

DOI: 10.13382/j.jemi.B2407448

基于自适应特征融合的抓取检测方法研究*

熊 煥^{1,2} 俞建峰^{1,2} 钱陈豪^{1,2} 蒋 毅^{1,2} 化春键^{1,2}

(1. 江南大学机械工程学院 无锡 214122; 2. 江苏省食品先进制造装备技术重点实验室 无锡 214122)

摘要:针对因角度训练标签冲突和可抓取区域与物体区域间的非一致性导致的现有抓取检测方法在复杂的非结构化抓取场景中抓取检测准确性不足的问题,本文提出了一种自适应特征融合抓取检测网络 AFFGD-Net。该网络首先采用基于分区法

的角度预测方法,将角度值编码为角度类别和偏移量两部分进行学习预测,冲突的角度值划分到同一类别,减少角度训练标签的冲突,偏移量用于补偿分类部分的精度损失,提升网络对抓取角度的预测准确率。其次,引入自适应感受野模块 (ARFB) 和注意力跳跃连接模块 ASCM, ARFB 增强网络对多尺度可抓取区域特征的表征能力,并通过自适应融合不同尺度特征,提升对多尺度物体的抓取检测能力, ASCM 通过自适应融合低层空间特征和高层语义特征以恢复可抓取区域的边缘特征,提高网络的抓取角度和抓取宽度预测准确率。最后,通过实验验证了所提网络的有效性。在 Cornell 数据集的图像划分和对象划分测试模式下, AFFGD-Net 的准确率分别达到 98.9% 和 97.7%, 在 Jacquard 数据集中准确率达到 95.2%。网络检测速度达到 111 FPS, 显示出良好的实时性。实验结果表明, AFFGD-Net 在抓取检测的准确性和实时性方面均优于现有方法, 验证了所提方法的有效性。

关键词: 抓取检测; 感受野模块; 特征融合; 注意力机制

中图分类号: TP391; TN957.52⁺9

文献标识码: A

国家标准学科分类代码: 520.2060

Research on grasp detection method based on adaptive feature fusion

Xiong Huan^{1,2} Yu Jianfeng^{1,2} Qian Chenhao^{1,2} Jiang Yi^{1,2} Hua Chunjian^{1,2}

(1. School of Mechanical Engineering, Jiangnan University, Wuxi 214122, China;

2. Jiangsu Key Laboratory of Advanced Food Manufacturing Equipment and Technology, Wuxi 214122, China)

Abstract: To address the problem of insufficient grasp detection accuracy of existing grasp detection methods in complex unstructured grasping scenarios due to the conflict of angle training labels and the non-consistency between graspable regions and object regions, this paper proposed an adaptive feature fusion grasp detection network, AFFGD-Net. The network firstly adopted the angle prediction method based on the partition method, which encoded the angle values into two parts, namely, angle category and offset for learning and prediction. The conflict angle values were divided into the same category to reduce the conflict of angle training labels, and the offset was used to compensate for the loss of accuracy in the classification part to improve the prediction accuracy of the network for grasp angle. Secondly, the adaptive receptive field block ARFB and attention skip connection module ASCM are introduced. ARFB enhanced the network's ability to characterise the features of multi-scale graspable regions, and improved the grasp detection ability of multi-scale objects by adaptively fusing features of different scales. ASCM recovered the edge features of the graspable regions by adaptively fusing the low-level spatial features and the high-level semantic features, which improved the network's grasp angle and grasp width prediction accuracy. Finally, the effectiveness of the proposed network was verified by experiments. The accuracy of AFFGD-Net reached 98.9% and 97.7% in the image segmentation and object segmentation test modes in the Cornell dataset, respectively, and 95.2% in the Jacquard dataset. The detection speed of the network reached 111 FPS, which showed good real-time performance. The experimental results showed that AFFGD-Net outperformed the existing methods in terms of both accuracy and real-time crawl detection, confirming the effectiveness of the proposed method.

收稿日期: 2024-04-18 Received Date: 2024-04-18

* 基金项目: 国家自然科学基金(51905215)、江苏省研究生科研与实践创新计划(KYCX23_2553)项目资助

Keywords: grasp detection; receptive field block; feature fusion; attention mechanism

0 引言

近年来,智能机器人技术在工业、农业、商业、军事和家庭服务等领域的应用得到了迅速发展^[1]。抓取是机器人应用场景中的常见任务,快速准确的抓取检测是实现机器人成功抓取的关键环节。面对复杂的非结构化抓取场景,越来越多的研究者开始将深度学习方法应用于解决抓取检测问题^[2]。

Morrison 等^[3-4]提出了一种生成式抓取检测方法 GGCNN,将图像中每个像素点均视为抓取点,通过网络模型以抓取图的形式直接预测输出各抓取点位置的抓取质量评分以及抓取角度和宽度两个位姿参数,实现了抓取检测。在 GGCNN 研究基础上,Wang^[5]等提出了一种有向三角形的抓取位姿表示方法,替换语义分割网络 Deeplabv3+^[6]的输出部分构建了抓取检测网络。该方法实现了针对三指机器人的抓取位姿检测,提高了生成式抓取检测方法的适用性。Cao 等^[7]提出了一种全卷积抓取检测网络,在研究中引入 SE 通道注意力机制^[8]并设计了残差挤压和激励网络(RSEN)以增强深度特征的提取能力。通过设计多尺度空间金字塔模块(MSSPM)和层级特征融合结构,网络获取了多尺度上下文信息。该方法增强了网络模型对复杂场景的适应性和鲁棒性,提高了抓取检测的准确率。尽管上述方法已经取得了不错的检测效果,但它们大多基于图像中待抓取物体的全局特征进行抓取位姿预测,忽略了可抓取区域与物体区域的非一致性问题^[9],非抓取区域的干扰特征可能会降低网络的抓取检测准确率。

在抓取检测网络的性能优化过程中,训练数据的标签质量对于模型学习抓取位姿起着至关重要的作用。Chalvatzaki 等^[10]在研究中发现,GGCNN 方法生成的角度图标签中存在角度值不连续性的问题,导致网络的角度训练标签出现冲突,网络无法有效区分和学习正确的抓取角度,抓取检测的准确率因此降低。像素级的标注能够为每个抓取位置提供准确的角度学习标签,减少冲突以提高网络检测准确率,但人工对数据集进行逐像素的重新标注极为耗时^[11]。王文俊等^[12]将抓取角度按照每 60°划分为了 3 个区间,采用 3 个检测头分别进行学习和预测,将互相冲突的角度标签分给不同的检测头进行学习,可有效减少学习标签冲突,检测准确率得到了提升。但该方法中每个检测头需要针对特定的角度区间进行优化,导致训练过程更加困难和耗时且训练数据不平衡还会导致模型只对特定区间角度的预测效果较好,泛化能力受限。

针对角度训练标签冲突和可抓取区域与物体区域的非一致性问题,提出一种基于自适应特征融合的生成式抓取检测方法。首先,提出一种基于分区法的角度预测方法,将抓取角度编码为角度类别和偏移量两个部分进行预测,偏移量用于补偿角度分类预测造成的精度损失,通过将冲突的角度值划分到同一类别,减少角度训练标签的冲突。该方法不用耗费大量时间对数据集进行重新标注,且无需采用多个检测头进行特定角度区间的优化学习,模型训练简单,能够避免因角度训练数据不平衡导致的模型泛化能力降低的问题。然后,提出了自适应感受野模块和注意力跳跃连接模块,利用注意力机制抑制非抓取区域的干扰特征,引导网络在多尺度特征融合和特征重构阶段关注可抓取区域特征,提升网络的抓取检测准确率。

1 问题描述

抓取检测是指获取待抓取物体的抓取位姿参数的过程,在图像坐标系中,一组平面抓取位姿通常可以用五维抓取参数进行表示:

$$g = (u, v, w, \theta, q) \quad (1)$$

式中: (u, v) 指抓取中心点坐标,抓取宽度 w 指抓取时夹爪需要开合的宽度, θ 指抓取角度, q 是指抓取质量,用于评估在该点进行抓取的成功率,一组抓取位姿的五维抓取表示如图 1 所示。



图 1 五维抓取表示

Fig. 1 Five-dimensional grasp representation

对于输入尺寸为 $C \times W \times H$ 的图像 I ,生成式抓取检测方法将图像中的每一个像素点位置都视为一个潜在抓取点,关联一组抓取位姿 g ,这些图像空间中的抓取位姿集合称为抓取图,表示为:

$$G = (\Theta, W, Q) \in \mathbb{R}^{3 \times W \times H} \quad (2)$$

式中: Θ 、 W 、 Q 分别指抓取角度图、抓取宽度图和抓取质量图,均是尺寸为 $W \times H$ 的图像,每个像素点位置都含有表示该点抓取位姿的参数 θ 、 w 和 q 的值。生成式抓取检测即是根据输入图像 I ,拟合输出抓取图 G 的过程,最后

通过式(3)计算获得图像空间中的最佳抓取位姿。

$$g^* = \operatorname{argmax}_g \mathbf{G} \quad (3)$$

2 抓取检测方法

2.1 抓取图标签生成

生成式抓取检测方法利用真值抓取图作为学习标签进行训练,为了训练抓取检测网络,对 Cornell 和 Jacquard 抓取数据集中的矩形框抓取位姿标签进行转化,生成抓取图标签,过程如图 2 所示。

定义抓取矩形中心 $1/3$ 区域为图像掩膜,针对抓取质量图标签 Q_T ,将掩膜区域填充为 1,表示在该区域位置

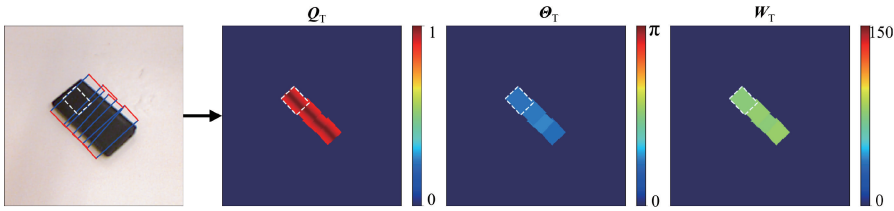


图 2 抓取图标签生成过程

Fig. 2 Process of generating grasp map label

2.2 自适应特征融合抓取检测网络

针对角度训练标签冲突和可抓取区域与物体区域的非一致性问题,提出自适应特征融合抓取检测网络 (adaptive feature fusion grasp detection network, AFFGD-Net),结构如图 3 所示。网络采用编码器-解码器结构,编码器部分用于特征提取,包括主干网络 Mobilenet-v3^[14]和自适应感受野模块 ARFB。解码器部分包括注意力跳跃连接模块 ASCM 和输出检测头 Head,用于接收编码器的输出特征进行特征重构后输出抓取检测结果。在特征提取阶段,首先 Mobilenet-v3 接收多模态的 RGB-D 图像输入,由第 1 个倒残差块输出的 2 倍降采样特征一个分支输入解码器,另一分支经后续 11 个倒残差块处理输出 4 倍降采样特征。然后将 Mobilenet-v3 输出的 4 倍降采样特征输入 ARFB 进行多尺度特征提取和融合,提升网络模型对不同尺度物体的抓取检测适应能力。在特征重构阶段,采用 ASCM 将 ARFB 的输出特征与 Mobilenet-v3 输出的 2 倍降采样特征进行融合,以恢复物体边缘特征,提高网络模型的抓取角度和抓取宽度预测准确性。最后将重构后的特征输入 4 个独立任务分支检测头,输出预测抓取图 $\mathbf{G} = (\boldsymbol{\theta}, \mathbf{W}, \mathbf{Q}) \in \mathbf{R}^{3 \times W \times H}$ 。

1) 基于分区法的角度预测方法

抓取角度预测准确性直接关系到网络的抓取检测性能,目前抓取检测中主要采用回归法和分类法两种方法进行抓取角度预测。回归法^[3]将抓取角度 θ 编码为 $\sin(2\theta)$ 和 $\cos(2\theta)$ 两个三角函数值的组合,利用网络分

进行抓取时的成功率为 1,其他区域填充为 0。为了引导网络模型学习最佳的抓取区域,强调物体中轴线处的最大抓取成功率特性,采用一维高斯核对生成的 Q_T 中的数值分布进行优化^[13]:

$$q_T(u, v) = q_{\min} + (1 - q_{\min}) \exp\left(-\frac{d^2}{2\sigma^2}\right) \quad (4)$$

式中: $q_T(u, v)$ 表示点 (u, v) 处的抓取质量真值, d 为点 (u, v) 处到矩形中轴线的垂直距离, q_{\min} 为截断阈值取 0.9。同样的,在掩膜区域填充抓取矩形表示的抓取角度 θ_T 、抓取宽度 w_T 的值,其他区域填充为 0,分别生成角度图标签 $\boldsymbol{\theta}_T$ 和宽度图标签 \mathbf{W}_T 。

别进行预测,将预测结果经过反三角函数处理后获得抓取角度预测值。分类法^[5]则是将连续的角度值按一定区间大小划分为离散的角度类别,网络预测出角度类别后即可获得角度预测值。

将密集重叠的抓取矩形标注转化为抓取图标签后,如图 2 所示,角度图标签 $\boldsymbol{\theta}_T$ 中存在角度值不连续的问题。利用回归法进行角度预测时,角度值不连续会导致网络的角度训练标签存在冲突,造成网络学习混乱,限制网络的检测性能。利用分类法进行角度预测时,采用较大的分类区间可以将冲突的角度值划分到同一类别,减少角度训练标签的冲突,但预测出的角度值存在较大的精度损失;若采用的分类区间较小则无法改善角度训练标签冲突的问题。

因此,提出基于分区法角度预测方法:将抓取角度 θ 编码为角度类别 c 和偏移量 r 两个部分进行预测,采用一个较大的分类区间,将冲突的角度值划分到同一类别,减少角度训练标签冲突,偏移量 r 用于补偿通过角度类别 c 计算出的角度值的精度损失。

Cornell 和 Jacquard 数据集的 $\boldsymbol{\theta}_T$ 中角度为 $[0, \pi]$ 的弧度值,根据预实验结果,按照每 $\pi/18$ 为一个区间划分角度类别标签和偏移量标签进行 $\boldsymbol{\theta}_T$ 的编码:

$$\mathbf{C}_T, \mathbf{R}_T = \operatorname{divmod}(\boldsymbol{\theta}_T, \pi/18) \quad (5)$$

式中: divmod 表示取商和余数的函数。如图 4 所示,将抓取角度图标签 $\boldsymbol{\theta}_T$ 编码为角度类别图标签 \mathbf{C}_T 和偏移量图标签 \mathbf{R}_T 后,角度训练标签的冲突问题得到明显改善,

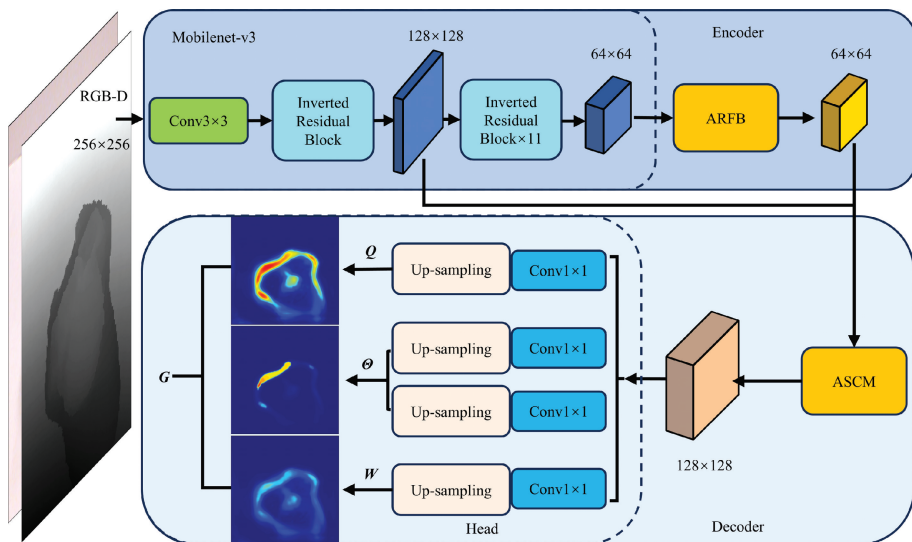


图 3 AFFGD-Net 结构

Fig. 3 AFFGD-Net structure

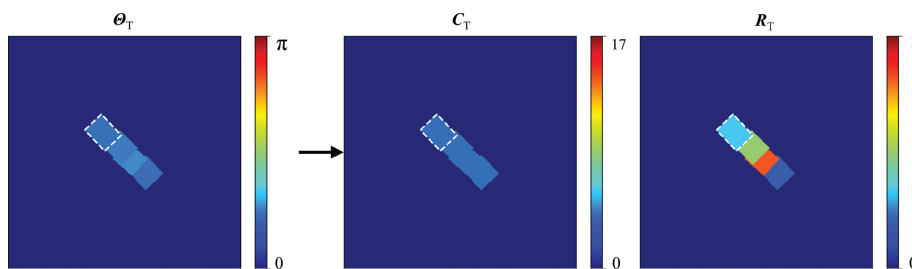


图 4 抓取角度图标签编码过程

Fig. 4 Process of encoding grasp angle map label

为了便于网络训练,将偏移量图 R_T 中的数值归一化到 $[0, 1]$ 范围内。

AFFGD-Net 采用基于分区法的角度预测方法,利用两个检测头分支分别进行抓取角度类别图 C 和偏移量图 R 两个部分的预测,整合后输出抓取角度图 Θ 。

$$\Theta = \frac{\pi}{18}(C + R) \quad (6)$$

2) 自适应感受野模块

多尺度特征融合策略在目标检测和语义分割^[15]等领域中广泛应用。感受野模块 (receptive field block, RFB) 通过融合不同感受野大小的全局多尺度特征,在针对多尺度物体进行检测时,取得了很好的效果。与目标检测需要关注物体全局特征信息不同,因为抓取检测中的可抓取区域往往不是全部的物体区域,要求网络更加关注不同尺度物体的局部可抓取区域特征。因此,在 RFB 中引入卷积块注意力模块 (convolutional block attention module, CBAM),引导网络更加有效地完成对多尺度特征的提取和融合。通过通道注意力模块 (channel

attention module, CAM) 和空间注意力模块 (spatial attention module, SAM) 分别生成通道注意力权重 M_c 和空间注意力权重 M_s 。与输入特征进行相乘,对不同通道和不同空间位置的特征进行加权,增加对重要特征的关注度, CBAM 结构如图 5 所示。

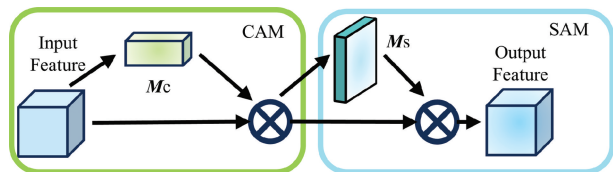


图 5 CBAM 结构

Fig. 5 CBAM structure

CAM 先对输入特征利用全局平均和最大值池化操作整合出各个通道的全局特征,经多层感知机 MLP 处理后将两部分特征相加,最后采用 sigmoid 函数激活后获得输入特征的通道注意力权重 M_c ^[16],过程如图 6 所示。

SAM 对输入特征在通道维度上进行全局平均和最大值池化操作整合出各个空间位置的全局特征并拼接,

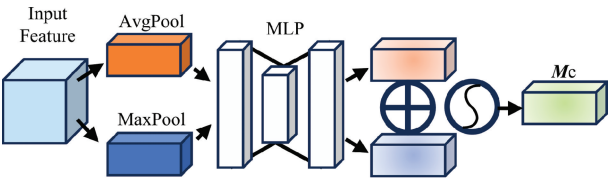


图 6 通道注意力权重生成过程

Fig. 6 Process of generating channel attention weights

经 7×7 卷积和 sigmoid 函数激活处理后,获得输入特征的空间注意力权重 M_s ,过程如图 7 所示。

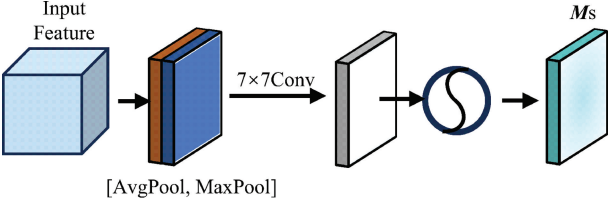


图 7 空间注意力权重生成过程

Fig. 7 Process of generating spatial attention weights

将 SAM 和 CAM 嵌入 RFB 中,提出自适应感受野模块(adaptive receptive field block, ARFB)。ARFB 接收主干网络输出的 4 倍降采样特征后输入 3 个并联分支,分别采用大小为 $\{1, 3, 5\}$ 的卷积核和空洞率为 $\{1, 3, 5\}$ 的 3×3 卷积提取不同感受野的多尺度特征,结合 SAM 自适应学习不同空间位置特征的注意力权重,通过对可抓取区域位置的特征进行加权,引导网络增强多尺度可抓取区域特征的表征能力。完成多尺度特征提取后,将 3 个分支的输出特征进行通道维度拼接,利用 CAM 和 1×1 卷积进行多尺度特征融合,通过 CAM 对不同通道的特征进行加权,调整 3 个分支输出特征的融合权重,自适应融合不同尺度的特征。ARFB 通过通道注意力机制和空间注意力机制抑制了非抓取区域的干扰特征,更有效地提取和融合了多尺度特征,提高了网络模型对多尺度物体的抓取检测适应能力。ARFB 中采用深度可分离卷积以减少模块参数量和降低计算复杂度,具体结构如图 8 所示。

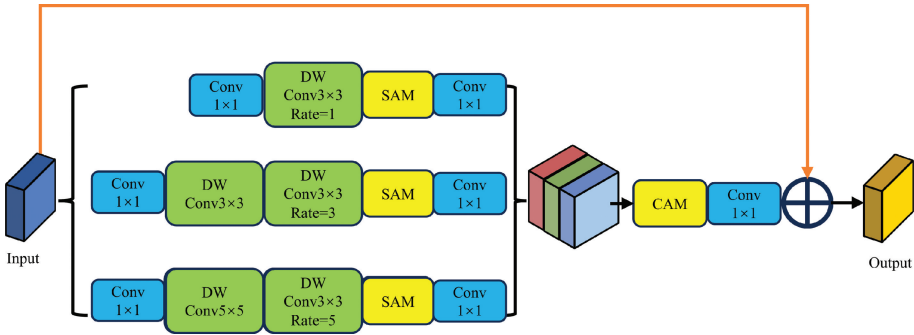


图 8 ARFB 结构

Fig. 8 ARFB structure

3) 注意力跳跃连接模块

物体边缘信息有助于抓取角度和抓取宽度的准确预测^[11],随着网络深度的增加,提取的特征中含有丰富的语义特征,但物体边缘等局部空间特征逐渐丢失^[17]。跳跃连接(skip connection)结构指将编码器(下采样路径)中的特征图与解码器(上采样路径)中对应的特征图进行连接的技术,有助于将低层特征和高层特征进行融合,恢复丢失的空间特征,提高网络的性能和准确性。针对抓取检测,由于可抓取区域与物体区域的非一致性,非抓取区域中存在大量的干扰特征,直接使用跳跃连接融合低层空间特征和高层语义特征进行物体边缘特征的恢复可能会导致重要特征的丢失,削弱特征的有效性,造成网络检测性能下降^[18]。因此,在跳跃连接结构中嵌入 CBAM,提出了注意力跳跃连接模块(attention skip connection module, ASCM),结构如图 9 所示。

ASCM 先将 ARFB 输出的特征进行 2 倍上采样和 $3 \times$

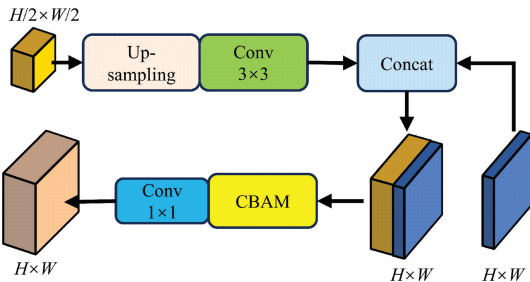


图 9 ASCM 结构

Fig. 9 ASCM structure

3 卷积处理,然后与 Mobilenet-v3 输出的 2 倍降采样特征进行拼接,利用 CBAM 生成的通道注意力和空间注意力权重对拼接好的特征进行加权,最后利用 1×1 卷积对加权后的特征进行融合。通过引入 CBAM,ASCM 能够有效抑制非抓取区域中的干扰特征,实现低层空间特征和高

层语义特征的自适应融合,恢复物体可抓取区域的边缘特征,提高网络模型对抓取角度和抓取宽度的预测准确性。

2.3 损失函数

AFFGD-Net 利用 4 个独立任务的检测头分别预测抓取质量图 \mathbf{Q} 、抓取宽度图 \mathbf{W} 、抓取角度类别图 \mathbf{C} 和偏移量图 \mathbf{R} 进行抓取检测,分别计算 4 个分支任务的损失,设计多任务损失函数用于网络参数优化。针对预测抓取质量图 \mathbf{Q} 、抓取角度偏移量图 \mathbf{R} 和抓取宽度图 \mathbf{W} 的 3 个回归任务,采用 Smooth L1 函数作为损失函数,三者的损失函数定义为:

$$Loss_{\mathbf{Q},\mathbf{W},\mathbf{R}} = \frac{1}{N} \sum_i \sum_y \text{SmoothL1}(y_{\mathbf{T}}^i - y^i) \quad (7)$$

$$y \in \{\mathbf{Q}, \mathbf{W}, \mathbf{R}\}$$

其中,SmoothL1 定义为:

$$\text{SmoothL1} = \begin{cases} 0.5(y_{\mathbf{T}}^i - y^i)^2, & \text{if } |y_{\mathbf{T}}^i - y^i| < 1 \\ |y_{\mathbf{T}}^i - y^i| - 0.5, & \text{其他} \end{cases} \quad (8)$$

式中: N 为样本数量, $y_{\mathbf{T}}^i$ 为第 i 个样本的抓取标签真实值, y^i 为网络输出的抓取预测值。针对抓取角度类别图 \mathbf{C} 的预测任务,采用交叉熵函数定义损失函数:

$$Loss_{\mathbf{C}} = -\frac{1}{N} \sum_i p_{\mathbf{T}}^i \log(p^i) \quad (9)$$

式中: N 为样本数量, $p_{\mathbf{T}}^i$ 为第 i 个样本的角度类别 one-hot 标签, p^i 为网络输出的对应样本的角度类别概率。最终,总的多任务损失函数为:

$$Loss_{\text{total}} = Loss_{\mathbf{Q},\mathbf{W},\mathbf{R}} + Loss_{\mathbf{C}} \quad (10)$$

3 实验结果和分析

3.1 实验设置与评价指标

实验数据为 Cornell 和 Jacquard 抓取数据集。Cornell 数据集规模较小,为了防止网络模型训练过拟合,采用数据集在线增强策略,训练时对输入图像和学习标签进行随机旋转和缩放。按图像划分和对象划分两种方式对 Cornell 数据集进行训练集和测试集的划分,划分后的训练集包括 708 张图像,测试集中包含 177 张。Jacquard 数据集规模较大,训练时不进行任何数据增强操作,按 9:1 的比例将其划分为训练集与测试集。数据集集中的图像分辨率均调整为 256×256 大小,以适应 AFFGD-Net 的输入。

采用 pytorch 深度学习框架进行 AFFGD-Net 的训练和测试。实验平台硬件为 Intel Core i7-12700K CPU, 32 G RAM, NVIDIA GeForce RTX 3080 GPU, 10 G 显存。网络训练采用 Adam 优化器进行参数优化,批大小设定为 8,迭代次数设定为 40 个 epoch,针对 Cornell 数据集,

一次迭代设置 1 600 的批次以进行充分训练。初始学习率设置为 0.001,采用 StepLR 学习率衰减策略,每训练 10 个 epoch 后,学习率衰减为原来的一半。

采用角度阈值和 Jaccard 系数标准^[19]作为抓取检测正确的判别条件,只要网络预测出的抓取矩形框与标签中的任何一个矩形框同时满足两者间的夹角小于 30° 和 Jaccard 系数大于 0.25 两个条件,则表示该抓取预测正确。Jaccard 系数定义如下:

$$J(g_p, g_t) = \frac{|g_p \cap g_t|}{|g_p \cup g_t|} \quad (11)$$

式中: g_p 和 g_t 分别表示预测抓取框和抓取框标签。为了计算 Jaccard 系数,将预测抓取矩形框的高度设为预测出的抓取宽度 w 的 1/2。

3.2 消融实验

为了验证提出的 AFFGD-Net 中各个模块的有效性,将 AFFGD-Net 中各个改进模块替换为原有结构设置基线网络,即基线网络中采用原始 RFB 和跳跃连接结构。同时,为了验证提出的基于分区法的角度预测方法的有效性,基线网络中采用了回归法进行抓取角度预测。通过在基线网络基础上依次修改抓取角度预测方法为提出的分区法,将原始 RFB 和跳跃连接替换为提出的 ARFB 和 ASCM 进行消融实验,实验数据为 Cornell 抓取数据集,按照图像划分和对象划分两种方式划分训练集和测试集,实验结果如表 1 所示。

表 1 消融实验结果

Table 1 Results of ablation experiment

实验序号	方法	准确率/%	
		图像划分	对象划分
1	基线	94.9	95.5
2	分区法	97.1	96.1
3	分区法+ARFB	98.3	97.1
4	分区法+ARFB+ASCM	98.9	97.7

对比实验 1 和实验 2,相较于基线网络,采用分区法进行抓取角度预测后,在图像划分测试模式下准确率提升了 2.2%,对象划分模式下准确率提升了 0.6%,表明提出的基于分区法的角度预测方法有效减少了网络的角度训练标签冲突问题,避免了网络学习混乱,抓取角度预测准确率得到了提高。对比实验 2 和实验 3,采用 ARFB 后,图像划分和对象划分测试模式下准确率分别提升了 1.2% 和 1%,说明相较于 RFB,ARFB 能够抑制非抓取区域的干扰特征,通过增强多尺度可抓取区域特征的表征能力和自适应融合不同尺度特征,提高了网络模型对多尺度物体的抓取检测能力。对比实验 3 和实验 4,采用 ASCM 后,图像划分和对象划分测试模式下准确率分别

提升了 0.6%,表明 ASCM 通过自适应融合低层空间特征和高层语义特征,有效恢复了物体可抓取区域的边缘特征,网络模型对抓取角度和抓取宽度的预测准确率得到了提升。实验结果表明,提出的基于分区法的角度预测方法以及 AFFGD-Net 中的 ARFB 和 ASCM 改进模块可有效提升网络的抓取检测准确率。

3.3 Cornell 数据集实验分析

AFFGD-Net 在 Cornell 数据集中的部分抓取检测结果如图 10 所示,第 1 行图中展示的是网络预测出的抓取

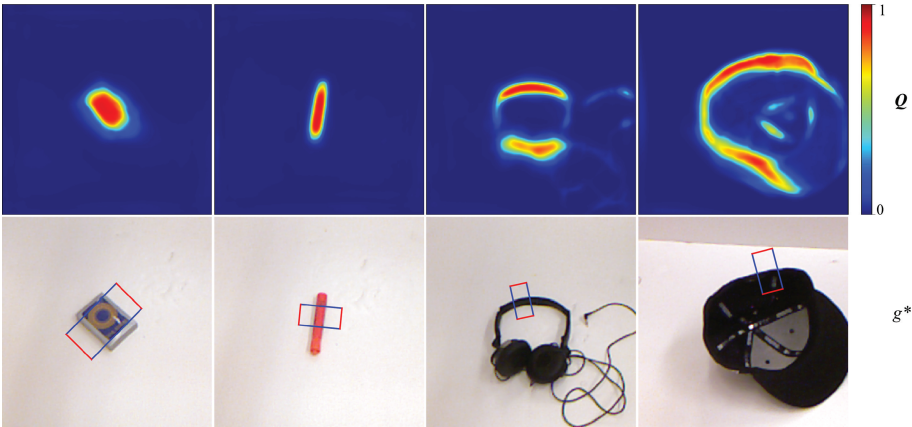


图 10 Cornell 数据集检测结果
Fig. 10 Detection results on Cornell dataset

AFFGD-Net 与其他抓取检测网络在 Cornell 数据集中的检测准确率对比如表 2 所示。

表 2 Cornell 数据集中检测准确率对比
Table 2 Comparison of detection accuracy on Cornell dataset

方法	准确率/%		速度/FPS
	图像划分	对象划分	
Res-SENet ^[7]	96.4	—	200
SGDN ^[5]	96.8	92.7	51.6
GR-ConvNet ^[20]	97.7	96.6	50
SE-ResUNet ^[21]	98.2	97.1	40
AFFGD-Net	98.9	97.7	111.6

Res-SENet、SGDN、GR-ConvNet 和 SE-ResUNet 等方法由于忽略了可抓取区域与物体区域的非一致性和角度训练标签冲突问题,非抓取区域的干扰加上网络无法学习到正确的抓取角度,限制了上述网络的抓取检测性能。得益于提出的基于分区法的角度预测方法避免了网络的学习混乱,ARFB 和 ASCM 抑制了非抓取区域的干扰特征,自适应融合生成有效特征,AFFGD-Net 在图像划分和对象划分测试模式下,取得了最高的 98.9%和 97.7%的检测准确率,由于采用了轻量化网络设计,检测速度可达 111 FPS,可进行实时抓取检测。

质量图 Q ,其中红色部分为物体的可抓取区域,值越接近于 1,表示抓取成功率越高。第 2 行显示的是通过获取抓取质量图 Q 中的最大值位置,结合角度图 Θ 和宽度图 W 中对应位置的角度值和宽度值得出的最佳抓取位姿 g^* 。由检测结果可以看出 AFFGD-Net 能够适应不同形状和尺度大小的物体进行抓取检测,面对帽子和耳机等可抓取区域与物体区域不一致的不规则物体时,仍然能够很好的关注到物体的可抓取区域,并预测出准确的抓取角度和抓取位姿。

为了进一步验证 AFFGD-Net 的性能和稳定性,测试了 AFFGD-Net 在不同角度阈值条件下的检测准确率,并与其它方法进行对比。如表 3 所示,AFFGD-Net 在角度阈值取 10° 时,图像划分和对象划分测试模式下准确率还能保持在 90% 以上,分别为 93.2%和 94.4%,证明 AFFGD-Net 具有较强的抓取检测性能和稳定性。

表 3 不同角度阈值下的准确率对比
Table 3 Comparison of accuracy at different angle thresholds

方法	数据划分	角度阈值				
		30°	25°	20°	15°	10°
FCGN ^[22]	图像划分	97.7	97.7	97.2	94.4	86.4
	对象划分	96.6	96.0	95.5	93.2	85.3
AFFGD-Net	图像划分	98.9	98.9	97.2	96.0	93.2
	对象划分	97.7	97.7	97.2	96.6	94.4

3.4 Jacquard 数据集实验分析

Jacquard 数据集中部分检测结果如图 11 所示,能够看出,抓取质量图 Q 中的值越接近于 1 的部分,均分布在物体的可抓取区域。检测结果表明 AFFGD-Net 面对 Jacquard 数据集中形状更复杂、更具挑战性的不同尺度大小物体时,依然能够关注到物体的可抓取区域进行准确的抓取检测。

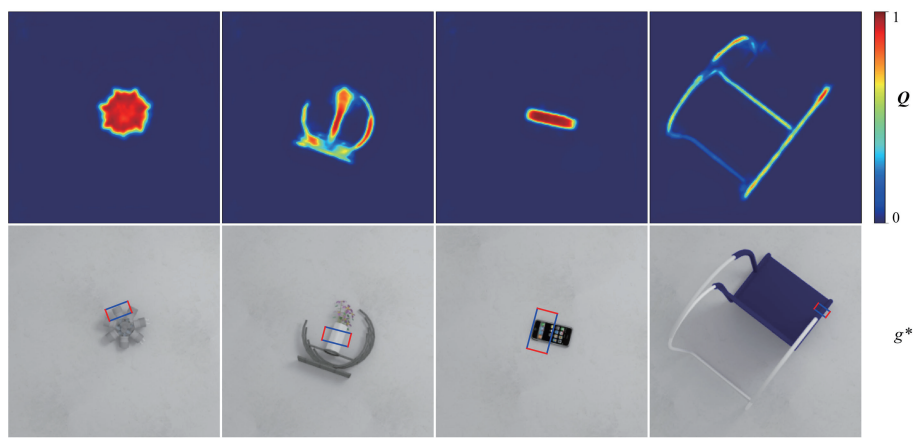


图 11 Jacquard 数据集检测结果
Fig. 11 Detection results on Jacquard dataset

为了测试 AFFGD-Net 在更大规模的数据集上的抓取检测效果,进行了 Jacquard 数据集实验。如表 4 所示为 AFFGD-Net 与其他方法在 Jacquard 数据集实验中的检测准确率对比结果,AFFGD-Net 取得了最高的 95.2% 的检测准确率,进一步证明了提出的 ARFB、ASCM 模块以及基于分区法的角度预测方法的有效性和 AFFGD-Net 良好的抓取检测泛化性能。

表 4 Jacquard 数据集中检测准确率对比
Table 4 Comparison of detection accuracy
on Jacquard dataset

方法	准确率/%
Jacquard ^[23]	74.2
GG-CNN2 ^[4]	84.0
GR-ConvNet ^[20]	94.6
Res-SENet ^[7]	94.8
AFFGD-Net	95.2

4 结 论

本文提出的自适应特征融合抓取检测网络 AFFGD-Net,通过创新的基于分区法的角度预测方法和在 ARFB 及 ASCM 中注意力机制的应用,有效地解决了抓取检测中的角度训练标签冲突和可抓取区域与物体区域间的非一致性问题。基于分区法的角度预测方法将抓取角度编码为角度类别和偏移量两个部分进行预测,偏移量用于补偿角度分类预测造成的精度损失,通过将冲突的角度值划分到同一类别,减少了角度训练标签的冲突。ARFB 和 ASCM 中利用注意力机制抑制非抓取区域的干扰特征,引导网络在多尺度特征融合和特征重构阶段关注可抓取区域特征。AFFGD-Net 在 Cornell 数据集中准确率

最高达到了 98.9%,在 Jacquard 抓取数据集中准确率达到到了 95.2%,检测速度达到 111 FPS。实验结果表明,所提网络有效提高了非结构化场景中的抓取检测准确率,相比现有方法具有更强的抓取检测性能和稳定性的同时具有较好的实时性。下一步工作是将 AFFGD-Net 应用于实际的机器人抓取系统,进行更多的实验验证和优化,推动技术的实用化进程。

参考文献

[1] TIAN H K, SONG K CH, LI S, et al. Data-driven robotic visual grasping detection for unknown objects: A problem-oriented review [J]. Expert Systems with Applications, 2023, 211: 118624.

[2] 夏浩宇,索双富,王洋,等. 基于 Keypoint RCNN 改进模型的物体抓取检测算法 [J]. 仪器仪表学报, 2021, 42(4): 236-246.

XIA H Y, SUO SH F, WANG Y, et al. Object grasping detection algorithm based on Keypoint RCNN improved model [J]. Chinese Journal of Scientific Instrument, 2021, 42(4): 236-246.

[3] MORRISON D, CORKE P, LEITNER J. Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach [J]. ArXiv preprint arXiv: 1804.05172, 2018.

[4] 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.

[5] WANG D X. SGDN: Segmentation-based grasp detection network for unsymmetrical three-finger gripper [J]. ArXiv preprint arXiv:2005.08222, 2020.

[6] CHEN L C, ZHU Y K, PAPANDREOU G, et al.

- Encoder-decoder with atrous separable convolution for semantic image segmentation [C]. Proceedings of the European Conference on Computer Vision (ECCV). 2018: 801-818.
- [7] CAO H, CHEN G, LI ZH J, et al. Residual squeeze-and-excitation network with multi-scale spatial pyramid module for fast robotic grasping detection [C]. 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021: 13445-13451.
- [8] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. Proceedings of the IEEE conference on computer Vision and Pattern Recognition, 2018: 7132-7141.
- [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 SH 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.
- [10] CHALVATZAKI G, GKANATSIOS N, MARAGOS P, et al. Orientation attentive robotic grasp synthesis with augmented grasp map representation[J]. ArXiv preprint arXiv:2006.05123, 2020.
- [11] WANG D X, LIU CH SH, CHANG F L, et al. High-performance pixel level grasp detection based on adaptive grasping and grasp aware network[J]. IEEE Transactions on Industrial Electronics, 2021, 69(11): 11611-11621.
- [12] 王文俊, 韩慧妍, 郭磊, 等. 基于角度约束与高斯化质量图的抓取检测方法研究[J]. 微电子学与计算机, 2022, 39(11): 37-44.
- WANG W J, HAN H Y, GUO L, et al. Research on grasp detection method based on angle constraint and gaussian quality map[J]. Microelectronics & Computers, 2022, 39(11): 37-44.
- [13] LI Y H, LIU Y, MA ZH Q, et al. A novel generative convolutional neural network for robot grasp detection on gaussian guidance [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-10.
- [14] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenet3 [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.
- [15] 田志新, 廖薇, 茅健, 等. 融合边缘监督的改进 Deeplabv3+水下鱼类分割方法[J]. 电子测量与仪器学报, 2022, 36(10): 208-216.
- TIAN ZH X, LIAO W, MAO J, et al. Improved Deeplabv3 + underwater fish segmentation method combining with edge supervision [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(10): 208-216.
- [16] 徐健, 陆珍, 刘秀平, 等. 注意力机制优化 RetinaNet 的密集工件检测方法研究[J]. 电子测量与仪器学报, 2022, 36(1): 227-235.
- XU J, LU ZH, LIU X P, et al. Attention mechanism optimization of RetinaNet for dense artifact detection[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(1): 227-235.
- [17] 程德强, 陈杰, 寇旗旗, 等. 融合层次特征和注意力机制的轻量化矿井图像超分辨率重建方法[J]. 仪器仪表学报, 2022, 43(8): 73-84.
- CHEN D Q, CHEN J, KOU Q Q, et al. Lightweight super-resolution reconstruction method based on hierarchical features fusion and attention mechanism for mine image [J]. Chinese Journal of Scientific Instrument, 2022, 43(8): 73-84.
- [18] ZHOU ZH N, ZHU X X, CAO Q X. AAGDN: Attention-augmented grasp detection network based on coordinate attention and effective feature fusion method [J]. IEEE Robotics and Automation Letters, 2023, 8(6): 3462-3469.
- [19] JIANG Y, MOSESON S, SAXENA A. Efficient grasping from RGBD images: Learning using a new rectangle representation [C]. IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2011: 3304-3311.
- [20] 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). 2017: 9626-9633.
- [21] YU SH, 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.
- [22] ZHOU X W, LAN X G, ZHANG H B, et al. Fully convolutional grasp detection network with oriented Anchor box [C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018: 7223-7230.
- [23] DEPIERRE A, DELLANDRÉA E, CHEN L. Jacquard:

A large scale dataset for robotic grasp detection [C].
2018 IEEE/RSJ International Conference on Intelligent
Robots and Systems (IROS). IEEE, 2018: 3511-3516.

作者简介



熊焕, 2022 年于江南大学获得学士学位, 现为江南大学硕士研究生, 主要研究方向为深度学习和抓取检测。

E-mail: xwh_year@126.com

Xiong Huan received his B. Sc. degree from Jiangnan University in 2022. Now he is a M. Sc. candidate of Jiangnan University. His main research interests include deep learning and grasp detection.



俞建峰 (通信作者), 2004 年于上海交通大学获得博士学位。现为江南大学机械工程学院教授, 主要研究方向为机器人运动控制、工业互联与智能传感、电机嵌入式驱动系统等。

E-mail: robotmcu@126.com

Yu Jianfeng (Corresponding author) received his Ph. D. degree from Shanghai Jiao Tong University in 2004. Now he is a professor in the School of Mechanical Engineering, Jiangnan University. His main research interests include robot motion control, industrial interconnection and intelligent sensing, and motor embedded drive system, etc.