Mechanical Engineering and Technology 机械工程与技术, 2026, 15(1), 73-81
Published Online February 2026 in Hans. https://www.hanspub.org/journal/met
https://doi.org/10.12677/met.2026.151008

Hans汉斯

LPGNet：基于多尺度特征的轻量化像素级抓取检测网络

宋立志*，李晨阳，付　旭，伊永烁

北京印刷学院机电工程学院，北京

收稿日期：2026年1月12日；录用日期：2026年2月5日；发布日期：2026年2月12日

摘　要

机器人抓取技术在仓储物流、工业分拣及服务机器人等领域具有广泛的应用前景。现有像素级抓取检测网络大多依赖参数量较大的主干网络，往往面临计算复杂度高、推理速度慢、难以部署于实时平台等问题。为此本文提出一种轻量化像素级抓取检测网络(Lightweight Pixel-level Grasp Network, LPGNet)。该网络以MobileNet为主干结构，引入轻量化多尺度特征增强模块(MDM-Lite)，通过编码器－解码器架构实现抓取区域、角度及宽度的像素级联合预测。在公开抓取数据集上对所提出方法进行了系统的实验验证，结果表明该方法在保证抓取检测精度的同时，具有更优的实时性能和部署潜力。

关键词

机器人抓取，像素级抓取检测，轻量化网络，多尺度特征

LPGNet: A Lightweight Pixel-Level Grasp Detection Network Based on Multi-Scale Features

Lizhi Song*, Chenyang Li, Xu Fu, Yongshuo Yi

School of Mechanical and Electrical Engineering, Beijing Institute of Graphic Communication, Beijing

Received: January 12, 2026; accepted: February 5, 2026; published: February 12, 2026

Abstract

Robotic grasping has broad application prospects in warehousing and logistics, industrial sorting,

*通讯作者。

文章引用: 宋立志, 李晨阳, 付旭, 伊永烁. LPGNet: 基于多尺度特征的轻量化像素级抓取检测网络[J]. 机械工程与技术, 2026, 15(1): 73-81. DOI: 10.12677/met.2026.151008

and service robotics. Most existing pixel-level grasp detection networks rely on backbones with a large number of parameters, and thus often suffer from high computational complexity, slow inference speed, and limited suitability for deployment on real-time platforms. To address these issues, this paper proposes a lightweight pixel-level grasp detection network, termed the Lightweight Pixel-level Grasp Network (LPGNet). LPGNet adopts MobileNet as the backbone and introduces a lightweight multi-scale feature enhancement module (MDM-Lite). With an encoder–decoder architecture, the network performs pixel-wise joint prediction of grasp regions, grasp angles, and grasp widths. Extensive experiments on public grasping datasets demonstrate that the proposed method maintains competitive grasp detection accuracy while achieving better real-time performance and deployment potential.

Keywords

Robotic Grasping, Pixel-Level Grasp Detection, Lightweight Network, Multi-Scale Features

1. 引言

随着人工智能与机器人技术的快速发展，机器人在复杂环境下完成自主操作任务的需求日益迫切[1]。在仓储物流自动分拣和工业零件装配等应用场景中，机器人需要在未知或半结构化环境下，快速、稳定地完成对目标物体的抓取操作，其性能直接影响到后续操作任务的成功率和系统整体的智能化水平。因此，研究高精度、高鲁棒性且具备实时性的机器人抓取检测方法具有重要的理论意义和工程应用价值。

传统的机器人抓取方法多依赖人工设计的几何特征或规则模型，对环境和物体形态具有较强的依赖性，难以适应复杂多变的实际场景[2]。近年来，深度学习技术通过引入卷积神经网络对图像特征进行学习，Morrison 等人[3]将深度图像输入到轻量级卷积网络 GGCNN 中，结合闭环控制实现机械臂对物体的实时抓取。Kumra 等人[4]结合残差思想提出 GRCNN 网络，提高了模型的泛化性能和特征提取能力。像素级抓取检测方法将抓取问题转化为图像到图像的预测任务，直接在像素层面上输出抓取区域、抓取角度及抓取宽度等信息，具有表达直观、候选抓取点密集等优点。

像素级抓取检测方法大多采用 ResNet [5]、Transformer [6]等参数量较大的深度网络作为主干结构。这类网络在提升特征表达能力的同时，也带来了较高的计算复杂度和存储开销，限制了其在实时系统和资源受限平台上的应用。因此，在保证抓取检测精度的前提下，如何设计结构紧凑、计算高效的轻量化像素级抓取检测网络，成为当前机器人抓取研究中亟待解决的问题。

针对现有像素级抓取检测方法在模型复杂度和实时性方面存在的问题，本文提出一种轻量化像素级抓取检测网络 LPGNet，以 MobileNet [7]作为主干网络，在显著降低模型参数量和计算复杂度的同时，实现对抓取区域、抓取角度和抓取宽度的联合预测。为了增强网络对不同尺度目标的感知能力，本文还设计了一种轻量多尺度特征增强模块 MDM-Lite，通过多尺度空洞卷积对抓取相关特征进行有效建模。

2. 抓取检测问题表达与系统建模

在实际应用中，许多机器人抓取任务可以近似建模为平面抓取问题。抓取姿态通常可表示为机械夹爪在二维平面上的位置和方向信息。其抓取参数由抓取中心点位置、抓取方向角度以及夹爪张开宽度等参数共同描述[8]。通过合理建模上述参数，机器人能够在二维图像空间中确定合适的抓取方式，并将其

映射到真实物理空间中执行抓取操作。

像素级抓取表示方法将抓取检测问题转化为图像到图像的预测任务，在输入图像的每一个像素位置上预测对应的抓取属性信息。该表示方式通常包含以下三类抓取属性图，抓取区域可行性图、抓取角度图、抓取宽度图。基于像素级的抓取检测问题可以形式化为一个多任务学习问题。给定输入图像 I，网络的目标是学习一个映射函数 F，输出对应的抓取属性图集合。

$$F(I)=\{G_r,G_\theta,G_w\}$$

其中，G_r 表示抓取区域可行性图，G_θ 表示抓取角度图，G_w 表示抓取宽度图。抓取区域可行性图用于监督抓取点位置的预测，抓取角度图用于约束抓取方向的学习，抓取宽度图则用于回归夹爪的张开宽度。

通过上述建模方式，抓取检测问题被转化为一个端到端，可训练的像素级预测任务，网络能够同时学习多种抓取属性之间的内在关联，从而生成更加一致和稳定的抓取结果。

3. 轻量化像素级抓取网络结构

3.1. 网络整体架构设计

LPGNet 采用编码器–解码器(Encoder-Decoder)式的整体框架，其核心思想是通过主干网络提取输入图像的多层次特征，再利用特征增强与解码结构逐步恢复空间分辨率，最终在像素层面输出抓取相关属性图。网络结构如图 1 所示，网络整体结构主要由四个部分组成：轻量化主干网络、特征增强模块、解码与特征融合模块以及多分支抓取预测头。

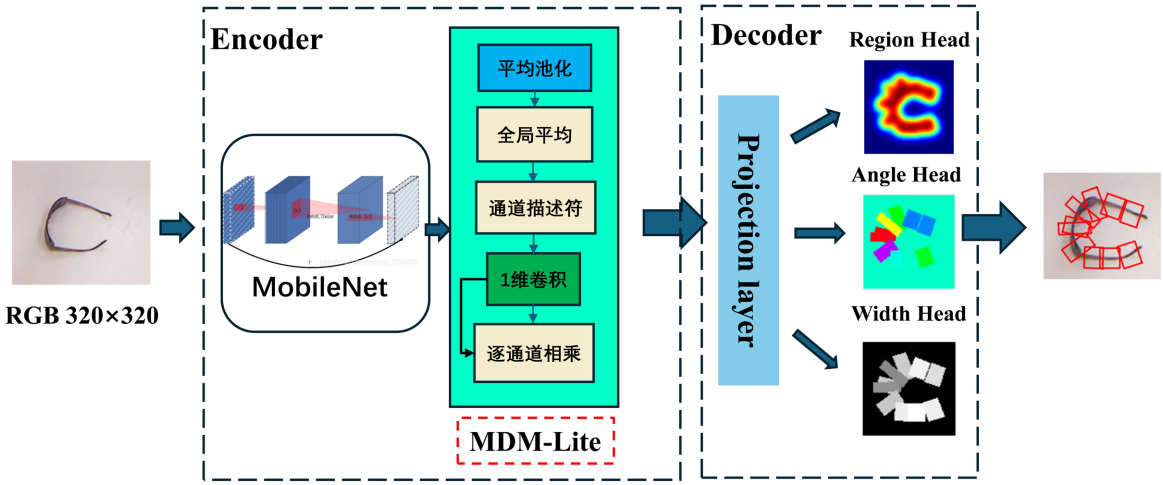

Figure 1. LPGNet network architecture
图 1. LPGNet 网络结构图

在网络前向传播过程中，输入的 RGB 图像首先经过主干网络进行特征提取，得到高层语义特征和低层细节特征。随后，高层特征通过特征增强模块进行多尺度增强，以提升其对抓取任务相关信息的表达能力。增强后的特征再与低层特征在解码阶段进行融合，通过逐步上采样恢复至原始分辨率。最后，网络通过多个并行的预测分支，分别输出抓取区域可行性图、抓取角度分布图以及抓取宽度图，从而完成像素级抓取检测任务。

3.2. 特征增强模块

由于轻量化网络的特征表达能力在一定程度上弱于深层主干网络。因此，LPGNet 在主干网络输出的

高层特征基础上引入特征增强模块 MDM-Lite，对抓取相关信息进行进一步建模。特征增强模块主要作用于高层语义特征，通过多尺度特征建模和通道级特征重标定，提升网络对不同尺度目标和关键抓取区域的感知能力。

MDM-Lite 模块以主干网络输出的高层特征为输入，通过并行的多分支结构提取不同尺度的上下文信息。在每个分支中，首先采用逐通道卷积(Depthwise Convolution)对输入特征进行空间信息建模，再通过归一化和非线性激活函数提升特征表达的稳定性。不同分支的输出在通道维度上进行拼接后，通过逐点卷积(Pointwise Convolution)进行特征融合，从而获得包含多尺度信息的增强特征表示。为保证网络训练过程的稳定性并避免特征退化，MDM-Lite 在模块末端引入残差连接，将输入特征与增强特征进行逐元素相加。

4. 实验设计与结果分析

为验证本文提出的轻量化像素级抓取检测网络 LPGNet 的有效性与实用性，本章在公开抓取数据集上对所提出方法进行系统实验。实验内容包括网络训练设置、评价指标说明、消融实验、以及与不同网络结构的对比分析，从抓取检测精度、模型复杂度和推理效率等多个角度对 LPGNet 的性能进行评估。

4.1. 数据集

本文选用 Cornell Grasp Dataset 作为实验验证数据集，数据集部分样本如图 2 所示。该数据集是机器人抓取领域广泛使用的公开数据集之一，包含多种日常物体的 RGB 或 RGB-D 图像，并为每幅图像提供了多个人工标注的可行抓取矩形。

Figure 2. Cornell grasping dataset
图 2. Cornell 抓取数据集

在实验中，按照像素级抓取检测的建模方式，将原始抓取矩形标注转换为抓取区域可行性图、抓取角度图以及抓取宽度图，作为网络的监督。数据集的训练集与测试集划分方式与主流研究工作保持一致，以便于与现有方法进行对比。

4.2. 评价指标与实验设置

为全面评估网络的抓取检测性能，本文采用以下评价指标。

抓取检测准确率(Accuracy)根据 Cornell 数据集的标准评价方式，当预测的抓取矩形与真实标注的交并比(IoU)大于设定阈值 0.25，

$$\mathrm{IOU}\left(G_P, G_t\right) = \frac{\left|G_P \cap G_t\right|}{\left|G_P \cup G_t\right|} > 0.25$$

且抓取角度误差小于给定范围时，认为抓取预测成功。

$$\mathrm{Angle}_p - \mathrm{Angle}_r > 30^\circ$$

训练在 PyTorch 框架下完成，输入图像统一调整为 320×320 分辨率。网络主干采用预训练权重进行初始化，以加快模型收敛速度并提升训练稳定性。损失函数由抓取区域、抓取角度和抓取宽度三部分损失加权组成。优化器类型、学习率及训练轮数等超参数设置与现有像素级抓取检测方法保持一致。

4.3. 结果分析

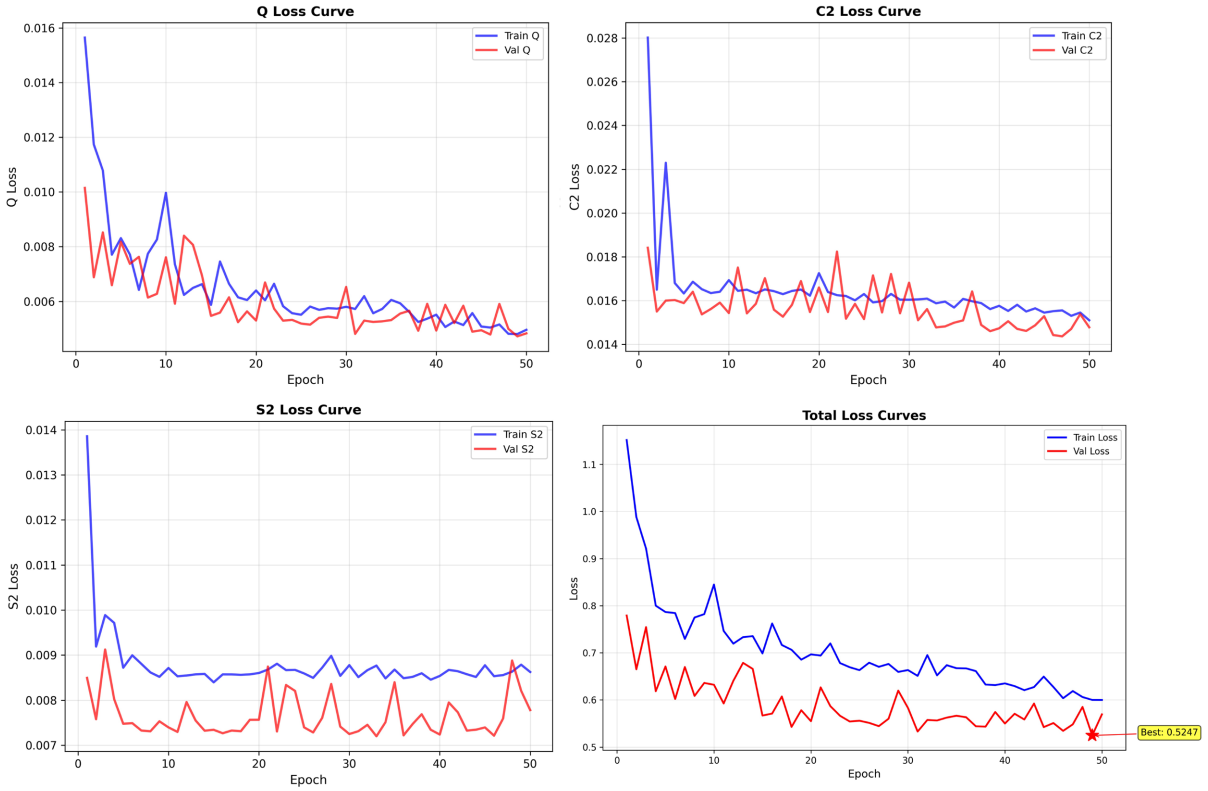

Figure 3. Loss curves of different tasks
图 3. 各任务损失曲线图

如图 3 所示，给出了 LPGNet 在 Cornell 数据集上的训练过程中各损失项随迭代轮数变化的曲线。可

以看出，网络在训练初期各损失项均呈现出快速下降趋势，表明模型能够有效学习抓取任务相关特征，并在较少的训练轮数内实现初步收敛。

从各任务损失的变化情况来看，抓取区域可行性损失下降速度最快，并在较早阶段达到稳定状态，表明网络能够较为容易地学习抓取点的空间分布特征。相比之下，抓取角度损失与抓取宽度损失的收敛速度相对较慢，其中抓取宽度损失由于属于连续回归任务，对局部几何信息较为敏感，因此在训练过程中表现出一定的波动性，但整体仍保持稳定下降趋势。随着训练的进行，总损失逐渐趋于平稳，且训练集与验证集损失变化趋势基本一致，未出现明显的震荡或反弹现象，说明网络训练过程稳定，不存在明显过拟合问题。

综上分析，所提出的 LPGNet 在多任务联合训练条件下，各损失项均能够稳定收敛，使得在轻量化网络结构下，仍能够同时兼顾抓取位置、方向与尺度等多种抓取属性的学习。

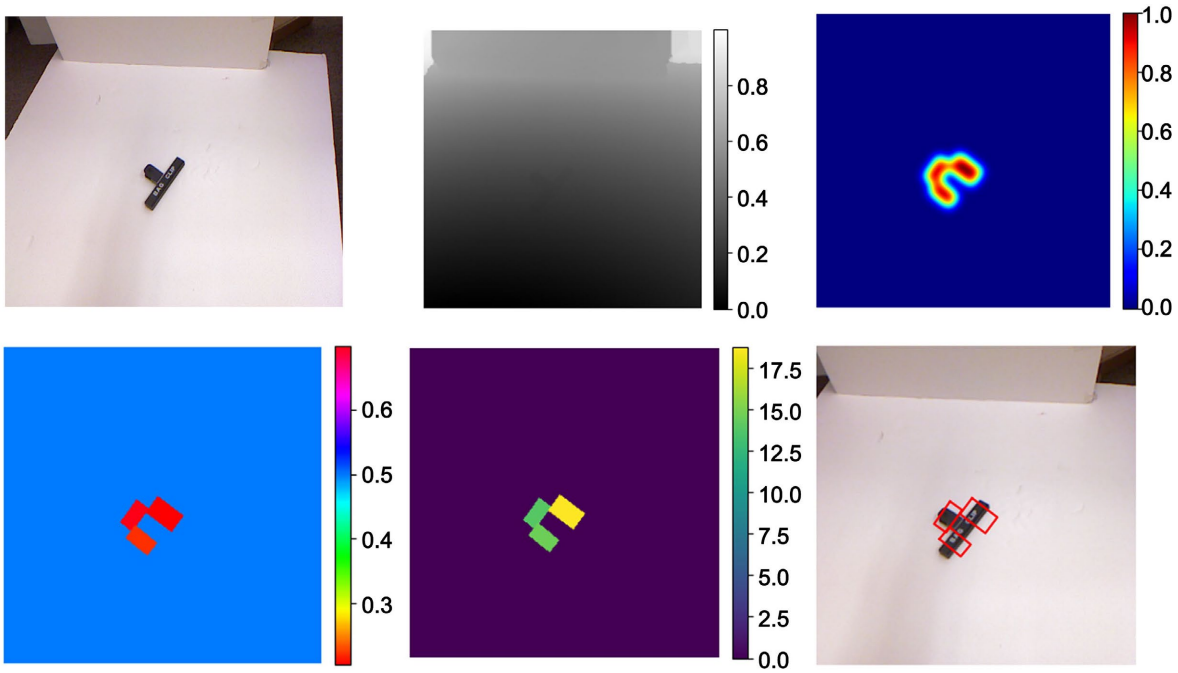

Figure 4. Grasp detection results
图 4. 抓取检测结果

如图 4 所示，给出了 LPGNet 在 Cornell 数据集样本上的像素级抓取检测结果可视化示例。图中依次展示了输入的 RGB 图像、归一化深度图、抓取质量图 Q、抓取角度图 θ、抓取宽度图 Wg 以及最终 RGB 图像上的预测效果。从抓取质量图 Q 可以看出，网络在目标物体表面形成了较为集中的高响应区域，该区域与人工标注的可行抓取位置高度一致，说明网络能够有效区分可抓取区域与背景区域，避免在无效区域产生错误响应。抓取角度图 θ 预测结果显示，在高抓取质量区域内，角度分布具有较强的一致性，未出现明显的角度跳变现象，说明网络在抓取方向建模过程中能够结合局部几何结构信息，学习到稳定合理的抓取方向。将抓取质量图叠加至原始 RGB 图像后可以直观观察到，预测的高置信度抓取区域与物体实际可抓取位置在空间上具有良好对齐性。

综合上述分析可以看出，所提出的 LPGNet 能够在单次前向推理过程中，同时准确预测抓取位置、抓取方向及抓取宽度等关键抓取属性，即使在轻量化网络结构约束下，LPGNet 仍能够保持对抓取关键特征的精确建模能力，体现了所提出网络结构在性能与效率之间的良好平衡。

4.4. 消融实验

为定量分析 LPGNet 各组成部分对检测精度与推理效率的影响，本文在 Cornell 数据集上开展消融实验。除被消融组件外，其余网络结构、输入分辨率(320 × 320)、训练策略与超参数保持一致；推理时间在同一硬件平台下统计，并在正式计时前进行若干次 warm-up 以消除缓存与编译开销影响。

为评估轻量化主干网络在保证检测精度的同时降低模型复杂度的有效性，本文在保持解码器与预测头不变的条件下，分别替换不同主干网络进行对比实验，结果如表 1 所示。

Table 1. Comparison of different backbone networks
表 1. 不同主干网络对比

主干网络	准确率(%)	总参数量(M)	FLOPs (G)	检测用时(ms)
ResNet18	98.38	20.6	6.5	22
ResNet34	98.50	30.7	10.8	30
ShuffleNetV2	97.90	11.5	3.7	13.8
MobileNetV2	98.31	12.3	4.2	14.2

在对比的四个主干网络中，ResNet34 取得了最高的准确率，但其参数量和 FLOPs 最大，检测用时达 30 ms。相比之下，本文选用的轻量化网络 MobileNetV2 表现出色，在参数量仅为 11.3M、FLOPs 为 4.2G 的情况下，依然保持了 98.31%的高准确率，且检测时间缩短至 14 ms。综合考虑性能与实时性要求，MobileNetV2 是该检测任务中兼顾精度与效率的最佳平衡选择。

为定量评估所提出 MDM-Lite 模块对 LPGNet 检测性能与计算开销的影响，本文开展了模块级消融实验。实验中除被消融的结构组件外，其余网络配置、输入分辨率及训练策略均保持一致，并在相同测试环境下统计模型的抓取检测准确率、参数量、计算量以及推理时延。相关结果汇总如表 2 所示。

Table 2. Ablation study of the proposed enhancement module
表 2. 改进模块消融实验

配置说明	准确率(%)	总参数量(M)	FLOPs(G)	检测用时(ms)
去除 MDM-Lite	97.70	10.8	3.6	13.3
去除多分支结构	98.05	11.6	3.9	13.6
去除空洞卷积	97.90	11.9	4.1	13.3
去除残差连接	97.95	12.3	4.2	14.1
完整 LPGNet	98.31	12.3	4.2	14.2

由表 2 可见，相较于不引入 MDM-Lite 的基线模型，完整模型获得了更高的检测准确率，同时仅引入有限的参数量与计算量增量，推理时延亦保持在可接受范围内，表明该模块在精度与效率之间实现了较优折中。进一步地，对 MDM-Lite 内部关键设计进行逐项移除均会导致性能不同程度下降，说明多分支特征融合、空洞卷积扩展感受野以及残差连接等结构对提升像素级抓取区域表征能力具有积极作用，从而验证了 MDM-Lite 结构设计的必要性与有效性。

4.5. 相关算法对比

为了与相关文献的结果进行比较，采用图像分割和对象分割方式对 Cornell 数据集进行划分。

1) 图像分割。将 Cornell 数据集中所有的图像随机划分为训练集和测试集。因此，在测试过程中可

能会出现与训练集中相同的物体，但该物体的放置姿势与训练集中不同，可用于测试模型对新位姿的检测性能。

2) 对象分割。将 Cornell 数据集中所有的图像按物体类别进行划分。因此，在测试过程中不会出现与训练集中相同的物体，可用于测试模型对新物体的泛化能力。

Table 3. Performance comparison of grasp detection methods
表 3. 不同方法抓取检测性能对比

模型	图像分割(%)	对象分割(%)	检测用时(ms)	总参数量
Fast Search [9]	60.50	58.30	5000	24,000,000
GG-CNN [3]	73.0	69.1	19	67,604
SAE [10]	73.90	75.60	1350	1,200,240
AlexNet,MultiGrasp [11]	88.0	87.1	76	28,400,000
Graspnet [12]	90.2	90.6	24	18,000,000
GR-ConvNet [13]	97.7	96.6	20	12,900,000
本文 LPGNet	98.31	97.84	14.2	12,305,080

表 3 给出了本文所提出的 LPG-Net 与多种代表性抓取检测方法在 Cornell 数据集上的性能对比结果，比较指标包括图像分割方式下准确率、对象分割方式下准确率以及单张图像的平均检测用时。

相比之下，本文提出的 LPGNet 在图像分割和对象分割两种评价方式下分别取得了 98.31%和 97.84%的抓取检测准确率，均明显优于对比方法。同时，其单张图像的平均检测用时仅为 14 ms，在保证高检测精度的同时实现了更快的推理速度。

5. 总结

针对现有像素级抓取检测方法在模型复杂度高、推理速度慢以及难以部署于实时或资源受限平台等方面存在的问题，本文围绕轻量化与高效抓取检测的研究目标，提出了一种基于多尺度特征增强的轻量化像素级抓取检测网络 LPGNet。该网络以轻量化主干网络为基础，通过引入多尺度特征增强模块与高效的解码融合结构，在降低模型参数量和计算复杂度的同时，有效提升了对抓取关键特征的建模能力。

Cornell 数据集上的实验结果表明，所提出 LPGNet 在图像分割与对象分割两种评价协议下均取得了较高的抓取检测准确率，并在推理速度方面优于多种代表性方法。实验分析进一步验证了该网络在抓取位置、抓取方向及抓取宽度等多任务联合建模方面的有效性，说明 LPGNet 在性能与效率之间取得了较为理想的平衡。

综合定量实验结果与定性可视化分析可以看出，本文提出的方法能够在保证抓取检测精度的同时显著提升推理效率，具有良好的实时性和工程应用潜力，为实际机器人抓取系统的部署提供了一种可行的解决方案。

参考文献

[1] 吕张成, 张建业, 陈哲钥, 等. 基于深度学习的工业零件识别与抓取实时检测算法[J]. 机床与液压, 2023, 51(24): 33-38.

[2] 周光亮. 复杂场景中物体位姿估计和抓取检测算法研究[D]: [博士学位论文]. 上海: 同济大学, 2023.

[3] Morrison, D., Corke, P. and Leitner, J. (2020) Learning Robust, Real-Time, Reactive Robotic Grasping. *The International Journal of Robotics Research*, **39**, 183-201. https://doi.org/10.1177/0278364919859066

[4] Lenz, I., Lee, H. and Saxena, A. (2015) Deep Learning for Detecting Robotic Grasps. *The International Journal of Robotics Research*, **34**, 705-724. https://doi.org/10.1177/0278364914549607

[5] 赵景波, 邱腾飞, 朱敬旭辉, 等. 基于 RP-ResNet 网络的抓取检测方法[J]. 计算机应用与软件, 2023, 40(3): 210-216.

[6] 陈鹏, 白勇, 陈旭, 等. 融合点云 Transformer 的多尺度抓取检测模型[J]. 计算机工程与应用, 2025, 61(22): 196-204.

[7] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L. (2018) Mobilenetv2: Inverted Residuals and Linear Bottlenecks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4510-4520. https://doi.org/10.1109/cvpr.2018.00474

[8] Dong, M. and Zhang, J. (2023) A Review of Robotic Grasp Detection Technology. *Robotica*, **41**, 3846-3885. https://doi.org/10.1017/s0263574723001285

[9] Jiang, Y., Moseson, S. and Saxena, A. (2011) Efficient Grasping from RGBD Images: Learning Using a New Rectangle Representation. 2011 *IEEE International Conference on Robotics and Automation*, Shanghai, 9-13 May 2011, 3304-3311. https://doi.org/10.1109/icra.2011.5980145

[10] Redmon, J. and Angelova, A. (2015) Real-Time Grasp Detection Using Convolutional Neural Networks. 2015 *IEEE International Conference on Robotics and Automation* (*ICRA*), Seattle, 26-30 May 2015, 1316-1322. https://doi.org/10.1109/icra.2015.7139361

[11] Asif, U., Tang, J. and Harrer, S. (2018) GraspNet: An Efficient Convolutional Neural Network for Real-Time Grasp Detection for Low-Powered Devices. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Stockholm, 13-19 July 2018, 4875-4882. https://doi.org/10.24963/ijcai.2018/677

[12] Fang, H., Wang, C., Gou, M. and Lu, C. (2020) Graspnet-1Billion: A Large-Scale Benchmark for General Object Grasping. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 11444-11453. https://doi.org/10.1109/cvpr42600.2020.01146

[13] Kumra, S., Joshi, S. and Sahin, F. (2022) Gr-Convnet V2: A Real-Time Multi-Grasp Detection Network for Robotic Grasping. *Sensors*, **22**, Article 6208. https://doi.org/10.3390/s22166208