沙伟, 张华. 面向机械臂分拣的筒子纱 6D 位姿估计网络研究[J]. 智能计算机与应用,2025,15(9):117-123. DOI:10.20169/j.issn.2095-2163.250919

面向机械臂分拣的筒子纱 6D 位姿估计网络研究

沙 伟[1,2], 张 华[1,2]

(1 浙江理工大学 机械工程学院, 杭州 310018; 2 浙江理工大学 浙江省现代纺织装备技术重点实验室, 杭州 310018)

摘 要: 为了提高筒子纱分拣任务中 6D 位姿估计的准确性与效率,提出一种新颖的位姿估计方法。该方法基于改进的 ResNeSt 网络实现向量场预测和语义分割,结合 EPnP 算法计算筒子纱 6D 位姿。ResNeSt 网络通过多分支特征提取与 Split-Attention 机制,有效聚合不同通道中的特征信息,用于构建编码器-解码器网络。单位向量误差的欧式距离损失及关键点到预测向量距离的正则化被用于网络训练。设计并制作了一组筒子纱数据集用于训练和测试。通过 2D Projection 指标和 ADD-S 指标对位姿估计结果进行评价。实验结果表明,所提方法显著提高筒子纱 6D 位姿估计精度和速度,减少了模型参数量,能够有效满足筒子纱的抓取和分拣任务的需求。

关键词: 筒子纱分拣; 位姿估计; 神经网络; 正则化; 数据集

中图分类号: TP391.4 **文献标志码**: A **文章编号**: 2095-2163(2025)09-0117-07

Research on thread roll 6D pose estimation network for robotic arm sorting

SHA Wei[1,2], ZHANG Hua[1,2]

(1 School of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2 Key Laboratory of Modern Textile Machinery Technology of Zhejiang Province, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: To enhance the accuracy and efficiency of thread rolls 6D pose estimation in the sorting task, a novel pose estimation method is proposed. This method implements vector field prediction and semantic segmentation based on the improved ResNeSt network, and uses the EPnP algorithm to compute the 6D pose of thread rolls. The ResNeSt effectively aggregates feature information in different channels through multi-branch feature extraction and Split-Attention mechanisms, and is used to build an encoder-decoder network. Euclidean distance loss of unit vector error and regularization of distance from keypoints to predicted vectors are used for network training. A thread roll dataset is designed and created for training and testing. The pose estimation results are evaluated using the 2D Projection metric and ADD-S metric. Experimental results demonstrate that the proposed method significantly improves the precision and speed of thread roll 6D pose estimation, reduces the model parameters, and effectively meets the requirements of thread roll grasping and sorting tasks.

Key words: thread roll sorting; pose estimation; neural network; regularization; dataset

0 引 言

在纺织行业中,筒子纱的分拣、搬运均需要人工参与,这种作业方式不仅效率低下,而且劳动强度较高。近年来,基于视觉引导的机械臂分拣筒子纱代替人工分拣的方案成为当下传统纺织行业的一个重要发展趋势[1]。然而,由于缺乏对筒子纱 6D 位姿检测的能力,一般的基于视觉引导的机械臂抓取任务仅针对置于平面上的物体,无法应对筒子纱堆叠场景[2]。筒子纱的 6D 位姿是指其相对于相机坐标系的三维旋转矩阵 $\boldsymbol{R}$ 和三维平移矩阵 $\boldsymbol{t}$。实现精确的筒子纱 6D 位姿估计的关键挑战有遮挡、光法线变化和弱纹理特征等[3-5]。Ulrich 等学者[6]通过对物体进行多视点采样建立模板库,将待测图像与模板库进行相似性匹配,可以适应弱纹理物体,然而这种方法空间搜索效率低,且采用稀疏视角采样无法保证位姿估计的精度。在传统方式中,利用手工制作的特征[7-9]建立输入图像与 3D 模型之间对应关系的姿态估计方法,由于无法获得深层次的上下文,在面对堆叠场景和光照变化时鲁棒性较差。

近年来,深度卷积神经网络在物体检测和语义分割中表现出强大特征提取能力,基于深度学习的

作者简介: 沙 伟(1994—),男,硕士研究生,主要研究方向:机器人视觉系统。Email:2101203142@qq.com; 张 华(1979—),男,博士,副教授,硕士生导师,主要研究方向:机器人运动控制,智能纺织装备控制。

收稿日期: 2023-12-21

位姿估计方法成为研究的热点。相较于从输入图像中直接回归物体姿态[10],使用卷积神经网络回归2D关键点,然后使用PnP算法[11]计算物体6D位姿的两阶段物体姿态估计方法具有稳定的算法可解释性。Rad等学者[12]通过检测物体图像包围盒顶点位置,在仅使用RGB图像下实现了较好的位姿估计效果,但物体被视为全局实体影响包围盒顶点检测精度。Peng等学者[13]提出PVNet回归物体向量场用于关键点投票的方法,对遮挡具有一定的鲁棒性,但是受限于ResNet[14]的单层特征和缺乏跨通道信息融合,物体位姿估计精度仍有提升空间。Zhang等学者[15]在多尺度上优化向量场回归,然而这种方法无法实现多分支结构通道间的信息交互。

基于以上分析,本文提出采用基于改进的ResNeSt[16]构建筒子纱6D位姿估计的编码器-解码器网络。ResNeSt将通道注意力和多分支结构融合在特征提取单元中,使模型参数量减少情况下,却增强了对弱纹理筒子纱特征提取能力。同时,考虑到像素点与关键点之间距离对关键点检测的影响[17],单位向量误差的欧式距离损失及关键点到预测向量的距离作为正则化被用于网络训练。最后,受限于纺织领域缺少用于筒子纱位姿估计的数据集,因此设计并制作了一组用于筒子纱位姿估计的数据集用来验证所提方法。

1 位姿估计网络

1.1 基于关键点检测的位姿估计方法概述

筒子纱位姿估计采用稀疏2D-3D关键点对应的方法[13],如图1所示。筒子纱的单张RGB图像输入到所提编码器-解码器网络进行语义分割和向量场预测,根据预测的向量场和语义分割掩码采用投票的方式检测筒子纱2D关键点,最后根据EPnP算法[10]计算出筒子纱的位姿矩阵$[\boldsymbol{R}\ \boldsymbol{t}]$。

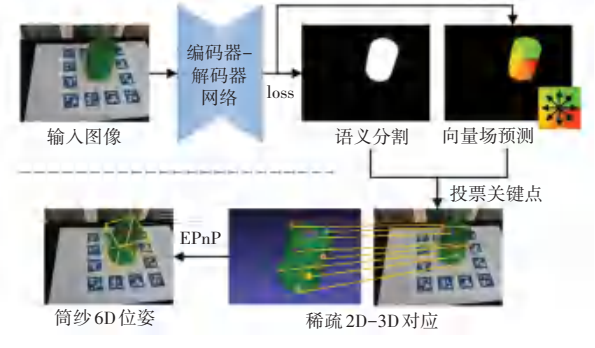

图1 筒子纱位姿估计方案

Fig. 1 Thread roll pose estimation scheme

关键点的向量场定义为筒子纱像素p相对于2D关键点k的单位方向向量$u_k(p)$,数学公式为:

$$u_k(p)=\frac{k-p}{\|k-p\|_2} \tag{1}$$

根据单位方向向量采用基于RANSAC投票机制生成关键点假设。具体来说,在预测的向量场中随机选取2个筒子纱像素对应预测的单位方向向量$v_k(p_1)$、$v_k(p_2)$,把两者所在直线的交点作为关键点k的假设h_k,如图2所示。重复此操作N次得到候选关键点假设集合$\{h_{k,i}\mid i=1,2,\cdots,N\}$。

图2 关键点假设

Fig. 2 Keypoint hypothesis

对于关键点k的一组假设集合,计算关键点假设$h_{k,i}$的投票得分$w_{k,i}$,数学公式为:

$$w_{k,i}=\sum_{p\in O}\mathbb{I}\left(\frac{(h_{k,i}-p)^{\mathrm{T}}}{\|h_{k,i}-p\|_2}v_k(p)\geqslant 0.99\right),$$
$$\mathbb{I}(\delta)=\begin{cases}1, & \text{若}\ \delta\ \text{为真}\\ 0, & \text{其它}\end{cases} \tag{2}$$

在投票过程中,具有更高得分的关键点假设代表该点将更有可能是正确关键点,因此选择具有最高得分的关键点假设来建立稀疏的2D-3D对应关系。而后用EPnP算法计算出6D位姿,具体为首先选取世界坐标系下控制点$\boldsymbol{C}_k^w$。对于世界坐标系下3D关键点$\boldsymbol{X}_k^w$可表示为:

$$\boldsymbol{X}_k^w=\sum_{n=1}^{4}\alpha_{kn}\boldsymbol{C}_n^w \tag{3}$$

其中$\sum_{n=1}^{4}\alpha_{kn}=1$,然后构建3D关键点$\boldsymbol{X}_k^w$在相机坐标系下的映射$\boldsymbol{X}_k^c$:

$$\boldsymbol{X}_k^c=\boldsymbol{R}\boldsymbol{X}_k^w+\boldsymbol{t}=\sum_{n=1}^{4}\alpha_{kn}(\boldsymbol{R}\boldsymbol{C}_n^w+\boldsymbol{t})=\sum_{n=1}^{4}\alpha_{kn}\boldsymbol{C}_n^c \tag{4}$$

其中,$\boldsymbol{C}_n^c$表示控制点$\boldsymbol{C}_n^w$在相机坐标系下的表示,这里,$n=1,2,3,4$。

接下来,利用相机内参K构建投影映射:

$$p_k=\boldsymbol{K}\boldsymbol{X}_k^c=\boldsymbol{K}\sum_{n=1}^{4}\alpha_{kn}\boldsymbol{C}_n^c \tag{5}$$

C_n^c 可通过式(5)求得，最后可以通过式(4)求得位姿[R　t]。

1.2　基于 ResNeSt 的像素级投票网络

所提方法基于改进的 ResNeSt50 构建像素级投票网络。ResNeSt 可以在不同的网络分支上应用通道软注意力来进行跨通道特征交互和学习多样化表示，从而提高语义分割和向量场的精度 。

ResNeSt 将类似 ResNeXt[18] 的多分支结构与 SKnet[19] 的软注意力机制结合。ResNeSt 模块结构如图 3 所示，将输入沿着通道维度划分为 K 个组(Cardinal)，每个组中又进一步分为 R 个分支(Split)，这样总共可以提取 $K \times R$ 个特征图组。对于组内一个分支在进行 3×3 卷积之后，使用拆分注意力模块(Split Attention)对各分支特征图组进行加权；然后对 K 个组输出的特征图进行拼接，最后进行 1×1 卷积后与输入进行相加。

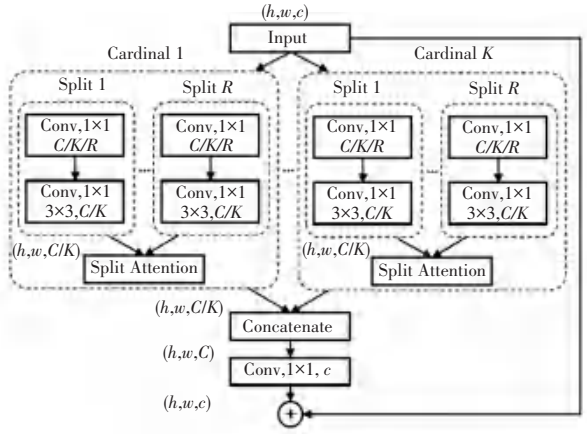

图 3　ResNeSt 模块结构图

Fig. 3　Block structure diagram of ResNeSt

拆分注意力机制的具体结构如图 4 所示，通过接收一个组内的 R 个输入并进行求和，再经由全局池化逐通道地聚合全局上下文，对此可以表示为：

$$s_c^k = \frac{1}{H \times W} \sum_i^H \sum_j^W \hat{U}_c^k(i, j) \tag{6}$$

其中，s_c^k 表示第 c 通道的全局信息；$\hat{U}_c^k(i, j)$ 表示来自第 k 组内全部分支求和后的第 c 通道特征图在 (i, j) 处的值，$k \in \{1,2,3,\cdots,K\}$；H 和 W 分别表示模块输出特征图高和宽。

对于一个组的输出是通过对组内 R 个分支输入特征图使用通道软注意力分别进行加权融合产生，具体公式如下：

$$V_c^k = \sum_{i=1}^{R} a_i^k(c)\, U_{R(k-1)+i} \tag{7}$$

$$a_i^k(c) = \begin{cases} \dfrac{\exp(G_i^c(s^k))}{\sum_{j=1}^{R} \exp(G_j^c(s^k))}, & R > 1 \\ \dfrac{1}{1 + \exp(-G_i^c(s^k))}, & R = 1 \end{cases} \tag{8}$$

其中，V_c^k 表示第 k 组输出中的第 c 通道的特征图；$a_i^k(c)$ 表示每个分支特征图组的权重；G_i^c 表示每个分支第 c 通道的权重映射函数；$U_{R(k-1)+i}$ 表示第 k 组内第 i 分支的特征图。

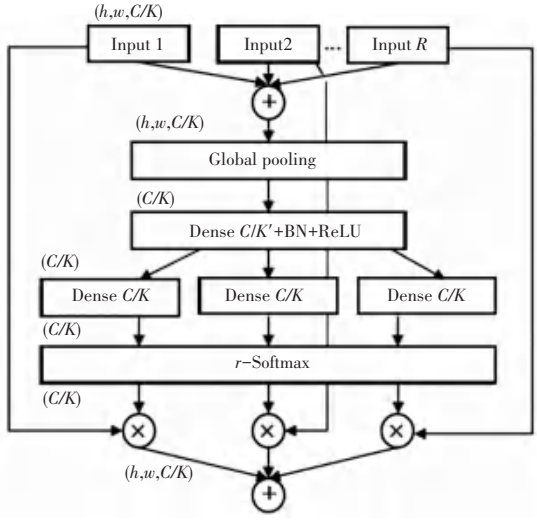

图 4　Split Attention 结构图

Fig. 4　Structure diagram of Split Attention

基于 ResNeSt50 构建的编码器-解码器像素级投票网络，如图 5 所示，用于筒子纱向量场和分割掩码的预测。在编码过程中将特征图下采样到输入 RGB 图像尺寸的 1/8 后，不再使用后续的 ResNeSt Block。在解码阶段，在特征图上反复执行跳跃连接、卷积和上采样操作，直到恢复到原输入图像尺寸。

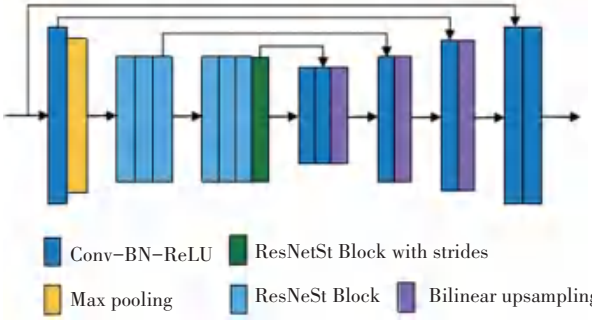

图 5　基于 ResNeSt 的像素级投票网络

Fig. 5　Pixel-level voting network based on ResNeSt

1.3　损失函数设计

对于向量场的预测任务，通过计算预测的单位方向向量与真实单位方向向量之间的欧式距离来定义向量场预测损失函数，数学公式具体如下：

$$L_{vf}=\frac{1}{KO}\sum_{k_i\in K}\sum_{p\in O}\|v_i(p)-u_i(p)\|_2 \tag{9}$$

其中，p 表示属于物体 O 的像素；K 表示物体关键点集合；$\boldsymbol{v}(p)$ 表示像素 p 的预测单位方向向量；$\boldsymbol{u}(p)$ 表示像素真实单位向量。

此外，像素到关键点距离对关键点假设的影响如图 6 所示。当 2 个像素的预测单位向量误差相同时，像素和关键点之间的距离影响关键点假设的偏差程度。因此关键点到预测单位向量的距离 d 作为正则化项被添加到向量场预测损失中，这本质上是增大了对远离关键点的单位向量预测的惩罚力度，正则化公式如下：

$$L_r=\frac{1}{KO}\sum_{k\in K}\sum_{p\in O}l(p,k)\cdot\frac{|v_p^y\cdot u_p^x-v_p^x\cdot u_p^y|}{\sqrt{(v_p^x)^2+(v_p^y)^2}} \tag{10}$$

其中，$l(p,k)$ 表示像素点 p 到关键点 k 的距离；$\boldsymbol{v}(p,k)=(v_p^x,v_p^y)$ 表示像素 p 对关键点 k 的单位方向向量的预测；$\boldsymbol{u}(p,k)=(u_p^x,u_p^y)$ 表示像素 p 对关键点 k 的真实单位方向向量。

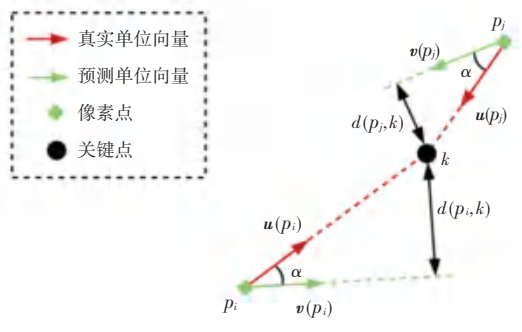

图 6 像素到关键点距离对关键点假设的影响

Fig. 6 The impact of pixel-to-keypoint distance on keypoint hypothesis

对于语义分割标签 $s(p)(s(p)\in[0,1],\forall O)$ 的预测，采用 Softmax 交叉熵损失函数：

$$L_{\text{seg}}=-\sum_{p\in O}\log(s(p)) \tag{11}$$

总的损失函数为：

$$L=\alpha L_{\text{seg}}+L_{vf}+\beta L_d \tag{12}$$

其中，α 和 β 表示损失函数的平衡权重系数。

2 数据集制作

由于没有可供筒子纱 6D 位姿估计任务使用的公开数据集，本文提出一种基于筒子纱 RGB 图像的数据集构建方案，并制作一组筒子纱数据集，用于验证所提方法，如图 7 所示。

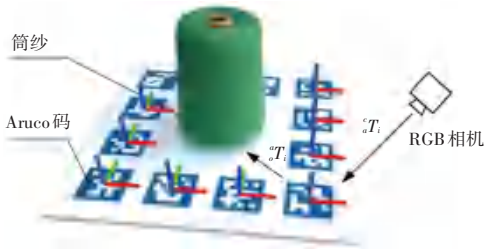

图 7 筒子纱数据集构建方案

Fig. 7 Scheme for creating the thread roll dataset

在筒子纱的周围放置 12 个 Aruco 标记，并分别确定筒子纱相对于每个 Aruco 标记的位姿 ${}^{a}_{o}T$；然后，使用相机获取筒子纱和 Aruco 标记的 RGB 图像，并用 OpenCV 检测可识别的 Aruco 标记相对于相机坐标系的位姿 ${}^{c}_{a}T$；最后，根据 Aruco 标记的位姿 ${}^{c}_{a}T$ 计算出筒子纱相对于相机的位姿 ${}^{c}_{o}T$，公式如下：

$${}^{c}_{o}T=\frac{1}{n}\sum_{i=1}^{n}{}^{c}_{a}T_i\,{}^{a}_{o}T_i \tag{13}$$

其中，${}^{c}_{o}T$ 表示筒子纱相对于相机的位姿；${}^{c}_{a}T_i$ 表示第 i 个 Aruco 标记相对于相机的位姿；${}^{a}_{o}T_i$ 表示筒子纱相对于第 i 个 Aruco 标记的位姿；n 表示可检测到的 Aruco 标记的数量。

对于筒子纱的掩码获取，通过固定筒子纱并使用三维扫描仪对筒子纱建模，如图 8 所示。随后根据前面计算出的筒子纱位姿通过对三维模型进行 2D 投影的方式获取筒子纱的掩码。

(a) 筒子纱 1 (b) 筒子纱 2 (c) 筒子纱 3 (d) 筒子纱 4

图 8 筒子纱三维模型

Fig. 8 Three-dimensional model of thread roll

制作的筒子纱数据集包括：分辨率为 640×480 的 RGB 图像、掩码、筒子纱位姿标签和筒子纱三维模型，共制作 4 种筒子纱类别，按照尺寸从小到大分别为筒子纱 1 到筒子纱 4，图像数量分别为 1 143、

1 201、1 240、1 203 和公开 LINEMOD 数据集中每个类别数量基本一致，可以满足训练和测试所需。

3　实验结果及分析

实验环境为 Intel Xeon Platinum 8255C CPU，GeForce 3090 GPU，Ubuntu18.04.5，Pytorch1.8.1 深度学习框架，CUDA11.1。采用 Adam 作为网络训练时的优化器，初始学习率为 0.001，每 20 轮迭代学习率衰减一半，批次大小设置为 16。同时，使用在线数据增强，包括随机裁切、缩放、旋转和颜色抖动等操作，按照 8∶2 划分训练集和测试集，关键点数量设置为 8，训练 240 轮。

3.1　评价指标

实验结果使用的评价指标为 2D Projection 指标和 ADD 指标[13]。其中，2D Projection 指标是如果物体模型点从网络估计的位姿到真实位姿的 2D 投影距离平均值小于指定像素（pixel）阈值，则认为姿势正确。ADD 指标计算在 3D 物体空间中的误差，如果预测的姿态和真实姿态之间的物体模型点的平均距离 ADD 小于指定倍数的模型直径 d，则认为预测正确，对于对称物体，使用最接近的模型点来计算平均距离 ADD-S。

3.2　实验结果分析

将本文所提方法和原 PVNet 方法分别使用筒子纱数据集进行训练和评价，实验对比结果见表 1、表 2。

从表 1 中可以看出，本文所提方法在 4 类筒子纱评估中，有 3 种不同阈值下的 2D Projection 精度均超过了原方法，尤其在 3 pixel 和 1 pixel 阈值下平均精度分别提升了 7.03% 和 12.17%，这表明本文所提方法在更高评价水平下准确率更高。由于使用基于改进的 ResNeSt50 的像素级投票网络能够更有效地提取弱纹理筒子纱的多分支特征，实现各通道间的信息交互。这保证了向量场的预测和与分割更准确，从而提高了筒子纱位姿估计的准确性。

表 2 中对比结果显示，本文所提方法相较于原 PVNet 方法在 ADD-S 指标下的精度得到提升，在 0.05d 和 0.02d 的阈值下，本文所提方法在 4 类筒子纱中的 ADD-S 指标均超过 PVNet 方法。尤其对于筒子纱 4 这类大尺寸且纹理较弱的筒子纱在 3 种阈值条件下的精度均有提升，这表明对大尺寸的弱纹理筒子纱进行位姿估计时具有更好的鲁棒性。

表 1　根据 2D Projection 指标比较筒子纱位姿估计结果

Table 1　Comparison of thread roll pose estimation results in terms of 2D Projection metric

训练数据集	PVNet			本文方法		
	1 pixel	3 pixel	5 pixel	1 pixel	3 pixel	5 pixel
筒子纱 1	24.45	94.75	99.56	46.72	98.69	99.56
筒子纱 2	10.79	93.36	99.59	17.42	95.85	100.00
筒子纱 3	9.68	87.90	97.18	21.37	94.35	98.79
筒子纱 4	3.40	74.71	95.40	11.49	89.98	98.28
平均值	12.08	87.68	97.93	24.25	94.72	99.16

表 2　根据 ADD-S 指标比较筒子纱位姿估计结果

Table 2　Comparison of thread roll pose estimation results in terms of ADD-S metric

训练数据集	PVNet			本文方法		
	0.02d	0.05d	0.10d	0.02d	0.05d	0.10d
筒子纱 1	45.41	93.01	100.00	64.63	99.13	100.00
筒子纱 2	54.77	95.43	100.00	61.82	97.10	100.00
筒子纱 3	44.35	84.27	100.00	46.77	89.51	100.00
筒子纱 4	34.48	73.56	97.70	50.00	91.38	100.00
平均值	44.75	86.57	99.43	55.81	94.28	100.00

3.3　遮挡实验

在分拣过程中会遇到遮挡的情况。为了测试所提方法在遮挡情况下的位姿估计效果，本实验将训练好的网络模型直接用来检测被部分遮挡的 4 类筒子纱，效果如图 9 所示。

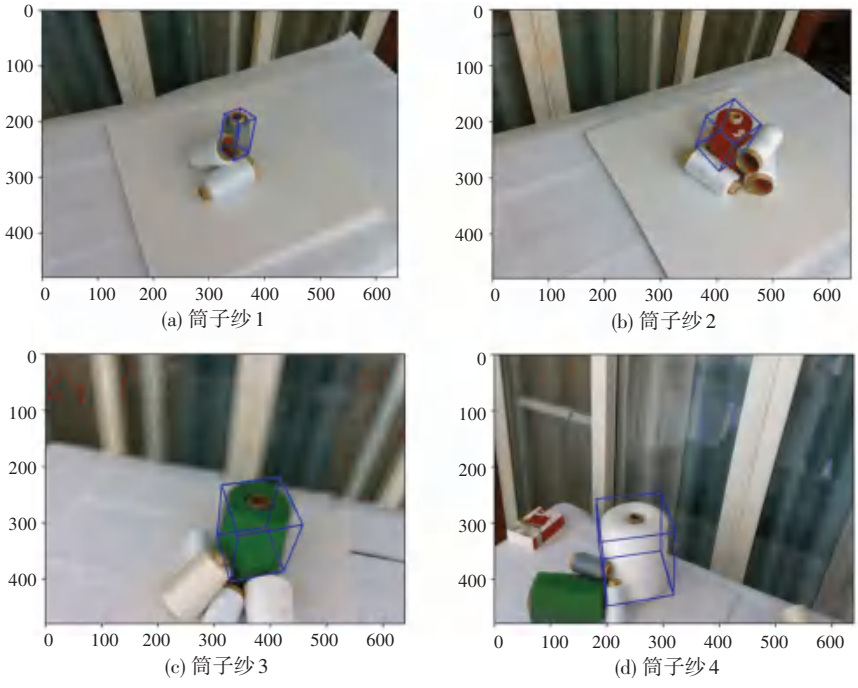

图 9 筒子纱截断场景测试可视化结果

Fig. 9 Visualization results of thread roll occlusion scene test

在面对筒子纱遮挡情况下,本文所提方法能够有效进行位姿估计,这是因为不可见部分的关键点检测可以通过可见部分的预测方向向量投票得到。此外,对于检测筒子纱 3 这种模糊和噪声多的图像也具有一定鲁棒性。

3.4 参数量与运行速度实验

分别输入 640×480(宽度×高度)的图像到所提网络和 PVNet 网络进行模型参数量和运行速度测试,实验结果见表 3。

表 3 在参数量与运行速度上对比

Table 3 Comparison in terms of parameters and running speed

对比方法	参数量/M	总耗时/ms	每秒帧数/FPS
PVNet	12.96	17.64	56
本文方法	3.92	15.17	65

本文所提方法在提升位姿估计精度同时,在网络参数量上仅为原网络的 30%,单张图像的总耗时减少了 2.47 ms,每秒帧数提高了 9。由此可知,本文所提网络在运行速度上更具有优势,能够满足在纺织过程中机械臂分拣筒子纱的实时性要求。同时由于参数量大幅降低,使得模型更容易部署在边缘计算设备上。

4 结束语

针对筒子纱分拣任务中弱纹理筒子纱 6D 位姿估计问题,本文基于改进的 ResNeSt50 的像素级投票网络,通过将多分支结构与通道注意力结合提升弱纹理筒子纱的特征提取质量,并提出单位向量误差的欧式距离损失和基于关键点到预测向量距离的正则化,用于网络训练。在构建的筒子纱数据集上进行实验。实验表明,本文所提方法能够提升弱纹理筒子纱的位姿精度。同时模型参数量大幅减少,模型的实时性有效提升,这使得筒子纱位姿估计网络能够更容易地部署在纺织车间中的边缘计算设备上。能够满足纺织工业中机械臂对筒子纱分拣,及上、下纱架等任务的需求。

参考文献

[1] 郭政良,马思乐,陈纪旸,等. 筒子纱包装自动整列系统的设计与实现[J]. 包装工程,2019,40(11):137-141.

[2] 任慧娟,金守峰,顾金芊. 基于颜色特征的筒纱分拣机器人识别定位方法[J]. 轻工机械,2020,38(4):58-63.

[3] ZHU Y, LI M, YAO W, et al. A review of 6D object poseestimation [C]//Proceedings of 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). Piscataway,NJ: IEEE, 2022:1647-1655.

[4] FAN Zhaoxin, ZHU Yazhi, HE Yulin, et al. Deep learning on monocular object pose detection and tracking: A comprehensive overview[J]. ACM Computing Surveys, 2022, 55(4):1-40.

[5] HOQUE S, ARAFAT M Y, XU S, et al. A comprehensive review on 3D object detection and 6D pose estimation with deeplearning[J]. IEEE Access, 2021, 9:143746-143770.

[6] ULRICH M, WIEDEMANN C, STEGER C. Combining scale-space and similarity - based aspect graphs for fast 3D

objectrecognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(10):1902-1914.

[7] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]//Proceedings of 2011 International Conference on Computer Vision. Piscataway, NJ: IEEE, 2011:2564-2571.

[8] ROTHGANGER F, LAZEBNIK S, SCHMID C, et al. 3D object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints[J]. International Journal of Computer Vision, 2006, 66(3):231-259.

[9] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.

[10] XIANG Yu, SCHMIDT T, NARAYANAN V, et al. PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes[J]. arXiv preprint arXiv,1711.00199,2017.

[11] LEPETIT V, MORENO-NOGUER F, FUA P. EPnP: an accurate $O(n)$ solution to the PnP problem[J]. International Journal of Computer Vision, 2009, 81(2):155-166.

[12] 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 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2017:3848-3856.

[13] PENG Sida, ZHOU Xiaowei, LIU Yuan, et al. PVNet: Pixel-wise voting network for 6DOF object pose estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6):3212-3223.

[14] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2016:770-778.

[15] ZHANG Yaoyin, WAN Lili, ZHU Yazhi, et al. VP-KLNet: Efficient 6D object pose estimation with an enhanced vector-field prediction network and a keypoint localization network [J]. Journal of Electronic Imaging, 2022, 31(5):17.

[16] ZHANG Hang, WU Chongruo, ZHANG Zhongyue, et al. ResNeSt: split-attention networks[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ: IEEE, 2022:2735-2745.

[17] ZHU Yazhi, WAN Lili, XU Wanru, et al. ASPP-DF-PVNet: Atrous spatial pyramid pooling and distance-filtered PVNet for occlusion resistant 6D object pose estimation [J]. Signal Processing: Image Communication, 2021, 95:116268.

[18] XIE Saining, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2017: 5987-5995.

[19] LI Xiang, WANG Wenhai, HU Xiaolin, et al. Selective kernel networks[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2019:510-519.