文章编号　1004-924X(2024)18-2792-11

基于空间信息聚合的遮挡目标抓取位姿检测

陈仁祥[1*]，邱天然[1]，杨黎霞[2]，张芷僮[1]，夏　亮[3]

(1. 重庆交通大学 交通工程应用机器人重庆市工程实验室，重庆 400074；
2. 重庆科技大学 工商管理学院，重庆 401331；
3. 重庆智能机器人研究院，重庆 400714)

摘要：针对机器人依靠视觉抓取时对遮挡目标抓取位姿检测准确率低的问题，提出基于空间信息聚合的遮挡目标抓取位姿检测方法。遮挡导致目标在相机视野中的本征特征改变，影响目标位置信息与形状结构特征。首先，使用坐标卷积代替传统卷积的方式进行特征提取，在输入特征图后新增坐标通道来提升网络对位置信息感知能力；其次，设计空间信息聚合模块，其采用并行结构增大局部感受野并沿空间方向对通道进行编码获取多尺度空间信息，再通过非线性拟合方式将信息聚合，使模型更好理解目标结构和形状；最后，抓取位姿检测网络输出抓取的质量、角度和宽度，并计算最佳抓取位置以建立最优抓取矩形框。在 Cornell Grasping 数据集、自建遮挡数据集、Jacquard 数据集验证，检测准确率分别达到 98.9%，94.7%，96.0%，在实验平台对目标的 100 次真实抓取实验中，成功率为 93%。所提方法在三个数据集上均取得了最高检测准确率，且在实际场景中检测效果更优。

关　键　词：抓取位姿检测；遮挡目标；空间信息聚合；坐标卷积

中图分类号：TP242　　**文献标识码**：A　　**doi**：10.37188/OPE.20243218.2792

Occluded target grasping detection method based on spatial information aggregation

CHEN Renxiang[1*], QIU Tianran[1], YANG Lixia[2], ZHANG Zhitong[1], XIA Liang[3]

(1. *Chongqing Jiaotong University, Chongqing Engineering Laboratory of Traffic Engineering Application Robot, Chongqing* 400074, *China*;
2. *School of Business Administration, Chongqing University of Science and Technology, Chongqing* 401331, *China*;
3. *Chongqing Intelligent Robot Research Institute, Chongqing* 4000714, *China*)
* *Corresponding author, E-mail: manlou. yue*@126. *com*

Abstract: Aiming at the problem of low accuracy of occlusion target grasping position detection when the robot relies on vision grasping, we proposed an occlusion target grasping position detection method based on spatial information aggregation. Occlusion led to the change of the target′s intrinsic features in the cam-

收稿日期：2024-05-16；**修订日期**：2024-07-22.

基金项目：国家自然科学基金资助项目(No.52475548)；重庆市教育委员会科学技术研究项目资助(No. KJZD-M202200701)；重庆市自然科学基金创新发展联合基金资助项目(No. CSTB2023NSCQ-LZX0127)；重庆市研究生联合培养基地项目资助(No. JDLH-PYJD2021007)；重庆市专业学位研究生教学案例库资助(No. JDALK2022007)

era′s field of view, which affected the target′s positional information and shape-structural features. First, coordinate convolution was used instead of traditional convolution for feature extraction, and a new coordinate channel was added after the input feature map to improve the network′s ability to perceive position information. Second, the spatial information aggregation module was designed, which adopted a parallel structure to increase the local sensing field and encoded the channels along the spatial direction to obtain multi-scale spatial information, and then aggregated the information through nonlinear fitting to make the model better understand the target structure and shape. Finally, the grasping position detection network outputted the grasping mass, angle and width, and calculated the optimal grasping position to establish the optimal grasping rectangular box. Validated on the Cornell Grasping dataset, the self-constructed occlusion dataset, and the Jacquard dataset, the detection accuracies reach 98.9%, 94.7%, and 96.0%, respectively, and the success rate is 93% in 100 real grasping experiments on the target in the experimental platform. The proposed method achieves the highest detection accuracy on all three datasets, and the detection effect is better in real scenes.

Key words: grasping detection; occluded target; spatial information aggregation module; CoordConv

1　引　言

依靠视觉进行目标抓取是机器人最重要的能力之一，抓取位姿检测指机器人通过视觉传感器获取目标信息，检测出目标物体抓取点与抓取姿势[1]。随着社会对自动化和智能化要求的不断提高，机器人的工作环境越来越复杂，抓取目标存在遮挡将导致目标抓取位姿检测准确率下降，影响机器人工作效率与安全性。因此，提升机器人对遮挡目标的抓取成功率具有重要的实际应用价值和研究意义。

传统方法使用目标3D模型或物理模型作为特征信息计算抓取参数，但对于未知目标或被遮挡目标难以适应[2]。Fang等[3]基于点云和深度图像进行六自由度(6DoF)抓取位姿检测，使用PointNet对不同物体进行分割。Yang等[4]提出基于3D重建的抓取位姿检测方法，包括形状重建网络和抓取建议网络。基于几何分析的6DoF抓取方法需完整3D点云信息，但实际工作环境中，视觉传感器较难完全获取抓取物体的3D模型，影响抓取效果。

平面抓取方法更能快速准确地完成抓取位姿检测，Lenz等[5]提出五维抓握表示方法，其中抓握位置由抓握矩形表示。Kumra等[6]提出生成残差卷积神经网络模型为未知物体生成抓取框，考虑了复杂场景下的抓取位姿检测。Yu等[7]利用具有通道注意力残差块开发检测网络，从RGB-D图像生成抓取姿势。Wang等[8]应用交叉窗口注意力来建模遥远像素间的长期依赖关系进行抓取，在复杂场景中进行了多目标抓取实验，但其模型参数量较大。徐胜军等[9]利用深度可分离卷积层、上下文聚合策略优化构造网络，在单目标抓取位姿检测时准确率与检测效率较高，但未考虑实际场景中的遮挡情况。当目标存在遮挡时，目标在相机视野中的目标位置信息和形状结构特征将改变[10]，导致抓取位姿检测网络特征提取不准确，影响模型抓取位姿检测准确率。

综上所述，针对机器人依靠视觉抓取时对遮挡目标抓取位姿检测准确率低的问题，提出基于空间信息聚合的遮挡目标抓取位姿检测方法。使用坐标卷积代替传统卷积的方式进行特征提取，在输入特征图后新增坐标通道来提升网络位置信息感知能力，并设计空间信息聚合模块采用并行结构增大局部感受野获取多尺度空间信息，同时沿空间方向独立处理每个通道信息并将其聚合，使模型更好理解目标结构和形状，从而提高遮挡目标位姿检测准确率。最后，通过实验验证了所提方法的可行性与有效性。

2　空间信息聚合的抓取位姿检测模型

2.1　抓取位姿检测表示方法

末端执行器与物体位置及角度关系表

示为[5]：

$$g=(x,y,\theta,h,w). \tag{1}$$

以抓取矩形框模拟二指平板末端执行装置，其中(x,y)表示抓取矩形框的中间点，θ表示抓取框相对于水平轴的夹角。h表示矩形框长度，w表示矩形框宽度，抓取位置矩形框表示如图1所示。

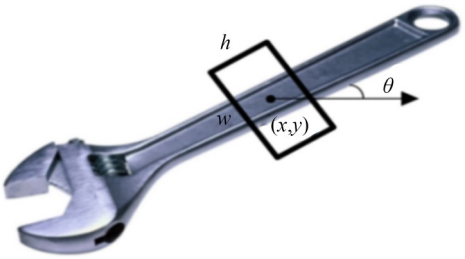


图1　矩形抓取框

Fig. 1　Rectangle grab box

2.2　空间信息聚合网络

所提出面向遮挡目标抓取位姿检测的空间信息聚合网络(Spatial Information Aggregation Networks, SIANET)如图2所示。SIANET是以残差卷积结构为基础[6]，针对遮挡目标抓取位姿检测的难点建立新结构所形成的网络，主要由特征提取、空间信息聚合模块、图像重建和输出结果四部分构成。特征提取过程中，遮挡导致目标位置信息受到干扰，在输入图像经过三个坐标卷积层进行特征提取，在输入特征图后新增坐标通道来提升网络位置信息感知能力。然后经过空间信息聚合模块，该模块采用并行结构，可以同时获取和处理来自不同分支的多尺度信息，从而增大了网络的感受野，使网络获取更全面、更丰富的信息。遮挡导致目标在相机视野中的形状结构特征改变，模块通过编码全局信息重新校准每个并行分支中的通道权重，以获取更精确的目标信息，有助于模型更好地理解目标的结构和形状。为避免因模型层数多导致的过拟合与梯度消失等问题，在残差层中使用跳跃连接使模型融合底层特征，从而更好地实现抓取特征融合。最后通过卷积转置运算对图像进行上采样，并输出抓取质量Q、抓取角度angle和抓取宽度width以及抓取矩形框。抓取质量取值为[0,1]，值越接近1表示抓取质量越高，抓取位置是抓取质量图像中具有最高抓取质量的位置。为保证网络训练中角度值的连续性，将角度值映射为三角函数，抓取角度angle由式(2)表示：

$$\theta_i=\frac{1}{2}\arctan\frac{\sin 2\theta}{\cos 2\theta}, \tag{2}$$

其中：θ_i表示最终抓取角度angle，θ表示抓取角度标量值，抓取宽度width表示抓取所需宽度。

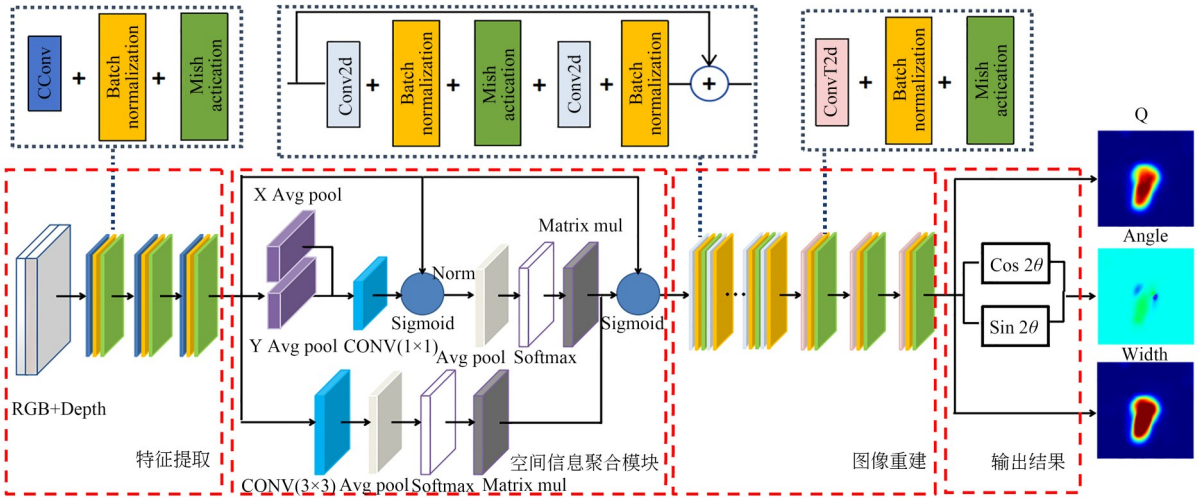


图2　模型结构

Fig. 2　Model structure

2.3　坐标卷积

在遮挡条件下，目标物体的部分特征可能被遮挡，传统卷积层在卷积核内部权重相同，只关注像素特征[11]，难以关注到目标间的位置关系，

而坐标卷积（CoordConv，CCONV）通过在输入特征图后新增坐标通道来提升网络位置感知能力。新增通道分别表示输入的X和Y坐标，从而使卷积过程可感知特征图的空间信息。使用坐标卷积进行特征提取时，能够提供更多位置信息，使网络能更好地学习图像中的位置关系。坐标卷积的计算公式为：

$$\boldsymbol{Y}_{i,j}=\sum_{k,l}\boldsymbol{X}_{k,l}\cdot\boldsymbol{K}_{i-k,j-l}\cdot P(k,l), \quad (3)$$

其中：$\boldsymbol{X}$表示输入特征图，$\boldsymbol{Y}$表示输出特征图，i和j表示输出特征图的坐标，k和l是卷积核的坐标，$P(k,l)$表示与输入数据的坐标位置相关的可学习参数，用于调整卷积核在不同位置的权重，能通过网络的训练过程学习得到，从而更好地适应输入数据的结构。坐标信息与原输入特征一起输入到卷积操作中，坐标卷积可表征特征图像素点的坐标，让卷积过程能一定程度感知目标位置来辅助检测并加强边界信息，有利于抓取位姿检测任务。

2.4　空间信息聚合模块

遮挡会导致目标在相机视野中的形状与结构产生变化，影响抓取位姿检测结果。因此，设计空间信息聚合模块（见图2），使用并行结构增大网络感受野，获得来自不同分支的多尺度信息和更精确的目标信息，并对通道间信息编码，将空间结构信息保留到通道中，以更好理解目标的结构和形状，提高检测准确率。并行结构也有助于避免网络过深的问题[12]。首先将输入特征划分为数个子特征用于学习不同语义：

$$\boldsymbol{X}=(\boldsymbol{X}_0,\boldsymbol{X}_i,\cdots,\boldsymbol{X}_{a-1}),\boldsymbol{X}_i\in\boldsymbol{R}^{C//\mathrm{a}\times H\times W}, \quad (4)$$

其中：$\boldsymbol{X}$表示输入特征，a表示划分子特征的数量，$\boldsymbol{C}$表示输入特征的通道数，H表示输入特征的高度，W表示输入特征的宽度。这样划分子特征可使网络学习不同语义信息，有助于区分物体的不同部分和形状，帮助网络检测出被遮挡物体的重要部分。

建立并行结构，使用X Avg pool与Y Avg pool全局平均池化操作分别沿着两个空间方向对通道进行编码，然后将两个编码的特征连接，再通过1×1卷积进行处理，并使用非线性Sigmoid函数拟合线性卷积上的2D二项分布，作为1×1分支。该分支沿空间方向对通道进行编码，有助于捕捉通道间的特征关系，提高模型对不同物体特征的感知能力，并降低遮挡带来的影响。

同时，使用单个3×3的卷积对输入特征图进行运算，作为3×3分支。目的是增大网络局部感受野，扩大特征空间，获取多尺度特征信息。最后，通过非线性拟合方式将获取的信息聚合，帮助网络更好理解遮挡目标结构和形状，从而提高遮挡目标位姿检测准确率。

2.5　遮挡目标抓取位姿检测流程

所提出空间信息聚合的遮挡目标抓取位姿检测流程如图3。主要步骤如下：（1）构建遮挡数据集并划分数据集；（2）构建SIANET抓取位姿检测网络并初始化参数；（3）将训练集与验证集数据输入到SIANET网络中，通过计算损失等方式对网络进行训练，更新网络参数并保存最终模型；（4）调用训练完成的模型，对目标进行抓取位姿检测；（5）通过预先进行的手眼标定工作，可以计算出机器人目标抓取位姿，实现目标抓取。

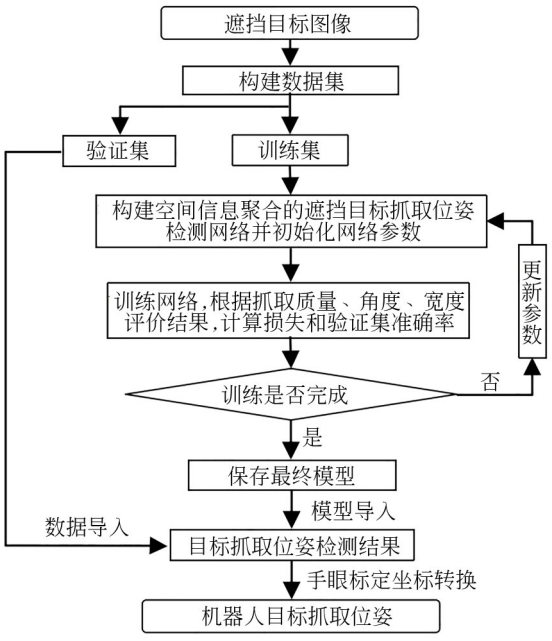


图3　抓取位姿检测流程

Fig. 3　Grasp pose detection process

3　实验与结果分析

训练环境所在平台为Ubuntu18.04，在

NVIDIA 1080ti上对网络进行50个Epochs，每个Epoch进行2 000次迭代的端到端训练。在训练过程中使用Adam（Adaptive Moment Estimation）优化算法，批量设定为8，初始学习率设置为0.001。

以抓取质量Q值最高的像素点作为图像中最佳抓取中心点位置坐标，然后计算对应角度与抓取宽度，并输出抓取矩形框。预测抓取框与真实抓取框参数在以下情况认为检测成功：(1)预测抓取矩形框与真实抓取矩形框之间的角度差小于30°；(2)预测抓取矩形框和真实抓取矩形框的Jaccard指数大于0.25，Jaccard指数可表示为：

$$\text{Jaccard}(A,B)=\frac{|A\cap B|}{|A\cup B|},\tag{5}$$

其中：A为预测抓取框面积，B为标定抓取框面积。$A\cap B$和$A\cup B$分别为两抓取框面积交集与并集。

3.1 数据集介绍

Cornell Grasp数据集[5]：该数据集被广泛用于训练和测试各种机器人抓取位姿检测算法，包含了从各个视角拍摄的240种日常物体的1 035幅图像，由彩色图像、深度信息、检测框标签等数据构成。Cornell Grasp数据集图片数量过小，使用旋转、裁剪和缩放等方式进行数据增强，按照9∶1的比例划分为训练集与验证集。

自建数据集：为评价分析SIANET对遮挡目标的检测性能，建立遮挡目标自建数据集进行实验。该自建数据集共有2 165幅图像，分为两个部分，第一部分数据为自摄数据，共205幅图像。搭建实验验证平台，使用平台配有的RealSense D435i深度相机对遮挡图像进行采集，实验平台与采集到的图像如图4所示，利用roLablImg软件对采集到的图片进行标注。

为保证数据多样性从而验证模型性能，第二部分数据以Cornell Grasping数据集为基础，首先将数据集中的图像添加干扰物品图片产生遮挡，其次根据干扰物品的添加位置对深度信息进行相应修改，最后对新生成的遮挡图片并进行重新标注，以保证与真实遮挡目标的数据相似性，人为将遮挡目标的被遮挡面积控制在10%～60%，图5为遮挡目标图片添加前后对比。

图4　实验平台与遮挡图像

Fig. 4　Experimental platform and occluded images

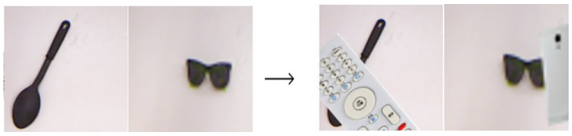


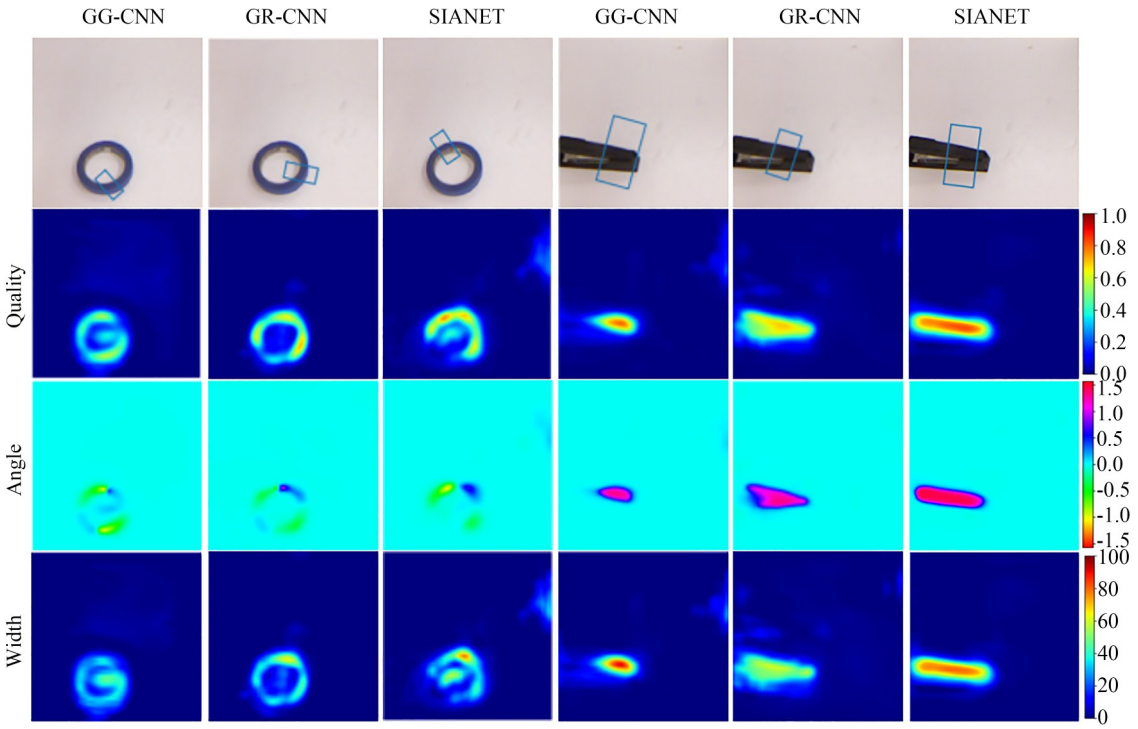
图5　遮挡图片前后对比

Fig. 5　Comparison before and after occlusion of images

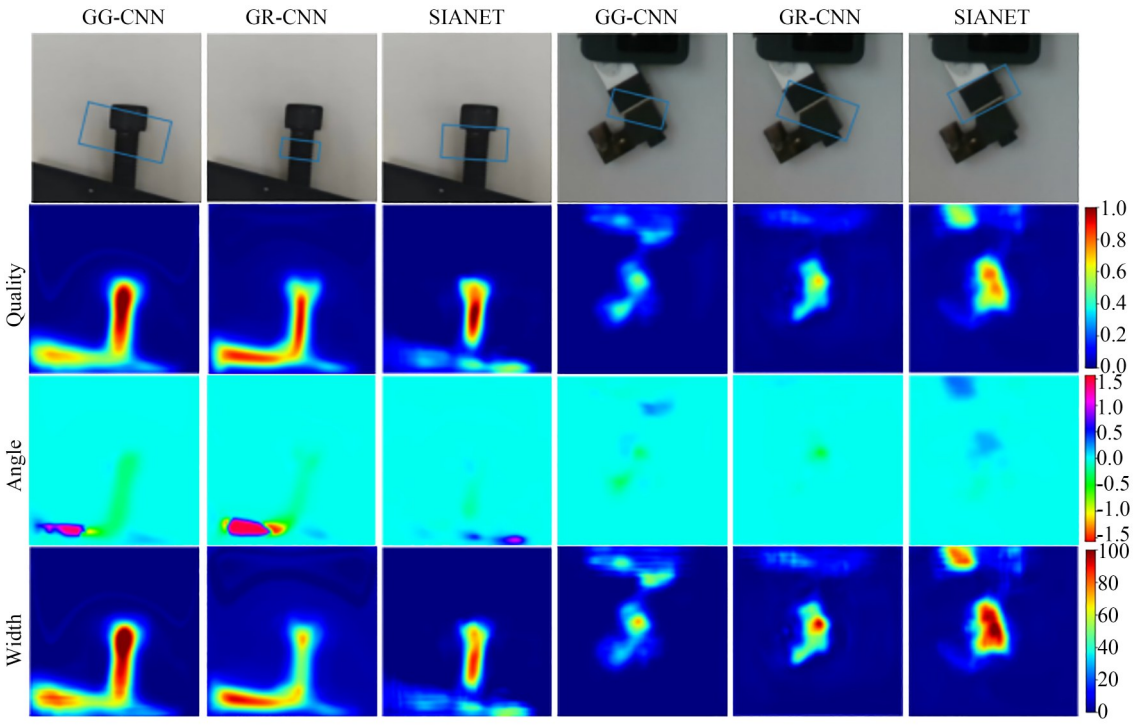
3.2 检测结果对比

为评价分析SIANET网络的性能，分别进行无遮挡目标（Cornell Grasp数据集）和有遮挡目标（自建数据集）实验。使用GR-CNN[6]，GGCNN[13]，SIANET在Cornell grasp数据集与自建数据集进行对比实验，如图6(a)和图6(b)所示。第一行表示网络输出的抓取位置，第二行至第四行表示抓取质量、角度和宽度图。

具有最高质量分数的像素将被作为抓取点。未存在目标遮挡时结果如图6(a)所示，根据该图，在对胶带进行检测时，GGCNN对抓取质量较高的位置检测较为准确，但对抓取角度的预测误差较大，而SIANET比GR-CNN有着更优的抓取框输出结果。在对订书机进行检测时与GR-CNN和GGCNN相比，SIANET目标位置信息与形状结构特征感知能力较强，目标质量评分较高区域形状与目标真实形状吻合，并得到最优抓取框结果。综合评价SIANET具有更优的检测效果。



(a) Cornell grasp数据集
(a) Cornell grasp data set



(b) 自建数据集
(b) Our data set

图 6　检测结果

Fig. 6　Detection results

根据图6(b),对于遮挡目标GGCNN难以辨别目标位置,从而导致抓取框预测位置不合理,这是因为GGCNN使用的是全卷积网络,且网络较简单且深度不足,在处理复杂的遮挡物体时,无法提取出足够丰富的特征。在目标被遮挡时,由于GR-CNN主要关注点是生成抓取位姿,没有充分考虑到遮挡物体的特殊性,在特征提取阶段使用普通卷积,它不会关注像素之间的位置关系,只关注像素的特征,使其在对抓取目标进行检测时,抓取质量分数较高的区域与目标实际形状差距较大,且抓取质量评分较低。SIANET具有最高的抓取质量分数,更优的抓取角度、抓取宽度与更合理的抓取矩形框结果,且抓取质量评分较高区域与目标实际可抓取部分形状基本一致,检测效果最好。

3.3 检测准确率对比

两种数据集均使用随机分割方式对数据集划分。将SIANET分别与GGCNN[13],SAE[5],GR-CNN[6],SE-ResUNet[7],TF-grasp[8]网络进行检测准确率对比实验,对比结果表1所示。

表1 检测准确率对比实验

Tab. 1 Comparison experiment of detection accuracy

模型	无遮挡Cornell Grasp准确率/%	有遮挡自建数据集准确率/%	FPS(frame/s)
GGCNN[13]	73.0	64.3	25
SAE[5]	73.9	65.5	0.6
GR-CNN[6]	97.7	90.2	25
TF-grasp[8]	97.9	91.1	13
SE-ResUNet[7]	98.2	90.9	27
SIANET	**98.9**	**94.7**	**30**

可以看出所提方法在无遮挡与有遮挡时均取得了较高准确率,分别为98.9%与94.7%,表现优于其他对比方法,且在有遮挡的数据集相比其他方法表现更优,在检测效率方面,SIANET的FPS(frame/s)最高,具有较优的检测效率。观察实验结果可以看出,对于遮挡目标,对比算法的性能均出现不同程度的下降,表明遮挡对抓取特征的有效提取产生负面影响。所提方法通过使用坐标卷积进行特征提取,提升网络位置信息感知能力,并设计空间信息聚合模块,建立不同分支,在增大感受野的同时沿空间方向对通道进行编码,有助于模型更好理解目标结构和形状。实验结果再次证明所提方法在遮挡目标位姿检测中的有效性。

3.4 消融实验与过程性证明

为分析各模块对遮挡目标抓取位姿检测效果的影响,对模型进行消融实验。消融实验结果如表2所示。

观察该表,相较于残差卷积结构,添加坐标卷积进行特征提取后,模型在输入特征图后新增坐标通道来提升网络位置信息感知能力,准确率提升1.7%。添加空间信息聚合模块后,模型使用大的局部感受野使网络收集多尺度空间信息,有助于模型更好理解目标的结构和形状,扩大特征空间,使准确率提升2.9%。在坐标卷积特征提取与空间信息聚合模块同时施加在网络后,准确率达到94.7%,说明坐标卷积与空间信息聚合模块对网络同时产生辅助作用。图7为网络添加模块前后过程性证明对比。

表2 自建数据集消融实验

Tab. 2 Dissolution experiment of the self-built dataset

网络结构	准确率/%
残差卷积结构	90.2
残差卷积结构+坐标卷积特征提取	91.9
残差卷积结构+空间信息聚合模块	93.1
SIANET	**94.7**

具有最高质量分数的像素将被作为抓取点,质量分数评分越高,代表这个位置抓取成功的可能性越大。观察图7,在未使用坐标卷积与空间信息聚合模块时,残差卷积结构质量分数较低,网络难以理解目标形状与位置;使用坐标卷积进行特征提取后,质量分数轻微下降,但网络能更清楚地感知目标位置,质量分数较高区域与实际

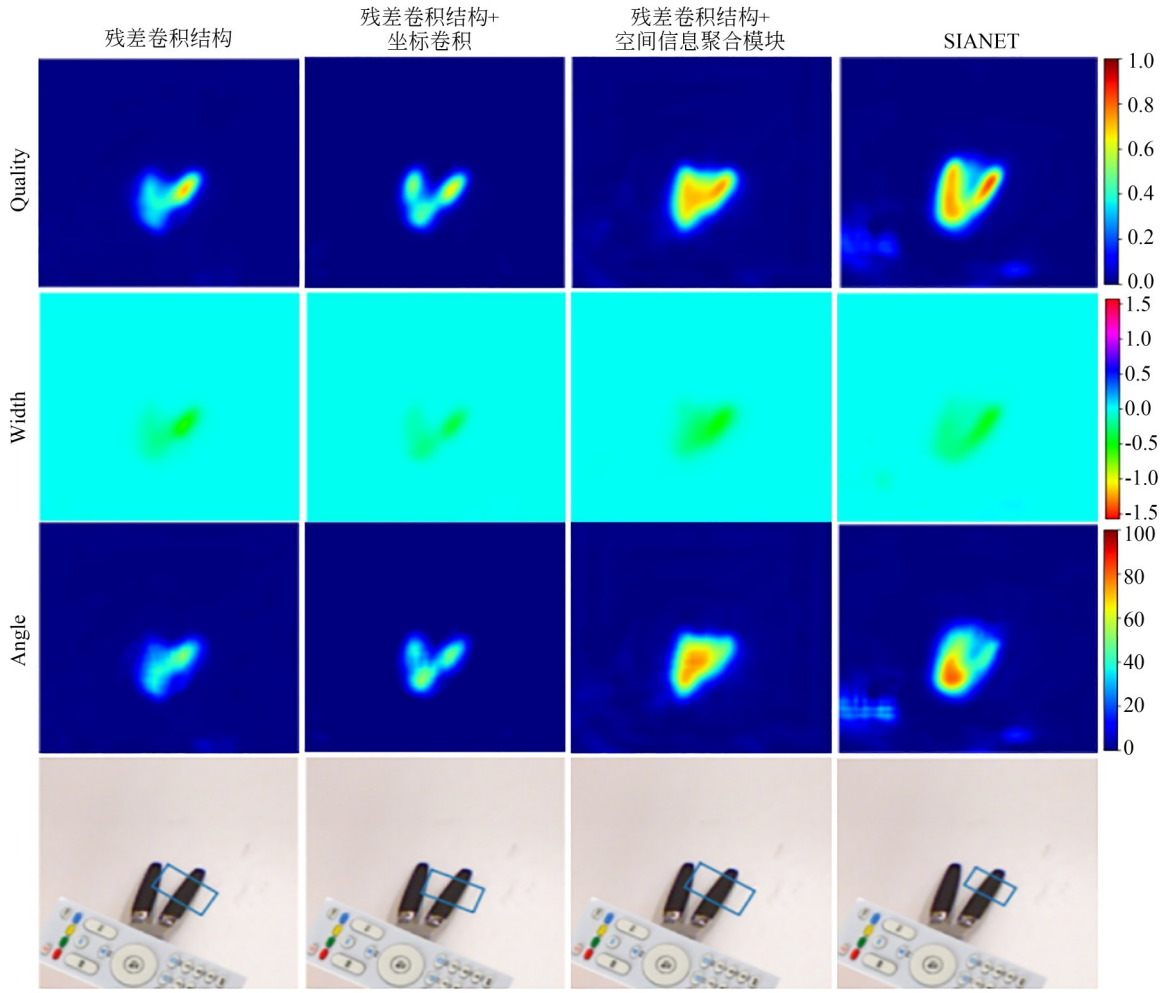


图 7　过程性证明实验

Fig. 7　Results of dissolution experiment

成功执行抓取位置更相符；加入构建的空间信息聚合模块后，质量评分较高区域形状与目标真实形状更加吻合，说明空间信息聚合模块有助于模型更好理解目标结构和形状；使用坐标卷积进行特征提取并加入空间信息聚合模块后，抓取质量分数最高，评分较高区域与目标真实形状最接近，并预测了更合理的抓取位置，说明 SIANET 同时发挥坐标卷积和空间信息聚合模块的作用，可有效改善目标在相机视野中本征特征发生改变导致检测准确率下降的问题。

3.5　Jacquard 数据集

为研究所提方法在其他数据集的适应能力，在 Jacquard 数据集[14]上进行对比实验，图 8 为 Jacquard 数据集中的图像。

Jacquard 数据集是一个大型的抓取数据集，专门用于机器人抓取任务的研究和评估，它基于

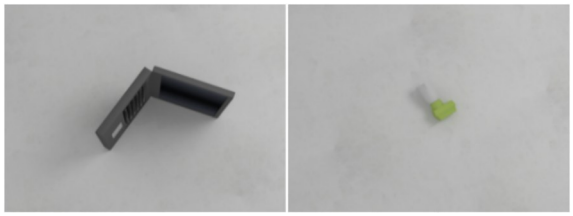


图 8　Jacquard 数据集图像

Fig. 8　Jacquard dataset images

CAD 模型创建，包含了超过 11 000 个不同物体的抓取数据，每个物体都有多个抓取姿态。在本实验中，使用在 Jacquard 数据集上经过验证且效果较优的几种方法与所提方法进行对比。选择对比结果表 3 所示。可以看出，在 Jacquard 数据集中，所提方法准确率达到 96.0%，网络总体性能更优，证明了 SIANET 模型在不同数据集上的有效性。

表 3　Jacquard 数据集对比实验

Tab. 3　Comparison experiment of Jacquard dataset

网络结构	准确率/%
GGCNN[13]	84.0
GR-CNN[6]	94.6
TF-grasp[8]	94.6
SE-ResUNet[7]	95.7
SIANET	96.0

3.6　实际场景抓取位姿检测

为验证真实场景下抓取位姿检测的准确性，搭建实验验证平台，执行装置选用 Robotiq 双指平行电动夹爪，使用 Intel RealSense D435i 深度相机作为图像传感器获得真实世界实验场景图像，采集待抓取目标 RGB 信息及深度信息[15]，将采集到的图像分别使用对比方法与 SIANET 方法进行抓取位姿检测，在图 9 中对检测结果进行展示。

观察实验结果可以看出，对比方法在进行存在遮挡的多目标抓取位姿检测时，抓取宽度和抓取角度都出现一定程度的偏差，且部分方法存在漏检现象。其中，SE-ResUNet 方法成功输出所有目标的抓取位姿，没有发生漏检情况，但抓取位置出现偏差。SIANET 在检测时，抓取角度与抓取宽度合理，输出抓取框效果较好，表明所提网络对于遮挡目标抓取位姿检测更有优势。可以看出，在第二组多目标抓取位姿检测实验中，GGCNN 与 TF-grasp 只对两个目标进行了检测，由于被遮挡的螺丝刀抓取评分低，导致未生成抓取检测框。其余对比方法对三个目标都生成了抓取框，但部分抓取框的抓取角度与位置不合理，SIANET 对视野内的所有目标进行了抓取位姿检测。

3.7　实际场景抓取实验

抓取实验中，需对机器人进行手眼标定，本次实验中采用眼在手上的标定方式。抓取物体、相机与机械臂之间的空间关系可表示为：

$$T_a^b = T_h^b \cdot T_c^h \cdot T_a^c. \tag{6}$$

机器人坐标系到标定板坐标系的转换矩阵记作 T_a^b。将机器人坐标系到夹爪坐标系的转换矩阵记作 T_h^b，该参数可以由机械臂控制端读取数据求出。将夹爪坐标系到相机坐标系的转换矩阵记作 T_c^h，该参数在标定过程中不会发生改变，需要在标定后被求出。将相机坐标系到标定板坐标系的转换矩阵记作 T_a^c。

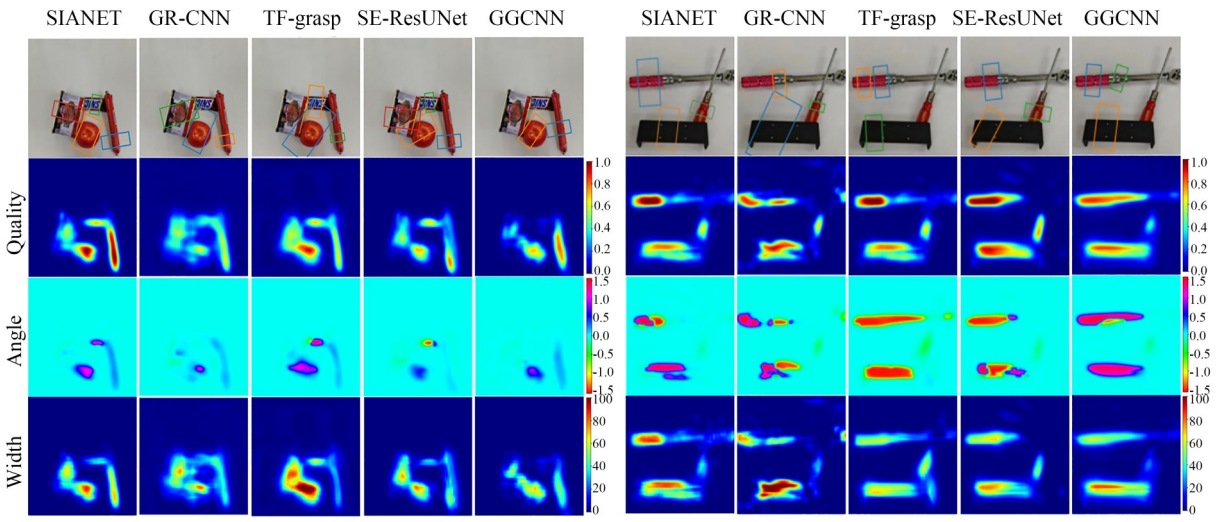


图 9　实际检测结果

Fig. 9　Actual test resultseffects

通过手眼标定与计算，可以将图像中检测到的抓取框转换成机械臂需要到达的位置表示[16]：

$$Z^C\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = K\begin{bmatrix} \boldsymbol{R} & T \\ 0 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{7}$$

其中：Z^C表示抓取目标在相机坐标系下距离相机中心的距离，可由深度信息得到，(x,y)表示像素坐标系中抓取中心点的坐标。K为相机内参，$\boldsymbol{R}$，T表示由手眼标定求出的相机坐标系与机械臂坐标系之间的位姿转换矩阵。X,Y,Z表示抓取位置在机械臂坐标系下的坐标位置。通过式(6)和式(7)可推导出机械臂抓取位置的坐标以及抓取角度和宽度。从而实现抓取功能。实验步骤为：(1)深度相机拍摄图片作为模型输入；(2)模型输出待抓取物体的抓取位置；(3)执行抓取动作。机械臂实际抓取过程如图10所示。

图10　实际抓取过程

Fig. 10　Actual grabbing process

在抓取实验验证中选择工业零件、螺丝刀、钳子、牙膏等目标作为测试对象。实验验证过程中，每次使用三种不同类型目标随机以不同的角度堆放在纯白色底板上，总共进行100次实际抓取，夹爪能将目标夹起不掉落，则认为抓取成功，实验结果如表4所示。SIANET的平均抓取成功率为93%，在实际抓取实验中表现最佳。

表4　实际抓取实验结果

Tab. 4　Actual capture of experimental results

模型	抓取成功率/%
GGCNN[13]	85
TF-grasp[8]	88
GR-CNN[6]	89
SE-ResUNet[7]	90
SIANET	93

4　结　论

针对机器人依靠视觉抓取时对遮挡目标抓取位姿检测准确率低的问题，提出基于空间信息聚合的遮挡目标抓取位姿检测方法，所提方法的优势在于：

(1)使用坐标卷积代替传统卷积的方式进行特征提取，在输入特征图后新增坐标通道提升了网络位置信息感知能力。设计空间信息聚合模块采用并行结构增大局部感受野获取多尺度空间信息，使模型更好地理解了目标结构和形状。

(2)所提方法在Cornell Grasping数据集、自建遮挡数据集、Jacquard数据集表现优于其他常见方法，检测准确率分别达到98.9%，94.7%，96.0%，且在实际遮挡场景下检测效果更好。证明了所提方法的有效性。

参考文献：

[1] 吴培良，刘瑞军，李瑶，等. 一种基于生成对抗网络与模型泛化的机器人推抓技能学习方法[J]. 仪器仪表学报，2022，43(5)：244-253.
WU P L，LIU R J，LI Y，*et al*. Robot pushing and grasping skill learning method based on generative adversarial network and model generalization [J]. *Chinese Journal of Scientific Instrument*，2022，43(5)：244-253.(in Chinese)

[2] 李文浩，谭宇飞，周登科，等. 融合多光学传感及变积分PID算法的搬运机器人[J]. 光学 精密工程，2023，31(23)：3504-3516.
LI W H，TAN Y F，ZHOU D K，*et al*. Transport robot integrating multi channel optical sensing and variable integral pid algorithm [J]. *Opt. Precision Eng.*，2023，31(23)：3504-3516.(in Chinese)

[3] FANG H S，WANG C X，GOU M H，*et al*. GraspNet-1Billion：a large-scale benchmark for general object grasping[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. *Seattle*，*WA*，*USA*. IEEE，2020：11441-11450.

[4] YANG D，TOSUN T，EISNER B，*et al*. Robotic grasping through combined image-based grasp proposal and 3D reconstruction[C]. 2021 *IEEE International Conference on Robotics and Automation (ICRA)*. *Xi'an*，*China*. IEEE，2021：6350-6356.

[5] LENZ I，LEE H，SAXENA A. Deep learning for detecting robotic grasps[J]. *The International Journal of Robotics Research*，2015，34(4/5)：705-724.

[6] KUMRA S, JOSHI S, SAHIN F. Antipodal robotic grasping using generative residual convolutional neural network[C]. 2020 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*). *Las Vegas*, *NV*, *USA*. IEEE, 2020: 9626-9633.

[7] YU S, ZHAI D H, XIA Y Q, *et al*. SE-ResUNet: a novel robotic grasp detection method[J]. *IEEE Robotics and Automation Letters*, 2022, 7(2): 5238-5245.

[8] WANG S C, ZHOU Z L, KAN Z. When transformer meets robotic grasping: exploits context for efficient grasp detection[J]. *IEEE Robotics and Automation Letters*, 2022, 7(3): 8170-8177.

[9] 徐胜军, 任君琳, 刘光辉, 等. 基于上下文聚合策略的轻量级编/解码抓取位姿检测[J]. 机器人, 2023, 45(6): 641-654.
XU S J, REN J L, LIU G H, *et al*. Lightweight encoding-decoding grasp pose detection based on a context aggregation strategy[J]. *Robot*, 2023, 45(6): 641-654. (in Chinese)

[10] 李明, 鹿朋, 朱龙, 等. 基于RGB-D融合的密集遮挡抓取检测[J]. 控制与决策, 2023, 38(10): 2867-2874, 中插16-中插19.
LI M, LU P, ZHU L, *et al*. Densely occluded grasping objects detection based on RGB-D fusion[J]. *Control and Decision*, 2023, 38(10): 2867-2874, 中插16-中插19. (in Chinese)

[11] 牛小明, 曾理, 杨飞, 等. 零部件光学影像精准定位的轻量化深度学习网络[J]. 光学 精密工程, 2023, 31(17): 2611-2625.
NIU X M, ZENG L, YANG F, *et al*. Lightweight deep learning network for accurate localization of optical image components[J]. *Opt. Precision Eng.*, 2023, 31(17): 2611-2625. (in Chinese)

[12] OUYANG D L, HE S, ZHANG G Z, *et al*. Efficient multi-scale attention module with cross-spatial learning[C]. *ICASSP* 2023 - 2023 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*). *Rhodes Island*, *Greece*. IEEE, 2023: 1-5.

[13] MORRISON D, CORKE P, LEITNER J. Learning robust, real-time, reactive robotic grasping[J]. *The International Journal of Robotics Research*, 2020, 39(2/3): 183-201.

[14] DEPIERRE A, DELLANDRÉA E, CHEN L M. Jacquard: a large scale dataset for robotic grasp detection[C]. 2018 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*). *Madrid*, *Spain*. IEEE, 2018: 3511-3516.

[15] 程德强, 张华强, 寇旗旗, 等. 基于层级特征融合的室内自监督单目深度估计[J]. 光学 精密工程, 2023, 31(20): 2993-3009.
CHENG D Q, ZHANG H Q, KOU Q Q, *et al*. Indoor self-supervised monocular depth estimation based on level feature fusion[J]. *Opt. Precision Eng.*, 2023, 31(20): 2993-3009. (in Chinese)

[16] 楚红雨, 冷齐齐, 张晓强, 等. 融入注意力机制的多模特征机械臂抓取位姿检测[J]. 控制与决策, 2024, 39(3): 777-785.
CHU H Y, LENG Q Q, ZHANG X Q, *et al*. Multi-modal feature robotic arm grasping pose detection with attention mechanism[J]. *Control and Decision*, 2024, 39(3): 777-785. (in Chinese)

作者简介：

陈仁祥(1983—),男,四川盐源人,博士,教授,博士生导师,2007年、2012年于重庆大学分别获得学士学位和博士学位,现为重庆交通大学机电与车辆工程学院副院长,主要从事机器视觉、智能测试技术与信号处理方面的研究。E-mail: manlou.yue@126.com

邱天然(1996—),男,贵州贵阳人,硕士研究生,2018年于杭州电子科技大学获得学士学位,主要从事计算机视觉与机器人控制方法研究。E-mail: 979895217@qq.com