分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2220320156

安徽工程大学
Anhui Polytechnic University

# 硕士学位论文

**题目**　基于双目视觉的机器人6DOF位姿测量算法研究

论文作者　何琴
校内导师　江明（教授）
校外导师　王鸿浩

专业学位类别　电子信息
研究方向（领域）　控制工程

论文提交日期：2025 年 4 月 29 日

分类号：TP391 单位代码：10363

密类级：公开 学号：2220320156

# 题目 基于双目视觉的机器人 6DOF 位姿测量算法研究

# 题目 RESEARCH ON 6DOF POSE MEASUREMENT ALGORITHM FOR ROBOTS BASED ON BINOCULAR VISION

论 文 作 者 ： 何琴

校 内 导 师 ： 江明(教授）

校 外 导 师 ： 王鸿浩

专业学位类别： 电子信息

研究方向（领域)： 控制工程

论文答辩日期： 2025 年 4 月 29 日

# 安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名： 

日期：　2025 年 5 月 15 日

# 安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 □，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 ☑。

学位论文作者签名： 

指导教师签名： 

日期： 2025 年 5 月 15 日

日期：2025 年 5 月 15 日

# 基于双目视觉的机器人 6DOF 位姿测量算法研究

## 摘要

在科技的发展和传统制造业加速转型升级的背景下，机器人智能抓取凭借其高效、精准等优势在工业生产中取得了广泛的应用，其中将视觉技术引入到机器人抓取中，是实现机器人智能化性能提升的发展趋势。目前，在复杂工业环境中，金属工件进行打磨、抛光等加工过程中会产生大量悬浮粉尘，干扰视觉系统成像效果。此外，金属工件表面的反光特性及加工遗留的划痕，进一步干扰视觉系统对目标工件的定位。这些因素影响视觉系统对目标工件的位姿测量，导致机器人抓取失败。针对上述问题，本文对基于机器视觉的机器人 6DOF 位姿测量算法展开研究，旨在通过获取较为精准的机器人抓取位姿，实现基于视觉引导的机器人稳定抓取。本文主要内容及实施方案归纳如下：

(1)提出一种基于双目视觉引导的 6DOF 机器人灵巧手抓取方案。该方案通过结合双目视觉获取较高精度的机器人 6DOF 抓取位姿，成功实现对目标工件的抓取。在机器人视觉定位时，若未对目标对象进行较为精准检测，则无法进行后续抓取位姿的计算，此时系统会自动保存问题图像并更新数据集，用于后续的模型训练。因此通过这种数据的迭代优化，机器人抓取目标工件的稳定性得以持续提升。

(2)工业领域中工件数据集获取途径有限，若依靠人工采集大量工件图像需投入大量人力成本。针对这一问题，提出一种基于生成对抗网络的工件数据集构建与增强方法。首先，采用相机方法获得原始数据集；其次，通过结合随机算法和传统数据增强技术对图像

进行多样化操作，不仅能实现对有限数据集的扩充，还能解决相机拍摄方向和角度单一的问题；然后，利用基于生成对抗网络的数据增强方法将工件拼接图进行背景风格转换，解决拼接图像中像素不一致的问题，使其更好地接近于相机采集的工件图像，提高检测模型的泛化性；最后，为提高数据增强模型的图像生成效果，本文对模型的生成器和判别器部分进行改进，以优化图像生成质量，为后续视觉系统对工件的识别定位提供丰富的图像数据集。

(3)针对金属工件反光、表面磨损及粉尘等因素影响工件的6DOF 抓取位姿的获取，导致预测抓取位姿与实际位姿出现偏差问题，提出一种基于深度学习的 6DOF 位姿测量方法。该方法采用关键点检测技术获取工件表面特征点的位姿信息，为基于双目视觉的6DOF 位姿测量提供准确的输入数据。其中，基于深度学习的多关键点检测模型以 CenterNet 为基础框架。为提高检测精度，于上采样阶段引入自适应细粒度通道注意力以更准确地强调有用特征，抑制无关或干扰性的特征，使模型能够更敏锐地捕捉工件的空间特征。与此同时，结合多尺度特征融合的方法解决检测时目标对象的特征损失问题，丰富空间信息，最终提升工件关键点检测精度。

基于上述技术研究，本文验证了数据增强算法在提升图像生成质量方面的效果，以及关键点检测算法在多类型工件数据集上的精度表现。实验结果表明，数据增强显著改善了图像生成质量，同时检测算法在复杂场景下获取了较高精度的位姿信息。此外，搭建基于双目视觉的机器人灵巧手抓取实验平台，验证位姿测量策略的有效性，开展了机器人的 6DOF 抓取位姿测量实验实现对工件的抓取。根据位姿测量实验结果显示，本文所提的机器人 6DOF 位姿测量策略获取的 $x$、$y$、$z$ 轴偏差范围为 0~0.012m，$x$、$y$、$z$ 轴的旋转偏差范

围为 0.003°~1.729°。实验数据验证了该策略在抓取位姿测量方面具有较高的抓取精度，证明了本文研究方法的可行性和实际应用价值。

**关键词：**机器人，双目视觉，生成对抗网络，6DOF 位姿测量，关键点检测

# RESEARCH ON 6DOF POSE MEASUREMENT ALGORITHM FOR ROBOTS BASED ON BINOCULAR VISION

## ABSTRACT

In the context of technological evolution and the rapid modernization of traditional manufacturing industries, robotic intelligent grasping has gained extensive application in industrial production due to its efficiency and precision. The integration of vision technology into robotic grasping represents a developmental trend aimed at enhancing the intelligent performance of robots. Currently, in complex industrial environments, metal workpieces undergoing processes like grinding and polishing generate substantial suspended particulates, which severely degrade visual imaging quality. Furthermore, the inherent reflectivity of metal surfaces and processing-induced scratches create additional interference with visual target localization. These combined factors compromise the 6DOF pose measurement accuracy of vision systems, ultimately leading to robotic grasping failures. Addressing these challenges, this paper conducts research on 6DOF pose measurement algorithms for robots based on machine vision, aiming to achieve stable grasping of robots based on vision guidance by obtaining more accurate robot grasping poses. The main contents and implementation plan of this paper are summarized as follows:

(1) A 6DOF dexterous hand grasping scheme for robots guided by binocular vision is proposed. This scheme successfully achieves the

grasping of target workpieces by combining binocular vision to obtain high-precision 6DOF grasping poses. During robot visual localization, if the target object is not accurately detected, the system will automatically save the problematic images and update the dataset for subsequent model training. Consequently, through this iterative optimization of data, the stability of the robot in grasping target workpieces is continuously enhanced.

(2) In the industrial field, obtaining workpiece datasets is often challenging, and relying on manual collection of a large number of workpiece images requires significant human resources. To address this issue, a method for constructing and enhancing workpiece datasets based on Generative Adversarial Networks (GANs) is proposed. First, the original dataset is acquired using a camera. Second, a combination of random algorithms and traditional data augmentation techniques is applied to diversify the images, not only expanding the limited dataset but also addressing the issue of single shooting directions and angles in camera-captured images. Then, a GAN-based data augmentation method is used to perform background style transfer on workpiece composite images, resolving pixel inconsistency issues in the composite images and making them more closely resemble camera-captured workpiece images, thereby improving the generalization ability of the detection model. Finally, to enhance the image generation quality of the data augmentation model, improvements are made to the generator and discriminator components of the model, optimizing the quality of generated images and providing a rich image dataset for subsequent visual systems in workpiece recognition and localization tasks.

(3) To address the issue of deviations between the predicted grasping pose and the actual pose caused by factors such as metal workpiece reflection, surface wear, and dust, which affect the acquisition of the

6DOF grasping pose, a deep learning-based 6DOF pose measurement method is proposed. This method utilizes keypoint detection technology to obtain the pose information of feature points on the workpiece surface, providing accurate input data for 6DOF pose measurement based on binocular vision. Among them, the deep learning-based multi-keypoint detection model is built upon the CenterNet framework. To enhance detection accuracy, an adaptive fine-grained channel attention mechanism is introduced during the upsampling stage to more precisely emphasize useful features while suppressing irrelevant or interfering features, enabling the model to more sensitively capture the spatial characteristics of workpieces. Concurrently, a multi-scale feature fusion approach is integrated to address feature loss issues during detection, enriching spatial information and thereby improving the accuracy of the detection model.

Based on the aforementioned technical research, this paper validates the effectiveness of the data augmentation algorithm in improving image generation quality and the accuracy of the keypoint detection algorithm on multi-type workpiece datasets. Experimental results demonstrate that data augmentation significantly enhances image generation quality, while the detection algorithm achieves high-precision pose estimation even in complex scenarios. Furthermore, a binocular vision-based robotic dexterous hand grasping platform was constructed to validate the effectiveness of the pose measurement strategy. A 6DOF grasping pose measurement experiment was conducted to achieve workpiece grasping. The experimental results of the pose measurement show that the proposed 6DOF pose measurement strategy achieves a positional deviation range of 0~0.012 m along the $x$, $y$, and $z$ axes and a rotational error range of 0.003°~1.729° for the $x$, $y$, and $z$ axes. Experimental data confirm that the proposed strategy achieves high grasping accuracy in pose measurement, demonstrating the feasibility and practical application value of the

proposed method.

He qin (Electronic Information)
Supervised by Professor Jiang Ming

KEY WORDS: robot, binocular vision, generative adversarial network, 6DOF pose measurement, keypoint detection

# 目录

# 第 1 章 绪论

## 1.1 研究背景和意义

随着现代工业领域对重复性、高精度及智能化的生产要求不断提高，诸如汽车零部件搬运，产品零件缺陷检测，电子元件装配等工作[1-3]，通过人工操作存在偏差，并且重复性工作也造成了人力的过度投入与资源浪费。特别在一些特殊的工作场景里，环境复杂且恶劣。以汽车制造车间的喷漆为例，喷涂过程中产生的漆雾含有大量有害挥发性物质，工人长期在这钟环境下工作，会严重影响身体健康。相比之下，工业机器人的应用不仅减少了恶劣环境对工人身体健康的影响，也提高了自动化水平，大幅提升了生产效率和工艺质量。

如今在工业自动化生产中以多关节的工业机器人为主，这类机器人可代替人类完成繁琐和危险的工作，具有工作效率高、负载能力强和工作时间长等优势，因此被广泛应用于不同领域下的搬运或抓取任务，图 1-1 展示了在不同生产场景下进行机器人抓取搬运任务的实际情况[4-6]。目前大部分机器人都是明确待抓取目标的 6DOF 位姿，按照事先规定好的轨迹工作。然而一旦出现目标偏离预设定抓取的位置和姿态，机器人会出现抓取失败情况。为提高机器人系统的环境适应性和作业成功率，人们通过增加传感器的方式来提高机器人的感知能力，准确获取目标对象 6DOF 位姿信息来提高其抓取成功率。在诸多传感器中视觉传感器获取的信息量相对较多，可以有效地获取物体的轮廓、颜色等特征信息，并人为地设计抓取点的位置。因此研究人员为机器系统赋予视觉感知能力，将人类视觉的高度抽象能力与计算机强大的数据分析处理能力相结合，并将其应用到工业自动化生产中[7]。

图 1-1 不同生产场景下的机器人搬运及抓取任务

机器视觉系统根据相机配置数量的差异，主要分为单目视觉、双目视觉和多目视觉三大类。其中，单目视觉系统在物体识别、定位、分类等任务中，由于仅能获取二维平面信息而无法重建目标物体的三维空间位置，在实际应用中对目标对象的 6DOF 位姿获取存在局限性[8,9]。为克服这一限制，结构光、激光、物理传感器等技术都已被应用来测量物体的三维信息，实现精确、快速的三维重建。然而，这些方法往往需要昂贵的专用设备，且存在便携性差、部署复杂等问题，难以实现大规模普及应用。相比之下，双目视觉和多目视觉系统的硬件设备相对便携，且三维重建的算法复杂性相对其它方式较低。而两者中双目视觉系统结构简洁，安装配置方法灵活多样，能够适应多样化的工业环境需求，得到较为精准的目标对象 6DOF 位姿信息，相较于多目视觉系统更具实用性和经济性，故双目视觉系统一直是基于视觉引导机器人抓取研究与应用的热点。

现阶段基于双目视觉 6DOF 位姿测量处理的目标对象大多是弱纹理或者无纹理的金属工件[10]，这类目标物体难以提供足够的特征信息，从而导致传统的 6DOF 位姿测量方法很难获取较高精度的位姿信息。尽管深度学习技术的崛起为位姿测量提供了新的解决思路，但仍然存在许多问题需要解决：(1)工业领域下进行视觉采集时工件常呈现反光影响图像采集质量；(2)生产过程中产生工件表面瑕疵(凹陷、划痕、气孔等)导致视觉信息缺失；(3)复杂工业环境下的光照变化和背景干扰。这些问题严重制约了机器人抓取的成功率和稳定性。因此研究基于双目视觉的机器人 6DOF 位姿测量算法对于提升工业机器人的智能化水平发展具有重要的现实意义和应用价值。

## 1.2 国内外研究现状

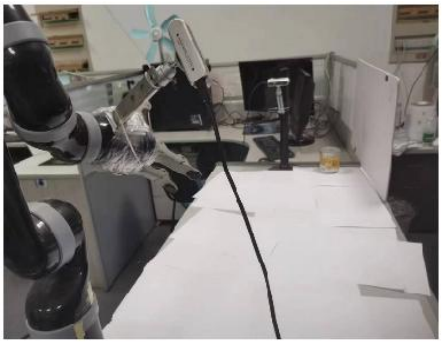
### 1.2.1 机器人抓取研究现状

在机器人抓取研究领域，众多学者通过深入探索，已取得了部分研究成果，为该领域的发展奠定了坚实基础。Jorgensen 等人实现了对可变形物体执行拾取和放置，主要使用结构光扫描仪来捕获要抓取的物体的点云，基于此点云以确定机器人的拾取和放置动作[11]。Hagelskjaer 等人通过视觉信息和抓取模拟来提供实际抓取时的不确定补偿，来实现机器人稳定抓取[12]。Kitagawa 等人为实现机器人能像人类一样执行双臂操作，还能够操作大型、重型以及不平衡的物体，

提出了一个基于少量经验学习且具备选择性双臂抓取功能的目标拾取系统，成功使机器人能够从单臂和双臂动作中选择并执行合适的抓取动作[13]。

当前在视觉定位机器人的应用上已有部分应用。如：瑞典 ABB Robotics 公司[14]推出的 1600-10/1.45 六轴工业机器人，性能可靠、操作简单、技术体系开放，可以快速精确地检测工件的姿态；加拿大 Kinova 公司[15]生产的 Jaco2 机械臂设计轻巧便捷，每个关节集成了高扭矩的无刷直流电机和 Harmonic Drive 减速器，内置的控制器方便用户轻松控制。沈阳新松机器人公司研发的平行 Delta 抓取机器人[16]如图 1-2(a)所示，采用并联的方式对物体进行抓取。博城机器人公司研发的环卫机器人[17]引入机器视觉技术来识别道路上不同类型垃圾。

(a)1600-10/1.45 六轴工业机器人



(b)Jaco2 机械臂

(c)新松并联机器人

(d)博城环卫机器人

图 1-2 抓取机器人类型

为提高机器人抓取智能化水平，王行洲[18]在工业机器人中引入多目视觉系统，提高机器人抓取的准确性。解霄鹏[19]提出了一种基于视觉反馈的机械臂位置模糊控制方法，该方法通过模拟人类抓取物品的行为模式，利用视觉传感器实时获取目标物体与机械臂末端的相对位置信息，从而控制机械臂的运动方向和速度。针对生产线复杂环境对搬运机械手抓取定位的干扰问题，孙文革[20]创新性地提出了基于智能视觉约束的生产线搬运机械手抓取定位控制方法。该方法通过建立视觉约束模型，优化机械手运动控制算法，有效提升了机械手在复

杂工况下的抓取精度和稳定性，为工业自动化生产提供了可靠的技术支持。

### 1.2.2 双目视觉研究现状

双目视觉是机器视觉的一种的重要分支，具备高识别精度和快速性等优势，在机器人领域有着重要应用[21]。上世纪的 80 年代，基于双目匹配的计算机视觉理论是由 Marr 教授[22]首次提出，通过完整地描述了计算机从场景中图像的采集到目标物体三维重建的过程，为双目立体视觉提供了理论基础。2010 年，Bahena 等人[23]通过双目视差对目标物体进行图像处理和边缘检测，提出了适用于车辆的减速带检测系统。2015 年，Benacer 等人[24]设计了一种基于 U-V 视差图分析的双目视觉检测算法，对不同环境、天气条件和光照下的道路和目标进行障碍检测。2018 年，冯英旺[25]设计出一种高精度的远距离双目被动测距系统，利用新型光轴校正方法对长焦相机获取的图像进行实时校正，并采用优化搜索策略的区域匹配算法和 SURF 特征匹配进行匹配，提高了目标测量精度。2019 年 Kalogeiton 等人[26]通过双目立体视觉实现了机器人对周边环境的精确定位，使其能够自主规划并沿着最佳路线行进，同时躲避任何局部障碍物。2020 年 Zhou 等人[27]提出了一种新型测量方法，首先根据特征点匹配结果得到部分有效点的深度值；然后，采用基于 MBC 的深度传播算法计算目标区域的深度信息；之后，根据相应的视差值计算单位像素的实际大小，并依据水平和垂直角度进行校正；最后实现物体的精确尺寸测量。

为了获取低亮度环境下目标对象的位姿信息，2021 年，李佳[28]对低亮度环境下图像采用构建校正映射表和 BM 立体匹配算法的方式获取校正畸变后的坐标，得出目标的深度距离。2022 年单福强[29]在双目视觉匹配研究中取得新进展，通过将灰度归一化互相关算法与极线约束相结合，实现了特征点搜索范围的优化，在提升匹配效率的同时保证了精度。2024 年 Borwonpob 等人[30]通过宽基线和非平行立体视觉的结合，完成多个空中机器人进行快速且运动高效的远程 3D 重建。此外，应用具备不同参数集的多个航空机器人，在不同深度范围内进行 3D 监控，同时进行水平和垂直立体的组合使用，提高深度估计的质量和完整性。同年，季娟娟[31]提出了一种基于双目视觉的集装箱吊放位姿测量方法，通过 HSV 色彩空间的特征轮廓提取和改进的 KCF 算法来完成目标的识别与实时跟踪，

并结合矩特征和三维位姿检测算法完成集装箱与吊具的位姿测量，实现了精准吊放。

### 1.2.3 6DOF 位姿测量研究现状

六自由度位姿测量简称 6DOF 位姿测量，其目的是找到图像中物体的六自由度(x、y、z 坐标轴和各个轴的旋转)位姿信息以及计算物体坐标系与相机坐标系之间的平移和旋转关系，如图 1-3 显示了物体与相机坐标系之间的关系。

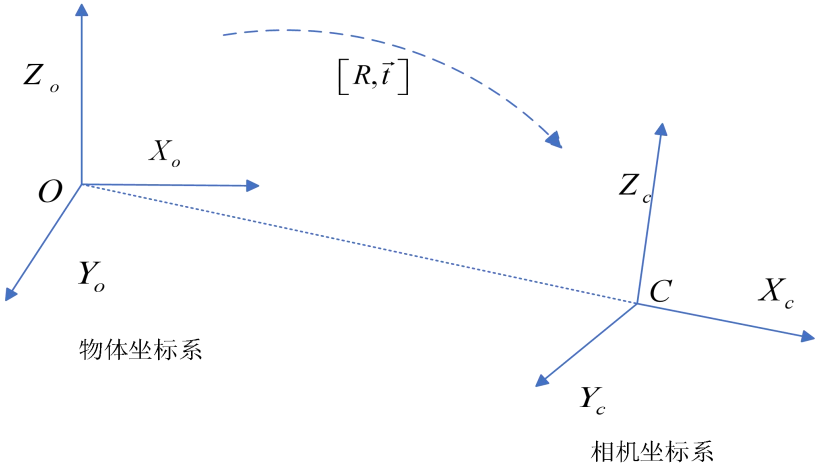



图 1-3 6DOF 位姿估计示意图

目前，对基于视觉的位姿定位测量技术发展较为成熟，方法繁多，主要方法包括基于特征对应的方法、基于模板匹配的方法以及基于深度学习的方法。

基于特征对应的方法常利用特征提取算法进行位姿测量任务，如 SIFT[32]、SURF[33]、ORB[34]等通过在多尺度下确定物体点信息，建立 2D-3D 对应特征关系，匹配当前的输入图像，再通过 PnP 算法[35]恢复目标物体的 6DOF 位姿。此方法适用于处理表面特征信息丰富的物体，但是当物体为光滑无纹理的金属工件时，会因为获取的特征信息不足而出现位姿测量结果不准确情况。

相比基于特征对应关系的方法，基于模板匹配的方法适用于弱纹理或无纹理的物体。Hinterstoisser 等人[36]提出了 LineMod 方法，该算法是对场景二维图像和模板二维图像的图像梯度进行计算，在对梯度方向进行量化处理后，求解出最符合的模板图像，并将其姿态值赋给场景二维图像。HodanT 等人[37]提出了一种新颖的架构，即通过 Kinect 体感相机(RGB Dsensors)对物体进行建模、检测和跟踪，同时优化 LineMod 方法，减轻了环境中的杂波和物体遮挡问题。

目前，基于深度学习的 6DOF 位姿估计方向涌现了大量的研究。刘鹏翔、梁冬泰等人[38]提出了基于生成对抗网络的单目相机 6D 位姿估计算法，通过对抗网络与 RGB-D 相机获取的深度图相结合，提高了目标位姿估计的准确性。Zhang 等人[39]考虑到工业中无纹理物体的位姿测量存在困难，提出了弱纹理物体 6DOF 姿态估计的边缘注意力网络 EANet，并采用端到端的方式回归物体姿势，以满足提高效率的目的。但有时依旧会存在对不同尺寸物体难以区分的情况，为此 Shi 等人[40]提出一种新颖、简化的姿态估计方法，通过运用贝叶斯辅助推理(BAI)方法补充深度网络，提高了姿势估计的准确性。而 He 等人[41]将视觉转换器和深度高分辨率网络与注意力机制相结合，捕获图像中的全局依赖关系，通过连续的全局特征交互使模型能够学习来自不同区域的判别特征，用于捕捉图像中物体的细节和姿势变化，提高姿态估计的准确性。

综上所述，传统的 6DOF 位姿测量方法可以获取较高的测量精度，但也存在着易受外界环境条件变化的干扰，稳定性有待提升。相比之下，基于深度学习的 6DOF 位姿估计方法在姿态检测精度和稳定性方面均展现出显著优势。但是深度学习算法在实际场景中仍存在一定局限性，比如部分检测网络在对目标对象进行识别定位前，需构建丰富的数据集以提高检测算法泛化性。因此在实际工业领域中，当数据集不足时会在一定程度上制约了基于深度学习的 6DOF 位姿估计方法的推广应用。

## 1.3 主要研究内容和结构

### 1.3.1 主要研究内容

针对工业复杂环境对视觉系统干扰导致机器人抓取任务失败问题，本文研究了基于双目视觉的机器人 6DOF 位姿测量算法，旨在通过提高抓取位姿精度，实现机器人对目标工件的精准抓取。本文从深度学习方法入手，通过数据增强技术和关键点检测算法实现对目标工件准确识别和位姿测量，成功完成基于双目视觉引导的 6DOF 机器人灵活手抓取任务，有效满足实际工程的需求。主要研究内容如下：

(1)以工业金属工件为研究对象，提出一种基于视觉引导的 6DOF 机器人灵巧手抓取方案，该方案适用于种类多样的金属工件的抓取。相较于传统机器人

抓取方法，在抓取位姿精度方面本文提出的方案有所提高，抓取过程更稳定。

(2)针对复杂环境下人工采集图像成本高问题，设计一种结合随机算法和数据增强技术的图像生成方法。通过基于生成对抗网络 CycleGan 实现对不同类型金属工件图像的背景迁移，实现对有限数据集的扩充。同时对 CycleGan 生成器和判别器的优化，提高图像生成的质量，为视觉检测模型构建金属工件数据集。

(3)针对工业环境中金属工件存在反光现象、表面磨损状况以及粉尘等因素严重影响视觉位姿测量，造成预测的抓取位姿与实际抓取位姿出现较大偏差问题，提出了一种基于深度学习的多关键点检测方法。该方法以 CenterNet 架构为基础，引入自适应细粒度通道注意力和多尺度特征融合模块，提高模型对目标工件的特征提取能力，实现高精度关键点检测，并凭借多个关键点位姿信息来进行工件的 6DOF 位姿测量。

(4)机器人抓取任务实验与分析。搭建基于双目视觉引导的机器人灵巧手抓取实验平台，结合加权最小二乘拟合算法获取的机器人抓取位姿信息来设计相关抓取实验，操控机器人灵巧手验证本文所提抓取策略的有效性。

### 1.3.2 论文组织结构

本文针对基于视觉引导的机器人抓取目标物体存在的问题，结合国内为研究现状，对现有技术进行研究并进行相应的算法优化，本文主要分为 6 章，具体安排如下：

第 1 章为绪论部分。介绍了本课题的研究背景与意义，以及本文研究内容的国内外研究现状，针对现在工业领域中机器人抓取目标工件存在的诸多问题，对机器人 6DOF 位姿测量的主要工作进行研究。

第 2 章为总体架构及相关理论基础。通过对机器人抓取中的难点进行分析并以此开展本论文研究的总体架构，并对涉及的相关基础理论和技术进行概述。主要研究包括图像数据增强技术、目标检测网络以及视觉引导机器人抓取相关算法。

第 3 章为基于生成对抗网络的图像生成。本文利用基于生成对抗网络的图像数据增强技术进行有限图像数据集的扩充，并通过分析 CycleGAN 在工业目标对象风格迁移中的不足来进行网络架构优化。首先在生成器方面，将 ResNet

网络替换成 DenseNet，引入多头自注意力机制，其次在判别器部分，将该部分更换成引入扩张卷积的多尺度判别器，最后进行对比实验和消融实验证明所提方法在图像生成的有效性。

第 4 章为基于深度学习的多关键点检测方法。首先，利用第 3 章图像生成模型生成的图像构建关键点检测数据集，并通过实验验证图像扩充后检测模型的关键点检测能力；其次，基于 CenterNet 框架设计关键点检测模型，用于获取工件关键点的位姿信息。为进一步提升检测精度，对模型的主干网络进行优化，同时通过消融实验和对比实验验证改进效果。

第 5 章为基于双目视觉的 6DOF 机器人位姿测量及抓取。首先利用第 4 章获得的关键点位姿信息进行双目视觉的位姿测量，计算出在双目视觉下工件位姿，然后结合三角测量算法获取机器人的抓取位姿，并进行多组实验进行位姿误差分析。最后控制带有灵巧手的工业机器人执行抓取任务，验证本文所提的基于双目视觉的机器人 6DOF 位姿测量策略的可行性。

第 6 章为总结与展望，对本文研究内容进行了总结，通过对本文所提的方法结果进行分析，并指出存在的不足和后续的工作方向。

# 第 2 章 总体方案设计及相关理论基础

## 2.1 引言

面对制造技术的不断发展，双目视觉引导的机器人抓取在日常生活和工业领域应用日益广泛。然而，在实际应用场景中，双目视觉机器人面临着复杂多变的作业环境，如金属工件进行打磨、抛光等加工过程中产生大量粉尘，这些粉尘会干扰双目视觉传感器的正常工作，导致工件局部关键特征信息难以被完整获取，从而显著增加了抓取任务的不确定性。此外，工业场景中的工件种类繁多、形态各异，依靠人工采集图像不仅成本高昂、效率低，且现有公开的工件数据集在样本多样性、场景覆盖度等方面都难以满足实际工业应用的需求。

针对上述难题，本章重点围绕工业复杂环境下双目视觉引导的机器人抓取中存在的问题展开深入分析，采用深度学习方法对复杂环境下的工件抓取检测问题进行研究，提出基于双目视觉的机器人 6DOF 位姿测量总体设计方案。该方案通过数据增强技术与关键点检测算法优化，实现了对目标工件的精准识别与稳定抓取，为工业机器人抓取技术的实际应用提供了可靠的技术支撑。

## 2.2 难点分析与总体方案设计

### 2.2.1 抓取任务难点分析

作为机器人末端执行机构的重要形态，多指灵巧手凭借其高自由度的运动特性优势，在工业自动化领域展现出显著的应用潜力。它通过多点接触机制，可适应不同形状物体的抓取需求，而这种独特的抓取方式不仅能够实现对异形物体的稳定抓持，还可精确控制施力大小和方向，从而确保对目标物体的抓取。

结合上述灵巧手抓取优势，本文通过双目视觉系统识别并检测目标金属工件，精确获取其位姿信息，进而通过机械臂控制系统将抓取位姿信息传递至机器人灵巧手来完成抓取任务。如图 2-1 所示，在实际工业环境进行抓取，金属工件在进行切割、打磨加工时，光线的动态变化导致每类工件都出现了局部反光现象，对抓取任务中的视觉检测与位姿的获取提出了更高的挑战。

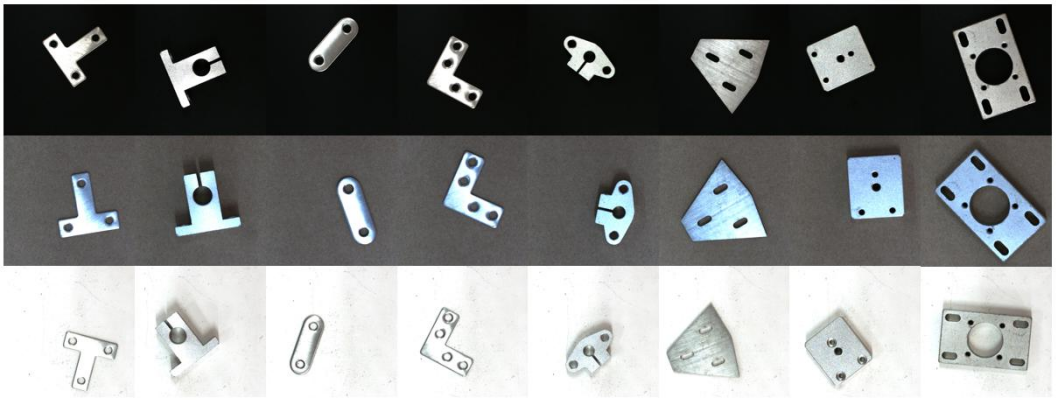


图 2-1 工件局部反光图

针对生产线上基于视觉引导的机器人对随机摆放的金属工件的抓取任务存在以下的难点：

(1)待抓取的金属工件拥有不同的形状、尺寸和材质，因此每类工件均需采集一定数量的图像用作视觉检测的数据依据。但采集丰富的工件图像获取途径少，大部分需要进行人工采集，则需要投入高昂的人工成本。

(2)金属工件在运输过程中存在摩擦、碰撞等情况，导致工件表面出现不同程度的磨损、划痕和腐锈，造成工件表面信息缺失。这些表面缺陷会影响视觉系统对工件特征的识别能力，影响后续机器人定位和抓取操作。

(3)由于工业领域存在粉尘、光线等外在因素的干扰，造成视觉系统对采集的工件图像质量偏低，图像质量的下降影响了工件的清晰度和细节呈现，导致检测系统出现误检或漏检等问题，使机器人抓取任务失败，影响生产效率。

### 2.2.2 总体方案设计

本文以弱纹理的金属工件为研究对象，以实现机器人灵巧手成功抓取目标工件为研究目标，提出了一种基于双目视觉引导的机器人抓取方案，着重研究了获取较高精度的机器人 6DOF 抓取位姿的方法。本文设计的以双目视觉引导的机器人抓取总体方案设计如图 2-2 所示，主要分为以下三个单元模块：

(1)图像数据增强模块：在机器人视觉系统中，高质量的图像数据集对于目标工件识别具有至关重要的作用。然而，人工采集图像不仅成本高昂，且由于工件种类的多样化，使得公开的工件数据集往往难以满足实际应用需求。为解决这一难题，本文提出了一种基于随机算法的数据增强技术。该技术通过背景风格迁移方法，有效生成了多样化状态下的工件图像，为后续视觉检测算法提

供了丰富的工件数据集支撑。

(2)目标对象关键点识别模块：目标工件的精准识别是实现视觉引导机器人抓取的关键，然而，工业现场复杂环境严重影响了视觉系统的成像质量，图像中目标工件存在镜面反光导致工件位姿信息的提取存在显著偏差。针对该难题，本研究提出了一种基于深度学习的工件关键点检测算法。该算法通过精确识别工件表面关键点的空间位姿，有效克服了图像中工件局部信息丢失带来的干扰，为后续机器人抓取位姿的计算提供了数据基础。

(3)机器人 6DOF 抓取模块：双目视觉系统对工件表面关键点检测后，首先采用“眼在手上”安装方式进行相机的内参标定、双目相对位姿获取、手眼标定及坐标转换等任务，根据获取的工件表面关键点位姿信息来计算不同视角下机器人的抓取位姿。其次将该位姿信息与实际抓取位姿进行多组实验对比，分析 6DOF 抓取位姿测量偏差。最后利用获取的抓取位姿信息控制机器人灵巧手进行工件抓取任务证明本文所提抓取策略的可行性。

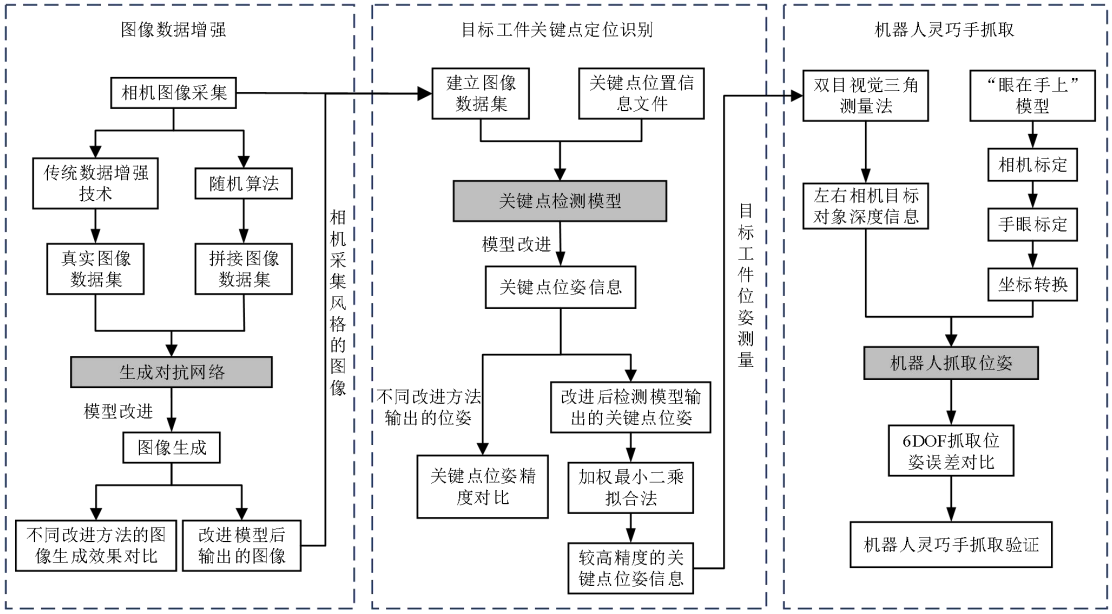



图 2-2 总体方案设计框架

## 2.3 图像数据增强方法

在深度学习中，随着训练样本量的不断增强，模型性能会实现一定程度的提升。然而，在实际应用场景中，构建大规模图像数据集往往需要大量人力资源的投入。因此为降低人工成本，研究者们尝试通过数据增强技术对原始数据进行多样化变换来产生新的训练样本[42]。这种方法不仅实现了训练数据集的扩

增，而且通过增加数据的多样性能提升模型的泛化能力，从而改善模型的整体性能。

### 2.3.1 传统图像数据增强方法

传统数据增强在图像中的使用主要有以下几种方式：

1、几何空间变换

图像的几何空间变换是指将图像中的坐标位置映射到新坐标系的过程，通过改变图像的位置、方向、尺寸或形状来实现空间转换。其中最常见的使用几何变换进行数据增强的方法包含平移、旋转、缩放和翻转[43]。

平移变换是将图像内的所有坐标点进行平移，实现它们所处空间位置的变化。该过程可以用式(2-1)表示，矩阵中的 $dx$ 和 $dy$ 分别表示 $x$、$y$ 轴的平移距离。

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2-1}$$

旋转变换本质上是围绕某特定的点(像常见的图像中心点)旋转图像的过程。在进行图像旋转变换时，其计算如式(2-2)所示。

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{2-2}$$

图像的缩放变换是对图像进行不同程度的缩放变换操作。实际运用时，通常将图像进行离散化处理和插值处理来影响图像特征信息。图像像素点变换如式(2-3)所示，其中 $scale$ 是缩放比例。

$$\begin{cases} x' = x * scale \\ y' = y * scale \end{cases} \tag{2-3}$$

翻转变换是一种基于镜像对称原理的图像几何变换方法，通过沿特定轴线对图像进行反射操作生成新的图像。在计算机视觉领域，翻转变换主要包含两种基本形式：水平翻转和垂直翻转。该技术就是将图像上下或左右的像素位置进行对调。

2、色彩空间变化

图像的色彩变换是通过调整图像色彩实现图像数据扩充的技术，能提升深度学习模型对色彩变化的识别能力。该技术主要基于 RGB 色彩空间进行调整，

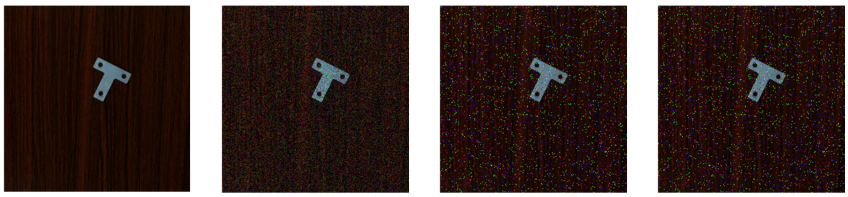
通过对色彩空间中的各个通道进行修改来实现图像色彩的精确控制。常见的色彩空间调整技术包括：亮度、对比度、饱和度调整、色相，如图 2-3 所示，在基于相机采集的原图基础上，通过组合这些调整技术可以生成具有丰富色彩变化的图像样本[44]。

(a)原图　(b)对比度调整　(c)亮度增强　(d)饱和度调整　(e)色相调整

图 2-3 色彩空间调整图像对比

3、随机噪声

随机噪声注入是一种重要的数据增强策略，可以用来增加数据集的多样性和数量，增强卷积神经网络对噪声干扰的适应能力和泛化性能[45]，如图 2-4 展示了在图像中增添随机噪声的效果图，其中包括常见的随机噪声包括：高斯噪声、椒盐噪声和均匀噪声。



(a)原图　(b)高斯噪声　(c)椒盐噪声　(d)均匀噪声

图 2-4 添加随机噪声图像

### 2.3.2 生成对抗网络

基于深度学习的数据增强方法主要是使用生成对抗网络(GAN)，该模型通过学习和模拟输入图像数据的特征分布，能够生成新的图像样本，属于半监督学习范畴，在图像生成、风格迁移和图像翻译等领域具有重要应用价值[46]。

GAN 由 Goodfellow[47]于 2014 年提出的，作为一种创新性的生成模型，它的出现引起了业界的广泛关注，并成为计算机视觉领域研究和应用的热点。基于零和博弈理论，生成对抗网络通过对抗训练机制学习图像数据特征，生成与实际采集图像近似的虚假图像。该图像生成模型主要由生成网络 G 与判别网络 D 构成，生成对抗网络示意图如图 2-5 所示。

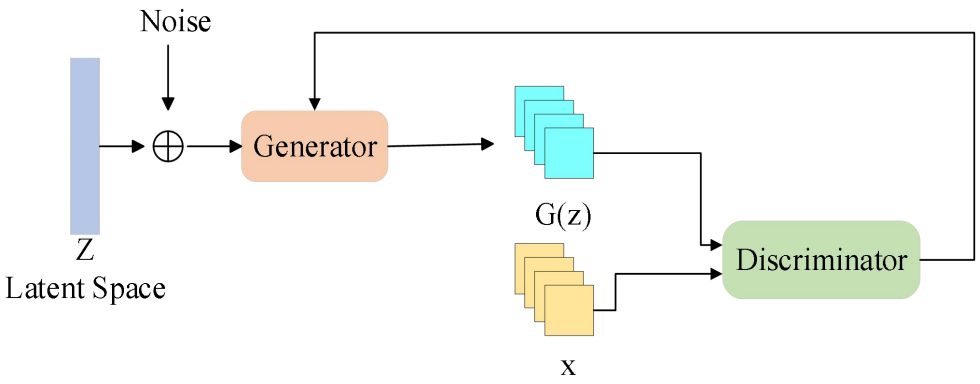



图 2-5 生成对抗网络示意图

上图 2-5 中噪声是从一个均匀分布或高斯分布中采样得到的向量，GAN 通过输入噪声将其映射入数据空间中，再凭借生成器输出生成的数据，生成与真实数据分布相近的数据。GAN 的核心在于生成器 G 与判别器 D 之间通过对抗博弈来学习真实数据的分布，生成接近于实际采集的图像效果。简而言之，生成器 G 主要生成与真实图像数据集类似的假图片，而判别器 D 则是严格地分辨真实图片和生成的假图片。为实现生成数据的高度逼真，生成器不断更新模型对训练图像数据集学习，逐步缩小生成的图像与实际采集的图像在数据分布上的差异。同时判别器也不断优化自身的参数，保证其更好的辨别生成器生成的假图片。GAN 网络的优化目标函数可表示为：

$$\min_{G}\max_{D} V_{GAN}(G,D)=E_{x\ P_{data}(x)}\left[\log D(x)\right]+E_{z\ P_{z}(z)}\left[f\left(G(z)\right)\right] \tag{2-4}$$

上式(2-4)中，$E$ 表示数据分布的期望值，$x$ 为实际采集的图像数据，服从真实数据 $P_{data}$ 的分布，而 $z$ 为输入的噪声数据，服从分布 $P_z$。显然，生成器 G 希望 $D\left(G\left(Z\right)\right)$ 尽可能地大，这时目标函数 $V_{GAN}\left(D,G\right)$ 会变小。与之相反，判别器 D 在判别过程中使 $D\left(G\left(Z\right)\right)$ 趋向于 0，此时的目标函数 $V_{GAN}\left(D,G\right)$ 就会越大。通过在训练过程中不断地对抗博弈来提高网络的性能，当且仅当生成的数据分布与真实数据分布趋于一致时，就能达到平衡状态。

## 2.4 基于深度学习的目标检测算法

基于深度学习的目标检测算法的两个基本研究方向是 Single-Stage 检测框架和 Two-Stage 检测框架，并且近些年有转变为基于 anchor 的检测框架和基于 anchor-free 的检测框架等两大类的趋势[48]。

### 2.4.1 基于 anchor 的目标检测算法

基于锚框的目标检测算法是计算机视觉领域最具代表性算法之一，覆盖了两阶段检测方法和一阶段检测方法两大类[49]。

目前，经典的一阶段目标检测算法应该是 YOLO 系列[50-51]和 SSD[52]，其中 SSD 算法是由刘伟在 2016 年提出，具有较快的检测速度以及较好的检测精度。SSD 算法的设计理念可以概括为：基于多尺度特征图检测、利用卷积的方法来检测和设置先验框[53]。而 YOLO 是一个简单且不复杂的对象检测模型，它通过用神经网络处理一次图片即可完成目标检测任务，包括目标类别的识别和位置信息的预测。尽管在算法计算速度上 YOLO 有着显著的提升，但其划分网格的方式比较粗糙，导致在小目标检测任务中的性能表现较为局限[54]。然而，YOLO 系列的提出为目标检测提供思路突破，实现了实时目标检测的可行性。

两阶段检测方法的经典算法 FasterR-CNN 是由 Shaoqing Ren 等人于 2015 年被提出，该算法是在 FastR-CNN 算法[55]基础之上进行改进的，通过 RPN 来生成锚框并且剔除部分冗余和无用的锚框从而减少了计算量，比原始模型 FastR-CNN 训练速度更快。FasterR-CNN 网络的结构如图 2-6 所示，主要分为：Conv layers、Region Proposal Networks(RPN)、Roi Pooling 和 Classification 四个部分。其中 RPN 是 Faster-RCNN 的核心部分，用于生成候选框。从图 2-7 可以看出，该网络结构采用双分支设计，实现对锚框正、负样本分类和回归锚框的偏移量，最终在“Proposal”层进行整合。

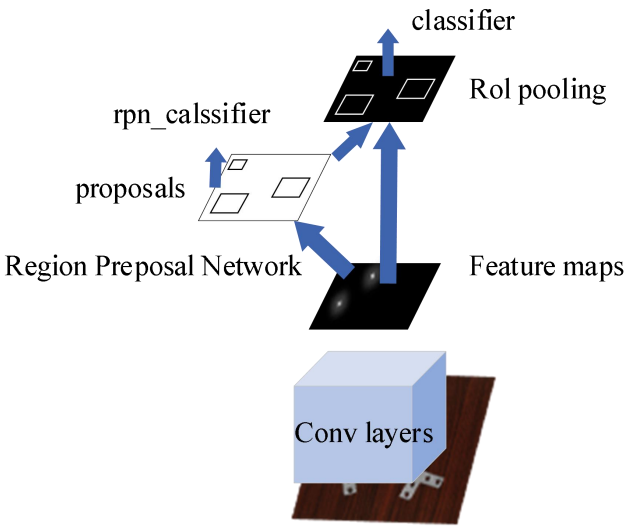



图 2-6 FasterR-CNN 网络结构图

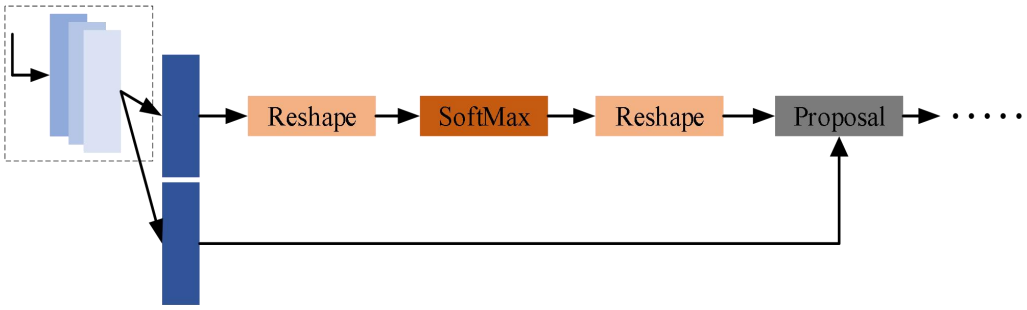



图 2-7 RPN 结构

### 2.4.2 基于 anchor-free 的目标检测算法

和传统的 anchor-based 的方法相比，anchor-free 目标检测算法摒弃了复杂的锚点机制，无需预先设置锚点的尺度、长宽比等人工超参数，从而简化了模型设计。在检测部分，此类算法直接预测目标的关键点或中心点，可以更灵活地处理不同大小和形状的目标，使检测更加的高效和精确[56]。

为解决传统目标检测算法中因引入锚点机制而导致的超参数敏感、计算复杂度高等问题[57]，Law 等人[58]提出了 CornerNet 目标检测算法。CornerNet 算法的网络结构如图 2-8 所示，网络使用一对角点的方式替代了传统的 anchor 方式，即将目标检测任务重新定义为对目标边界框左上角和右下角两个关键点的预测问题，并在网络中引入了角点池化层，以增强网络对角点特征的提取能力，从而更精确地定位目标边界。但 CornerNet 在角点匹配过程需要消耗大量计算资源，降低了检测速度，在对目标对象关键点检测过程中存在误检、漏检现象。

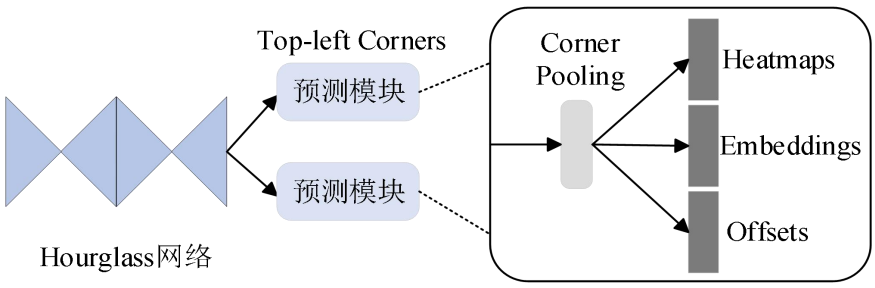



图 2-8 CornerNet 网络结构

Zhou 等人[59]于 CornerNet 的思想上提出 ExtremeNet 算法，进一步推动了基于关键点的目标检测方法的发展。ExtremeNet 算法的网络结构如图 2-9 所示，与 CornerNet 不同的是，该算法通过检测目标最上、最下、最左、最右四个极值点和中心点来实现对目标物体的识别。具体而言，该算法通过预测目标的四

个极值点和中心点，并利用纯几何匹配算法对这些点进行组合。当获取的这几个关键点坐标位置越接近于实际点位置，则归为一组并归纳为候选目标框。最后，通过验证组合中是否存在中心极值点来筛选出更为精准的目标框。

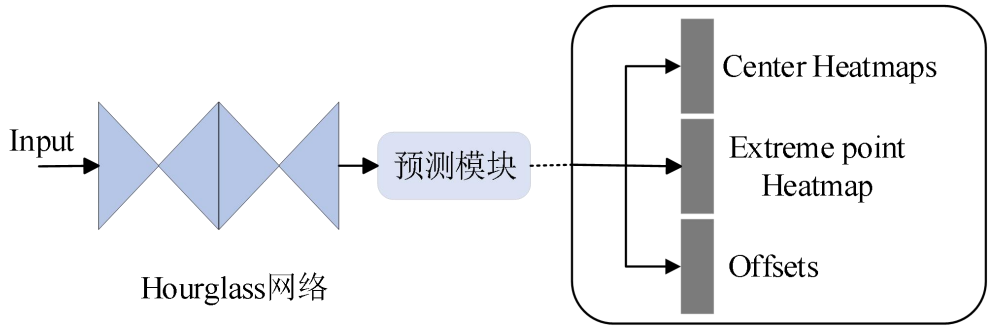



图 2-9 ExtremeNet 网络结构示意图

Zhou 等人[60]针对 CornerNet 在关键点分组配对过程中计算效率低下的问题，提出 CenterNet 关键点检测算法，其结构如图 2-10 所示。该算法将目标检测任务重新定义为目标中心点的检测问题，即通过关键点预测定位目标的中心点，并将该点处的图像特征直接回归目标的边界框尺寸和位置信息，从而避免了 CornerNet 中复杂的关键点配对过程。与其它关键点检测算法相比，CenterNet 使用高分辨率特征图，减少了模型训练时存在的特征点信息损失，提升网络的检测能力。

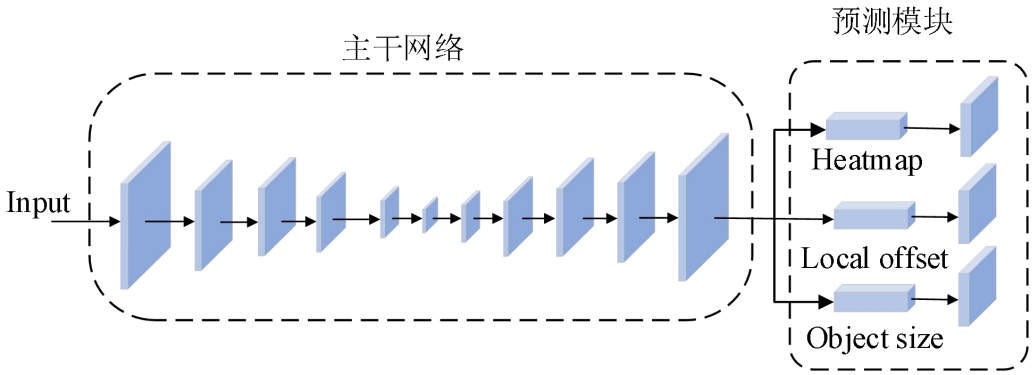



图 2-10 Centernet 网络结构示意图

上述基于 anchor-free 的目标检测算法避免了复杂的锚框超参数设计(如尺度、长宽比等)以及耗时的锚框匹配过程，显著简化了检测流程。相比其它 Anchor-free 算法，CenterNet 算法模型简单，在检测精度与速度之间实现了更好的平衡，可以满足视觉检测算法在复杂工业环境中获取较高精度的目标对象位姿的需求。因此，本文将 CenterNet 算法应用于金属工件的多关键点检测算法的研究中，进一步提升算法在复杂环境干扰下的检测性能。

## 2.5 双目视觉系统与相机标定

### 2.5.1 双目视觉系统模型

双目视觉系统通过配置两个具有固定位置关系的视觉传感器，同步采集同一场景的左右相机采集的图像，结合两个相机的内外参数和相对位姿，成功获取场景中目标对象的深度信息，实现对目标的三维重建。而双目视觉测量就是模拟人类视觉的测量方法，通过双眼从不同视角来观测客观的三维空间，由于几何光学的作用，两只眼睛所看到的图像会存在一定的差异，即所谓的视差[61]。

双目视觉系统模型如图 2-11 所示。两个相机之间的间距称为基线距离 $T$，且焦距均为 $f$。由于左右相机空间位置的差异，所以在进行图像采集时，图像中目标对象上的同一点 $P$ 在左、右相机中的成像位置会存在差异，左右相机上点 $P$ 的像素坐标分别为 $p^l(x_l, y_l)$ 和 $p^r(x_r, y_r)$，视差 $d$ 则为左右相机采集的图像出现重叠时 $p^l$ 和 $p^r$ 之间的距离。

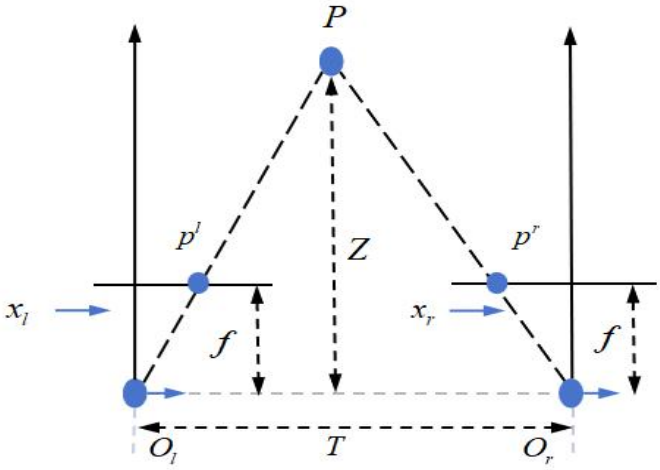



图 2-11 双目视觉系统模型

其中 $O_l$ 和 $O_r$ 是左右两个相机的位置，$P$ 是空间中一点，$P$ 在相机 $O_l$ 中的成像点是 $p^l$，在相机 $O_r$ 中的成像点是 $p^r$，线段 $x_l$ 和 $x_r$ 分别为 $p^l$ 和 $p^r$ 到相机成像面边界的距离，则点 $P$ 在左右相机中的视差 $d$ 为：

$$d = |x_l - x_r| \tag{2-5}$$

假设 $p^l$ 和 $p^r$ 两点之间的距离为 $s$，$T$ 为左右两相机之间的基线距离，则由上图 2-11 可知距离 $s$ 为：

$$s = T - \left( x_l - x_r \right) \tag{2-6}$$

由三角形相似原理可得：

$$\frac{T - \left( x_l - x_r \right)}{Z - f} = \frac{T}{Z} \tag{2-7}$$

其中，$f$ 为相机的焦距，$Z$ 为空间点 $P$ 到两个相机基线的距离，则解得距离 $Z$ 为：

$$Z = \frac{fT}{x_l - x_r} \tag{2-8}$$

获得 $P$ 点相对于视觉系统的深度信息后，利用二维相机获得的图像信息，可获得 $P$ 点的三维坐标。当获得工件表面 $P_1$、$P_2$、$P_3$ 三个点的三维坐标时，可将其转化为两个向量 $\overrightarrow{P_1P_2}$、$\overrightarrow{P_1P_3}$，通过向量积可求得工件的三维位姿。

### 2.5.2 双目相机标定

双目视觉系统的特征点匹配需要以相机标定作为前提。标定过程是通过建立空间点的三维坐标与其在图像平面上的投影坐标之间的映射关系，从而确定相机的内外部参数。这一过程对于提高视觉测量精度至关重要，因此吸引了众多学者的深入研究。其中 Zhang 等人[62]提出了基于单平面棋盘格的标定方法，该方法巧妙地结合了自标定法和传统标定法，使用二维棋盘格作为标定物，取代了传统的三维标定物体，这种设计不仅简化了标定物的制造和存储过程，还提高了使用的便捷性，同时精度相比自标定法更高，因此工业界和学术界都广泛接受该标定方法。

对于标定板上任意一点 $P$ 有：

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = sK \begin{pmatrix} r_1 & r_2 & r_3 & t \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ 0 \\ 1 \end{pmatrix} = sK \begin{pmatrix} r_1 & r_2 & t \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ 1 \end{pmatrix} \tag{2-9}$$

上式(2-9)中，$K$ 为相机的内参数矩阵，相机的位姿 $T = \left[ R, t \right]$，其中 $R = \left[ r_1, r_2, r_3 \right]$。如果令 $H = \left[ h_1, h_2, h_3 \right] = sK \left[ r_1, r_2, t \right]$，将其中各个分量进行提取可以得到：

$$\begin{cases} h_1 = sKr_1 \\ h_2 = sKr_2 \\ h_3 = sKt \end{cases} \tag{2-10}$$

记 $\lambda = 1/s$，可以得到：

$$\begin{cases} r_1 = \lambda K^{-1} h_1 \\ r_2 = \lambda K^{-1} h_2 \\ r_3 = \lambda K^{-1} h_3 \end{cases} \tag{2-11}$$

根据旋转矩阵的正交特性，可知 $r_1, r_2, r_3$ 三个向量均为单位向量且两两正交，将 $r_1^T r_2 = 0$ 和 $\|r_1\| = \|r_2\|$ 将其带入式(2-11)中得到：

$$\begin{cases} h_1^T K^{-T} K^{-1} h_2 = 0 \\ h_1^T K^{-T} K^{-1} h_1 = h_2^T K^{-T} K^{-1} h_2 \end{cases} \tag{2-12}$$

由式(2-12)可知，为了推导方便，定义对称矩阵 $B = K^{-T} K^{-1}$ 为一个 3×3 矩阵，其中各元素用 $B_{ij}$ 表示，即：

$$b = \begin{pmatrix} B_{11} & B_{12} & B_{22} & B_{13} & B_{23} & B_{33} \end{pmatrix} \tag{2-13}$$

通过理论坐标和实际坐标之间的关系和式(2-13)再经过一定的推导可以得到式(2-14)。

$$h_1^T B h_j = v_{ij}^T b \tag{2-14}$$

其中，

$$v_{ij} = \begin{pmatrix} h_{1i}h_{1j} & h_{1i}h_{2j} + h_{2i}h_{1j} & h_{2i}h_{2j} & h_{3i}h_{1j} & h_{3i}h_{2j} + h_{2i}h_{3j} & h_{3i}h_{3j} \end{pmatrix}^T \tag{2-15}$$

根据 $v_{ij}$ 的定义，可以将式(2-14)改写为：

$$\begin{pmatrix} v_{12}^T \\ v_{11}^T - v_{22}^T \end{pmatrix} b = 0 \tag{2-16}$$

当有 $n$ 幅图像时可以得到式(2-17)所示的方程组。

$$\begin{pmatrix} v_{12}^T \\ \left(v_{11}^T - v_{22}^T\right)^T \\ M \\ \left(v_{11}^n\right)^T \\ \left(v_{11}^n - v_{22}^n\right)^T \end{pmatrix} b = 0 \tag{2-17}$$

在求出矩阵 $B$ 以后，使用 cholesky 分解，可以得到内参数的求解表达式(2-18)。

$$\begin{cases} v_0 = -\left(B_{12}B_{13} - B_{11}B_{23}\right) / \left(B_{11}B_{22} - B_{12}^2\right) \\ t = B_{33} - \left(B_{13}^2 + v_0\left(B_{12}B_{13} - B_{11}B_{23}\right)\right) / B_{11} \\ f_x = \sqrt{t / B_{11}} \\ f_x = \sqrt{tB_{11} / \left(B_{11}B_{22} - B_{12}^2\right)} \\ s = B_{12}f_x^2 f_y / t \\ u_0 = sv_0 / f_y - B_{13}f_x^2 / t \end{cases} \tag{2-18}$$

外参数可以根据单应性矩阵的定义得到：

$$\begin{cases} r1 = \lambda K^{-1}h_1 \\ r2 = \lambda K^{-1}h_1 \\ r3 = \lambda K^{-1}h_1 \\ t = \lambda K^{-1}h_1 \end{cases} \tag{2-19}$$

由于本文运用双目视觉系统，当完成单个相机的标定外，还需要获取对左右相机之间的相对位置。因此通过单目标定可得到左右相机在世界坐标中的位姿 $T_{wl} = \left[R_{wl}, t_{wl}\right]$ 和 $T_{wr} = \left[R_{wr}, t_{wr}\right]$，进而可以直接通过坐标变换得出：

$$\begin{pmatrix} x_l \\ y_l \\ z_l \end{pmatrix} = T_{wl}^{-1}T_{xr}\begin{pmatrix} x_r \\ y_r \\ z_r \end{pmatrix} = R_{wl}^{-1}R_{wr}\begin{pmatrix} x_r \\ y_r \\ z_r \end{pmatrix} + R_{wl}^{-1}t_{wr} - t_{wl} \tag{2-20}$$

由上式(2-20)可知：

$$\begin{cases} R_{lr} = R_{wl}^{-1}R_{wr} \\ t_{lr} = R_{wl}^{-1}t_{wr} - t_{wl} \end{cases} \tag{2-21}$$

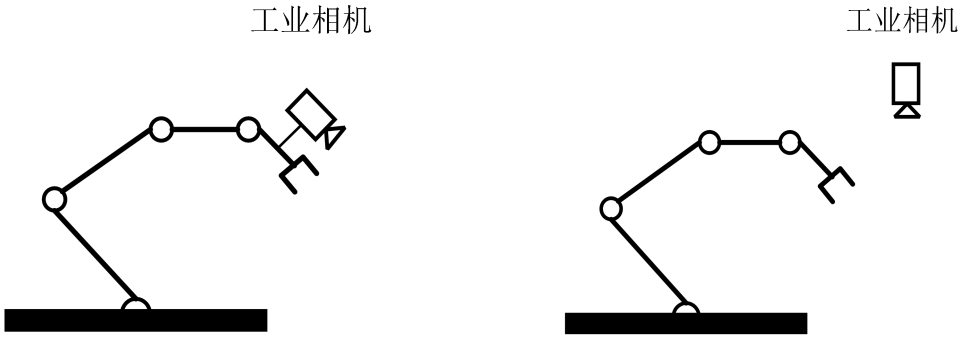
其中 $R_{lr}$、$t_{lr}$ 分别是左相机到右相机的旋转和平移，右相机中坐标转换到左相机坐标系下变换矩阵为 $T_{lr} = \left[R_{lr}, t_{lr}\right]$。

### 2.5.3 手眼标定

1、手眼标定原理

根据相机的安装位置，机器人手眼视觉模型可以分为眼到手(eye-to-hand)和眼在手(eye-in-hand)[63]。如图 2-12(a)所示，眼在手上模型通常是将相机安装在

机械臂上，此时相机会随着机械臂的运动而同步移动。相反，眼在手外模型则是将相机安装在外部固定支架上，如 2-12(b)图所示。这两种模型都是将相机与机械臂之间的不变量确定下来，并借助标定技术计算出两者之间的转换矩阵，从而获得机械臂末端坐标系与相机坐标系之间的坐标转换关系。




(a)眼在手上模型(eye-in-hand)　　(b)眼在手外模型(eye-to-hand)

图 2-12 模型安装图

在上述两种模型中，eye-in-hand 模式下的相机可以随着机械臂的运动而移动，能够动态调整视角，适应不同的任务需求。这种灵活性在复杂场景下的抓取中尤为重要，因此本文选用眼在手上模式进行标定，其中不同坐标系的转换关系如式(2-22)。

$$\left(T_{\text{bace}}^{tool}\right)_{(i)} T_{cam}^{base} \left(T_{cal}^{cam}\right)_{(i)} = \left(T_{base}^{tool}\right)_{(i+1)} T_{cam}^{base} \left(T_{cal}^{cam}\right)_{i+1} \tag{2-22}$$

式中：$\left(T_{bace}^{tool}\right)_i$——第 i 组位置中，机械手末端与基座的坐标转换关系；

$\left(T_{bace}^{tool}\right)$——相机与机器人基座的坐标转换关系；

$\left(T_{cal}^{cam}\right)_i$——第 i 组位置中，标定板和相机的位姿转换关系。

对式(2-22)进行转换，其结果如式(2-23)所示：

$$\left(T_{bace}^{tool}\right)_{(i-1)}^{-1} \left(T_{bace}^{tool}\right)_i T_{cam}^{bace} = T_{cam}^{bace} \left(T_{cal}^{cam}\right)_{i+1} \left(T_{cal}^{cam}\right)_i^{-1} \tag{2-23}$$

记 $A = \left(T_{bace}^{tool}\right)_{(i+1)}^{-1} \left(T_{bace}^{tool}\right)_i$，$B = \left(T_{obj}^{cam}\right)_{i+1} \left(T_{obj}^{cam}\right)_i^{-1}$，$X = T_{cam}^{bace}$，则式(2-23)可写为：

$$AX = XB \tag{2-24}$$

其中，$A$ 表示在机器人运行两次后，这两次末端执行器的位姿关系；$B$ 表示这两次采集的图像中标定板的位姿关系；$X$ 便是手眼标定后获取到的相机坐标系与机器人末端执行器之间的关系。

通过上式(2-24)得出 $X$ 值后，将相机坐标系下的物体位姿转换到机器人坐标

系下，计算公式如式(2-25)所示：

$$T_{obj}^{bace} = T_{cam}^{base} T_{obj}^{cam} \tag{2-25}$$

2、手眼标定实验

本文针对工程实际应用场景搭建 eye-in-hand 工业机器人视觉系统，其手眼标定实验装置如图 2-13 所示。在进行手眼标定实验时，首先需要将圆点标定板固定在抓取平台上，相机固定在机器臂的双目支架上面。然后不断变换工业机器人末端的位姿，使得工业相机以不同姿态采集位置固定不变的圆点标定板图像，同时记录此时工业机器人示教器上中相对应的位置信息。最后过利用 Halcon 求解机器人的基座标基坐标与相机坐标的关系。

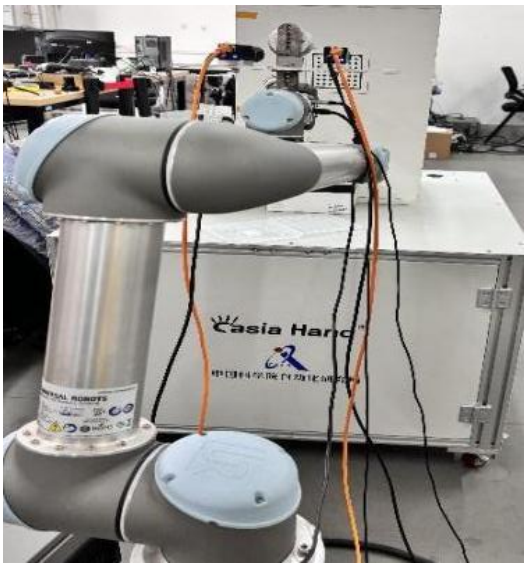


图 2-13 手眼标定实验装置图

在进行手眼标定前需获取左右相机内参外参，本文通过 2.3 节的相机标定获取了左右相机内参外参，具体参数如表 2-1 所示。同时并该参数作为输入数据进行手眼标定实验和计算左右相机之间的相对位姿关系，为机器人 6DOF 位姿测量实验提供数据支持。其中左相机相对于右相机的位姿具体数值见表 2-2。

表 2-1 左右相机参数

| | *f/mm* | k | Sx | Sy | Cx | Cy | W | H |
|---|---|---|---|---|---|---|---|---|
| 左相机 | 0.0121667 | -496.459 | 4.79942e-06 | 4.8e-06 | 619.02 | 541.538 | 1280 | 1024 |
| 右相机 | 0.0121487 | -480.042 | 4.79902e-06 | 4.8e-06 | 622.877 | 531.058 | 1280 | 1024 |

表 2-2 左相机相对于右相机的位姿

| 参数变量 | X(m) | Y(m) | Z(m) | A(°) | B(°) | C(°) |
|---|---|---|---|---|---|---|
| 数值 | 0.121671 | 0.000254675 | 0.0186007 | 359.358 | 346.966 | 359.772 |

为确保标定实验的精度满足要求，共采集了多组标定板图像数据。实验过程中拍摄的标定板图像如图 2-14 所示，共 16 张圆点标定板图像。基于采集的图像数据和位姿信息，采用 Halcon 视觉处理软件进行系统标定，获得的标定结果见表 2-3。

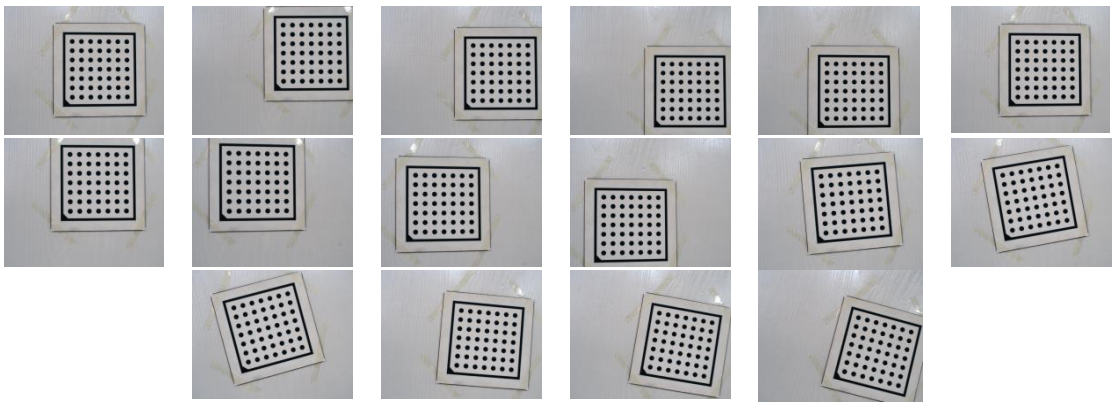


图 2-14 圆点标定板实验采集图像

表 2-3 手眼标定结果

| 标定变量 | *X*(m) | *Y*(m) | *Z*(m) | *A*(°) | *B*(°) | *C*(°) |
|---|---|---|---|---|---|---|
| 变量值 | 0.0672135 | -0.105518 | 0.00509049 | 0.131666 | 358.447 | 359.458 |

手眼标定的均方差和最大误差值结果如表 2-3 所示。

表 2-4 手眼标定误差

| | 均方差 | 最大误差 |
|---|---|---|
| 平移向量(mm) | 0.151 | 0.245 |
| 旋转矩阵(°) | 0.058 | 0.145 |

### 2.5.4 坐标系的建立与转换

立体空间中的点 $P_w$ 投影到二维平面上点 $P$ 需要经历一系列坐标变换过程，主要包括四大坐标系的之间的转换。

(1)由世界坐标系到相机坐标系：点 $P_w\left(x_w,y_w,z_w\right)$ 到 $P_c\left(x_c,y_c,z_c\right)$。空间一点 $P_w$ 转换到 $P_c$ 有：

$$P_c = RP_w + t \tag{2-26}$$

$$R(\alpha,\beta,\gamma)=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\alpha & -\sin\alpha\\ 0 & \sin\alpha & \cos\alpha\end{bmatrix}\begin{bmatrix}\cos\beta & 0 & \sin\beta\\ 0 & 1 & 0\\ -\sin\beta & 0 & \cos\beta\end{bmatrix}\begin{bmatrix}\cos\gamma & -\sin\gamma & 0\\ \sin\gamma & \cos\gamma & 0\\ 0 & 0 & 1\end{bmatrix} \tag{2-27}$$

$$t=\left(t_x,t_y,t_t\right) \tag{2-28}$$

其中 $R$ 为旋转矩阵，$t$ 为平移向量，$(\alpha,\beta,\gamma)$ 则分别表示绕 $X_C$、$Y_C$、$Z_C$ 轴的旋转角度，$\left(t_x,t_y,t_t\right)$ 分别表示沿 $X_C$、$Y_C$、$Z_C$ 轴方向的平移距离。通过齐次坐标表示为：

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}=\begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{2-29}$$

(2)由相机坐标系到图像坐标系：$P_c\left(x_c,y_c,z_c\right)$ 到 $p(x,y)$。依据针孔模型可得：

$$\begin{cases} x=\dfrac{f\cdot x_t}{z_c} \\ y=\dfrac{f\cdot y_c}{z_c} \end{cases} \tag{2-30}$$

其中 $(x,y)$ 为理想情况下图像坐标系下对应的投影点的坐标，齐次坐标表示则有：

$$z_c\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}=\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{2-31}$$

(3)图像坐标系到像素坐标系：$p(x,y)$ 到 $(u,v)$。图像坐标系与像素坐标系下的点有如下关系：

$$\begin{cases} u=\dfrac{x}{d_x}+u_0 \\ v=\dfrac{y}{dy}+v_0 \end{cases} \tag{2-32}$$

其中 $dx$ 代表 $x$ 轴方向一个像素的宽度，$dy$ 代表 $y$ 轴方向上一个像素的高度。$(u_0,v_0)$ 称为图像平面的主点。用齐次坐标表示为：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}=\begin{bmatrix} 1/dx & 0 & u_0 \\ 0 & 1/dx & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2-33}$$

则相机模型的投影关系如式(2-34)：

$$Z_c\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}f/dx & 0 & u_0 & 0\\0 & f/dy & v_0 & 0\\0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}R & t\\0 & 1\end{bmatrix}\begin{bmatrix}x_w\\y_w\\z_w\\1\end{bmatrix}=M_1M_2\begin{bmatrix}x_w\\y_w\\z_w\\1\end{bmatrix} \tag{2-34}$$

其中，$M_1$为相机的内部参数，$M_2$为相机的外部参数。

## 2.6 本章小结

本章介绍了基于双目视觉引导的机器人抓取的整体方案设计，首先阐述了数据增强技术，从传统图像数据增强方法到基于生成对抗网络的图像生成技术，系统性地剖析了其原理与应用。其次，在目标检测算法部分讲述了基于 anchor 和基于 anchor-free 的检测算法。最后介绍了机器人抓取的相机标定、手眼标定以及坐标转换，在手眼标定实验中选择眼在手上模型，通过该实验，获取机械臂末端与相机之间的转换关系。

# 第 3 章 基于生成对抗网络的图像生成

## 3.1 引言

在工业环境中，光照条件复杂多变，常出现光线不足、过度曝光等问题，导致采集的图像可见度和画面质量明显降低问题。而进行相机采集图像时，金属工件的材质特性易产生局部反光现象，这些因素叠加使得图像丢失局部细节和降低全局对比度，不仅会极大地影响到产线工作人员的主观感受，还会降低视觉算法性能。此外采集大量高质量图像数据集需要投入高昂的人力成本，而金属工件的类型多样、尺寸丰富，使得公开数据集往往难以满足实际需求。因此为了获取多样化的工件图像，本文引入数据增强技术实现对有限数据集的扩增，为视觉算法提供更丰富的训练样本。

目前，对于图像的数据集扩充主要包括基于传统方法和基于生成对抗网络的生成式扩充两种[64]。传统扩充方法通过对图像的像素灰度值和像素位置进行线性调整，虽然能够在一定程度上增加数据量，但其仅能基于现有图像进行变换。相比之下，生成对抗网络能够能够生成具有高度真实感的图像样本，其生成效果更贴近实际视觉成像特征。

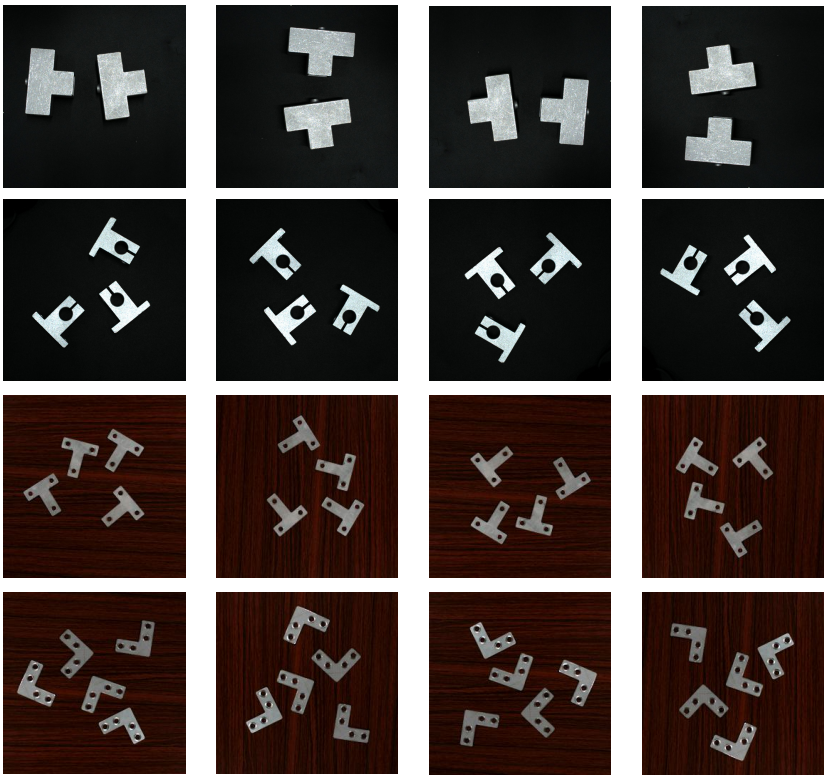
基于上述技术特点，本文提出了一种基于生成对抗网络 CycleGAN 的图像数据集扩充方法，该方法将随机算法与数据增强技术相结合，使得模型能够利用有限的数据集来进行图像风格迁移，生成与真实相机采集图像相似的图像。此外，本文对 CycleGAN 模型进行了改进，通过优化网络结构，提升模型在工件图像风格转换任务中的性能表现。使其生成图像的质量和细节表现更加接近真实场景，为抓取任务中的目标检测提供了高质量的图像支持。

## 3.2 基于传统图像增强算法的图像生成

基于生成对抗网络的数据增强模型需要以一定数量的数据集作为输入，因此在相机拍摄的有限图像数据集的基础上，采用传统的数据增强方法对原始图像进行变换操作，构建工件图像数据集用于后续实验训练。本节将以不同类型金属工件图像数据为例，进行基于传统图像扩充技术给图像处理。

### 3.2.1 基于旋转变换的图像增强

旋转增强作为一种广泛应用的数据扩充技术，其核心在于通过沿任意轴向对图像进行随机角度的旋转来生成多样化的训练样本。当图像进行某一角度旋转时存在局部空缺，图像算法会对该空缺部分进行填充，但生成的图像与实际拍摄图像存在像素值差异。针对这一问题，本文采用了 90°、180°和 270°这三个特定角度的旋转增强策略，使得生成的图像背景更接近实际拍摄效果，其效果如图 3-1 所示。可以直观地观察到图像在不同旋转角度下图像不仅呈现出显著的形态和结构差异，而且金属工件也展示了多样化的空间位置状态。这种处理方式不仅有效扩充了图像数据集的规模，更重要的是保证了增强后图像的逼真度，扩充了一定量的图像数据集。



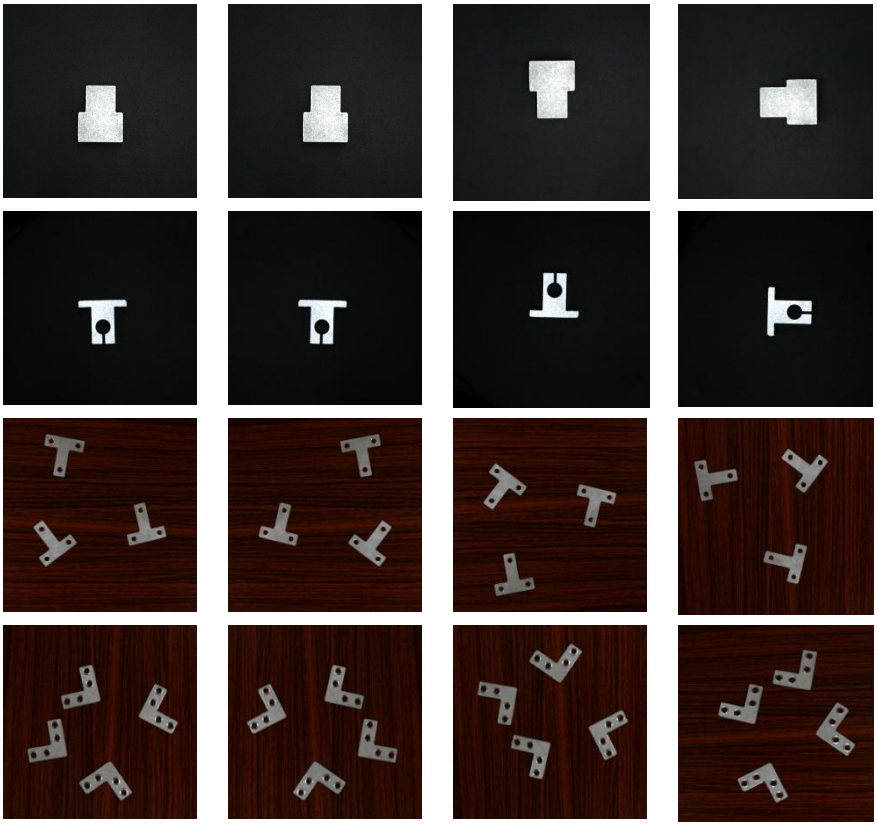
(a)原图　(b)旋转 90°　(c)旋转 180°　(d)旋转 270°

图 3-1 图像多角度旋转

### 3.2.2 基于翻转变换的图像增强

常见的翻转是通过将图像在水平或垂直方向上翻转 180°来生成新的训练样本，除了这基本翻转外还存在对角线翻转、任意角度翻转等更为复杂的翻转技

术，如图 3-2 展示图像多方向翻转效果图，可以看出经多方向翻转处理后，图像在保持原有特征的同时获得了多样化的视角表达，这种处理方式能提升了模型对图像方向的识别能力。



(a)原图　　(b)左右翻转　　(c)上下翻转　　(d)对角线翻转

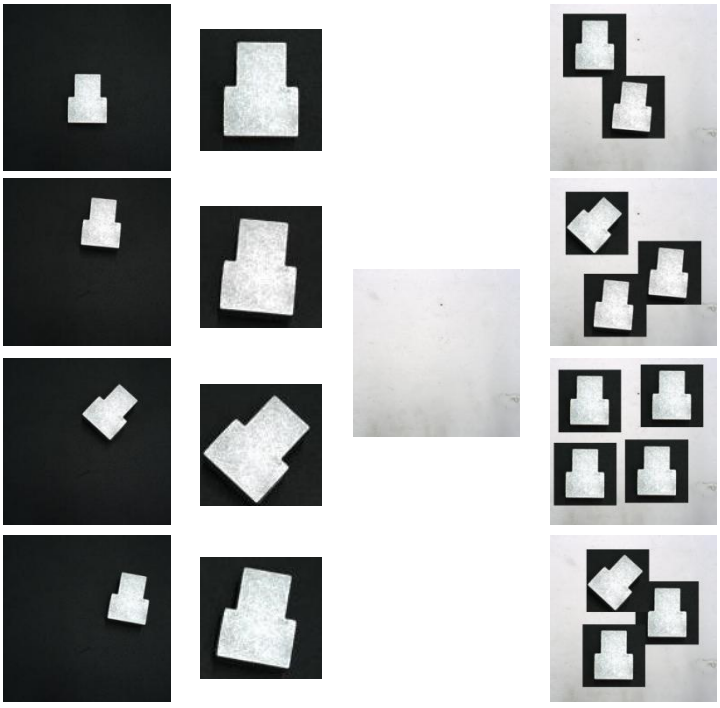
图 3-2 图像多方向翻转

### 3.2.3 基于随机算法的图像增强

裁剪也是一种常用的数据扩充技术，其核心在于从原始图像中提取具有不同尺寸和形状的子区域，以此生成新的图像。本文主要应用裁剪与拼接技术的结合来获取随机摆放状态下的工件图像。其中裁剪部分应用了模板匹配算法，精准定位并裁剪原始图像中包含目标对象的区域，具体而言，模板匹配算法通过在原始图像中搜索与给定模板具有最高相似度的区域，从而确定目标对象的位置，实现对该区域的精确裁剪。而拼接部分则结合了随机算法，首先对裁剪的工件图进行随机选取，接着对选取的图像进行随机位置的拼接合成，最终得到了图 3-3 中(d)拼接图像，具体裁剪效果如图 3-3 所示。

通过上述三种方法，对相机采集的有限数据集进行了多维度变换，有效扩

充了金属工件图像数据集。具体而言，基于传统图像增强算法，对工件目标图像进行角度旋转、翻转及裁剪拼接等几何变换，显著增加了数据样本的多样性，为后续 CycleGAN 模型的图像生成提供了高质量的训练基础。



(a)原图　(b)裁剪图　(c)背景图　(d)拼接图

图 3-3 结合随机算法图像拼接

## 3.3 基于生成对抗网络 CycleGAN 的图像生成

### 3.3.1 CycleGAN 网络结构

2017 年由 Zhu 等人[65]提出了一种无监督生成对抗网络 CycleGAN，通过引入了循环一致性的思想，成功实现了无监督的图像翻译任务。传统基于深度学习的图像翻译方法通常依赖于大量精确配对的训练数据，然而在实际应用场景中，配对数据集往往很难获取。在艺术风格迁移任务中，难以获得足够数量的莫奈真迹及其对应的真实场景图像。但 CycleGAN 的设计突破了这一限制，实现了在非配对数据集上进行有效的训练，丰富了图像翻译技术的应用。

CycleGAN 整体结构如图 3-4 所示，模型包涵两个生成器网络 $G$、$F$，两个判别器网络 $D_X$ 、 $D_Y$ 。CycleGAN 的核心目标是在保持图像内容特征不变的前提下，实现源域 $X$ 和目标域 $Y$ 之间的图像风格迁移。在本文研究中，生成器网络 $G$ 负责将源域 $X$ 中的拼接的工件图像 $x$ 转换为相机采集的真实图像风格 $Y$ 的

图像，而判别器网络 $D_Y$ 则用于评估生成的图像与目标域 $Y$ 的真实图像在风格特征上的相似度。通过这种对抗性训练机制，CycleGAN 能够在不依赖配对数据的情况下，实现高质量的图像风格转换。

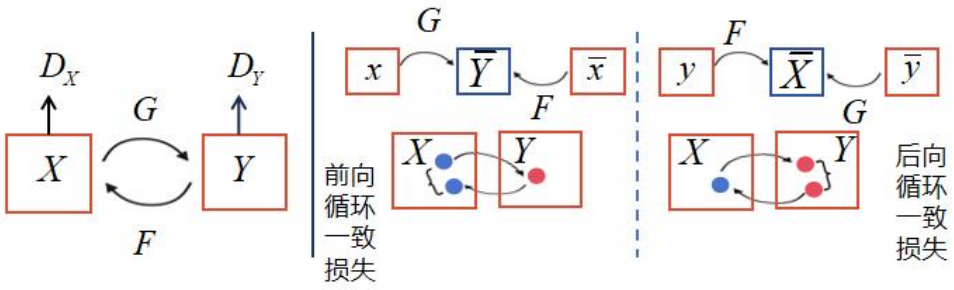



图 3-4 CycleGAN 工作原理图

在本文的 CycleGAN 模型的架构设计中，原域 $X$ 中的图像 $x$ 是自建的工件数据集中的拼接图像，而目标域 $Y$ 中的样本 $y$ 是由视觉系统采集的真实图像。CycleGAN 模型通过不断的学习将 $x$ 转换成目标域 $Y$ 中的样本 $y$，实现具有像素值差异的拼接图像转换为相机采集图像。对于判别器网络 $D_Y$ 来说，它通过不断学习来辨别相机采集的工件图和由生成器生成的虚假图像，将真实图像 $y$ 和生成的虚假图像 $G(x)$ 进行区分。通过不断的训练学习，生成器网络 $G$ 不断优化其参数，尽可能的生成与目标域样本高度相同的图像，以此来“欺骗”判别器网络。同样的，判别器网络也在持续提升其辨别图像真伪的能力。通过这种过方式，生成器网络 $G$ 逐步提高生成样本的质量，最终能够产生与目标域样本在视觉特征上难以区分的逼真图像。

### 3.3.2 CycleGAN 损失函数

1、对抗性损失函数。

在 CycleGAN 框架中，对抗性损失函数作为核心训练机制，在生成器和判别器的训练中发挥着关键作用。该损失函数通过构建生成器与判别器之间的对抗机制，驱动生成器不断优化输出的图像质量，使其在视觉特征分布上逐步逼近目标域的真实图像。在整个过程中的损失函数为式(3-1)所示：

$$\begin{aligned} L_{adv} &= L_{adv}(G, D_Y, X, Y) + L_{adv}(F, D_X, X, Y) \\ &= \mathrm{E}_{y\ p_{data}(y)} \log(D_Y(y)) + \mathrm{E}_{x\ p_{data}(x)} \left( \log\left(1 - D_Y(G(x))\right)\right) \end{aligned} \tag{3-1}$$

式(3-1)中，$E_{y\sim P_{datd(y)}}$ 是 $Y$ 域中选取的样本图像;$E_{x\sim P_{datd(x)}}$ 是 $X$ 域中选取的样本图像。

在实际训练过程中，以其中一对生成器网络 $G$ 和判别器网络 $D_Y$ 网络为例，生成器网络 $G$ 希望生成的图像越接近 $Y$ 域中相机采集的图像样本越好，则上式(3-1) 中 $D_Y\left(G\left(x\right)\right)$ 尽可能得大，$\log\left(1-D_Y\left(G\left(x\right)\right)\right)$ 越小越好，这时 $L\left(G,D_Y,X,Y\right)$ 会变小。对于判别器网络 $D_Y$ 而言，为了能够辨别生成器生成的假图片，它希望自己的性能更好，那么 $D_Y\left(y\right)$ 就会越大，$D_Y\left(G\left(x\right)\right)$ 就会越小，$L_{GAN}\left(G,D_Y,X,Y\right)$ 就越大越好。而另一对生成器网络 F 和判别器网络 $D_X$ 原理与之类似，它们之间的对抗性损失函数如下：

$$L_{GAN}\left(F,D_x,Y,X\right)=\mathrm{E}_{x\ p_{data}(y)}\log\left(D_x\left(x\right)\right)+\mathrm{E}_{y\ p_{data}(y)}\left(\log\left(1-D_X\left(F\left(y\right)\right)\right)\right) \quad (3\text{-}2)$$

2、循环一致性损失函数

CycleGAN 模型中引入循环一致性损失函数的核心目的在于确保图像风格迁移过程中保持双向映射的稳定性与一致性。具体而言，当源域 $X$ 中的工件拼接图像 $x$ 通过生成器 G 转换为目标域 Y 的相机采集图像风格后，再通过生成器 $F$ 转换回源域 $X$ 时，重建图像应与原始图像 $x$ 保持高度相似。同理，目标域 Y 中的相机采集图像 $y$ 经过生成器 $F$ 转换后再通过生成器 $G$ 重建时，也应满足 $G\left(F\left(y\right)\right)\approx y$ 的约束条件。在本文采用该模型进行训练时，这种双向一致性约束有效防止了模式崩溃现象的发生，确保了图像内容在风格迁移过程中的完整性，使得模型生成的图像越接近于真实采集的图像，或者近似于工件拼接图像。该模型的循环一致性损失函数可以通过式(3-3)表示：

$$L_{cyc}\left(G,F\right)=\mathrm{E}_{x\ p_{data}(x)}\left[\left\|F\left(G\left(x\right)-x\right)\right\|_1\right]+\mathrm{E}_{y\ p_{data}(y)}\left[\left\|G\left(F\left(y\right)-y\right)\right\|_1\right] \quad (3\text{-}3)$$

式中，“$\left\|\cdots\right\|_1$”符号—$L_1$ 距离。

3、总损失函数。

CycleGAN 模型的损失函数就是两个对抗性损失和循环一致性损失的和，如式(3-4)所示：

$$L\left(G,F,D_X,D_Y\right)=L_{adv}\left(G,D_Y,X,Y\right)+L_{adv}\left(F,D_X,X,Y\right)+\lambda L_{cyc}\left(G,F\right) \quad (3\text{-}4)$$

式中，参数 $\lambda$ 是对抗性损失和循环一致性损失的权重比。

## 3.4 基于 CycleGAN 网络图像生成算法优化

CycleGAN 作为一种突破性的生成对抗网络架构，成功解决了传统图像转换任务中需要成对训练数据的限制，仅需分别提供源域和目标域的图像数据集即可实现有效训练。但是由于 CycleGAN 对图像数据集的特征提取能力不足，在风格迁移实验结果中存在输出结果图像质量不高、分辨率低的问题[66]，现就 CycleGAN 模型目前仍然存在的问题进行分析并进行改进。

### 3.4.1 存在的问题

在采用 CycleGAN 训练模型时，发现 CycleGAN 模型在工件图像生成任务中存在以下生成图像的质量问题，具体如下所述：

(1)CycleGAN 生成器采用了残差网络结构，这种结构虽然在一定程度上缓解了深层网络中的梯度消失问题，有助于减少特征信息的丢失。但随着网络层数的加深，解码器部分仍面临着高维特征向低维特征传递效率低下的问题。这种局限性导致模型在进行风格转换时难以充分保留原始图像的细节特征，具体表现为生成图像可能出现伪影、模糊和结构失真等质量问题[67]。因此在进行金属工件图像背景风格迁移时，由于 CycleGAN 模型对图像特征提取能力不足，导致生成的图像与真实图像相比存在一定的差距。

(2)CycleGAN 模型中采用的 PathGAN 判别器结构，能够有效捕获图像在不同分辨率下的细节特征。然而，该结构基于局部感受野的设计机制存在一定局限性，其关注范围局限于图像的局部区域，难以建立全局特征的关联性，这种局部化的特征处理方式可能会制约判别器的整体性能表现。

### 3.4.2 生成器网络改进

CycleGAN 的生成器架构由编码器、残差块和解码器三大模块构成，实现了近似于相机采集图像的生成。在图像处理过程中，输入图像数据集首先通过卷积网络进行特征提取，随后将提取的特征与随机噪声融合，经由多个残差块进行特征优化，最终通过反卷积操作生成目标图像。其中卷积层承担编码器的功能，通过下采样操作提取图像特征信息并将其降为低维数据。而反卷积层则

为解码器，通过多层上采样操作将编码器提取的特征信息重构还原为原始尺寸的图像。在编码器与解码器之间，由大量残差块构成的深度残差神经网络发挥着风格转换的关键作用。图 3-5 为模型的整个生成器网络结构的结构图。

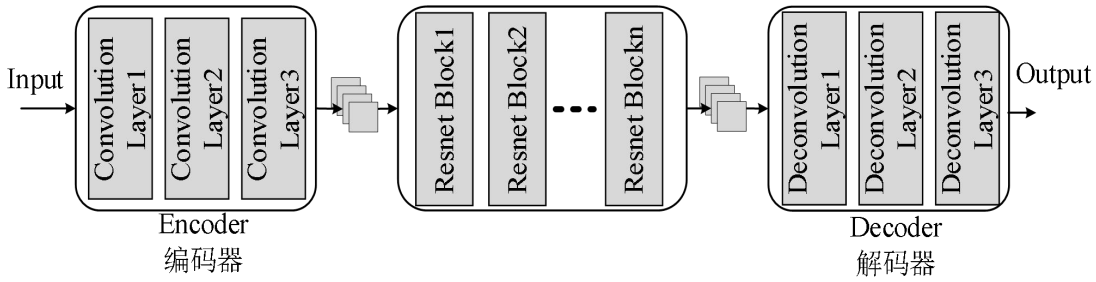



图 3-5 生成器网络结构

本文通过在原始 CycleGAN 模型生成器部分进行结构优化与改进，旨在提升模型对图像数据集的特征提取能力，同时有效保留相机采集图像的风格特征，从而实现工件拼接图像向真实采集图像的风格迁移。考虑到拼接图像与真实图像在像素值分布上存在显著差异，因此本文于生成器加入多头自注意力模块和入密集连接模块，通过多层次特征学习机制增强模型的特征表达能力，提升工件拼接图像向真实图像背景风格迁移的效果。

1、多头自注意力机制

注意力机制的工作原理与人类的视觉功能相似，能够在图像的全局视图中聚焦于关键区域，同时为该区域分配更多的计算资源[68]。与传统的注意力机制不同，自注意力机制(Self-attention Mechanism)不仅能够捕捉输入数据的外部信息，还能够获取数据内部的关联性，从而更全面地学习输入数据的结构特征，其具体结构如图 3-6 所示。将自注意力机制与生成对抗网络(GAN)相结合，能够显著增强生成对抗网络对输入数据全局空间关系的建模能力。具体而言，该机制通过学习图像中不同区域之间的关联性，并基于注意力权重对各区域像素进行加权求和，从而在降低计算复杂度的同时，提升了模型对图像全局信息的捕捉能力。

自注意力包含查询向量 Query、键向量 Key 和值向量 Value，其工作原理是通过计算查询向量与键向量之间的内积相关性，生成注意力分布，即每个位置对序列其他位置的相关性权重。随后，利用这些注意力权重对值向量进行加权求和，从而得到序列的新的特征表示。自注意力机制的计算过程可通过下式(3-5)所示：

$$Attention\left(Q,K,V\right)=soft\max\left(\frac{QK^{T}}{\sqrt{d}}\right)V \tag{3-5}$$

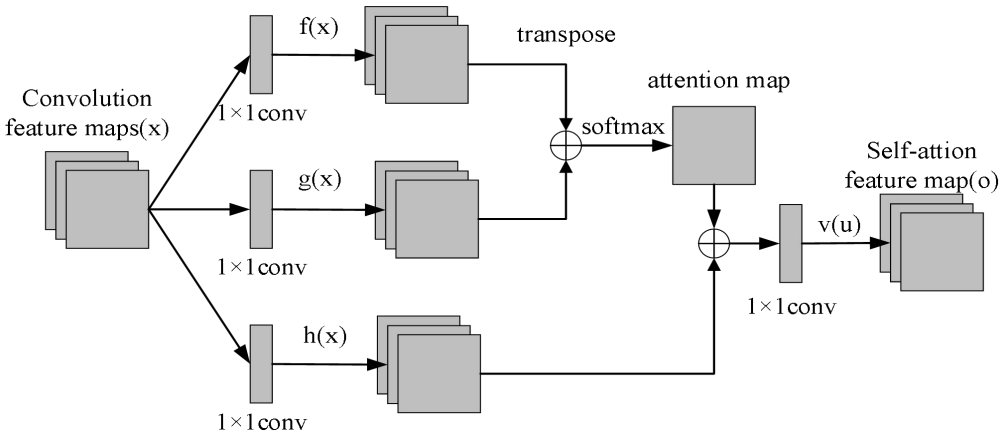



图 3-6 自注意力机制网络架构

多头自注意力机制由上述多个自注意力模块构成，其结构如图 3-7 所示，每个自注意力计算单元被称为一个“头”，且不同头之间相互独立[69]。多头机制的核心思想是通过并行学习查询向量、键向量和值向量在不同子空间中的表示，从而捕捉输入数据的多样化特征。具体而言，多头机制将线性映射后的向量 Query、Key 和 Value 按照预设的头数在通道维度上均匀分割，并对每一部分分别执行自注意力计算(如式(3-5))，从而形成多个独立的注意力子空间。多头注意力可表示为式(3-6)：

$$MultiHead\left(Q,K,V\right)=Concat\left(head_{1}\ldots,head_{n}\right)W^{O} \tag{3-6}$$

其中，

$$head_{i}=Attention\left(Q_{i},K_{i},V_{i}\right) \tag{3-7}$$

$Q_{i},K_{i},V_{i}$ 为 $Q$， $K$， $V$ 在通道维度分割出的子向量。通过计算不同子向量间的独立注意力权重，输出多个相互独立的注意力计算结果，随后进行特征叠加，并通过线性变换层进行特征融合，最终得到多头自注意力机制的输出结果。

基于上述分析，本研究在 CycleGAN 模型中引入了多头自注意力模块，使得模型能够学习输入数据在不同语义维度上的上下文依赖关系。在图像风格迁移任务中，源域与目标域之间往往存在复杂的非线性分布差异。而多头自注意力机制通过增加模型的表达能力，能够更好地适配源域与目标域之间的语义差异，从而增强了对两域分布差异的建模能力。具体而言，在对工件拼接图像向视觉系统采集的图像进行风格迁移时，CycleGAN 网络模型能够更准确地捕捉

两域之间的特征对应关系，尽可能地将拼接图像转换为真实图像。这一改进为图像风格迁移任务提供了更强大的特征提取和转换能力，有助于提升风格迁移的效果。

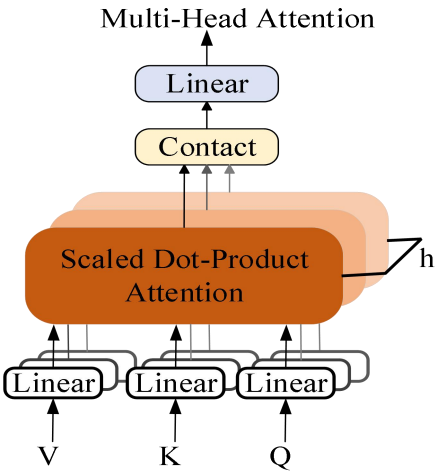



图 3-7 多头自注意力结构

2、DenseNet

CycleGAN 生成器中的转换器使用 ResNet 残差网络，其结构如图 3-8 所示。ResNet 通过使用恒等映射连接方式[70]，将输入特征直接跨越中间的卷积层并与输出特征进行累加，从而有效缓解了深层网络中可能出现的梯度消失和特征丢失问题。当相邻两层通道数相同时映射关系可以表示为 $H(x)=F(x)+x$，不同时表示为 $H(x)=F(x)+wx$， $w$ 为卷积操作用于调整通道维度。

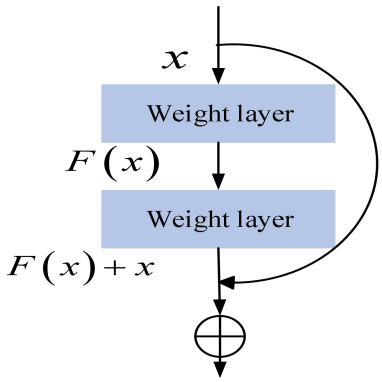
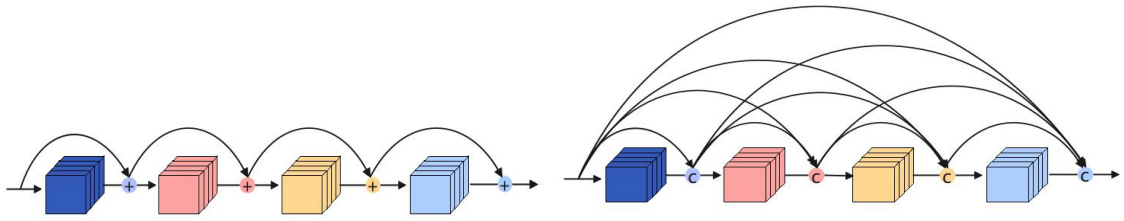



图 3-8 ResNet 网络连接映射

ResNet 通过引入跳跃连接机制，突破了传统网络的层级限制，有效缓解了网络层数增加导致的梯度爆炸和梯度消失问题，但其在图像特征提取的精准性和网络性能提升方面仍存在一定局限性。而 DenseNet 在 ResNet 的基础之上进行了优化，其核心改进在于采用了密集连接机制。在 DenseNet 每一层均直接与

所有后续层相连，即每个卷积层都能够接收来自之前所有层的输出信息作为额外输入，从而最大限度地保证了网络中信息传递效率，这种密集连接方式不仅实现了特征的高效复用，还缓解了模型训练时存在的梯度消失和模式退化问题，图 3-9 展示了 ResNet 与 DenseNet 的内部的连接方式结构对比。与图 3-9(a)中原残差网 ResNet 网络结构方式不同，DenseNet 内部层与层之间采用密集连接方式，特征图大小相同。因此 DenseNet 相对于 ResNet 来说在特征传递方面更具优势，能减少网络中的参数量并缓解梯度消失和模式退化问题。基于 DenseNet 在特征传递和参数效率方面的显著优势，本文将生成器转换器中的 ResNet 模块替换为 DenseNet 模块，以进一步提升模型的性能。



(a)原残差网 ResNet 结构　　(b)具有密集连接的卷积神经网络 DenseNet

图 3-9 ResNet 与 DenseNet 结构对比

改进后的 CycleGAN 的生成器网络结构如图 3-10 所示，通过结合多头自注意力机制，将其加入转换器于解码器之间，使得生成器能够更好地捕捉图像的多层次信息，从而在处理图像时更加聚焦于重要的细节特征。同时，将转换器中的 ResNet 模块替换为 DenseNet 模块，这一改进不仅有效缓解了梯度爆炸和梯度消失问题，还通过密集连接机制实现了更深层次的特征表示，提升了网络的表征能力。将这两种技术结合起来，改进后的 CycleGAN 在进行图像风格迁移时保持网络的稳定性，并提高对图像的特征提取能力，使网络能够学习到更接近源数据集的数据分布特征，从而生成与真实相机采集图像高度相似的视觉效果。

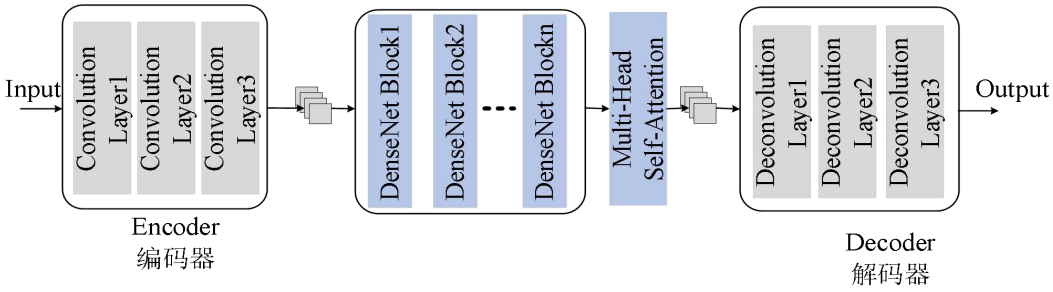



图 3-10 改进后的 CycleGAN 生成器网络结构图

### 3.4.3 判别器网络改进

CycleGAN 模型采用双 PatchGAN 判别器结构，其在风格迁移任务中表现出显著的性能优势，尤其是在高分辨率图像处理和细节清晰化方面具有突出表现[71]。两个判别器的网络架构完全相同，如图 3-11 所示，分别用于判别真实图像和生成的图像。

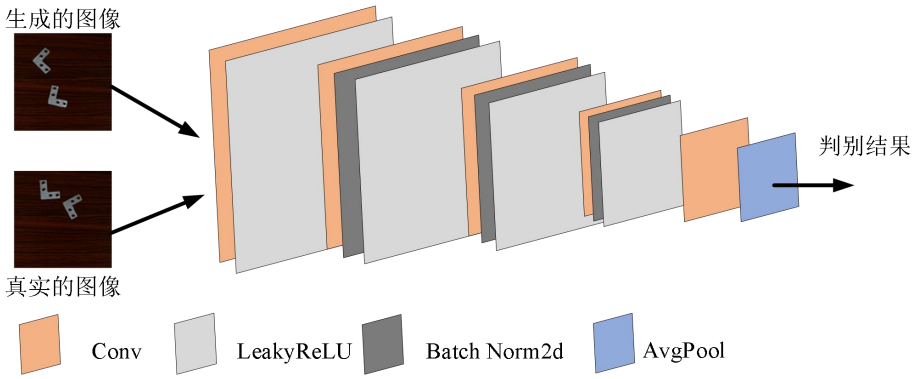



图 3-11 PatchGAN 判别器网络结构

尽管 CycleGAN 中采用的 PatchGAN 判别器在提升生成图像质量方面取得了一定成效，能够对图像的局部细节进行更精细的判别，有助于生成器学习到更细致的图像特征，但其在捕捉图像全局信息方面仍存在显著局限性。因此本文使用多尺度判别器，通过同时关注图像的全局结构和局部细节信息，进一步优化了判别器的性能。如图 3-12 所示，该多尺度判别器由三个判别器构成，其内部网络结构相同，每个判别器处理不同尺度的输入图像：原始分辨率图像、2 倍下采样图像以及 4 倍下采样图像。这种多尺度设计使得模型能够自适应地捕捉不同层次的特征，在处理不同大小分辨率图像时，使得判别器能够关注整体结构或局部细节特征，通过加权相加得到最后判别输出。

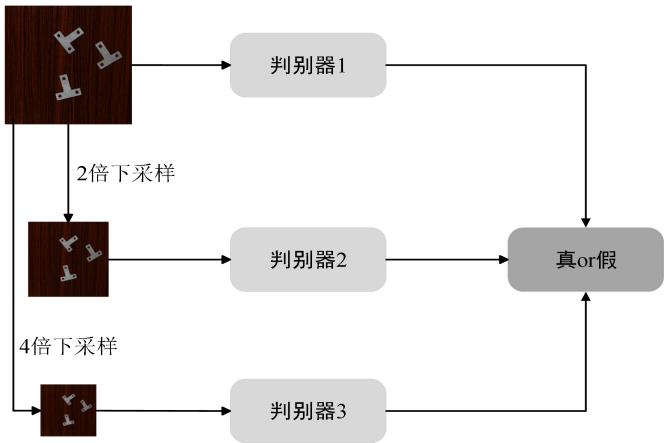



图 3-12 多尺度判别器网络结构图

为了更有效地学习图像数据的分布特征及其隐含特征，本研究对判别器结构进行了优化，将原判别器中第二层和第三层的标准卷积替换为扩张卷积[72]，图 3-13 展示了加入扩张卷积的判别器结构，使得每个卷积核能够覆盖更广泛的区域，为模型提供更大的感受野，从相同数量的参数中获得更多的图像信息。扩张卷积的原理如图 3-14 所示，当扩张率为 1 时，扩张卷积与 3×3 普通卷积并无两异；当扩张率为 2 时，卷积核的有效感受野扩展至 7×7；当扩张率为 4 时，感受野进一步扩展至 15×15。这种设计使得模型能够在不同尺度上捕捉图像的全局和局部特征，从而提升判别器的特征提取能力。

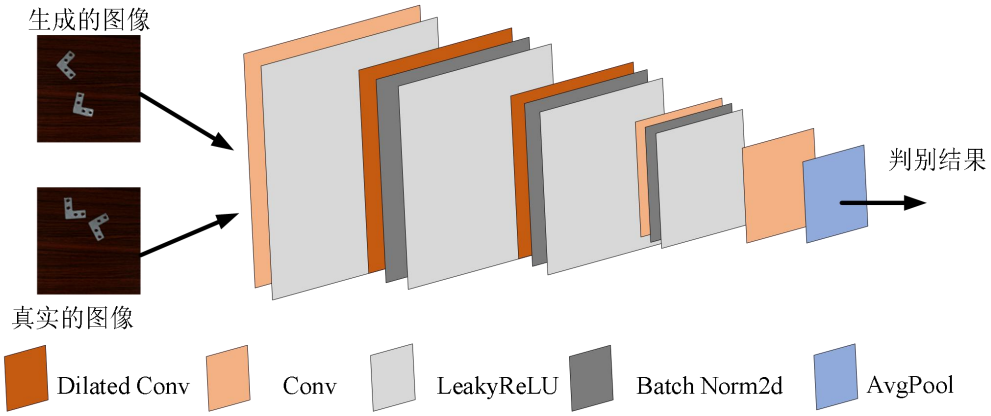
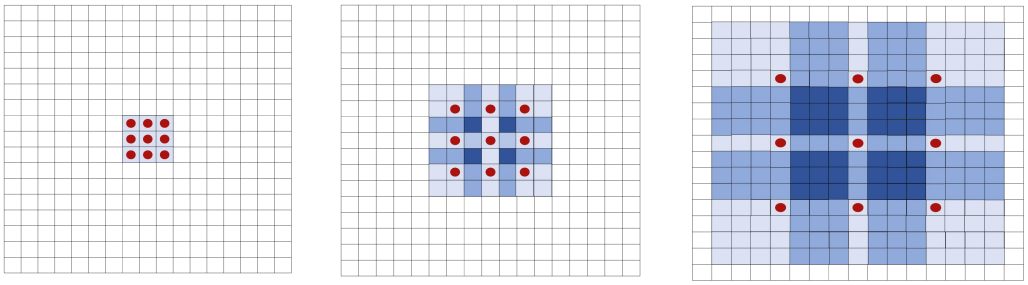



图 3-13 改进后 PatchGAN 判别器结构



(a)扩张率=1　　(b)扩张率=2　　(c)扩张率=4

图 3-14 扩张卷积原理图

## 3.5 实验内容与结果分析

### 3.5.1 图像生成效果评价指标

峰值信噪比(PSNR)和结构相似性(SSIM)是评价生成图像质量的重要定量指标，而为了更全面地衡量生成模型的表现，本文还引入了 FID 作为评价指标，该指标能够有效反映生成图像与真实图像在特征空间中的分布差异。表 3-1 展示这三种评价指标的特点。基于这些指标，本节在不同数据集上对改进的

CycleGan 模型进行了训练测试，以验证其性能优势。

表 3-1 图像生成效果评价指标

| 指标名称 | 峰值信噪比(PSNR) | 结构相似性(SSIM) | FID |
|---|---|---|---|
| 方法 | 比较原始图像和生成图像之间的均方误差 | 通过亮度、对比度和结构三个方面来衡量两张图片之间的相似性 | 衡量生成图像和可见光 RGB 图像之间特征表示的均值和协方差的差距 |
| 取值范围 | 通常在 20-50dB | [-1，1] | 0-500 |
| 特点 | 值越高，表示图像质量越好 | 值越接近 1，表示图像结构越相似 | 值越低，表示生成图像与真实图像的分布越接近 |

(1)SSIM 是一种用于评估两幅图像相似度的指标，能够有效衡量图像失真前后的一致性，同时也广泛应用于评估模型生成图像的真实性。其式如(3-8)所示：

$$SSIM = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2\mu_y^2 + c_1\right)\left(\sigma_x^2\sigma_y^2 + c_2\right)} \tag{3-8}$$

其中 $x$ 和 $y$ 代表两张图像，$\mu_x$ 和 $\mu_y$、$\sigma_x^2$ 和 $\sigma_y^2$ 分别为它们的像素均值、方差，$\sigma_{xy}$ 则是图像 $x$ 和 $y$ 的像素协方差，$c_1 = \left(k_1 L\right)^2$ 和 $c_2 = \left(k_2 L\right)^2$ 是用于保持稳定性的常数，$L$ 是像素值范围。当 SSIM 等于 1 时，表示两张图像完全一致。

(2)PSNR 用于计算不同图像的峰值误差，主要用于评估图像的失真程度，其计算式如(3-9)所示：

$$PSNR = 10\log_{10}\frac{\left(Max\ Value\right)^2}{MSE} \tag{3-9}$$

其中 *Max Value* 是像素的最大可能强度，MSE 是均方误差。

(3)FID 分数是一种基于特征空间分布距离的图像质量评价指标。其核心思想是通过 Inception v3 模型的倒数第二个全连接层提取图像的高维特征向量，这些向量能够有效捕捉图像的视觉特征信息。FID 的计算方法如式(3-10)所示：

$$FID = \left\|\mu_r - \mu_g\right\|^2 + T_r\left(\sum\nolimits_r + \sum\nolimits_g - 2\left(\sum\nolimits_r \sum\nolimits_g\right)^{\frac{1}{2}}\right) \tag{3-10}$$

其中，$r$ 表示真实图像的分布，g 表示生成图像的分布，$\mu_r$ 和 $\mu_g$ 是真实图像和生成图像的特征表示的平均值，$\Sigma_r$ 和 $\Sigma_g$ 分别表示真实图像和生成图像的特征表示的协方差矩阵，$T_r$ 表示计算矩阵的迹，$\left|*\right|$ 表示两个向量之间的欧几里得

距离。由式(3-10)可知，FID 值越小，表明生成图像与真实图像在特征空间中的分布越接近，两者的相似性越高，从而反映出生成图像的质量越好。

### 3.5.2 实验准备

本文所有实验均于在 Windows 操作系统下运行，GPU 为 NVIDIA GeForce RTX 3060，Python 版本为 3.9，Pytorch 版本是 1.12。

本实验使用生成对抗网络 CycleGan 实现的非成对图像数据集进行风格迁移，实现拼接图像转换为相机采集图像的操作。构建的金属工件数据集总共有七类，其中具有相机拍摄风格的真实图像共有 6600 张，由随机算法生成的拼接图像也是 6600 张，所有图像统一定义为 256×256 像素大小。每类数据集均划分为两个部分，其中 80%的工件图像被选取作为训练数据集，支持模型在训练过程中学习到图像的数据信息，而剩余的 20%则被用于验证，以此检验模型是否具备良好的性能。实验中每个数据集为四个文件：trainA、trainB、testA 和 testB，每张图片均为 256×256 像素大小。其中 A 类数据集是由相机采集的真实工件图像，B 类数据集是结合随机算法生成的拼接图像，通过这两种不同背景风格的图像数据集作为输入，实现工件背景的风格迁移，实现由随机算法生成的拼接图像转换为近似于相机采集的真实图像，丰富后续检测模型的输入数据。

### 3.5.3 消融实验

为评估模型改进方法在 CycleGAN 风格迁移任务中的有效性，于不同工件数据集上进行了消融实验，通过对输入图像到图像风格转换后的效果图进行比较，并结合 PSNR、SSIM、FID 等图像评价指标，分析算法在不同数据集上的性能表现。本文比较了引入多头自注意力模块、替换 ResNet 网络以及实现多尺度判别等情况下模型图像生成效果，具体图像生成效果对比结果如图 3-15。

由图 3-15 的效果对比图可以看出，原始 CycleGAN 模型在将相机采集图像与拼接图像进行风格转换时，生成的不同类型工件图像存在明显的边缘弯曲、模糊以及目标工件缺失等问题。而通过引入多头自注意力模块、替换 ResNet 网络以及实现多尺度判别等改进措施，显著提升了图像生成质量。以工件四为例，原始 CycleGAN 模型在生成目标工件图像时出现了额外的工件特征，而改进后

的 CycleGAN 模型能够更精准地捕捉关键区域特征，生成图像在细节保留、轮廓完整性和纹理自然度方面均表现出显著提升。此外，改进后的模型生成图像与目标域风格的契合度更高，在视觉特征上与原始相机采集图像的相似度也显著增强。表 3-2 进一步展示了改进前后模型在不同工件数据集上的图像质量评估结果，验证了改进措施的有效性。

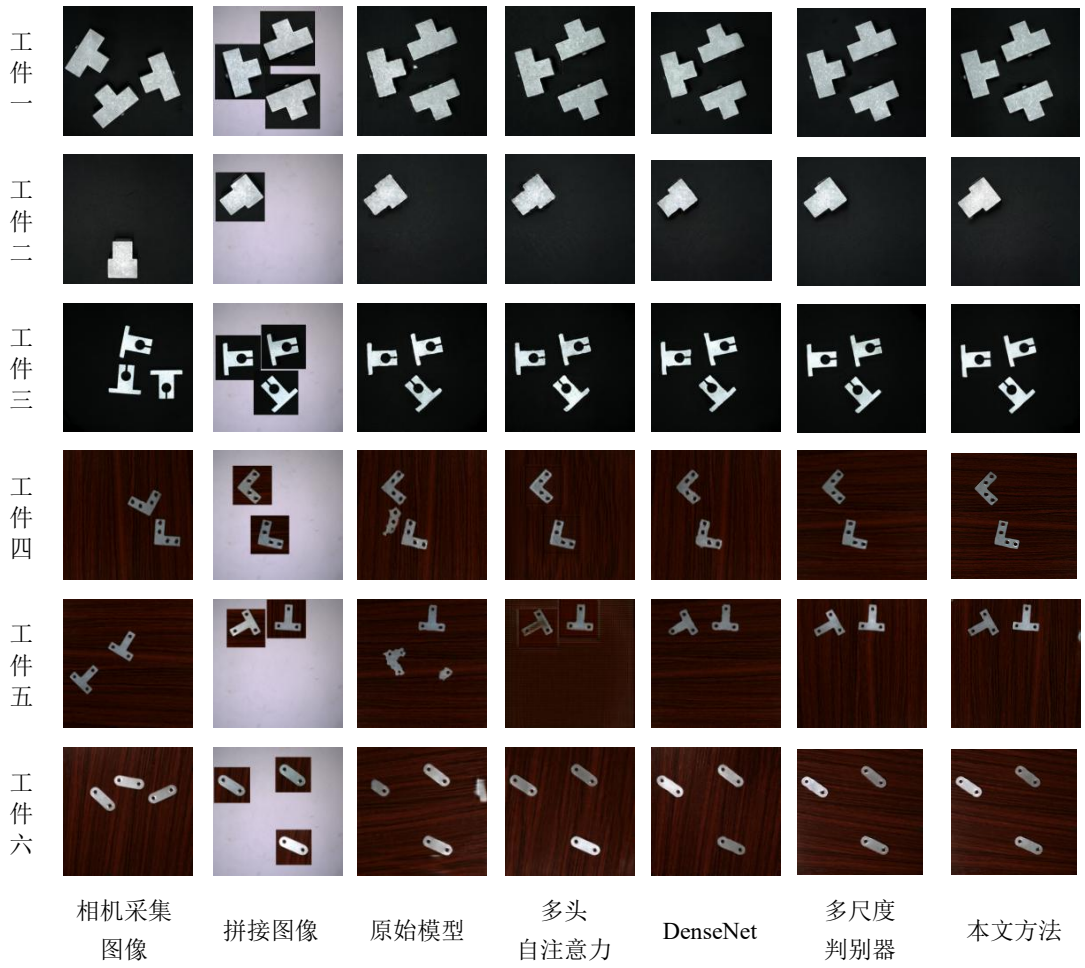



图 3-15 消融实验结果对比图

通过对图像评价指标数据表 3-2 展开深入分析可知，这些改进方法在 SSIM 和 PSNR 评价指标数据上总体上大于原始模型生成的图像，具体而言，PSNR 值越大证明生成的图像质量更好，而 SSIM 值越接近于 1 证明图像越相似。由于本文主要对比生成图像中的工件是否完整的还原拼接图像中的工件，所以在图像质量评价指标 SSIM 和 PSNR 上得出的数值偏低。在特征分布相似性方面，改进模型的 FID 数值均小于原始模型生成的图像，这些数值的减小显示改进后的图像生成效果比原始模型更好，即改进后生成的假图片与视觉采集的真图片在特征分布上具有更高的相似度，符合本文生成真实图像的需求。综合三个评

价指标的结果，本文方法在三个评价指标上相较于原始模型有了不同程度的提升，证明改进后的 CycleGan 模型在增强图像风格迁移效果方面的有效性。

表 3-2 评价指标数据对比

| | | PSNR↑ | SSIM↑ | FID↓ |
|---|---|---|---|---|
| 工件一 | CycleGan | 28.85247 | 0.47845 | 76.46506 |
| | 多头自注意力 | 28.72757 | 0.48795 | 54.89069 |
| | DenseNet | 28.72320 | 0.45167 | 68.04776 |
| | 多尺度判别器 | **28.9826** | 0.52362 | 35.93233 |
| | 本文方法 | 28.94182 | **0.52569** | **35.58428** |
| 工件二 | 原始模型 | **29.61371** | 0.41876 | 97.42829 |
| | 多头自注意力 | 28.32963 | 0.42784 | 63.60937 |
| | DenseNet | 28.29041 | 0.41333 | 99.22901 |
| | 多尺度判别器 | 28.35993 | **0.45597** | 60.98541 |
| | 本文方法 | 28.29205 | 0.45356 | **58.06638** |
| 工件三 | 原始模型 | 28.77390 | 0.33506 | 75.20807 |
| | 多头自注意力 | 28.72928 | 0.30186 | 98.79448 |
| | DenseNet | 28.72408 | 0.30505 | 99.25033 |
| | 多尺度判别器 | **28.87342** | **0.37365** | 64.86293 |
| | 本文方法 | 28.74596 | 0.33397 | **63.17014** |
| 工件四 | 原始模型 | 28.31457 | 0.221917 | 81.128446 |
| | 自注意力 | 28.34143 | 0.25619 | 79.65404 |
| | DenseNet | 28.33664 | 0.234932 | 80.54185 |
| | 多尺度判别器 | 28.39950 | 0.25427 | 28.36812 |
| | 本文方法 | **28.44558** | **0.26977** | **24.07848** |
| 工件五 | 原始模型 | 28.35618 | 0.224703 | 101.891 |
| | 自注意力 | 28.38036 | 0.253217 | 90.739 |
| | DenseNet | 28.32552 | 0.216388 | 88.327 |
| | 多尺度判别器 | 28.36339 | 0.261234 | 28.36116 |
| | 本文方法 | **28.3891** | **0.26368** | **25.4107** |
| 工件六 | 原始模型 | 28.17202 | **0.25736** | 53.14695 |
| | 自注意力 | 28.17201 | 0.24437 | 38.96504 |
| | DenseNet | 28.13081 | 0.19913 | 50.39646 |
| | 多尺度判别器 | 28.18992 | 0.24349 | 36.85125 |
| | 本文方法 | **28.19149** | 0.25236 | **33.78672** |

### 3.5.4 对比实验

为了验证所改进模型的有效性，将在工件数据集上与图像风格迁移模型UGATIT[73]、CUT[74]和 SimDCL[75]进行对比实验，验证改进模型在图像风格迁移上的效果提升，实验结果如图 3-16 所示：

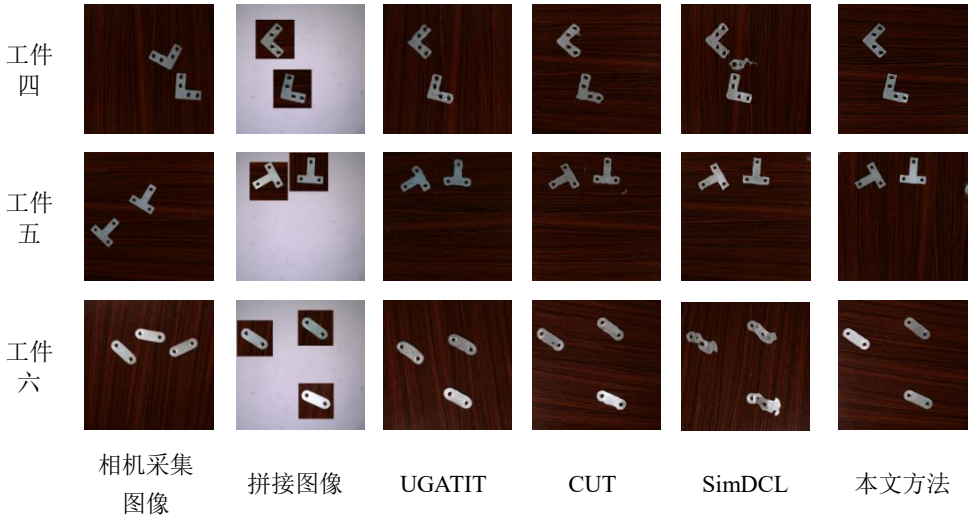



图 3-16 对比实验结果图

表 3-3 对比实验数据分析对比

| | | PSNR↑ | SSIM↑ | FID↓ |
|---|---|---|---|---|
| 工件四 | UGATIT | 28.35520 | 0.25027 | 143.19590 |
| | CUT | 28.34433 | 0.23022 | 118.96807 |
| | SimDCL | 28.35322 | 0.23234 | 83.69356 |
| | 本文方法 | **28.44558** | **0.26977** | **24.07848** |
| 工件五 | UGATIT | 28.36675 | 0.22432 | 144.11285 |
| | CUT | 28.37517 | 0.230704 | 191.32215 |
| | SimDCL | 28.36791 | 0.24293 | 65.22858 |
| | 本文方法 | **28.3891** | **0.26368** | **25.4107** |
| 工件六 | UGATIT | 28.18023 | 0.21531 | 70.46232 |
| | CUT | 28.05742 | 0.2291 | 66.16053 |
| | SimDCL | 28.15282 | 0.20934 | 229.71737 |
| | 本文方法 | **28.19149** | **0.25236** | **33.78672** |

从上面的实验结果对比图 3-17 来看，所提的 CycleGan 改进方法生成的目标样式图像中目标工件细节更加清晰，边缘轮廓保持完整，未出现明显的形变，整体视觉效果更加接近于真实相机拍摄的图像。相比之下，UGATIT、CUT 和 SimDCL 则存在目标工件模糊扭曲问题。从图像生成效果评价指标上也可以看

出，本文 CycleGan 改进方法在图像生成效果方面均优于其它风格迁移模型。其中较低的 FID 值表明生成图像与真实图像的分布更为接近，SSIM 和 PSNR 提高说明结构相似度更接近原图和生成的失真少。这些实验数据充分验证了所提改进方案的有效性，为基于生成对抗网络的图像风格迁移任务提供了新的解决思路。

## 3.6 本章小结

本章构建了基于视觉引导的机器人抓取过程的金属工件数据集，解决了依靠人工采集丰富的图像造成人工成本投入过高问题。尽管基于传统的数据增强技术生成大量接近于相机采集的图像存在局限性，但通过基于生成对抗网络的数据增强的方法，生成的图像成功地贴合相机采集图像目标对象各异、位置多变、方向不同等特点。而后通过对网络模型的优化，进一步提高生成的图像质量。本章的工作为后续检测模型的训练提供了更充分和可靠的数据支持，对于提高多关键点检测算法的准确性具有重要意义。

# 第 4 章 基于深度学习的多关键点检测方法

## 4.1 引言

通过第 3 章节基于生成对抗网络的图像扩充方法，为机器人抓取位姿的获取构建高质量、多样化的数据集。然而，在复杂工业环境中，获取的图像数据集中工件表面存在反光、磨损和腐锈等情况，这些因素会干扰视觉系统对位姿信息的准确获取，导致机器人未能成功抓取目标工件。针对该问题，本文提出了一种基于关键点检测的位姿测量方法，即通过获取视觉系统左右相机图像中目标工件表面关键点位姿信息，进行坐标转换实现工业机器人对目标工件的 6DOF 抓取位姿的求取。

本章以无锚框检测模型 CenterNet 作为关键点检测的基础框架，获取双目视觉系统中左右相机中目标工件表面多关键点位姿信息。而为获取较高精度的位姿信息，对 CenterNet 原始网络模型架构进行了改进，引入自适应细粒度通道注意力和加权双向特征金字塔网络结构，加强模型对重要细节信息的学习，从而提升算法的检测性能。通过在多组工件测试集上进行实验，验证改进后 CenterNet 网络模型在多关键点检测性能。

## 4.2 目标检测算法 CenterNet

CenterNet 是 2019 年提出的目标检测网络算法[60]，该模型不需要区域建议以及 ROI 等组件，通过直接预测目标中心点和边界框尺寸来实现目标检测，所以 CenterNet 在速度和精度上都有提升。

### 4.2.1 CenterNet 主干网络

多深层聚合网络 DLANet、沙漏网络 Hourglass 和残差网络 ResNet 是 CenterNet 中的主干网络，相比较于传统的主干特征提取网络来说，DLA-34 是一种深度残差网络，其核心在于采用了多层级深度融合的 DLA 结构。这种结构通过实现不同层级特征的高效融合，在保证实时性的同时显著提升了检测精度，因此本文选取 DLA-34 作为 CenterNet 结构的主干特征提取网络。

DLA 是一种常图像分割和目标检测任务的深度学习架构，其核心在于通过多层次特征聚合来提升视觉识别性能[76]。在视觉识别任务中，随着神经网络深度的增加，模型对视觉系统采集的图像中目标对象的单一层次的特征往往难以准确的获取，因此需要通过聚合多层网络信息来进行检测，提高对目标对象的类别与位置信息的识别准确率。DLA 的结构如图 4-1 所示。

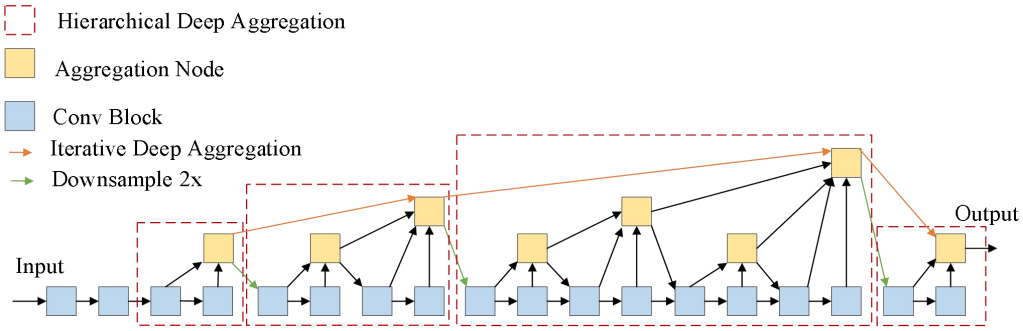



图 4-1 DLA 的结构

由上图 4-1 的 DLA 结构图可以看出，该网络结合了迭代深度聚合(IDA)和分层深度聚合(HDA)。其中，IDA 能够融合不同分辨率和尺度的特征图，而 HDA 将不同阶段的浅层和深层特征进行组合，形成更加丰富的特征表达。其中，HDA 的优势在于能够直接将浅层组合特征传递到深层，从而更好地保留浅层的细节信息，这对于处理多尺度和多分辨率的输入数据尤为重要。图 4-1 中红色边框标注的 HDA 结构展示了其树形层次特征传播机制，这种设计能够更高效地传递特征和梯度信息；绿色线表示下采样操作，用于降低特征分辨率；橘色线则代表 IDA 的迭代连接方式，实现深层与浅层特征的深度融合。总体而言，DLA 通过 HDA 实现浅层与深层特征的高效融合，并借助 IDA 的迭代机制不断优化特征表达，最终输出具有丰富语义信息和精确空间信息的特征表示。这种设计使得 DLA 在处理复杂视觉任务时表现出色。

### 4.2.2 CenterNet 的边界框表示

边界框的表示形式是目标检测算法中定位对象的核心要素。如图 4-2 所示，CenterNet 采用了一种简洁高效的表示方法：以中心点坐标和宽高作为物体的基本描述[77]。具体而言，该方法通过预测中心点坐标的偏移量，最终回归得到目标框的完整表示形式$(x, y, w, h)$，其中$(x, y)$表示目标中心点的坐标，$w$ 和 $h$ 分别表示目标的宽度和高度。这种表示方式不仅简化了边界框的回归过程，还提

高了检测算法对目标位置和尺寸的预测精度。

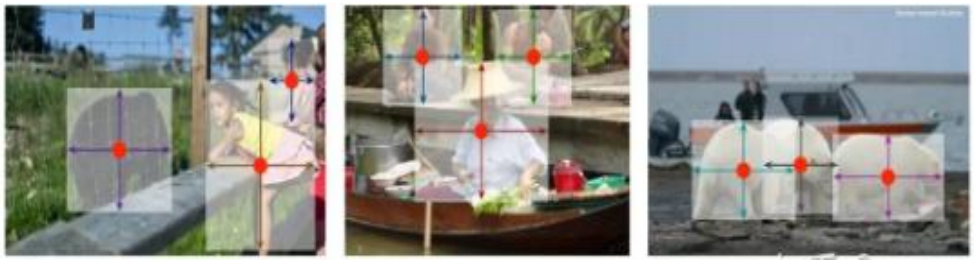


图 4-2 CenterNet 边界框表示示意图

### 4.2.3 正负样本分配

CenterNet 采用高斯热图实现样本的区分，以高斯核中心点作为正样本而中心点附近的点则被视为困难样本，高斯核范围之外的区域则被归类为负样本。给定中心点 $p$，其计算式如式(4-1)所示：

$$p=\left(\frac{x_1+x_2}{2},\frac{y_1+y_2}{2}\right) \tag{4-1}$$

其中，$(x_1,x_2)$、$(y_1,y_2)$是目标框的左上和右下角点的位置。对于类别中每个经过网络下采样后的低分辨率中心点坐标，设为 $\tilde{p}=\left[\frac{p}{R}\right]$，其中 R=4，然后将坐标通过高斯核分布到热图上，高斯核的计算公式为：

$$Y_{xyc}=\exp\left(-\frac{(x-\tilde{p}_x)+(y-\tilde{p}_y)^2}{2\sigma_p^2}\right) \tag{4-2}$$

其中，$\sigma_p$ 是一个与目标尺寸相关的标准差，c 是类别数目。因此，$\hat{Y}_{x,y,c}=1$ 表示在当前中心点位置的物体类别是 $c$，$\hat{Y}_{x,y,c}=0$ 则表示当前位置不存在类别 $c$ 目标。这种基于高斯热图的样本分配方法能够有效区分正负样本，并提高模型对目标中心点的定位精度。

### 4.2.4 CenterNet 损失函数

CenterNet 的损失函数如式(4-3)所示，主要包括三部分：热力图损失、尺寸损失、中心点偏移损失。

$$L_{\det}=L_k+\lambda_{size}L_{size}+\lambda_{off}L_{off} \tag{4-3}$$

其中，$L_k$、$L_{size}$、$L_{off}$分别表示 Heatmap 中心点损失、目标尺寸损失和偏移值损失。在模型训练阶段，通过回归计算尽可能地找到损失的最小值，从而得到预测的权重结果。以下将分别阐述各损失函数的计算方式：

1、Heatmap 中心点损失

在 CenterNet 中，Heatmap 中心点损失通过热图计算来实现目标类别的回归。对于每个热图，模型会在目标中心点的位置生成一个关键点，当同一类别的两个高斯分布发生重叠时，采用像素最大值点的方式来确定关键点位置。Heatmap 中心点损失函数如下：

$$L_k = \frac{-1}{N}\sum_{xyc}\begin{cases}\left(1-\hat{Y}_{xyc}\right)^{\alpha}\log\left(\hat{Y}_{xyc}\right) & Y_{xyc}=1\\ \left(1-\hat{Y}_{xyc}\right)^{\beta}\log\left(\hat{Y}_{xyc}\right)^{\alpha}\log\left(1-\hat{Y}_{xyc}\right) & otherwise\end{cases} \tag{4-4}$$

其中，超参数$\alpha$和$\beta$取值为 2 和 4，用于平衡难易样本，$N$是图像中物体的个数。$\sum$符号的下标中 $x$、$y$ 表示所有热图上的所有坐标点，预测值和标记的真值分别用$\hat{Y}_{xyc}$和$Y_{xyc}$表示。

2、目标尺寸损失

尺寸回归对应包含目标宽高的损失函数，假设目标 $k$ 的类为 $c$，它的包围框为$\left(x_1^k, x_2^k, y_1^k, y_2^k\right)$，中心点为$p_k = \left(\frac{x_1^k + x_2^k}{2}, \frac{y_1^k + y_2^k}{2}\right)$，对每个目标 $k$ 的大小进行回归，最终回归到$s_k$预测尺寸$s_k = \left(x_2^k - x_1^k, y_2^k - y_1^k\right)$，尺寸损失函数如下：

$$L_{size} = \frac{1}{N}\sum_{k=1}^{N}\left|\hat{S}_{pk} - s_k\right| \tag{4-5}$$

其中，$N$是图像中物体的个数，$s_k$、$\hat{S}_{pk}$分别为真实框和预测框的尺寸面积。

3、偏移值损失

当提取的特征重新映射到原始图像时，下采样操作会引入步幅相关的误差。为了补偿这种误差，算法为每个中心点预测一个局部偏移量，用于校正坐标偏差。对于类别 $c$ 的所有中心点，模型共享相同的偏差预测值，$p$ 和 $\tilde{p}$ 分别为图像中目标真实和预测的中心点坐标，$R$ 为输出步幅，采用传统 L1 损失函数。监督学习仅在关键点位置 $p$ 起作用，所有其他位置都被忽略。偏差损失函数如下：

$$L_{off} = \frac{1}{N}\sum_{p}\left|\hat{O}_{\tilde{p}} - \left(\frac{p}{R} - \tilde{P}\right)\right| \tag{4-6}$$

## 4.3 多关键点检测模型优化

在工业环境不均匀光照、粉尘的干扰下，将待检测的图像作为输入，经过一系列的模型训练后，一些目标工件对象的细节特征会在这个过程中存在丢失，进而在后续的检测模块出现漏检的问题，影响检测网络的检测性能。因此对解决同一张图像中目标尺度存在差异而造成的局部目标信息丢失、特征提取不充分的问题，对 CenterNet 算法进行改进，在主干网络部分嵌入自适应细粒度通道注意力模块，增强网络对目标物的关注以及对无关背景的抗干扰能力；在原有的模型上加入 BiFPN 结构，捕捉多尺度的特征信息，加强对重要信息的学习，提升算法对工件表面多关键点的检测精度。图 4-3 展示了 Centernet 网络模型改进后的结构。

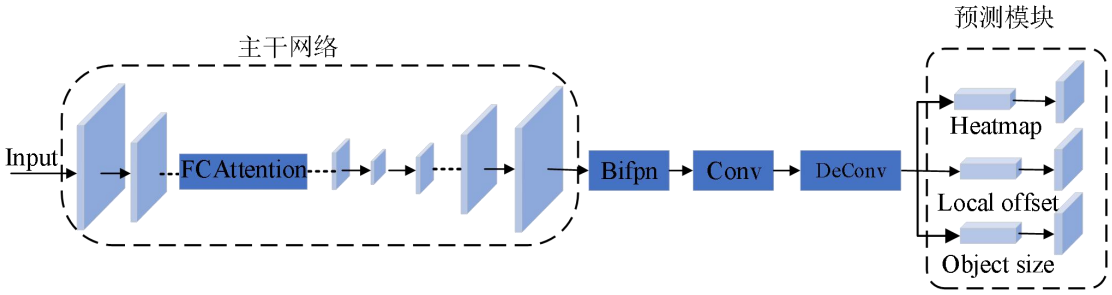



图 4-3 改进的 CenterNet 网络模型

### 4.3.1 自适应细粒度通道注意力

自适应细粒度通道注意力(Adaptive fine-grained channel attention, FCAttention)[78]是一种能有效整合全局和局部信息、合理分配权重的通道注意力机制，它能够自适应地关注不同通道上的特征信息，使得网络模型能够更准确地强调有用特征，抑制无关或干扰性的特征。与通道注意力采用全连接层来捕获全局信息不同，FCAttention 通过利用相关矩阵来捕捉全局和局部信息之间的相关性，促进网络之间的交互，够更有效地整合全局和局部信息。同时这种机制使得网络能够在不同粒度上进行特征权重分配，提升目标对象检测结果的准确性。

自适应细粒度通道注意力模型结构见图 4-4，首先，给定一个特征图 $F \in {}^{C\times H\times W}$，其中 $C$、$H$ 和 $W$ 分别代表通道数、长度和宽度，该模型通过对输

入的特征图进行全局平均池化，以获取全局的通道信息，$U$ 的第 $n$ 个通道表示的计算公式为：

$$U_n = GAP(F_n) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_n(i,j) \tag{4-7}$$

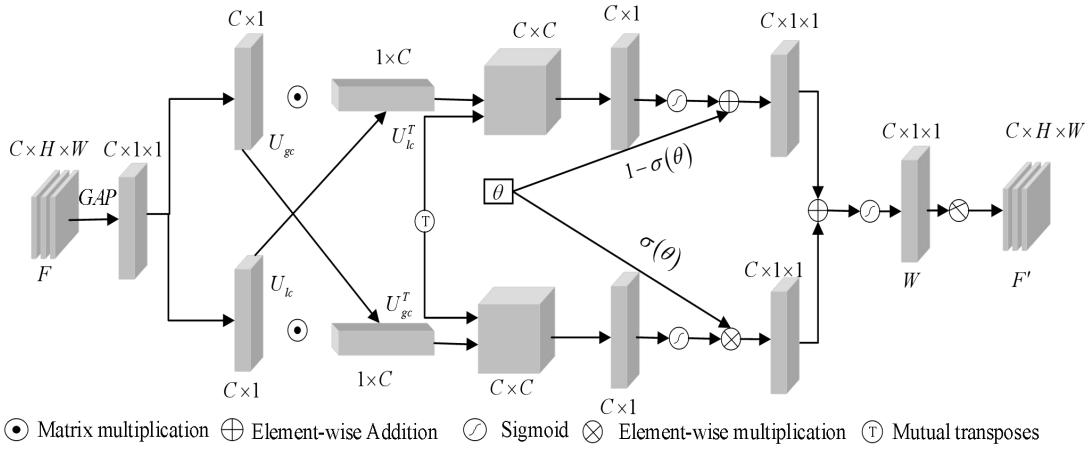



图 4-4 自适应细粒度通道注意力模型结构

式(4-7)中 $F_n(i,j)$ 是第 $n$ 个通道特征图在位置处的值，$GAP(x)$ 是全局平均池化函数，通过该函数可以将特征图 F 的形状从 $C \times H \times W$ 转换为 $C \times 1 \times 1$。之后，为了在获得局部通道信息的同时保证较少的参数数量，通过矩阵 $B$ 进行局部通道交互，$B = [b_1, b_2, b_3, \ldots, b_k]$，其公式如下：

$$U_{lc} = \sum_{i=1}^{k} U \ \ b_i \tag{4-8}$$

其中 $U_{lc}$ 表示局部信息，$k$ 是相邻通道的数量。为了进一步获得全局通道信息并加强表示全局信息的能力，利用对角矩阵 $D$ 来捕获所有通道之间的依赖关系作为全局信息，设置对角矩阵 $D = [d_1, d_2, d_3, \ldots, d_c]$，通过式(4-9)获取全局信息。

$$U_{gc} = \sum_{i=1}^{c} U \ \ d_i \tag{4-9}$$

其中 $U_{gc}$ 表示全局信息，$c$ 为通道数。通过结合对角矩阵获得的全局信息与通过带矩阵获得的局部信息，促进全局信息和局部信息之间的有效交互。其中通过交叉相关操作在各种粒度上捕捉两者之间的相关性，形式如下：

$$M = U_{gc} \ \ U_{lc}^{T} \tag{4-10}$$

$M$ 表示相关矩阵。从相关矩阵及其转置中提取行和列信息作为全局和局部

信息的权重向量，进行准确分配特征权重并减少计算复杂性。接着通过一个可学习的因子实现动态融合，如下所示：

$$U_{gc}^{w}=\sum_{j}^{c}M_{i,j},i\in 1,2,3\ldots c \tag{4-11}$$

$$W=\sigma\left(\sigma\left(\theta\right)\right)\times\sigma\left(U_{gc}^{w}\right)+\left(1-\sigma\left(\theta\right)\right)\times\sigma\left(U_{lc}^{w}\right) \tag{4-12}$$

其中$U_{gc}^{w}$和$U_{lc}^{w}$是融合后的全局通道权重和局部通道权重，符号$\theta$表示的是sigmoid激活函数。这种方法允许避免局部信息和全局信息之间的冗余交叉相关操作，同时进一步促进它们之间的相互作用。最后选择性地强调了有信息量的特征，抑制了较不有用的特征，并实现了对图像相关特征的更精确的权重分配。将获得的权重与输入特征图相乘，其结果如下式(4-13)所示：

$$F^{*}=W\otimes F \tag{4-13}$$

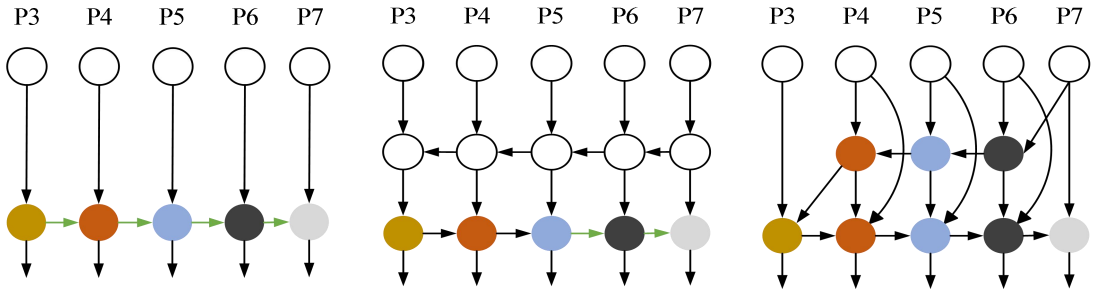
式(4-13)中$F$是输入特征图，$F^{*}$是最终输出的特征图。

鉴于以上启发，本研究在 Centernet 模型的基础上，创新性地在主干网络 DLA-34 中嵌入 FCAttention 模块，该模块的引入旨在网络能够更有效地聚焦于图像中的有用特征，同时抑制无关或干扰性的特征，从而提升网络模型对工件表面关键点的检测精度。考虑到实际工业场景中视觉采集图像的复杂性，比如目标工件放置于带有条纹、污渍及划痕等干扰因素的复杂背景下，这些背景噪声会干扰检测模型对工件的识别，影响关键点的检测。为此，本研究引入的自适应细粒度通道注意力机制，在工件表面关键点检测网络中，通过自适应的方式分配特征权重，能够在不同的任务中灵活调整关注的特征，从而更好地反映输入图像的细节和上下文信息。通过这种方式，Centernet 网络模型能够更有效地捕捉目标工件在不同维度上的重要特征，从而在复杂工业环境中实现更精确的关键点检测。

### 4.3.2 多尺度特征融合

传统的特征金字塔网络(Feature Pyramid Network,FPN)能够有效地融合底层特征图的位置信息，然而，由于高层特征图在 2 倍上采样过程中语义信息被削弱，导致其表达能力受到限制。PANet 应用了简洁的双向融合结构，在 FPN 基础上添加了了自下而上的融合通道，显著提高了模型预测精度[79]。在此基础上，

BiFPN(加权双向特征金字塔网络)进一步优化了特征融合机制。从图 4-5(c)中可以看出，不同于 PANet，BiFPN 移除了一些对网络贡献小的节点，同时额外额外增加了一条从输入节点到输出节点的直接连接，增强了特征传递的效率。而在融合过程中，考虑到各级特征对融合后特征的贡献度不同，BiFPN 为每一个输入点赋予一个权重，这样有效解决了 FPN 和 PANet 在特征融合过程中未能充分考虑各级特征贡献度的问题，进一步提升了特征金字塔的表达能力和检测性能。



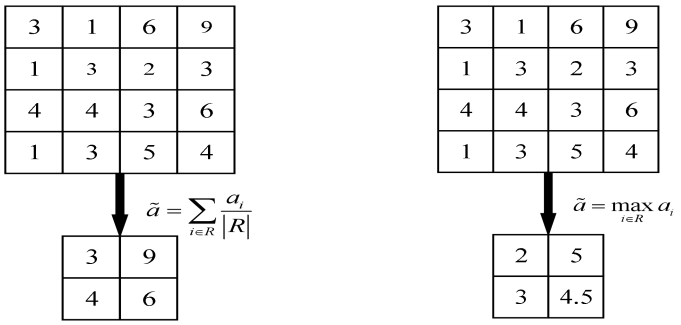

(a)FPNet　　(b)PANet　　(c)BiFPN

图 4-5 不同特征融合网络结构图

在 CenterNet 网络进行工件数据集训练和预测时，主干网络在生成多分辨率特征图的过程中仍然存在特征信息丢失的问题。原始 CenterNet 模型的定位结果只能通过操作具有语义信息的高级特征层来获得。因此，在进行金属工件表面关键点检测时，会丢失一些低层次的目标信息，忽略大尺度跨度的目标信息，在一定程度上影响了检测结果的准确性。因此本文在主干网络后引入双向加权特征融合模块(BiFPN)，将不同尺度特征图上下文信息有效融合，在保留足够细节的同时，提高随机摆放的工件表面关键点检测精度。

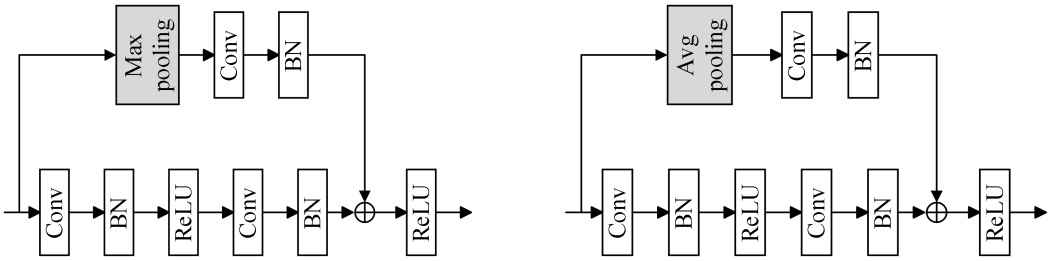
为增强特征层的特征保留能力，本研究将原始模型残差分支中的最大池化操作替换为平均池化。传统最大池化方法仅提取每个池化窗口内的最大值作为特征表示，虽然能够突出显著特征，但会丢失非最大值所包含的金属工件的纹理细节特征。此外，最大池化将局部区域特征压缩为单一值，导致空间信息的损失。相比之下，平均池化通过计算池化区域的特征均值，能够更全面地保留目标对象的细节特征，从而提高检测模型的检测能力，实现对多关键点较为精准的定位。如图 4-6 所示为不同池方式对比，其中 $a$ 代表一个大小为 4×4 的激活区域，局部区域为 R，经过池化后的输出为 $\tilde{a}$ 。所以为保留目标对象的细节信息，在下采样中使用平均池化(如图 4-7)，提升模型对全局特征信息提取的能

力。




(a)平均池化　　(b)最大池化

图 4-6 不同池化方式对比




(a)改进前　　(b)改进后

图 4-7 下采样改进结构

## 4.4 实验内容与分析

### 4.4.1 关键点检测精度评价指标

本文采用关键点预测坐标与真实坐标之间的欧式距离的平均值作为评价指标，以此来评估检测模型在关键点定位方面的准确性。此外，当检测到的关键点与标记的实际位置之间的欧式距离 $d$ 不大于某个误差值 $R_{ture}$，就判定该关键点坐标检测准确。

若 $(x_t, y_t)$ 为数据集中关键点的真实像素坐标，$(x_p, y_p)$ 为检测出的关键点像素坐标，其欧式距离 $d$ 和关键点的距离误差平均值 $L$ 计算方式如下式(4-14)、(4-15)所示：

$$d=\sqrt{\left(x_p-x_t\right)^2+\left(y_p-y_t\right)^2} \tag{4-14}$$

$$L=\frac{\sum_{i=1}^{N}\left(\frac{1}{n}\sum_{j=1}^{n}d_{i,j}\right)}{N} \tag{4-15}$$

上式 $d_{i,j}$ 表示图中第 $i$ 个图的第 $j$ 关键点的预测坐标与真实坐标距离，$N$ 表

示样本数据集的数量，$n$ 为图中关键点数量。若 $d < R_{ture}$，则认为关键点检测正确。为了统计数据集关键点检测的正确率，定义 $\delta(d < R_{true}) = 1$，$\delta(d \geq R_{true}) = 0$，则关键点检测的准确率为：

$$AP_I = \frac{\sum_{i=0}^{N} \delta(d_i < R_{ture})}{N} \tag{4-16}$$

$AP_1 ... AP_n$ 定义为 $n$ 个关键点的准确率，则其所有关键点的平均准确率 $mAP$ 为：

$$mAP = \frac{\sum_{I=1}^{n} AP_I}{n} \tag{4-17}$$

### 4.4.2 数据集标注

为保证对关键点检测阶段的图像输入，本文基于工业场景中所使用的各类工件建立数据集。其中部分图像为实际拍摄获取，其它部分为第 3 节基于生成对抗网络模型生成，本文所收集的工业金属工件图像数量共计为 1800 张，其中共包含六种工件，为验证本文所提出的改进后的多关键点检测方法的广泛适用性，将不同种类工件数据集进行训练测试，运用关键点检测评价指标计算检测效果。为使图像能够用于训练，本文对收集到的图像进行了人工标注，使用 labelme 标注工具为各类型工件进行水平包围框以及工件表面多个关键点标注，其标注结果为工件类别、包围框的两个坐标参数、5 个关键点标签及坐标参数，包围框的 4 个坐标参数分别为 $X_{\min}$、$Y_{\min}$、$X_{\max}$ 和 $Y_{\max}$，其中，$X_{\min}$、$Y_{\min}$ 代表包围框左上角点的坐标，$X_{\max}$、$Y_{\max}$ 代表包围框右下角点的坐标，其标注效果如下图 4-8 所示。

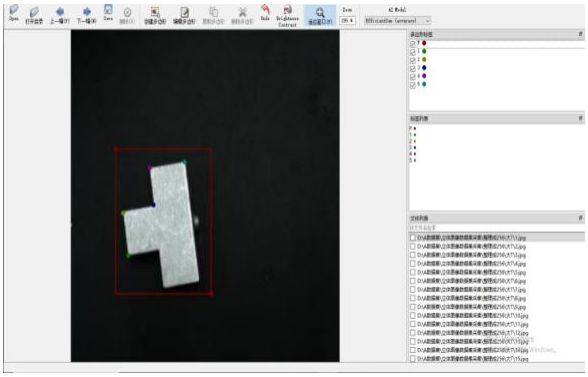


图 4-8 工件关键点标注图

### 4.4.3 数据增强图像关键点检测

数据增强是一种通过模拟真实场景、扩充数据集规模并丰富数据特征多样性的技术手段，旨在提升模型的检测精度、泛化能力以及整体性能。本文采用自建数据集的方法生成数据样本，为确保其规范性、可用性以及是否能够有效支持深度学习模型的训练与预测，需对其进行系统性测试。本节以 Centernet 算法为测试模型，以图像数据增强模型生成的若干图片为待训练数据集，并利用距离误差平均值 $L$ 和平均准确率 $mAP$ 作为模型关键点检测性能的评估指标，下表 4-1 给出了本文数据增强方法前后检测模型的结果对比。其中第一组数据集为相机采集的原始数据集，图像按照训练集：验证集：测试集＝8：1：1 的比例进行划分；第二组实验使用了数据增强方法生成的图像；第三组实验数据集是由前两组实验数据集的结合构建的，其中相机采集的图像数量与数据增强模型 CycleGan 生成的图像数量相同。上述每组实验是在 256×256 像素的输入图像分辨率下进行的。

表 4-1 实验结果

| 组数 | $L$ | $mAP$ |
|---|---|---|
| 第一组 | 1.147 | 0.546 |
| 第二组 | 1.361 | 0.422 |
| 第三组 | **0.983** | **0.563** |

根据实验结果可知，在以生成对抗网络 CycleGan 生成的金属工件图像作为训练数据集的第二组实验中，对测试集进行工件表面多个关键点位置进行精度验证时，其预测的点位置误差值高于第一组。这是由于 CycleGan 生成的图像虽然在视觉上与实际采集的图像具有较高的相似性，但图像中的细微结构与纹理等细节信息与实际相机采集的图像仍存在差距无法精准还原，从而对相机采集图像的关键点检测产生干扰，而第三组实验中同时存在相机采集图像和 CycleGan 生成的图像，其距离误差平均值 $L$ 为 0.983 像素、检测成功率 $mAP$ 为 56.3%。相较于其它组实验，第三组实验的数据集在关键点检测任务中展现出了更为优异的检测效果，体现出更高的鲁棒性与准确性。以上数据说明了本文所构建数据集的真实性与可靠性，不仅为降低人工采集图像的成本提供了有效途径，还通过在有限工件图像数据集中增加数据增强模型生成的图像提升视觉检测模型的检测性能。

### 4.4.4 消融实验

为验证所提多关键点检测模型进行改进方法的有效性，在相机采集的工件图像数据集上进行消融实验。表 4-2 呈现了在三种不同类型工件的测试数据集上进行的消融结果，该实验以 Centernet 作为原始模型，并利用距离误差平均值 $L$ 和平均准确率 $mAP$ 作为模型关键点检测性能的评估指标，而图 4-9 展示在不同工件测试集上原始网络和改进后的检测模型的关键点预测结果，图中每个不同颜色的点表示了工件表面不同的关键点位置。

表 4-2 不同方法检测结果

| | CenterNet | | CenterNet+FCA | | CenterNet+BiFPN | | 本文方法 | |
|---|---|---|---|---|---|---|---|---|
| | *L* | *mAP* | *L* | *mAP* | *L* | *mAP* | *L* | *mAP* |
| 工件一 | 1.328 | 0.406 | 1.177 | 0.444 | 1.256 | 0.407 | **1.091** | **0.469** |
| 工件二 | 1.281 | 0.525 | 0.953 | 0.583 | 0.875 | 0.667 | **0.833** | **0.698** |
| 工件三 | 1.147 | 0.546 | 0.874 | 0.633 | 1.078 | 0.576 | **0.803** | **0.658** |

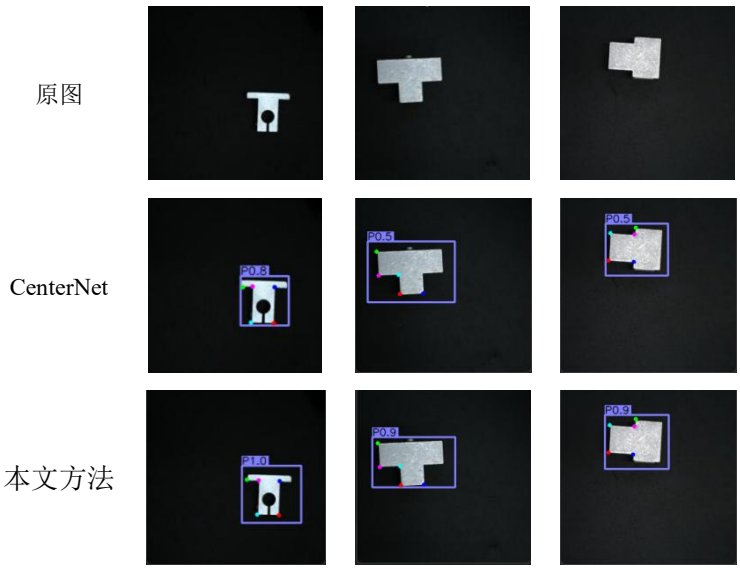



图 4-9 检测结果对比图

通过对表 4-2 的分析可以得出，进行不同类型工件进行多关键点检测时，不同模块的引用对原网络的检测性能均有一定程度的提升。其中 FCAttention 机制的加入对网络检测性能的提升明显高于 BiFPN，而将 FCAttention 机制与 BiFPN 相结合的改进策略，使得该检测模型性能最优。而在图 4-9 中可以看出，本文改进后的关键点检测模型获取点的位置绝大部分与实际位置更接近。通过上述实验结果表明，本文所提方法对 Centernet 实现对工件表面多关键点的检测具有优越的准确度，改进后的模型具有良好的性能。

### 4.4.5 对比实验

为了更进一步验证本文算法的检测性能，本文进行了不同模型的对比实验。采用相同的工件数据集和参数的设置，在 Yolov7 关键点检测网络上进行了训练和测试，并与用本文方法改进后的 CenterNet 进行了比较，对比结果如表 4-3 所示。

表 4-3 不同目标检测模型实验结果

| | CenterNet | Yolov7 | 本文方法 |
|---|---|---|---|
| $L$ | 1.328 | 1.495 | **1.091** |
| $mAP$ | 0.406 | 0.397 | **0.469** |

由表 4-3 可以看出，在对目标金属工件的多关键点检测任务中，相较于自上而下的目标检测一阶段算法 Yolov7，改进后的 CenterNet 关键点检测模型对关键点位置信息的获取更为准确。而在同一个工件数据集下，改进后的 CenterNet 算法在距离误差的平均值 $L$ 上比原始 CenterNet 算法降低了 17.8%，同时在 $mAP$ 值上提高了近 8.7%。因此，从关键点检测精度上看，本文优化算法具有明显的优势。

## 4.5 本章小结

本章针对金属工件的多关键点检测，采用了 CenterNet 算法作为工件表面关键点识别基础研究算法，并对于 CenterNet 的基础网络架构进行了改进，实现对多关键点检测精度的提升。主要包括对 CenterNet 的主干网络 DLA-34 的改进，对上采样部分引入自适应细粒度通道注意力，以及在主干网络后加入 BiFPN，以此融合各个层级特征信息，解决了目标特征损失与消失的问题，丰富了空间信息，进一步提高了工件的多关键点检测精度。改进后的 CenterNet 关键点检测算法在自建的工件数据集上进行了验证性实验，在不同数据集上检测的关键点位姿和实际关键点位姿之间的平均距离误差值均小于原始模型。结果表明：本文提出的改进的 CenterNet 在检测的准确度较高，从而验证本文设计的抓取目标检测网络模型在目标检测任务上的有效性。本章的工作为后续双目视觉 6DOF 位姿测量提供了可靠的数据支持，对于实现机器人灵巧手成功抓取具有重要意义。

# 第 5 章 基于双目视觉的 6DOF 机器人位姿测量及抓取

## 5.1 引言

通过前面章节的研究，进行了整个抓取策略的数据增强、关键点检测、手眼标定等工作，并在自建的工件数据集上进行算法的验证。基于这些研究基础，本章首先介绍了机器人抓取实验所涉及的硬件平台和软件系统组成，然后为验证本文所提的基于双目视觉引导的 6DOF 机器人灵巧手抓取策略的有效性，在工件数据集上进行多组对比实验，评估 6DOF 位姿测量算法的精度。最后通过实际金属工件的识别与抓取实验，进一步验证提出抓取策略的可行性。

## 5.2 实验平台搭建

### 5.2.1 实验平台硬件

本研究构建的硬件平台主要包括 UR5 机械臂，灵巧手、双目相机，计算机，机器人控制器等，如图 5-1 所示。当机器人开始工作时，双目视觉系统首先获取目标工件在左右相机中的图像信息，并将数据实时传输至计算机，随后计算机对目标工件表面关键点位姿进行识别与定位，同时基于预先标定的手眼转换将工件在相机中的位姿通过计算机转换到机械臂坐标系下，最终，计算机将获取得到的抓取位姿发送至运动控制器，驱动机器人执行抓取动作，成功抓取目标工件。

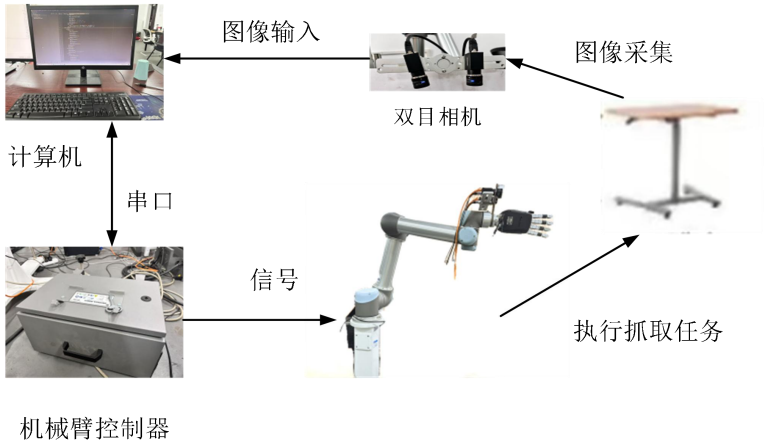



图 5-1 机器人抓取流程

1、UR5 简介

UR5 机械臂是丹麦优傲机器人公司的一款经典机械臂，其主要参数如表 5-1 所示。

表 5-1 UR5 机械臂主要参数

| 项目 | 运动副 | 最大有效载荷/kg | 最大延伸/mm | 关节活动范围/° |
|---|---|---|---|---|
| 参数 | 6 转动副度 | 5 | 850 | 正负 360 |

2、五指灵巧手

末端执行器是机器人在抓取作业中直接执行操作任务的关键部件，通常安装于机械臂末端，实现对不同类型的金属工件的抓取任务。末端执行器有很多种，本文选择中国科学院自动化研究所研发的 Casia Hand 系列仿人灵巧手来与机械臂配合使用。灵巧手如图 5-2 所示，该灵巧手与 UR5 机械臂组合，使抓取具有更高的准确性与实时性，抓取过程较为稳定。

图 5-2 五指灵巧手

3、双目相机与镜头选型

为了完成金属工件的识别与检测以及机械臂灵巧手的抓取工作，需要相机完成图像的采集。本文使用两个海康威视工业相机，其型号为 MV-CS050-10GC，该设备具有高动态范围、优异成像质量和低功耗等特点，具体参数如表 5.2 所示。工业相机镜头采用 M0814-MP2，镜头的焦距为 8mm。

表 5-2 相机参数

| 相机参数 | 数值 |
|---|---|
| 型号 | MV-CS050-10GC |
| 尺寸 | 29mm×29mm×42mm |
| 重量 | 约 100g |
| 分辨率 | 2448×2048 |
| 动态范围 | 72dB |
| 信噪比 | 40dB |

### 5.2.2 实验平台软件搭建

本文的软件系统主要包括有：Microsoft Visual Studio2019，OpenCV4.4.0，python3.9 和相机驱动，其具体参数与 3.5 节使用 Windows 系统一样。本文采用 VS 联合 C++编程实现了实验平台的可视化界面，采集的工件图像的处理由 HALCON 进行多样化操作。双目图像采集界面如图 5-3 所示，图像分辨率大小为 1280×1024。

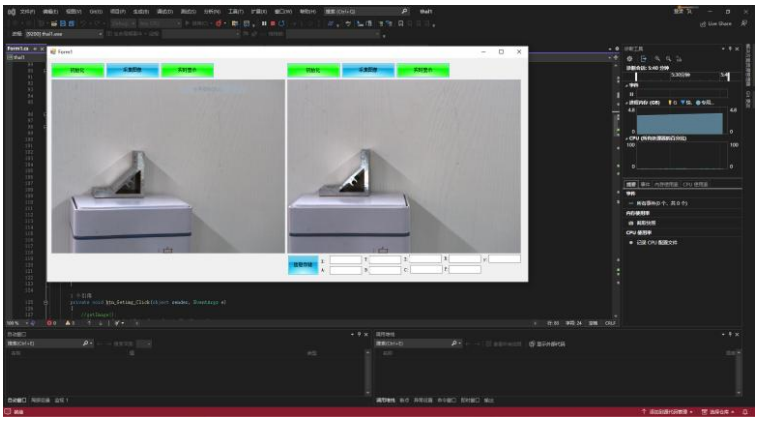


图 5-3 双目图像采集界面

## 5.3 关键点识别与位姿测量实验

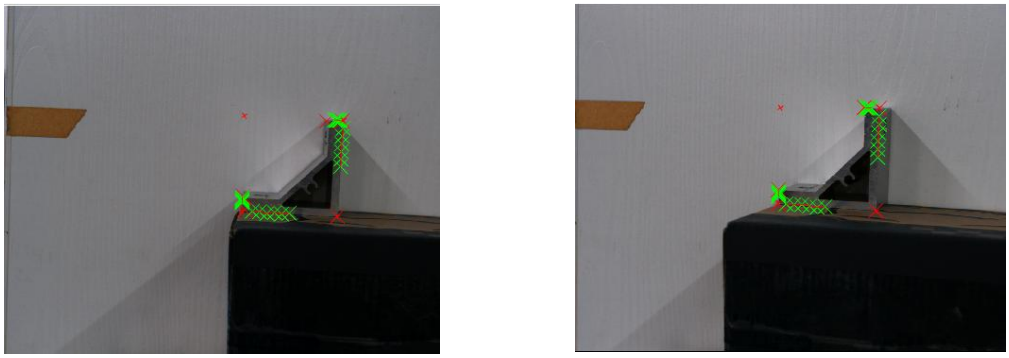
### 5.3.1 多关键点坐标识别及获取

通过搭建的机器人灵巧手抓取平台，对需要进行抓取的目标工件进行一定数量的图像采集，利用第 3、4 章方法对图像中的工件对象进行多个关键点的检测。其中所有关键点均为目标对象的同一平面上的特征点，避免不同平面关键点位姿对位姿测量的影响。当用双目相机拍摄图像进行关键点检测后，需要对其进行预测点信息保存，同时记录采集该图像时机械臂末端位姿，为后续位姿测量实验提供数据支持。

由于多关键点检测网络预测的工件表面点的位姿存在一定误差，直接利用相似三角形的性质求解位姿估计问题的方法可能导致预测点与实际关键点之间的误匹配，降低位姿求解的精度。为解决该问题，本文运用一种加权最小二乘拟合法。该算法是一种常用的直线拟合方法，能显著降低异常点对多点直线拟合的干扰。因此对于神经网络预测中存在的潜在偏差情况，本文在进行关键点检测任务中引入加权最小二乘拟合算法，可以消除数据中的异常值对多点拟合

直线的影响，达到对目标对象表面多关键点的位姿信息获取。

通过将检测模型预测的坐标作为初始输入，对目标对象边缘若干点进行识别获取，然后使用一种加权最小二乘拟合法方法迭代地选择并更新每个点的位姿信息，从而剔除来自异常关键点的位姿信息，实现对左、右相机采集的图像中物体边缘进行直线拟合(如图 5-4)，在此基础上通过两直线的交点位姿来更新物体表面多关键点信息，最终进行准确的位姿估计。



(a)左相机图像　　(b)右相机图像

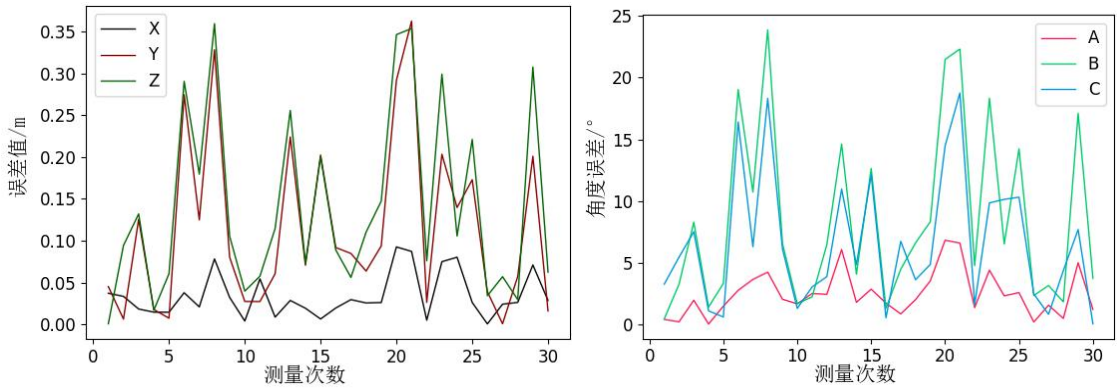
图 5-4 工件边缘直线拟合图

### 5.3.2 目标对象 6DOF 位姿测量实验

在进行机器人抓取实验之前，需对本文所提方案的 6DOF 位姿测量精度进行验证。实验采用对比分析法，将基于双目视觉系统获取的机器人抓取位姿(实际位姿)与成功抓取时的理论位姿进行比对，获取机器人抓取位姿的误差值。实验具体过程为：首先目标工件固定不动，通过移动机械臂采集若干组图像；接着使用本文所提基于深度学习的多关键点检测模型输出所需的物体表面关键点信息，同时以第 2 章中双目视觉系统进行相机标定和手眼标定的数据作为依据，计算出机械臂抓取工件时的位姿信息；然后为获取更加精准的的关键点位姿信息，在检测模型所得的工件表面关键点位姿信息的基础上，结合边缘点和角点共线的特点，采用加权最小二乘拟合算法，使得目标工件的边缘特征点被加权以减小异常点位置数据的影响，进行不同直线两两相交来获取工件表面关键点的位姿信息；最后依据这些位姿信息获取机器人灵巧手的抓取位姿，并对比分析拟合方法优化前后的抓取位姿偏差。

本文共采集了三十组图像进行实验，将基于检测模型预测的关键点信息获取的抓取位姿与机器人理论位姿进行对比，计算抓取位姿的误差值。从图 5-5

中可以看出，在使用加权最小二乘拟合算法前，机器人抓取位姿在 $x$、$y$、$z$ 轴的误差值范围为 0~0.35m。$x$、$y$、$z$ 轴的旋转误差值范围为 0.0447°~25°。为了更加直观看出具体误差数值，本文于表 5-3、5-4 展示了其中 14 组数据。通过对比对照组位姿，发现在机器人的抓取位姿误差值偏大，这主要由于机器人位姿测量受双目视觉标定、机器人手眼标定以及一些连接件的安装精度导致，另外视觉检测的微小偏差也会对后续位姿测量任务造成干扰，导致预测的机器人抓取位姿与对照组的抓取位姿存在偏差。




(a)x、y、z 轴的误差对比　　(b)x、y、z 轴旋转误差对比

图 5-5 拟合前抓取位姿误差对比

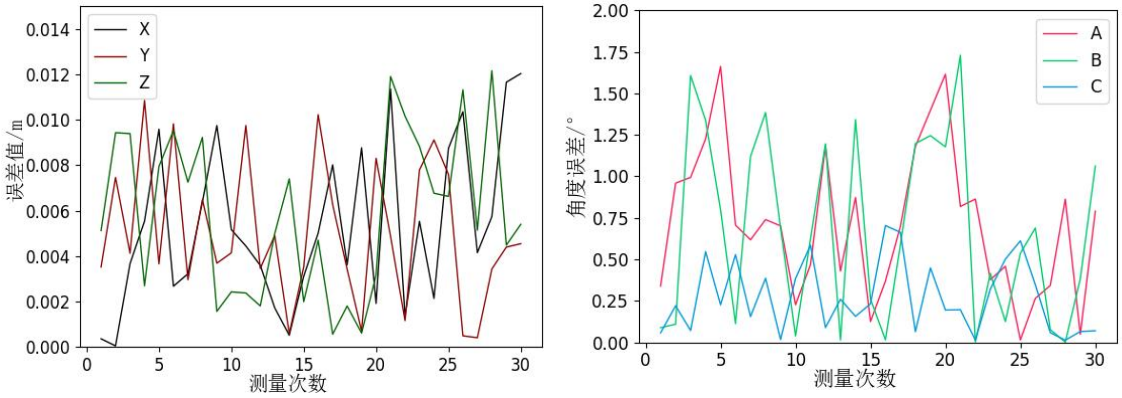
表 5-3 拟合前 $x$、$y$、$z$ 轴平移的误差

| 实验组 | 空间坐标(m) | $\Delta x$ (m) | $\Delta y$ (m) | $\Delta Z$ (m) |
|---|---|---|---|---|
| 对照组 0 | [0.0,0.0,0.0] | -- | -- | -- |
| 测试组 1 | [0.03739,0.0453,-0.00082] | 0.03739 | 0.0453 | 0.00082 |
| 测试组 2 | [-0.03373,-0.07848,-0.04744] | 0.03373 | 0.07848 | 0.04744 |
| 测试组 3 | [-0.01849,-0.12571,-0.13213] | 0.01849 | 0.12571 | 0.13213 |
| 测试组 4 | [-0.01472,-0.01795,-0.01703] | 0.01472 | 0.01795 | 0.01703 |
| 测试组 5 | [-0.01465,0.00748,0.06057] | 0.01465 | 0.00748 | 0.06057 |
| 测试组 6 | [0.03796,-0.2743,-0.29053] | 0.03796 | 0.2743 | 0.29053 |
| 测试组 7 | [0.02101,-0.12486,-0.17965] | 0.02101 | 0.12486 | 0.17965 |
| 测试组 8 | [0.07835,-0.32816,-0.35947] | 0.07835 | 0.32816 | 0.35947 |
| 测试组 9 | [0.03237,0.08038,0.10556] | 0.03237 | 0.08038 | 0.10556 |
| 测试组 10 | [-0.00413,-0.02753,-0.03988] | 0.00413 | 0.02753 | 0.03988 |
| 测试组 11 | [-0.03988,-0.02742,0.0577] | 0.03988 | 0.02742 | 0.0577 |
| 测试组 12 | [-0.00881,0.06057,0.11466] | 0.00881 | 0.06057 | 0.11466 |
| 测试组 13 | [0.02871,-0.22406,-0.25567] | 0.02871 | 0.22406 | 0.25567 |
| 测试组 14 | [0.01941,0.07073,0.07292] | 0.01941 | 0.07073 | 0.07292 |

表 5-4 拟合前 $x$、$y$、$z$ 轴旋转的误差

| 实验组 | 空间坐标(°) | $\Delta A$ (°) | $\Delta B$ (°) | $\Delta C$ (°) |
|---|---|---|---|---|
| 对照组 0 | [180.0,180.0,180.0] | -- | -- | -- |
| 测试组 1 | [0.403622,359.551,356.742] | 0.4036 | 0.449 | 3.258 |
| 测试组 2 | [359.79,356.746,5.44824] | 0.21 | 3.254 | 5.4482 |
| 测试组 3 | [358.043,351.722,7.51098] | 1.957 | 8.278 | 7.511 |
| 测试组 4 | [0.0280945,358.593,1.09735] | 0.0281 | 1.407 | 1.0974 |
| 测试组 5 | [1.47833,3.31944,359.393] | 1.4783 | 3.3194 | 0.607 |
| 测试组 6 | [357.241,340.973,16.3996] | 2.759 | 19.027 | 16.3996 |
| 测试组 7 | [356.353,349.278,6.28682] | 3.647 | 10.722 | 6.2868 |
| 测试组 8 | [355.764,336.127,18.3159] | 4.236 | 23.873 | 18.3159 |
| 测试组 9 | [2.03353,6.5326,353.891] | 2.0335 | 6.5326 | 6.109 |
| 测试组 10 | [358.336,358.324,1.29183] | 1.664 | 1.676 | 1.2918 |
| 测试组 11 | [2.49768,2.22292,3.02715] | 2.4977 | 2.2229 | 3.0271 |
| 测试组 12 | [2.43254,6.45631,356.129] | 2.4325 | 6.4563 | 3.871 |
| 测试组 13 | [353.938,345.39,10.964] | 6.062 | 14.61 | 10.964 |
| 测试组 14 | [1.78597,4.06766,355.163] | 1.786 | 4.0677 | 4.837 |

通过加权最小二乘拟合法获取较为精准的关键点信息后，对比机器人抓取的理论位姿，机器人抓取的实际位姿误差值的折线图绘制于图 5-6 中，三十组实验中 $x$、$y$、$z$ 轴的误差值范围为 0~0.012m。$x$、$y$、$z$ 轴的旋转误差值范围为 0.003°~1.729°。为了更加直观看出具体误差数值，本文于表 5-5、5-5 展示了 14 组数据对比。尽管仍然受双目视觉标定、机器人手眼标定以及一些连接件的安装精度影响，通过精加工后的关键点位姿输入获取了较高精度的机器人抓取位姿。




(a)x、y、z 轴的误差对比　　(b)x、y、z 轴旋转误差对比

图 5-6 拟合后位姿误差对比

表 5-5 拟合后 $x$、$y$、$z$ 轴平移的误差

| 实验组 | 空间坐标(m) | $\Delta x$ (m) | $\Delta y$ (m) | $\Delta Z$ (m) |
|---|---|---|---|---|
| 对照组 0 | [0.0,0.0,0.0] | -- | -- | -- |
| 测试组 1 | [-0.00036,0.00352,0.00512] | 0.00036 | 0.00352 | 0.00512 |
| 测试组 2 | [-4.87586e-05,0.00746,0.0094] | 0.00005 | 0.00746 | 0.0094 |
| 测试组 3 | [0.00367,0.00412,-0.00939] | 0.00367 | 0.00412 | 0.00939 |
| 测试组 4 | [0.00553,0.01085,-0.00269] | 0.00553 | 0.01085 | 0.00269 |
| 测试组 5 | [0.00958,0.00366,0.00795] | 0.00958 | 0.00366 | 0.00795 |
| 测试组 6 | [0.00958,0.00982,0.00952] | 0.00958 | 0.00982 | 0.00952 |
| 测试组 7 | [0.00322,0.00296,-0.00725] | 0.00322 | 0.00296 | 0.00725 |
| 测试组 8 | [0.00638,0.00649,-0.00922] | 0.00638 | 0.00649 | 0.00922 |
| 测试组 9 | [0.00974,-0.00369,-0.00157] | 0.00974 | 0.00369 | 0.00157 |
| 测试组 10 | [0.00516,0.00414,-0.00243] | 0.00516 | 0.00414 | 0.00243 |
| 测试组 11 | [0.00445,0.00974,-0.00238] | 0.00445 | 0.00974 | 0.00238 |
| 测试组 12 | [-0.00359,0.00346,-0.00181] | 0.00359 | 0.00346 | 0.00181 |
| 测试组 13 | [-0.00172,-0.00491,-0.00499] | 0.00172 | 0.00491 | 0.00499 |
| 测试组 14 | [0.00052,0.00058,-0.0074] | 0.00052 | 0.00058 | 0.0074 |

表 5-6 拟合后 x、y、z 轴旋转的误差

| 实验组 | 空间坐标(°) | $\Delta A$ (°) | $\Delta B$ (°) | $\Delta C$ (°) |
|---|---|---|---|---|
| 对照组 0 | [180,180,180] | -- | -- | -- |
| 测试组 1 | [0.339797,0.0891948,359.942] | 0.3398 | 0.0892 | 0.058 |
| 测试组 2 | [0.95936,359.89,359.779] | 0.9594 | 0.11 | 0.221 |
| 测试组 3 | [0.993775,358.393,359.928] | 0.9938 | 1.607 | 0.072 |
| 测试组 4 | [1.22592,358.663,359.453] | 1.2259 | 1.337 | 0.547 |
| 测试组 5 | [1.66143,359.201,359.773] | 1.6614 | 0.799 | 0.227 |
| 测试组 6 | [0.706123,0.11356,359.471] | 0.7061 | 0.1136 | 0.529 |
| 测试组 7 | [0.618187,358.88,359.844] | 0.6182 | 1.12 | 0.156 |
| 测试组 8 | [0.739883,358.616,359.613] | 0.7399 | 1.384 | 0.387 |
| 测试组 9 | [0.702008,359.311,359.982] | 0.702 | 0.689 | 0.018 |
| 测试组 10 | [359.773,0.0399114,359.615] | 0.227 | 0.0399 | 0.385 |
| 测试组 11 | [0.46825,359.377,359.414] | 0.4682 | 0.623 | 0.586 |
| 测试组 12 | [1.18123,358.804,0.0894691] | 1.1812 | 1.196 | 0.0895 |
| 测试组 13 | [359.57,359.985,0.260521] | 0.43 | 0.015 | 0.2605 |
| 测试组 14 | [0.872944,358.658,0.157178] | 0.8729 | 1.342 | 0.1572 |

通过对比分析拟合前后的位姿误差数据，可以明显观察到采用加权最小二

乘拟合算法后的位姿测量精度得到显著提升。实验结果表明，在工业领域常见的复杂场景下，当视觉系统采集到的金属工件图像因局部反光而导致关键特征信息丢失时，本文所提抓取策略有效改善反光干扰对工件位姿测量的影响，实现了机器人对金属工件较高精度的 6DOF 抓取位姿测量。

## 5.4 基于双目视觉机械臂的抓取实验

为验证本文所提抓取策略对金属工件抓取的的有效性，本节进行机器人的抓取实验。首先通过双目视觉系统采集图像信息，利用关键点检测网络获取目标工件关键点的位姿信息；然后结合第 2 章中视觉系统进行的相机标定和手眼标定结果进行计算，采用双目三角测量法进行三维重建，并通过坐标转换计算出机器人灵巧手的精确抓取位姿；最后，将计算得到的位姿数据转换为机器人控制指令，传输至机器人控制器中，控制器根据接收到的指令规划机械臂运动轨迹，驱动机械臂运动至目标位置，完成机器人灵巧手抓取任务。图 5-7 为机器人灵巧手对不同类型工件的抓取过程。

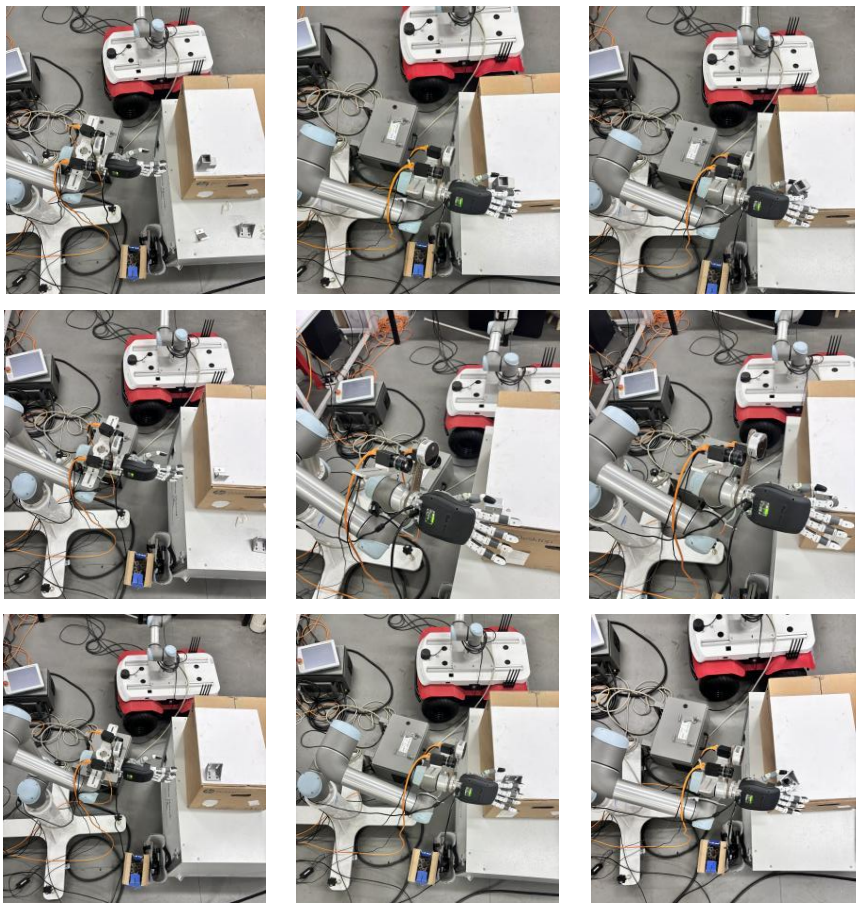


图 5-7 不同类型工件抓取过程

## 5.5 本章小结

本章首先介绍了抓取实验的硬件平台和软件环境，并列出详细参数；然后进行了基于双目视觉的金属工件 6DOF 位姿测量实验，对不同位姿状态下机器人进行了多组误差对比与分析，确定 6DOF 位姿测量的误差范围。实验获取了较高精度的 6DOPF 机器人抓取位姿，表明了本文所提策略适用于机器人对金属工件的抓取任务，稳定性和准确率俱佳，最后用获取的机器人抓取位姿控制机器人灵巧手完成抓取任务。

# 第 6 章 总结与展望

## 6.1 工作总结

本文针对基于双目视觉的机器人的 6DOF 位姿测量算法开展研究，主要研究了图像数据增强技术、关键点检测、边缘拟合以及目标对象的位姿测量等，同时针对本文所提的一种基于视觉引导的机器人抓取策略搭建了基于双目视觉的 6DOF 机器人抓取平台，通过相机标定、手眼标定以及坐标转换等获取的数据设计了目标对象 6DOF 位姿测量实验，验证本文抓取策略的有效性，最终通过机器人灵巧手实现对金属工件的抓取。主要完成了以下工作：

(1)对基于生成对抗网络的图像数据增强技术进行了相关研究。为降低人工采集图像成本，利用 CycleGan 不需要进行数据配对、生成多样化、数据利用率高和循环一致性损失提高模型稳定性等优点，实现对有限图像的扩增。在进行图像数据增强模型训练前，为构建数据集，应用了传统图像处理方法，对有限数据集进行一定数量的扩增操作。在此基础上结合随机算法生成了在随机摆放状态下的金属工件图像，构建 CycleGan 网络模型的工件数据集。为促使生成对抗网络 CycleGan 生成较高质量的图像，本文从模型的网络架构层面入手，通过在生成器中引入多头自注意力模块、替换转换器部分以及在判别器部分实现增设多尺度判别，引导生成器生成更贴近真实相机采集效果的图像，为后续机器人的定位检测提供高质量、丰富多样的数据集。实验结果表明，以图像生成质量评价指标作为数据依据，相比于模型改进前，在不同数据集的测试数据结果上，本文提出的方法生成的图像质量均有较大提升。

(2)研究了机器人抓取任务中工件表面多关键点信息获取方法。在 CenterNet 初始模型架构上，通过加入数据增强模型生成的图像数据集来提升模型的鲁棒性和泛化能力。随后为了提高模型对目标工件的检测性能，在 CenterNet 的主干网络中加入 FCAttention 机制，该机制它能够自适应地关注不同通道上的特征信息，使得网络模型能够更准确地强调有用特征，抑制无关或干扰性的特征。考虑到模型训练过程中出现目标特征损失与消失问题，本文利用 BiFPN 模块能融合各个层级特征信息的特性来进一步提高 CenterNet 网络对多关键点的检测精度。

实验数据显示，改进后 CenterNet 网络在多种工件数据集上的检测性能均取得了提升。对比初始模型，改进后的 CenterNet 模型在关键点检测方面，距离误差平均值最大降幅可达到了 0.344 像素，检测成功率提高了 11.2%。

(3)在成功获取工件多关键点信息的基础上，为验证基于视觉引导的 6DOF 机器人灵巧手抓取方案可行性，本文搭建了一个基于双目视觉的 6DOF 机器人抓取平台。依托该平台，开展了机器人 6DOF 位姿测量实验，同时应用加权最小二乘拟合算法获取更高精度的机器人抓取位姿。实验结果表明，在金属工件抓取位姿测量中，本文所提方法的获取了较高精度的机器人 6DOF 抓取位姿。为进一步验证抓取方法的可靠性，进行了视觉引导的机器人抓取实验，最终机器人凭借较高精度的 6DOF 抓取位姿，成功且顺利地完成了对目标工件的抓取任务。

通过以上工作，证明本文提出的基于双目视觉的机器人的 6DOF 位姿测量算法在实验中有着较高精度的位姿测量结果，同时使得基于双目视觉引导的机器人成功抓取金属工件，为复杂环境中的 6DOF 机器人抓取提供思路。

## 6.2 展望

基于双目视觉的机器人 6DOF 位姿测量及抓取是一个包含许多细节的课题，尽管本文所提方法获得了较高精度的抓取位姿，但由于个人能力和时间限制，该研究还存在着以下不足之处，需要后续进一步更深入地进行研究。

(1)本研究的 6DOF 机器人抓取主要聚焦于单个工件的抓取任务，然而在实际的生产环境中，会存在金属工件遮挡情况。因此未来的研究方向将根据生产中存在的抓取问题，实现在目标工件的局部部位被遮挡下更加准确识别和定位，满足工业领域的高效生产需求。

(2)本文仅通过应用与双目视觉下的三角测量算法和多次迭代来获取图像中场景的深度信息，没有与其它位姿估计算法进行对比实验，后续研究应通过与其它算法进行对比选出可以达到最佳匹配效果的位姿估计算法。

# 参考文献

[1] 邵青伟. 工业机器人在汽车智能制造生产线中的应用研究[J].汽车测试报告,2023,(09):52-54.

[2] Farag E H, Khamesee B M ,Toyserkani E . Eddy Current Sensor Probe Design for Subsurface Defect Detection in Additive Manufacturing[J]. Sensors, 2024,24(16): 5355.

[3] 蒋萌. 基于双目视觉的目标识别定位及机器人抓取研究[D].江苏:南京航空航天大学,2018.

[4] 陈佳兴. 无序堆叠物体的机器人视觉检测与抓取技术[D].山东:青岛科技大学,2024.

[5] 吴晓辉. 基于 3D 视觉的机械臂及手爪协同控制方法研究[D].辽宁：沈阳大学,2024.

[6] 陶邦升. 基于深度学习的机器人智能抓取系统研究[D].上海:上海海洋大学,2024.

[7] 杨雄. 基于视觉的机器人抓取过程的智能学习方法研究[D].四川:西南科技大学,2023.

[8] 高正浩. 基于双目视觉的目标识别与定位方法研究[D].天津:河北工业大学,2023.

[9] 周子祺. 基于双目视觉的方形药瓶定位及抓取系统研究[D].浙江:浙江科技大学,2024.

[10] 曹熙淦. 基于 3D 视觉的机械零件位姿估计方法研究[D].甘肃:兰州理工大学,2023.

[11] Jorgensen T B, Jensen S H N, Aanæs H, et al. An Adaptive Robotic System for Doing Pick and Place Operations with Deformable Objects[J]. Journal of Intelligent & Robotic Systems, 2019,94: 81-100.

[12] Hagelskjaer F , Kramberger A , Wolniakowski A ,et al.Combined Optimization of Gripper Finger Design and Pose Estimation Processes for Advanced Industrial Assembly[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),Macau,China, IEEE, 2019: 4-8.

[13] Kitagawa S, Wada K, Hasegawa S, et al.Few-experiential learning system of robotic picking task with selective dual-arm grasping[J].Advanced Robotics, 2020(4):1-19.

[14] 张晨.基于 3D 视觉的机器人小物品抓取技术研究[D].福建:华侨大学,2023.

[15] 王惠.基于机器视觉的透明物体检测及抓取[D].江苏:中国矿业大学,2023.

[16] 侯维广.基于深度学习的机器人智能抓取研究[D].辽宁:沈阳化工大学,2023.

[17] 孙旭东.基于 YOLOv5s 的环卫机器人双目视觉系统设计及实现[D].陕西:西安理工大学,2024.

[18] 王行洲,秦立宇,陈亮宏,等.多目视觉系统在电芯自动上料设备中的应用[J].机械研究与应用,2024,37(1), 150-152.

[19] 解霄鹏,王佳玮,高波,等.基于视觉反馈的机械臂位置模糊控制[J].应用科技,2023,50(06):111-116+131.

[20] 孙文革.视觉校对约束下生产线搬运机械手抓取定位控制[J].机械制造与自动化,2024,53(05):203-208.

[21] 崔赫.基于深度学习和双目视觉的工业零件抓取检测算法研究[D].四川:西南交通大学,2023.

[22] Marr D, Nishihara H K. Representation and recognition of the spatial organization of three-dimensional shapes[J]. Proceedings of the Royal Society of London. Series B. Biological Sciences, 1978, 200(1140): 269-294.

[23] Bahena J M R, Cruz C A, Lopez A Z, et al. Speed booms detection for a ground vehicle with computer vision[C]//Proceedings of the 12th WSEAS international conference on Mathematical and computational methods in science and engineering. 2010: 258-264.

[24] Benacer I, Hamissi A, Khouas A. A novel stereovision algorithm for obstacles detectionbased on UV-disparity approach[C]//2015 IEEE International Symposium on Circuitsand Systems (ISCAS).Lisbon,Portugal, 2015: 369-372.

[25] 冯英旺. 双目被动测距系统设计及算法研究[D]. 江苏:南京理工大学, 2019.

[26] 李铭.基于双目视觉的物体识别与定位研究[D].湖北:武汉工程大学,2023.

[27] Zhou Y , Li Q , Chu L ,et al. A measurement system based on internal cooperation of cameras in binocular vision[J].Measurement Science and Technology, 2020, 31(6):065002.

[28] 李佳. 基于车载双目相机的行人测距技术研究[D]. 辽宁:沈阳工业大学, 2021.

[29] 单福强.基于双目视觉的非合作目标位姿测量技术的研究[D].北京:中国科学院大学(中国科学院西安光学精密机械研究所),2022.

[30] Sumetheeprasit B, Rosales Martinez R, Paul H, et al. Long-range 3D reconstruction based on flexible configuration stereo vision using multiple aerial robots[J]. Remote Sensing, 2024, 16(2): 234.

[31] 季娟娟,王佳,高少薄,等.基于双目视觉的集装箱吊放位姿测量研究[J].计算机仿真,2024,41(05):296-302.

[32] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal ofcomputer vision, 2004, 60 (2): 91-110.

[33] Bay H, Tuytelaars T, Gool L V. Surf: Speeded up robust features[C]// 9th European Conference on Computer Vision,Graz,Austria, 2006: 404-417.

[34] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]//2011 International conference on computer vision,Barcelona, Spain, 2011: 2564-2571.

[35] 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.

[36] Hinterstoisser S, Cagniart C, llic S, et al. Gradient response maps for real-time detection of textureless objects[J]. lEEE transactions on pattern analysis and machine intelligence, 2011,34(5):876-888.

[37] Hodaň T, Zabulis X, Lourakis M, et al. Detection and fine 3D pose estimation of texture-less objects in RGB-D images[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015: 4421-4428.

[38] 刘鹏翔,梁冬泰,梁丹,等.基于生成对抗网络的单目相机 6D 位姿估计[J].机械设计与研究,2020,36(06):72-78.

[39] Zhang Y, Liu Y, Wu Q, J, et al. EANet: Edge-Attention 6D Pose Estimation Network for Texture-Less Objects[J]. IEEE Transactions on Instrumentation and Measurement,2022,71:1-13.

[40] Shi D , Rahimpour A , Ghafourian A ,et al.Deep Bayesian-Assisted Keypoint Detection for Pose Estimation in Assembly Automation[J].Sensors, 2023, 23(13):6017.

[41] He R, Wang X, Chen H, et al. Vhr-birdpose: vision transformer-based hrnet for bird pose estimation with attention mechanism[J]. Electronics, 2023, 12(17): 3643.

[42] 刘萌萌.基于深度学习的微小元件姿态识别算法研究[D].河南:郑州大学,2020.

[43] 陶贤鹏.基于 WGAN-GP 的水稻图像数据增强方法研究[D].黑龙江:黑龙江八一农垦大学,2023.

[44] Gowda S N, Yuan C. ColorNet: Investigating the importance of color spaces for image classification[C]//Computer Vision – ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, 2019: 581-596.

[45] Momeny M, Latif A M, Agha Sarram M, et al. A noise robust convolutional neural networkfor image classification[J]. Results in Engineering, 2021, 10: 100225.

[46] 韩雨晴.基于数据增强技术的古籍文字识别方法研究[D].福建:厦门理工学院,2021.

[47] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.

[48] 杜青松.基于深度学习的航拍车辆检测算法研究[D].四川:西南交通大学,2021

[49] 宋晓琳.基于深度学习的行人检测算法研究[D].北京:北京邮电大学,2023.

[50] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time objectdetection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016: 779-788.

[51] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprintarXiv:1804.02767, 2018.

[52] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 2016: 21-37.

[53] 黄尧.基于改进 YOLOv5s 的道路目标检测算法研究[D].辽宁:沈阳师范大学,2023.

[54] 王科理.高速铁路供电安全检测监测系统图像智能识别方法研究[D].北京:中国铁道科学研究院,2023.

[55] Ren S , He K , Girshick R ,et al.Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J].IEEE Transactions on Pattern

Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[56] 徐妍,孙维东,赵伶俐,等.利用无锚框型深度网络的 SAR 船舶检测[J].测绘科学, 2024, 000(8):11.

[57] 李露,陈科研,刘辰阳,等.一种无锚旋转框遥感图像船只目标检测算法[J].北京航空航天大学学报,2023:1-9.

[58] Law H, Deng J. Cornernet: Detecting objects as paired keypoints[C]// Proceedings ofthe European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 734-750.

[59] Zhou X, Zhuo J, Krahenbuhl P. Bottom-up object detection by grouping extreme and center points[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019:850-859.

[60] Zhou X, Wang D, Krähenbühl P. Objects as points[J]. arxiv preprint arxiv:1904.07850, 2019.

[61] 端木慧娉.基于双目视觉的协作机器人灵巧抓取关键技术研究[D].上海:上海师范大学,2024.

[62] Zhang Z. System and method for calibrating a camera with one-dimensional objects,US: 2007.

[63] 何林杰.基于双目视觉的工业机器人目标识别与定位技术研究[D].四川:电子科技大学,2022.

[64] 张世童.面向海底管道检测的水下航行器路径规划研究[D].天津:天津大学,2021.

[65] Zhu J Y, Park T, lsola P, et al, Unpaired image-to-image translation using cycle-consistentadversarial networks[C]//Proceedings of the lEEE International Conference on ComputerVision,2017:2223-2232.

[66] 张玉鑫.基于 CycleGAN 的风景图片风格迁移研究[D].甘肃:兰州理工大学,2024.

[67] 李钢.基于改进 CycleGAN 和稳定扩散模型的陶瓷图案生成算法研究[D].江西:景德镇陶瓷大学,2024.

[68] 徐曼.基于 CycleGAN 的 HDR 重建方法研究[D].浙江:浙江理工大学,2022.

[69] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision, Virtual, 2021: 10012-10022.

[70] 王奇坤.基于深度学习和结构导向的单目图像深度估计[D].湖北:武汉大学,2022.

[71] 刘志.基于改进 CycleGAN 的汉字风格迁移研究[D].河北, 河北经贸大学,2024.

[72] 杨博,潘少伟,李宗强,等.基于多尺度和上下文信息的高分辨网络岩心图像分割[J].科学技术与工程, 2024, 24(25):10659-1066.

[73] Kim J, Kim M, Kang H, et al. U-GAT-IT: Unsupervised Generative Attentional Networkswith Adaptive Layer-Instance Normalization for Image-to-Image Translation[J]. ArXiv,abs/1907.10830, 2020.

[74] Park T, Efros A A, Zhang R, et al. Contrastive Learning for Unpaired Image-to-Image Translation[C]//European Conference on Computer Vision (ECCV), Springer, 2020: 319-345.

[75] Han J, Shoeiby M, Petersson L, et al. Dual Contrastive Learning for UnsupervisedImage-to-Image Translation[C]. Computer Vision and Pattern Recognition(CVPR),2021.

[76] 晏鹏.基于深度学习的自动驾驶场景单目 3D 目标检测方法[D].江苏:南京邮电大学,2023.

[77] 于瀛.基于 CenterNet 的水下目标检测方法研究[D].辽宁:大连海事大学,2023

[78] Sun H, Wen Y, Feng H, et al. Unsupervised bidirectional contrastive reconstruction and adaptive fine-grained channel attention networks for image dehazing[J]. Neural Networks, 2024, 176: 106314.

[79] 邢春雨.基于 CenterNet 的口罩佩戴检测算法研究[D].安徽,安徽理工大学,2022.

# 攻读学位期间发表的学位论文及参加的科研项目

## 一、读硕期间发表的学术论文

[1] He Q, Wan, G Y, Jiang M,et al.Research on 6DOF Pose Measurement Algorithm of Industrial Metal WorkpieBinocular Vision[C]//2024 39th Youth Academic Annual Conference of Chinese Association of Automation(YAC), EI 收录.

[2] 邢凯盛,何琴,万国扬,等.用于机器人的立体视觉 6DOF 位姿测量方法[J].中国惯性技术学报(已录用).

# 致谢

时光荏苒，转眼间我的研究生生涯即将画上句号。回首这段求学之路，我深感幸运与感恩。无论是学术上的探索，还是生活中的点滴，都让我深刻体会到坚持与努力的意义。在此，我谨向所有帮助、支持与陪伴我的人致以最诚挚的谢意。

首先，衷心感谢我的导师江明教授。在研究生三年的学习生涯中，您严谨的治学态度、渊博的学识、敏锐的学术思维、精益求精的工作作风以及诲人不倦的师者风范，深深感染并激励着我，成为我毕生学习的楷模。无论是科研上的悉心指导，还是求职过程中的宝贵建议，您都给予了我无私的帮助与支持。师恩如山，学生铭记于心。

感谢万国扬老师。在科研学习过程中，您给予了我极大的帮助与支持。从最初对视觉抓取课题的陌生到逐渐上手操作，您不辞辛劳地为我答疑解惑，耐心指导。每当我因实验数据不理想而沮丧时，您总是及时鼓励我，并提供新的思路与方向，让我重拾信心。感谢您一直以来的悉心指导与包容，这段时间给您添麻烦了这段时间给您添麻烦了。

感谢柏受军老师在我们科研生活的关心，特别是在论文修改的关键阶段，您不辞辛劳地陪伴我们一遍遍打磨稿件。尽管有时我们未能按时完成任务，但您始终以极大的耐心和责任感督促我们，帮助我们不断改进。非常感谢柏老师的幸苦付出，在此在这里由衷的说一声感谢！

其次，感谢张键、黄致远、陈金城师兄，陶秀文师姐，我的同门汪倩倩、束浩林、卓越、居鸿源、黎景涛、成志新，还有师弟程卢超和滕明遥的帮助，让我的研究生生活多姿多彩，是你们的陪伴激励我不断向前！

最后，感谢我的家人们对我的关心和支持，让我在无助时有了依靠，愿我的家人们永远身体健康，平安喜乐。

三年转瞬即逝，给我的人生留下了浓墨重彩的一笔，我会牢记这三年的时光，牢记老师同学们的关心教导，牢记家人们的理解和支持，谢谢大家！

尚德教学 唯实惟新