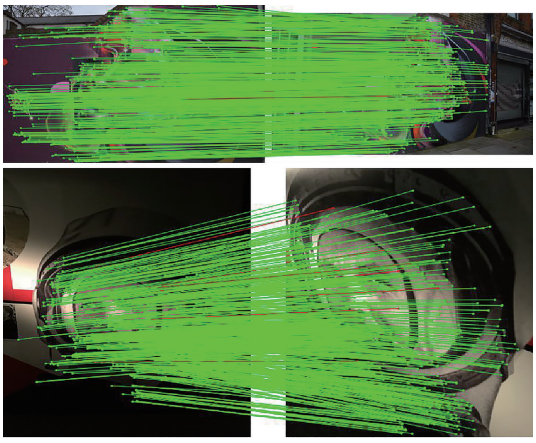

图 3　LoFTR 局部特征匹配效果

Fig. 3　The effect of local feature matching by LoFTR

使用 FPN 标准卷积结构从图像对 I^A、I^B 中提取原始图像维度 1/8 的粗粒度特征图 $\tilde{F}^A$ 和 $\tilde{F}^B$,以及原始图像维度 1/2 的细粒度特征图 $\hat{F}^A$ 和 $\hat{F}^B$。

FPN 标准卷积网络被认为是一种高效的 CNN 特征提取方法,该方法具有局部空间不变性,高效地使用降采样策略,从而减少 LoFTR 模块的输入长度。为了提取与位置和上下文相关的局部特征,将 $\tilde{F}^A$ 和 $\tilde{F}^B$ 展平为一维向量并添加位置编码,然后输入 LoFTR 模块中的自注意层(Self-attention)、互注意层(Cross-attention)进行处理,并交错重复 N_c 次输出易于匹配的特征 $\tilde{F}^A_{tr}$ 和 $\tilde{F}^B_{tr}$,计算两者之间的得分矩阵 $\boldsymbol{S}$ 如下:

$$\boldsymbol{S}(i,j)=\langle \tilde{F}^A_{tr}(i),\tilde{F}^B_{tr}(j)\rangle \tag{1}$$

式(1)中,<·,·>表示内积,i,j 为图像对中某一点的坐标,$\tilde{F}^A_{tr}$ 和 $\tilde{F}^B_{tr}$ 为输出的特征。

使用 SuperGlue 中的最优传输层(OT)进行匹配,同时在 $\boldsymbol{S}$ 的两个维度上应用 Softmax 得到一个置信矩阵 $\boldsymbol{P}_c$,如式(2)所示:

$$\boldsymbol{P}_c(i,j)=\mathrm{Softmax}(\boldsymbol{S}(i,\cdot))j\cdot\mathrm{Softmax}(\boldsymbol{S}(\cdot,j))i \tag{2}$$

在式(2)的基础上可以选择置信度高于 θ_c 的匹配,并进一步执行最近邻(MNN)准则,得到粗粒度匹配 M_c,如式(3)所示:

$$M_c=\{(\tilde{i},\tilde{j})\mid\forall(\tilde{i},\tilde{j})\in\mathrm{MNN}(\boldsymbol{P}_c),\boldsymbol{P}_c(\tilde{i},\tilde{j})\geqslant\theta_c\} \tag{3}$$

式(3)中,对于每一个粗粒度匹配$(\tilde{i},\tilde{j})\in M_c$,首先在细粒度特征映射 $\hat{F}^A$ 和 $\hat{F}^B$ 上确定$(\hat{i},\hat{j})$的位置,然后裁剪出大小为 $w\times w$ 的局部窗口,将其输入一个较小的 LoFTR 模块(N_f 层),生成以 $\hat{i}$ 和 $\hat{j}$ 为中心的局部特征图 $\hat{F}^A_{tr}(i)$ 和 $\hat{F}^B_{tr}(j)$。

将 $\hat{F}^A_{tr}(i)$ 的中心向量与 $\hat{F}^B_{tr}(j)$ 中的所有向量计算相关,并生成一个 Heatmap,通过计算概率分布的期望,可以得到在 I^B 上具有亚像素精度的最终位置 $\hat{j}$。最终产生细粒度的匹配项 M_f。在 Linemod[17] 数据集下 LoFTR 的匹配效果如图 4 所示,Linemod 数据集将在仿真实验处详细介绍。

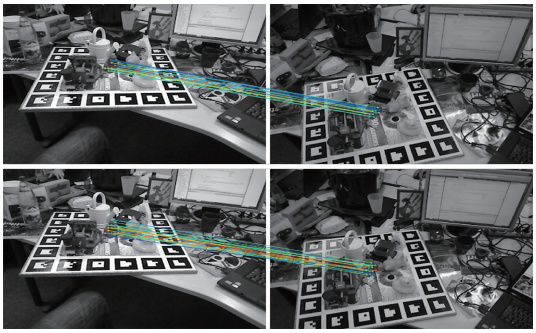

图 4 Linemod 的匹配效果

Fig. 4 Matching effect of Linemod

1.2 ICP 精准配准

点云配准是通过在两个点云之间创建相应的点对来创建变换模型的过程。现有的点云配准算法中应用最广泛的是 ICP 算法,该算法是基于最近点云之间的距离创建配准关系的模型。如果配准的初始姿态匹配较好,该算法可以获得精确的配准结果,否则会倾向于局部最优的状态。在通过上一节获得最终的匹配预测之后,本文使用传感器的内部矩阵获取深度信息,将匹配点对转换为包含三维几何信息的点云 $A=\{a_1,a_2,\cdots,a_n\}$ 与 $B=\{b_1,b_2,\cdots,b_n\}$,并作为 ICP 配准算法的初始点云来实现精确的物体姿态估计,具体流程如图 5 所示。

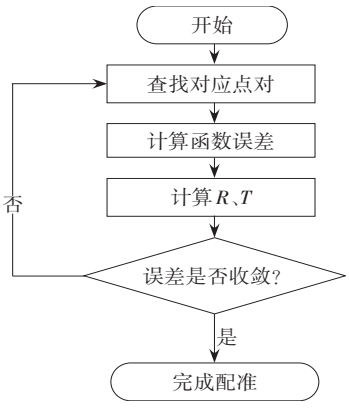

图 5 ICP 精细配准流程

Fig. 5 ICP fine registration process

ICP 算法的核心是迭代计算位姿变换矩阵,使源点集 A 与目标点集 B 之间的误差函数最小,从而得到最优解。点云 A 中的点 a_i 找到与点云 B 的欧氏距离最短的点 b_i,将 a_i 和 b_i 作为相应的点对,计算变换矩阵,然后计算配准误差 $E(\boldsymbol{R},\boldsymbol{T})$,如式(4)所示。本文的目标则是旋转矩阵 $\boldsymbol{R}$ 和平移向量 $\boldsymbol{T}$,使得 $E(\boldsymbol{R},\boldsymbol{T})$ 最小。

$$E(\boldsymbol{R},\boldsymbol{T})=\frac{1}{N}\sum_{i=1}^{n}\|b_i-(\boldsymbol{R}a_i+\boldsymbol{T})\|^2 \tag{4}$$

式(4)中,N 为点云数量,a_i 是点云 A 中的一个点,b_i 是点云 B 中的一个点,接下来将阐述具体计算过程。

根据点云坐标信息,计算两个点云对应的重心,如式(5)所示:

$$\begin{cases}\mu_a=\dfrac{1}{N}\sum_{i=1}^{n}a_i\\ \mu_b=\dfrac{1}{N}\sum_{i=1}^{n}b_i\end{cases} \tag{5}$$

式(5)中,μ_a、μ_b 分别代表点云 A 和 B 的重心。计算旋转矩阵和平移向量及其误差函数,使得其值最小。

预先设定阈值 ε 和最大迭代次数 $N_{\max}$,将上一步的变换矩阵作用于点云 A,得到新的待配准点云 A',计算 A'和 B 的距离误差 d,如式(6)所示。若 $d<\varepsilon$ 或者当前迭代次数大于 $N_{\max}$,则结束算法,否则将 A'设置为新的要进行配准的点云,重复上述步骤,直到满足算法结束的条件。

$$d=\frac{1}{N}\sum_{i=1}^{N}\|A'-B\|^2 \tag{6}$$

针对点云配准和特征提取网络的特点,本文采用的损失函数包括 2 个部分:两点云之间倒角距离、源点云和目标点云提取特征后经过平均池化得到点云平均特征。

本文采用两个点云之间的倒角距离 L_{AB} 损失来刻画两个点云之间的空间距离,如式(7)所示:

$$L_{AB}=\frac{1}{N}\sum_{a_i\in A}\min_{b_i\in B}\|A-B\|_2+\frac{1}{N}\sum_{b_i\in B}\min_{a_i\in A}\|B-A\|_2 \tag{7}$$

在两个点云提取特征后,经过平均池化得到点云平均特征,计算两个平均特征的欧几里得距离 L_D,如式(8)所示:

$$L_D=\|f(A)-f(B)\|_2 \tag{8}$$

其中,$f(\cdot)$为平均池化得到的点云平均特征,所以,最终的损失函数如式(9)所示:

$$L_{\text{tatal}}=L_{AB}+L_D \tag{9}$$

2　仿真实验与结果分析

2.1　数据集

为了验证本文所提出的结合局部特征匹配和点云配准的姿态估计网络的可行性与适用性,本文在服务器(操作系统为 Ubuntu 18.04 LTS,CPU 为 Intel Xeon 4210 R 2.4GHz,GPU:NVIDIA GeForce A40)上对其进行验证。

本文使用公共数据集 Linemod 对网络进行验证。Linemod 数据集具有场景杂乱、光照变化强烈和物体纹理弱的特点,这些特点使得在 Linemod 数据集中验证物体 6D 姿态的算法非常困难,具有挑战性。因此,使得该数据集广泛应用于物体 6D 姿态估计。Linemod 数据集由单目标数据集和多目标数据集组成,每个数据集包含 13 个子集,每个子集中有 1 300 张物体图像和物体的 3D 模型及其对应的虚拟 3D 控制点文件,并添加了 6D 姿态和实例语义分割掩码。Linemod 数据集是验证 6D 姿态估计算法中最典型的基准数据集之一,在很多算法中都被使用过[18]。

另外,本文还采用了 Occlusion Linemod 数据集,该数据集是通过对 Linemod 数据集的每个场景添加注释来创建的,每个图像都有不同的遮挡程度。一般用来验证有遮挡环境下的物体 6D 姿态估计效果。

2.2　评价指标

使用物体 6D 姿态估计中最常用的平均距离 d_{ADD} 作为度量标准来评价配准效果,它主要测量真实姿态和估计姿态之间的平均距离,在 Linemod 数据集中用于评价非对称物体。d_{ADD} 的计算方法如式(10)所示。当平均距离小于误差阈值时,认为估计的 6D 姿态是正确的。本实验将误差阈值设置为对象模型最大直径的 10%。

$$d_{\mathrm{ADD}} = \frac{1}{N}\sum_{x\in M} \| (\boldsymbol{R}x + \boldsymbol{T}) - (\hat{\boldsymbol{R}}x + \hat{\boldsymbol{T}}) \| \tag{10}$$

式(10)中,$[\boldsymbol{R},\boldsymbol{T}]$表示物体真实姿态,$[\hat{\boldsymbol{R}},\hat{\boldsymbol{T}}]$表示估计的姿态。$M$ 表示坐标系中的对象三维模型,x 是属于模型 M 的点,N 是 M 中点云的数量。由于对称物体旋转角度的不确定性,采用平均最近点距离 $d_{\mathrm{ADD\text{-}S}}$ 作为对称物体的评价准则。$d_{\mathrm{ADD\text{-}S}}$ 计算三维物体真实模型上最近点与预测模型之间的平均距离,具体定义如式(11):

$$d_{\mathrm{ADD\text{-}S}} = \frac{1}{N}\sum_{x_1\in M}\min_{x_2\in M} \| (\boldsymbol{R}x_1 + \boldsymbol{T}) - (\hat{\boldsymbol{R}}x_2 + \hat{\boldsymbol{T}}) \| \tag{11}$$

2.3　结果分析

为验证本文姿态估计的效果,将本文算法与目前物体 6D 位姿估计方法中较先进且常用的模型进行了比较。使用的模型分别为文献[3]中的 BB8 模型、文献[12]中的 GDR-NET 模型及 PVNet[19] 模型,比较结果如表 1 所示。

表 1　Linemod 上不同算法的 $d_{\mathrm{ADD\text{-}S}}$

Table 1　$d_{\mathrm{ADD\text{-}S}}$ scores of different algorithms on Linemod

方　法	BB8 模型	GDR-NET 模型	PVNet 模型	Ours 模型
Ape	40.4	38.2	43.6	43.1
Benchvise	89.1	84.1	99.9	89.3
Camera	84.2	82.3	86.8	86.1
Can	92.6	90.7	95.5	94.2
Cat	46.3	50.0	79.3	68.2
Drill.	68.2	68.1	96.4	74.5
Duck	32.1	34.2	52.5	47.2
Egg.	96.4	93.9	99.2	97.1
Glue	91.2	91.1	95.7	93.3
Hole	79.5	77.8	82.0	80.4
Iron	90.8	88.2	98.9	91.0
Lamp	83.7	80.3	99.3	87.3
Phone	84.6	73.0	92.4	87.6
Mean	75.3	73.2	86.3	80.0

因为 PVNet 网络充分使用了 Linemod 数据集中的真实数据进行训练,因此该网络使用的数据集细节更贴近物体,使得该模型的效果远好于 BB8 模型、GDR-NET 模型以及本文模型,但本文模型在 Ape、Camera、Can、Glue 等物体类别中都获得了与 PVNet 模型相近 $d_{\mathrm{ADD\text{-}S}}$ 值,本文模型相较于 BB8 模型,$d_{\mathrm{ADD\text{-}S}}$ 提高了 4.7%,相较于 GDR-NET 模型,$d_{\mathrm{ADD\text{-}S}}$ 提高了 6.8%。但在 Cat、Duck 等弱纹理物体上,本文的模型效果非常优异,比 BB8 模型分别提高了 21.9%、15.1%,比 GDR-NET 分别提高了 18.2%、13%,说明本文模型在物体纹理及几何细节较弱时表现优异。其原因是本文中所采用的 LoFTR 局部特征匹配方法在设计上汲取了 Transformer 的优势,并借助自注意层和互注意层获得了两幅图像的特征描述符,从而达到了更为准确的效果。此种方法带来的全局接受域特性有效提升了网络性能,让网络在纹理较弱、细节少的区域同样可以实现密集匹配。Linemod 数据集结果的可视化如图 6 所示,根据旋转矩阵 $\boldsymbol{R}$ 和平移向量 $\boldsymbol{T}$ 移动模型中的点,并将其投影到 RGB 图像上。

图 6　Linemod 数据集可视化

Fig. 6　Visualization of Linemod dataset

因为 RGB-D 相机在黑暗或强光条件下捕捉图像,只影响 RGB 图像,而深度图像没有影响,为了验证本文的网络在光照变化下的鲁棒性,本文通过改变同一场景中 RGB 图像的亮度值进行实验。

在光照变化条件下的 6D 姿态估计可视化结果如图 7 所示,可以看出:该网络可以较好地解决光照变化条件下姿态估计困难的问题。

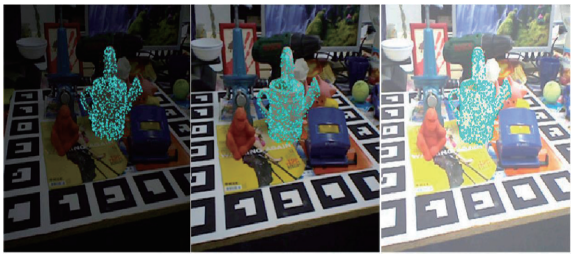

图 7 光照下 6D 姿态估计结果

Fig. 7 Results of 6D pose estimation under illumination

表 2 显示了本文网络在 Occlusion Linemod 数据集的实验结果。本文网络估计的 6D 姿态结果表明:在 Linemod 数据集上 $d_{\text{ADD-S}}$ 值为 80.0%,在 Occlusion Linemod 数据集上 $d_{\text{ADD-S}}$ 值为 78.0%,$d_{\text{ADD-S}}$ 指标有轻微下降,经过分析是因为注意力机制的网络生成的注意力图得分高的像素点由于物体存在遮挡,使得对应关系存在偏差。虽然 PVNet 数据集进行了真实样本训练表现最优越,但本文模型的 $d_{\text{ADD-S}}$ 值比 BB8 模型高了 3.6%,比 GDR-NET 模型高了 5.2%。证明本文模型在有严重的遮挡和背景杂波环境下对物体 6D 姿态估计的性能相较于其他模型表现更优异。

表 2 Occlusion Linemod 上不同算法的 $d_{\text{ADD-S}}$

Table 2 $d_{\text{ADD-S}}$ scores of different algorithms on Occlusion Linemod

方 法	BB8 模型	GDR-NET 模型	PVNet 模型	Ours 模型
Ape	39.4	38.0	42.3	42.1
Benchvise	87.3	83.8	96.7	88.9
Camera	83.2	81.6	85.4	85.1
Can	91.9	90.4	94.5	93.9
Cat	45.3	49.5	78.1	67.9
Drill.	67.1	67.8	94.9	73.8
Duck	31.0	33.7	51.7	47.1
Egg.	95.7	93.5	98.3	96.5
Glue	90.6	91.0	94.2	93.0
Hole	78.5	76.9	81.3	80.1
Iron	89.7	88.1	97.6	90.4
Lamp	83.4	79.6	98.4	86.9
Phone	84.2	72.4	91.0	87.3
Mean	74.4	72.8	85.0	78.0

本文还对网络在噪声下的物体 6D 姿态估计精度做了验证实验。众所周知,当 RGB-D 相机在真实环境中捕捉图像时,噪声对深度图像的影响非常大。深度图像中每个像素的数据是从物体到相机的距离,以 mm 为单位。本文通过在 Linemod 数据集的深度图像的每个像素中添加随机噪声,随机噪声数范围为(-25,25),以此来验证本文的 6D 姿态估计网络对噪声的鲁棒性。表 3 显示了该网络在随机噪声增加条件下的性能,可以看出,该网络在噪声情况下也有较好的精度。

表 3 随机噪声下网络精度

Table 3 Network accuracy under random noise

噪音范围/mm	0	5	10	15	20	25
精度/%	99.4	99.4	99.3	99.2	99.1	99.0

3 结论与展望

在复杂非结构化环境中面对未知物体,还存在物体 6D 姿态估计困难的问题,主要原因在于现有的物体 6D 姿态估计方法比较依赖物体丰富的纹理或几何细节,且存在遮挡或光照复杂的场景下姿态估计网络泛化性不足的问题。针对此难题,本文设计了结合局部特征匹配与点云配准的姿态估计网络,用于未知物体的 6D 姿态估计。利用 LoFTR 局部特征匹配算法获取足够多的特征点,并得到像素级的匹配点集。通过 RGB-D 的深度信息得到相应的 3D 点云,并将其作为 ICP 精准配准算法的初始点云来解决 ICP 精准配准算法过度依赖初始配准点云的问题,避免了点云配准时基于 RGB-D 数据陷入局部最优的问题。实验表明:本文网络在两个基准数据集 Linemod 和 Occlusion Linemod 上都表现出较好的性能,可以高效率、高精度地进行物体 6D 姿态估计,网络泛化性好,能够准确地估计严重遮挡、背景杂波和光照差等条件下的物体姿态,且对随机噪声也具有较好的鲁棒性。

本文网络可以用来估计任何未知物体的 6D 姿态,但仅在数据集上做了仿真实验,在未来的研究中,可以在真实环境中进行实验验证,并且在局部特征匹配过程中仍需消耗一定的时间,网络中某些参数的阈值设置问题也需要进一步优化,这些问题将在未来工作中继续改进。

参考文献(References):

[1] 邢广鑫, 许钢, 荣桂兰, 等. 动态环境下改进 ICP 算法的 RGB-D SLAM 研究[J]. 重庆工商大学学报(自然科学版), 2020, 37(3): 81-87.
XING Guang-xin, XU Gang, RONG Gui-lan, et al.

Improvement of the ICP algorithm of RGB-D SLAM in dynamic environment[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2020, 37(3): 81-87.

[2] 王太勇, 于恩霖. 基于三维关键点投票的物体位姿估计方法[J]. 天津大学学报(自然科学与工程技术版), 2024, 57(3): 291-300.
WANG Tai-yong, YU En-lin. Object pose estimation method based on 3D key points voting[J]. Journal of Tianjin University (Science and Technology), 2024, 57(3): 291-300.

[3] RAD M, LEPETIT V. BB8: A scalable, accurate, robust to partial occlusion method for predicting the 3D poses of challenging objects without using depth[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 3848-3856.

[4] ZAKHAROV S, SHUGUROV I, ILIC S. DPOD: 6D pose object detector and refiner[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 1941-1950.

[5] 何弦, 李佳宸, 金立, 等. 三维模板跟踪的基准合成数据集构建及算法评估[J]. 计算机学报, 2022, 45(3): 585-600.
HE Xian, LI Jia-chen, JIN Li, et al. A synthetic dataset and performance evaluation for 3D template tracking[J]. Chinese Journal of Computers, 2022, 45(3): 585-600.

[6] 王连庆, 钱莉. 基于 3D 标定块的机器人与 3D 相机手眼标定研究[J]. 激光与光电子学进展, 2021, 58(24): 2433001.
WANG Lian-qing, QIAN Li. Research on robot hand-eye calibration method based on three-dimensional calibration block[J]. Laser & Optoelectronics Progress, 2021, 58(24): 2433001.

[7] 任笑圆, 蒋李兵, 钟卫军, 等. 基于视觉的非合作空间目标三维姿态估计方法[J]. 电子与信息学报, 2021, 43(12): 3476-3485.
REN Xiao-yuan, JIANG Li-bing, ZHONG Wei-jun, et al. A vision-based method for 3D pose estimation of non-cooperative space target[J]. Journal of Electronics & Information Technology, 2021, 43(12): 3476-3485.

[8] CHARLES R Q, HAO S, MO K, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 77-85.

[9] 栗仁武, 张凌霄, 高林, 等. 基于点云的类级别物体姿态估计[J]. 智能科学与技术学报, 2022, 4(2): 246-254.
LI Ren-wu, ZHANG Ling-xiao, GAO Lin, et al. Category-level object pose estimation from depth point cloud[J]. Chinese Journal of Intelligent Science and Technology, 2022, 4(2): 246-254.

[10] CHEN K, DOU Q. SGPA: Structure-guided prior adaptation for category-level 6D object pose estimation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 2753-2762.

[11] 曾一芳, 钱伟中, 王旭鹏, 等. 基于关键点的类别级三维可形变目标姿态估计[J]. 计算机应用研究, 2022, 39(2): 587-592.
ZENG Yi-fang, QIAN Wei-zhong, WANG Xu-peng, et al. Category-oriented 3D articulated objects pose estimation based on key points[J]. Application Research of Computers, 2022, 39(2): 587-592.

[12] WANG G, MANHARDT F, TOMBARI F, et al. GDR-net: Geometry-guided direct regression network for monocular 6D object pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 16606-16616.

[13] LEE T, LEE B U, KIM M, et al. Category-level metric scale object shape and pose estimation[J]. IEEE Robotics and Automation Letters, 2021, 6(4): 8575-8582.

[14] BESL P J, MCKAY N D. A method for registration of 3-D shapes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14(2): 239-256.

[15] 宋涛, 曹利波, 赵明富, 等. 三维点云中关键点的配准与优化算法[J]. 激光与光电子学进展, 2021, 58(4): 0415008.
SONG Tao, CAO Li-bo, ZHAO Ming-fu, et al. Registration and optimization algorithm of key points in three-dimensional point cloud[J]. Laser & Optoelectronics Progress, 2021, 58(4): 0415008.

[16] SUN J, SHEN Z, WANG Y, et al. LoFTR: detector-free local feature matching with transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 8918-8927.

[17] HINTERSTOISSER S, LEPETIT V, ILIC S, et al. Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes[C]//Asian Conference on Computer Vision. Berlin. Heidelberg: Springer, 2013: 548-562.

[18] WANG C, XU D, ZHU Y, et al. DenseFusion: 6D object pose estimation by iterative dense fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 3338-3347.

[19] PENG S, LIU Y, HUANG Q, et al. PVNet: Pixel-wise voting network for 6DoF pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4556-4565.

责任编辑:代小红