基于改进 YOLOv11-seg 的高密集场景工业零件单目测量方案

程煜博[1)]

[1)](西安交通大学计算机学院　西安　710049)

摘　要　工业生产中高密集场景下的零件检测与测量是实现自动化和智能化生产的关键环节。然而，由于零件密集度高、尺寸小以及遮挡问题，以往的工业零件目标检测与测量方法在精度和效率上存在局限性。为此，本文提出了一种基于改进 YOLOv11-seg 的单目视觉测量方案，旨在实现高密集场景中的工业零件精确检测与测量。我们通过引入额外的检测头、改进的 NMS 方法以及加入 SEAM 注意力机制，显著提高了 YOLO-seg 模型在小目标、高密度场景下的识别能力与检测精度。通过引入 Ghost 卷积，在尽量不影响精度的情况下有效降低了模型的参数量，便于模型在移动端的部署。同时，我们针对螺栓和垫片两种常见的工业零件，设计了一套基于 OpenCV 的几何处理的方法，实现了像素级的工业零件尺寸测量。我们还以摆放的桌面边长为基准，将像素个数转化为毫米级精度的尺寸信息。我们通过选取不同尺寸的零件、在不同摆放密度和姿态下对方案进行了测试，结果表明我们所提方案能够实现毫米级的测量精度，并能够在移动端处理器上高效运行。

关键词　工业零件测量；YOLO；SEAM 注意力；单目测量；高密集实例分割

Monocular measurement solution for industrial components in high-density scenes based on improved YOLOv11-seg model

CHENG Yu-Bo[1)]

[1)](School of Computer Science, Xi'an Jiaotong University, Xi'an, 710049)

Abstract　Component detection and measurement in high-intensity scenarios in industrial production is a key component in realizing automated and intelligent production. However, due to the high component density, small size, and occlusion problems, previous target detection and measurement methods for industrial components have limitations in terms of accuracy and efficiency. To this end, this paper proposes a monocular vision measurement scheme based on an improved YOLOv11-seg, aiming to achieve accurate detection and measurement of industrial components in highly dense scenes. We significantly improve the recognition capability and detection accuracy of the YOLO-seg model in small-target, high-density scenes by introducing an additional detection head, an improved NMS method, and incorporating the SEAM attention mechanism. By introducing the Ghost convolution, the number of parameters of the model is effectively reduced without affecting the accuracy as much as possible, which facilitates the deployment of the model on the mobile side. Meanwhile, we designed a set of OpenCV geometrical processing based methods for two common industrial components, namely bolts and gaskets, to realize pixel-level dimensional measurement of industrial components. We also converted the number of pixels into dimensional information with millimeter-level accuracy by using the edge length of the placed desktop as a benchmark. We tested our scheme by selecting components of different sizes and under different placement densities and attitudes,

———————————————

程煜博，本科，西安交通大学计算机试验班2101.

and the results show that our proposed scheme is able to achieve millimeter-level measurement accuracy and can run efficiently on mobile processors.
Keywords industrial components measurement；YOLO；SEAM attention；monocular measurement；highly dense instance segmentation

1 绪 论

随着工业生产向着自动化和智能化的方向快速发展，高密集场景下的工业零件检测与测量在生产效率提升和品质控制中扮演着至关重要的角色。这些应用场景通常伴随复杂的视觉特征，包括零件密集摆放、尺寸小、遮挡频繁等问题，给传统的检测与测量方法带来了巨大挑战。现有基于计算机视觉和深度学习的检测算法虽取得了一定成效，但在面对高密度、小尺寸和复杂遮挡的情况下，往往难以在检测精度、计算效率和系统鲁棒性方面兼顾，无法满足工业领域对实时性和精确度的需求。

近年来，目标检测领域的快速发展催生了许多新型方法，其中基于深度学习的 YOLO（You Only Look Once）系列算法由于其高效性和较好的性能得到了广泛应用[1]。然而，标准 YOLO 模型在复杂工业场景下对小目标和高密度目标的检测仍存在不足，难以直接满足工业零件测量任务的需求。因此，改进和优化现有检测算法，使之更加适用于高密集场景，成为一个亟需解决的问题。

本文提出了一种基于改进 YOLOv11-seg 的单目视觉测量方案[2]，旨在针对上述挑战提供更精确和高效的检测与测量方法。为提升模型在高密集场景中的性能，我们引入了额外的检测头和改进的非极大值抑制（NMS）方法，以增强对小目标的检测能力和遮挡处理能力。同时，通过集成 SEAM 注意力机制[3]，补偿遮挡面部的响应损失，从而改善物体遮挡的检测效果，使其在复杂场景下依然能够准确识别目标。为降低模型复杂度，便于在移动设备上部署，我们在模型中引入了 Ghost 卷积[4]，优化参数量的同时尽量保持模型精度。此外，针对工业场景中常见的螺栓和垫片两种零件，我们结合 OpenCV 几何处理方法[5]，实现了对零件的像素级尺寸测量，并以摆放的桌面边长为基准，将像素个数转化为毫米级精度的尺寸信息。

本研究旨在提供一种在高密集场景下表现优异的工业零件检测与测量解决方案，为工业自动化生产提供更高效、更精准的技术支撑。实验结果表明，所提出的改进方案能够在移动端处理器上高效运行，实现毫米级的测量精度，为工业生产的智能化升级提供了有力保障。

2 高密集场景工业零件单目测量方案

图 2-1 展示了本文提出的基于改进 YOLOv11-seg 的高密集场景工业零件单目测量方案。

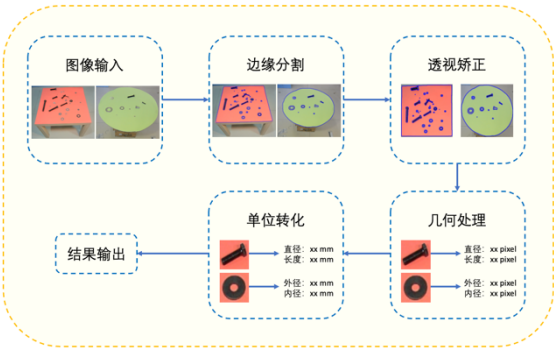

图 2-1 本文提出的工业零件测量方案总流程图

2.1 改进的 YOLOv11-seg 模型架构

YOLO（You Only Look Once）是一种实时目标检测算法，与传统的目标检测方法相比具有显著的优势。YOLO 将目标检测问题转化为单一的回归问题，通过单一神经网络直接从图像中预测目标类别和位置，这种设计不仅显著提高了检测速度，使其能够以实时帧率处理视频数据，还提升了全局感知能力，减少了对局部特征的过度依赖，从而在复杂场景中表现出更好的鲁棒性。此外，YOLO 的端到端架构便于训练和优化，对硬件资源需求相对较低，这使其成为实时场景应用（如移动设备）中的理想选择。相较于 R-CNN 系列等分阶段检测方法[6]，YOLO 在速度与精度之间取得了卓越的平衡，尤其适用于实时性要求高的任务。YOLOv11 是 YOLO 系列的一个版本，采用了改进的主干和颈部架构，通过引入 C3k2 模块替换原有的 C2f 模块实现 CSP 模块的优化[7]，在空间金字塔池化（SPP）模块后引入 C2PSA 自注意力模块增强了特征提取能力[8]，针对效率和速度进行了优化。在 COCO 数据集上达到了高于以往的 mAP 和更少的参数量[9]。

YOLO-seg 是基于 YOLO 目标检测架构的一种轻量级实例分割模型，不仅能快速识别图像中的目

标，还通过增加分割分支实现像素级的目标掩码生成。YOLO-seg 在处理多目标分割任务时具有优势，可以精准区分不同对象的轮廓，完成检测和分割的双重任务。凭借其端到端的训练架构和相对较小的计算资源需求，YOLO-seg 在工业检测等领域展现出广阔的应用前景。其设计在保证检测效率的同时显著提升了分割精度，是在资源受限条件下实现高精度实例分割的理想模型。

然而，工业零件场景由于零件密集度高、尺寸小以及相互遮挡等情况的存在，标准的 YOLO 架构检测精度较低。我们通过引入额外的检测头、改进的 NMS 方法、引入 Ghost 卷积以及加入 SEAM 注意力机制，显著提高了 YOLO-seg 模型在小目标、高密度场景下的识别能力与检测精度，同时保证模型在移动端的运行速度。图 2-2 展示了我们改进的 YOLOv11-seg 模型架构。

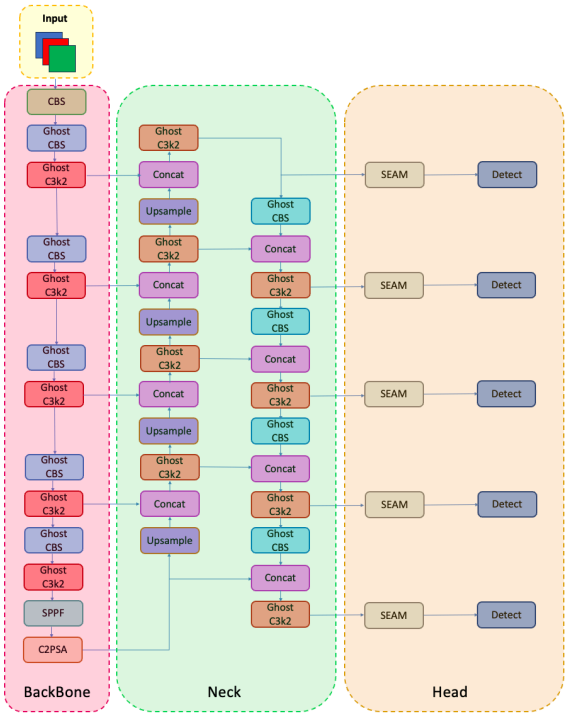

图 2-2 改进的 YOLOv11-seg 模型架构

2.1.1 P2 和 P6 检测头

YOLO 的多尺度检测能力在经典版本中存在一定局限性，尤其是针对小目标的检测效果往往不足。标准的 YOLO 架构通常在特征金字塔（Feature Pyramid）中使用三个检测头（P3、P4 和 P5）[10]，分别处理特征图中的不同尺度信息。然而，由于一些小目标往往存在于更高分辨率的层中，而大目标信息则可能在更低分辨率的层中，因此直接加入 P2 和 P6 检测头可以进一步丰富模型的多尺度特征表达能力。

• P2 层的作用：P2 层特征图分辨率高，适合检测小目标的细节信息。在物体检测场景中，小目标检测经常依赖较高的分辨率，而 P2 层的加入可以更细致地捕获小目标的边缘和细节，从而提高对小物体的检测精度。

• P6 层的作用：P6 层分辨率低，但具备较大的感受野，适合大物体的检测。对于大物体或图像中尺寸较大的物体，P6 层可以在低分辨率特征图上直接定位和分类，减少复杂度的同时增强对大目标的敏感度。

除此之外，我们增加了 Backbone 和 Neck 部分的卷积深度，以加强对于高分辨率输入图像的特征提取能力。

2.1.2 SEAM 注意力模块

不同种类目标之间的遮挡会造成对齐错误、局部混叠和特征缺失。我们通过引入 SEAM 注意力模块，实现多尺度检测、强调图像中的有效零件区域并相反地弱化背景区域。SEAM 注意力模块先经过逐通道分离的卷积学习到不同通道的重要性并减少参数量，再使用两层全连接网络融合各个通道的信息，使得网络可以加强所有通道之间的连接，最后将 SEAM 模块的输出作为 Attention 与原始特征相乘。该模块能够通过学习到的遮挡目标与未遮挡目标之间的关系来弥补遮挡场景下的损失。我们将 SEAM 模块置于 Neck 部分与 Head 部分的连接处。

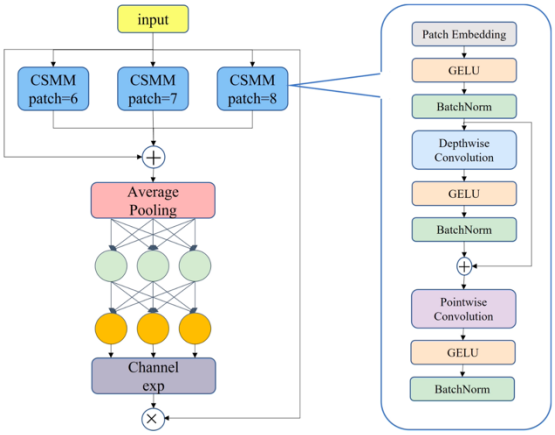

图 2-3 SEAM 注意力模块

2.1.3 递减的 NMS 阈值策略

传统 NMS 使用固定的交并比（IoU）阈值来判断是否保留候选框（YOLOv11 使用 CIoU[11]）。由于小目标通常更容易出现高重叠区域，这种方式在小目标检测中通常表现不佳。同时，当目标存在同种

类之间遮挡、密集分布或尺度变化较大时，传统 NMS 容易导致误检或漏检。我们采用了一种动态调整 NMS 的策略，对于我们依次选择保留的第 i 个候选框，根据以下公式逐渐减少 NMS 阈值：

$$nms_threshold_{i+1} = \eta \cdot nms_threshold_i$$

其中 η 是超参数，用于逐步减小 NMS 的阈值。

我们动态调整的阈值可以更好地适应目标的分布和尺度变化，初始较高的阈值能够保留更多的候选框，防止误删，逐步降低则可以精细地抑制重叠框。

2.1.4 Ghost 卷积

标准卷积需要通过大量的卷积核对输入特征图进行逐点计算，计算复杂度高，并且很多计算会导致生成的特征具有高度冗余性。Ghost 卷积假设标准卷积生成的特征图中很多信息是冗余的，可以通过更简单的操作生成一部分特征。通过标准卷积生成较少数量的基础特征称为 Ghost 基特征，使用简单的线性变换从这些基特征中生成更多的冗余特征。

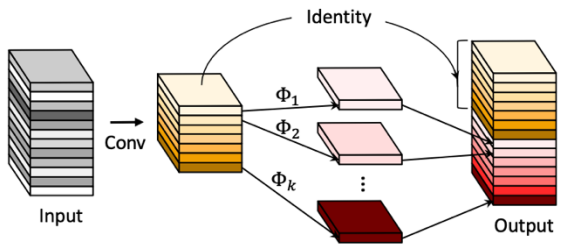

图 2-4 Ghost 卷积

我们采用的 Ghost 卷积先使用标准的卷积生成一半的特征图，再使用 5×5 的逐通道卷积生成另一半的特征图[12]，最终将得到的特征图拼接。我们将原有模型中的卷积模块替换成 Ghost 卷积模块，实验证明引入 Ghost 卷积能够减少 26.7%的参数量，并且模型的检测精度几乎不变。

2.2 几何处理

2.2.1 单应矩阵透视矫正

在工业测量任务中，由于相机拍摄角度或光学畸变，导致图像中平面对象的形状发生变形，如矩形桌面被拉伸为梯形，同时零件的尺寸和形状也会发生变化。为了恢复真实的几何关系，通常需要进行透视矫正。我们基于单应矩阵（Homography Matrix）实现对于桌面的透视矫正[13]。

单应矩阵是一种描述两个平面间射影变换关系的矩阵。在计算机视觉中，单应矩阵可通过点对之间的关系进行估计，用于图像对齐与变换。假设在图像平面上，原始点为$p = [x, y, 1]^T$，变换后对应点为$p' = [x', y', 1]^T$，它们之间满足以下关系：

$$p' = Hp$$

其中，H 是 3×3 的单应矩阵，可表示为：

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

变换后坐标通过齐次归一化得到：

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}$$

$$y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}$$

我们基于对于桌面的分割，使用 Douglas-Peucker 算法进行多边形拟合，从而获取到四个角点。设原始图像中桌面的四个角点坐标为：

$$P_i = [x_i, y_i, 1]^T, i = 1,2,3,4$$

目标图像中的理想矩形坐标为：

$$Q_i = [u_i, v_i, 1]^T, i = 1,2,3,4$$

通过约束$Q_i = HP_i$，可得到一组线性方程组，转化为矩阵形式：

$$Ah = 0$$

其中 h 是单应矩阵的向量形式。通过 SVD 求解最小二乘解即可得到 H。

利用得到的单应矩阵 H，对原始图像进行变换，采用逆映射计算目标图像像素值。目标图像坐标(u, v) 对应原图中的点(x, y):

$$p = H^{-1}p'$$

从而得到矫正后的图像。

2.2.2 最小二乘拟合

在工业实践中，我们通常需要螺栓的长度和直径信息。为了获得用像素表示的螺栓尺寸信息，我们将分割的螺栓边缘像素基于最小二乘法进行直线拟合。最小二乘法（Least Squares Method）是一种通过最小化数据点与模型之间的误差平方和，来获得模型参数的优化方法。在直线拟合问题中，我们假设一组二维数据点$(x_i, y_i)(i = 1,2\ 3, \dots, n)$，目标是找到一条直线$y = ax + b$，使得该直线与数据点的拟合误差平方和最小。

我们定义目标函数为误差平方和：

$$S(a, b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

为了找到使误差平方和 $S(a, b)$最小的参数a和b，对a和b分别求偏导数，并令其等于零：

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{n} x_i(y_i - (ax_i + b))^2 = 0$$

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{n} (y_i - (ax_i + b))^2 = 0$$

解出a和b即可得到直线的最优拟合参数。

我们将螺栓的边缘点向拟合的直线垂直投影，取投影点之间的最大距离即为螺栓的长度，取边缘点与直线垂直距离的 30%中位数的平均值为螺栓的直径。以此获得以像素值为单位的螺栓尺寸测量。

2.2.3 霍夫圆拟合

在工业实践中，我们通常需要垫片的内径和外径信息。垫片可以视为由两个相同圆心、不同直径的圆形构成。我们通过霍夫圆拟合的方法分别对两个圆形进行拟合，拟合得到的两个圆的直径分别作为垫片的内径和外径。

霍夫变换（Hough Transform）是一种用于检测几何形状的方法[14]，可用于拟合二维平面上的圆。其原理是将数据点从图像空间映射到参数空间，通过参数空间中的峰值确定最佳拟合圆的参数。我们将将数据点(x_i, y_i) 映射到三维参数空间(a, b, r)。对于半径r，点(x_i, y_i)的可能圆心(a, b)满足：

$$a = x_i - r\cos\theta$$
$$b = y_i - r\sin\theta$$

我们将参数空间(a, b, r)离散化为网格点，对于每个数据点，计算可能的(a, b, r)，并在累加器中为对应位置累加票数。最后累加器中票数最高的位置对应最优参数(a^*, b^*, r^*)，从而获得最佳拟合圆的方程：

$$(x - a^*)^2 + (y - b^*)^2 = (r^*)^2$$

以此，我们获得了以像素值为单位的垫片内径和外径测量。

2.3 像素个数转化为长度单位

在我们的数据集中，零件分别摆放在固定大小的正方形桌面和圆形桌面上。我们以桌面的尺寸为基准，计算出每个像素值在现实平面上的长度，从而计算出以长度单位表示的零件尺寸。

3 实验结果与分析

3.1 数据集

我们将黑色、银色两种颜色，尺寸不同的螺栓和垫片以不同的姿态和密度分别放置在边长为 0.55m 的正方形桌面和直径为 0.6m 的圆形桌面上，桌面为任意颜色，并且零件之间存在着不同程度的遮挡。除此之外，我们还放置了不同种类的除螺栓垫片之外的干扰物，来模拟工业场景零件中可能混杂其他杂物的情况。表 3-1 展示了数据集中不同零件的数量统计情况。

表 3-1 工业零件测量数据集统计数

数据集	图像数量	螺栓数量	垫片数量
训练集	792	4475	6264
验证集	88	530	705
测试集	88	514	742

3.2 P2 和 P6 检测头

为了验证我们对 YOLOv11-seg 模型加入 P2 和 P6 检测头的有效性，我们对其进行了消融实验。我们选取标准的 YOLO 模型、标准 YOLO 模型+P2 检测头、标准 YOLO 模型+P6 检测头以及标准 YOLO 模型+P2 和 P6 两种检测头共四种模型进行实验。由于垫片内外圈是同心圆的特点，我们将内外圈的分割结果分别进行标注并统计。我们分别选取 mAP50 和 mAP50-95 作为评测指标。表 3-2 和表 3-3 展示了实验的结果：

表 3-2 P2 和 P6 检测头的消融实验(mAP50)

	垫片外圈	垫片内圈	螺栓
YOLO	0.79	0.61	0.962
YOLO+P2	0.876	0.748	**0.972**
YOLO+P6	0.784	0.623	0.96
YOLO+P2+P6	**0.895**	**0.76**	0.97

表 3-3 P2 和 P6 检测头的消融实验(mAP50-95)

	垫片外圈	垫片内圈	螺栓
YOLO	0.415	0.283	0.689
YOLO+P2	0.499	0.381	0.723
YOLO+P6	0.405	0.299	0.686
YOLO+P2+P6	**0.501**	**0.395**	**0.728**

从实验结果来看，增加 P2 检测头能够明显增加模型对小目标的检测能力，单独增加 P6 检测头的效果不显著，但是同时增加 P2 和 P6 检测头的模型在实验中取得了最好的结果。

3.3 SEAM 注意力模块

为了验证 SEAM 注意力模块的有效性，我们以 YOLO+P2+P6 模型为基础，对加入 SEAM 注意力模块前后的模型检测效果进行了实验。表 3-4 和表 3-5 展示了实验的结果：

表 3-4 加入 SEAM 注意力模块前后的实验结果(mAP50)

	垫片外圈	垫片内圈	螺栓
无 SEAM 模块	0.895	0.76	0.97
加入 SEAM 模块	**0.904**	**0.762**	**0.971**

表 3-5 加入 SEAM 注意力模块前后的实验结果(mAP50-95)

	垫片外圈	垫片内圈	螺栓
无 SEAM 模块	0.501	0.395	0.728
加入 SEAM 模块	**0.534**	**0.396**	**0.749**

从实验结果可以看出，SEAM 注意力模块对于不同种类目标的检测效果均有提升。其中，对于垫片内圈的提升不明显，考虑到是因为垫片内圈目标较小且较少有遮挡；对于垫片外圈和螺栓的提升更为明显，尤其是对于 mAP50-95 这一指标有较大的提升。

3.4 Ghost 卷积

为了验证使用 Ghost 卷积模块替换标准卷积模块能够有效减少模型的参数量，我们统计了使用标准卷积模块和 Ghost 卷积模块的模型的参数量并进行对比。此外，为了验证 Ghost 模块能够有效提高模型在移动端部署的推理速度，我们在 ONNX 运行时环境上部署模型[15]，并使用 CPU 进行推理。图 3-1 对比展示了使用标准卷积模块和 Ghost 卷积模块的模型的参数量和移动端推理的 fps：

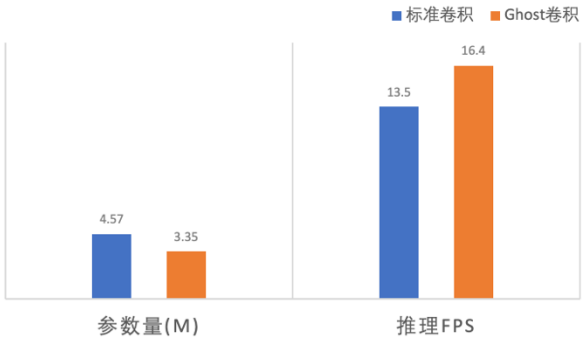

图 3-1 模型参数量和移动端推理 fps 对比

从实验结果可见，相比于标准卷积，Ghost 卷积能够减少 26.7%的参数量，在 ONNX 运行时环境下部署时增加 21.5%的推理 fps。这说明将原有标准卷积模块替换为 Ghost 卷积模块能够有效实现模型的轻量化，便于模型在移动端设备进行部署和高效推理。

与此同时，为了验证 Ghost 卷积模块对模型精度的影响，我们对采用标准卷积的模型和采用 Ghost 卷积的模型进行了检测精度的实验，实验结果展示在表 3-6 和表 3-7:

表 3-6 标准卷积和 Ghost 卷积模型检测精度实验结果(mAP50)

	垫片外圈	垫片内圈	螺栓
标准卷积	0.904	0.762	0.971
Ghost 卷积	0.907	0.759	0.965

表 3-7 标准卷积和 Ghost 卷积模型检测精度实验结果(mAP50-95)

	垫片外圈	垫片内圈	螺栓
标准卷积	0.534	0.396	0.749
Ghost 卷积	0.536	0.388	0.745

从实验结果来看，使用 Ghost 卷积替代原有的标准卷积并不会造成显著的精度损失。由于 Ghost 卷积减少了特征图的数量，实验的某些指标上有轻微的浮动，但是相比于参数量的减少和推理时 fps 的增加，精度上的变化并不显著。这说明引入 Ghost 卷积对于模型性能的提升是有意义的。

3.5 递减的 NMS 阈值策略

为了验证使用递减的 NMS 阈值策略能够减少密集场景下漏检和误检的可能性，我们在模型推理时分别采用了普通的 NMS 策略和递减的 NMS 阈值策略，并对这两种策略下检测 Box 的 Acc 指标进行比较，结果如表 3-8 所示：

表 3-8 普通 NMS 策略和递减的 NMS 阈值策略模型检测精度实验结果(Acc)

	垫片外圈	垫片内圈	螺栓
普通 NMS 策略	0.981	0.928	0.964
递减的 NMS 阈值策略	**0.99**	**0.935**	**0.982**

从实验结果可以看出，使用递减的 NMS 阈值策略的模型在检测 Box 的 Acc 指标上有更好的表现，这说明在高密集场景的工业零件分割任务上递减的 NMS 阈值策略能够有效减少漏检和误检的可能性，更加精细地筛选重叠框。

3.6 尺寸测量结果

我们分别选取了 20 种不同尺寸的螺栓和 20 种不同尺寸的垫片，通过人工测量的方式获得其真实的尺寸信息，并通过本文提出的方案对其进行尺寸测量。我们使用本文提出的方案分别测量螺栓的长度、直径和垫片的外径、内径，与人工测量得到的真实值相比计算相对误差，并对不同相对误差的区间进行计数，绘制出频数直方图如图 3-2 所示:

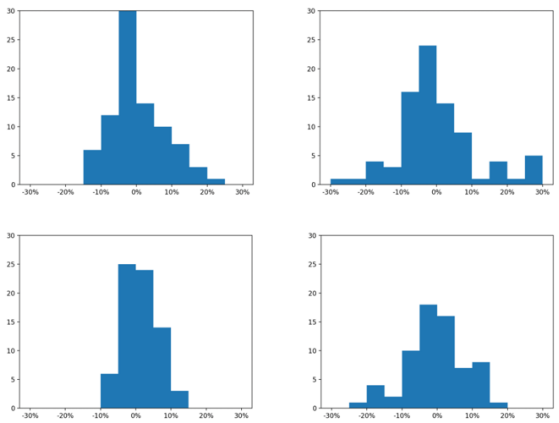

图 3-2 测量相对误差频数分布直方图

其中，左上图是螺栓长度测量的相对误差频数分布直方图，右上图是螺栓直径测量的相对误差频数分布直方图，左下图是垫片外径测量的相对误差频数分布直方图，右下图是垫片内径测量的相对误差频数分布直方图。

此外，我们统计了本文提出的方案测量结果的平均相对误差。螺栓长度测量的平均相对误差为 6.43%，螺栓直径测量的平均相对误差为 8.64%，垫片外径测量的平均相对误差为 4.88%，垫片内径测量的平均相对误差为 12.60%。

从实验结果来看，本文提出的方案的测量相对误差大致呈正态分布，对于大部分的测量指标有较高的准确性，对于微小目标（如垫片内圈）能够实现高准确度的检测和一定精度的分割和测量。

4 结论与展望

本文提出了一种基于改进 YOLOv11-seg 的单目视觉测量方案，旨在实现高密集场景中的工业零件精确检测与测量。实验结果表明我们所提方案能够具有较高的检测准确率、实现毫米级的测量精度，并能够在移动端处理器上高效运行。未来我们将继续跟随 YOLO 版本架构的迭代对我们的模型进行升级优化，并基于我们已有的方案加入对更多种类的工业零件进行尺寸测量的能力。

参考文献

[1] Redmon, Joseph et al. “You Only Look Once: Unified, Real-Time Object Detection.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015): 779-788.

[2] Ultralytics. ultralytics/ultralytics. Github, 2024, https://github.com/ultralytics/ultralytics

[3] Yu, Ziping et al. “YOLO-FaceV2: A Scale and Occlusion Aware Face Detector.” Pattern Recognit. 155 (2022): 110714.

[4] Han, Kai et al. “GhostNet: More Features From Cheap Operations.” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019): 1577-1586.

[5] OpenCV. "Open Source Computer Vision Library." OpenCV, https://opencv.org/. Accessed 26 Nov. 2024.

[6] Girshick, Ross B. et al. “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.” 2014 IEEE Conference on Computer Vision and Pattern Recognition (2013): 580-587.

[7] Wang, Chien-Yao et al. “CSPNet: A New Backbone that can Enhance Learning Capability of CNN.” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2019): 1571-1580.

[8] He, Kaiming et al. “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition.” IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (2014): 1904-1916.

[9] Lin, Tsung-Yi et al. “Microsoft COCO: Common Objects in Context.” European Conference on Computer Vision (2014).

[10] Lin, Tsung-Yi et al. “Feature Pyramid Networks for Object Detection.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 936-944.

[11] Zheng, Zhaohui et al. “Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression.” ArXiv abs/1911.08287 (2019): n. pag.

[12] Chollet, François. “Xception: Deep Learning with Depthwise Separable Convolutions.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 1800-1807.

[13] Hartley, Richard, and Andrew Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.

[14] Paul V. C. Hough, Method and Means for Recognizing Complex Patterns, U.S. Patent 3,069,654, 1962.

[15] "ONNX: Open Neural Network Exchange." ONNX, Open Neural Network Exchange, https://onnx.ai/.