

基于改进YOLOv8的零件图纸裁剪

唐晓龙, 覃文俊, 杨建逾

上海理工大学机械工程学院, 上海

收稿日期: 2025年5月25日; 录用日期: 2025年6月18日; 发布日期: 2025年6月25日

摘要

零件图纸不仅代表着零件的制造要求, 同时也是原创性设计的体现, 图纸比对可以很好地保护其原创性。图纸裁剪是图纸比对技术的先行军, 通过裁剪可以高效率地进行信息提取和比对。为了解决传统分割方法在零件图纸分割领域中的不足, 本文提出了一种改进的YOLOv8的零件图纸裁剪方法, 引入了CBAM注意力机制, 能够高效地将图纸图像裁剪为标题栏区域、技术要求区域、其余粗糙度区域和图元区域, 并利用轮廓检测方法将图元区域进一步裁剪为若干视图。实验结果表明本文的改进模型比传统的区域生长方法裁剪速度更快, 比原始的模型的mAP和召回率分别上升了5.5%和7.6%。

关键词

机械零件图纸, 区域生长, YOLOv8, 轮廓检测, 图像分割

Part Drawing Cropping Based on an Improved YOLOv8

Xiaolong Tang, Wenjun Qin, Jianyu Yang

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: May 25th, 2025; accepted: Jun. 18th, 2025; published: Jun. 25th, 2025

Abstract

Part drawings not only represent the manufacturing requirements of parts, but also the embodiment of original design, and drawing comparison can well protect the originality. Drawing cutting is the pioneer of drawing comparison technology, and information extraction and comparison can be carried out efficiently through cutting. In order to solve the shortcomings of the traditional segmentation method in the field of part drawing segmentation, this paper proposes an improved YOLOv8 part drawing cropping method, which introduces the CBAM attention mechanism, which can efficiently crop the drawing image into the title block area, the technical requirement area, the rest roughness

area and the element area, and further crop the element area into several views by using the contour detection method. Experimental results show that the improved model in this paper has a faster cropping speed than the traditional regiometric growth method, and the mAP and recall rates are increased by 5.5% and 7.6%, respectively, compared with the original model.

Keywords

Mechanical Part Drawings, Region Growing, YOLOv8, Contour Detection, Image Segmentation

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着制造业的迅速发展,产品设计和生产所需的图纸数量激增。这些图纸承载了技术信息,代表企业的核心竞争力和创新成果。为了保护专利和知识产权,图纸比对技术应运而生。图纸裁剪作为图纸比对技术中的关键环节,对提高比对精度和效率至关重要。通过将复杂图纸有效裁剪为多个独立区域,后续比对过程变得更加清晰和系统化。图纸裁剪实际上是图像分割的一种特定应用,可以提供精简且可靠的图像特征信息,进而有效地提高后续视觉任务的处理效率,以便进行进一步分析和比对[1][2]。在现代制造业中,机械零件的设计和制造流程越来越依赖于计算机辅助设计,图纸分割是图纸以电子形式存储与传输和智能识别的必由之路[3]。

机械零件图纸包含了丰富的几何信息和制造细节。传统的图纸分割方法常常依赖于手动标注或基于简单阈值的处理,这在处理大量图纸的分割任务时显得力不从心。尽管近期基于图像的语义分割方法在分割自然图像方面显示出巨大潜力且被广泛应用于医疗诊断、自动驾驶和智能识别等领域,但将这些方法应用于机械图纸的准确分割仍然是一个重大挑战[4][5]。因此,开发一种鲁棒的图纸分割方法显得尤为重要。

卷积神经网络、循环网络、对抗网络等各种类型的深度神经网络正在有效地解决许多具有挑战性的计算机视觉任务[6]。Buric [7]将深度卷积网络作为特征提取器,使用 U-Net 模型自动定位和识别狗的眼病,有效地分割和诊断了这些疾病。Yong [8]利用卷积神经网络,进行二进制图像分割,为全面的海滩测绘工作提供新的方法。Juhong A [9]利用 U-Net 网络模型,引入了多头注意力机制,对 MPI-CT 图像进行分割,显示出极好的结果。郭宏志[10]采用改进的蜜蜂觅食算法,引入自适应领域收缩策略,在多阈值图像分割应用中体现出了更好的效果。

上述文献表明了把优化算法应用于图像分割领域,尤其是医学影像方面具有巨大的优势,尽管神经网络广泛应用于医学领域的图像分割,并且效果显著。马宇超[11]搭建了基于 Compact 主干网络的目标检测网络用来检测并分割机械零件图纸中的标注部分,并利用卷积神经网络完成标题栏、图元部分的分割,速度快,但分割精度不高,且分割的图纸较为简单。由于目前使用神经网络裁剪机械零件图纸的国内外研究较少,所以本文提出了一种改进的 YOLOv8 的零件图纸裁剪方法,旨在实现对机械零件图纸的准确分割。

2. 图纸图像区域划分

机械零件图纸一般包括标题栏区域、技术要求区域、粗糙度区域和图元区域,其中图元区域一般又

包含若干个视图区域，如图 1 所示。图纸裁剪是指通过裁剪的方法，提取感兴趣区域，将标题栏区域、技术要求区域裁剪出来，剩下的部分便是图元区域，流程如图 2 所示。可以细化比对任务，优化比对流程，提高比对效率。本文对比了两种裁剪方法，为图纸的信息提取提供了一个新的方法。

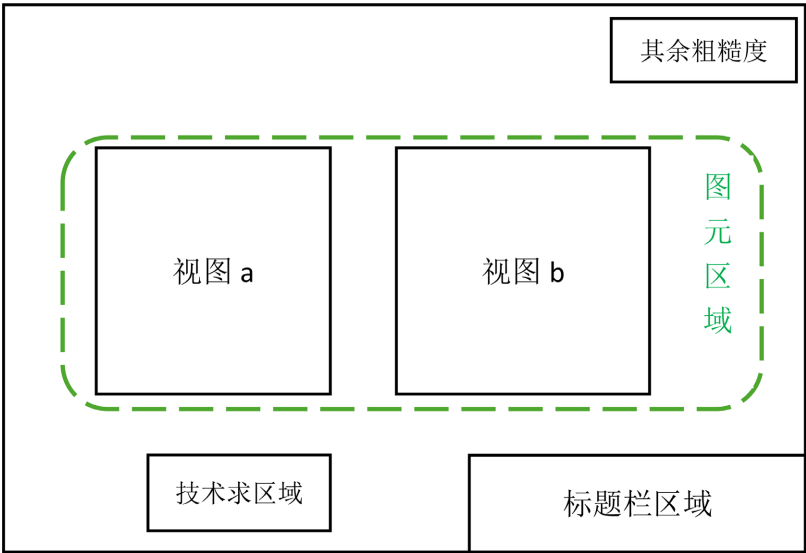


Figure 1. Division of mechanical parts drawing areas
图 1. 机械零件图纸区域划分

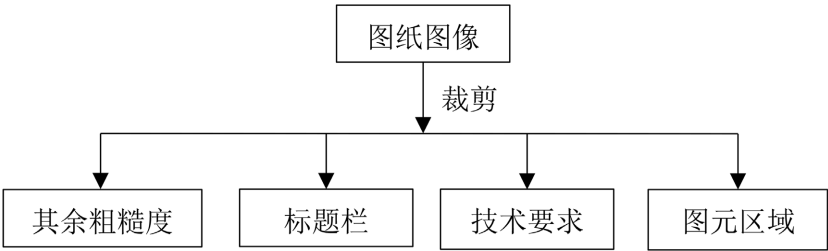


Figure 2. Cutting process
图 2. 裁剪流程

3. 基于区域生长的裁剪方法

基于区域生长的图纸裁剪方法是最直观的裁剪方法。它的基本思想是从一个或多个初始种子点开始，根据预先设定的相似性准则，将与种子相邻且满足条件的像素逐步加入同一个区域，直到没有新的像素能够加入为止，具体的流程如图 3 所示。

首先需要在图像中选择一个或多个代表性像素作为种子点，也可以选择某些区域作为种子点。种子点就是在图像中进行区域生长的初始点，可以根据图像的构图特征来自行设置，也可以借助自动化算法来确定。然后需要定义一个或多个相似性阈值(如像素灰度、颜色、纹理特征等)，也就是生长规则，来衡量相邻像素与种子点的相似性。当相邻像素与种子点之间的差异小于该阈值时，认为它们属于同一区域，就将其加入到当前区域中，并把该像素作为新的生长点继续扩展区域。当所有与当前区域相邻的像素均不满足相似性条件时，区域生长过程终止，并生成该区域的掩模，此时该区域已经完整分割出来，对应的图纸区域也会被裁剪出来。

种子点的选取如图 4(a)所示，将红色和绿色矩形框的四条边线作为种子点向四周生长，生长效果如

图 4(b)所示, 会将感兴趣区域完全包裹并返回矩形框的坐标信息, 然后根据这些坐标信息生成对应区域的掩模实现裁剪效果。

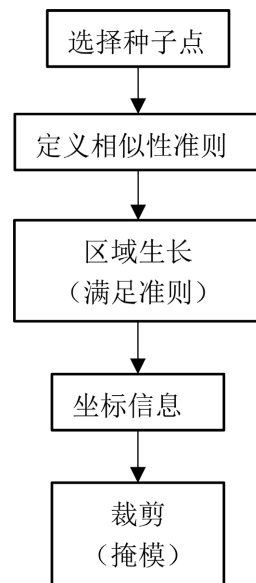


Figure 3. Regional growth process

图 3. 区域生长流程

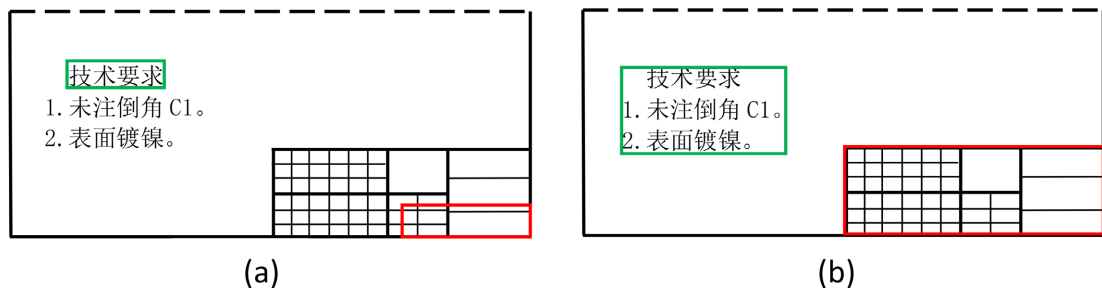


Figure 4. Regional growth effect

图 4. 区域生长效果

传统的区域生长方法能够有效裁剪出具有连通性和内部相似性较强的区域, 对于图像的不同区域的特征表现较为敏感, 且算法直观、实现简单, 易于理解。但是该方法对于种子点的选取要求较高, 而且往往需要根据不同的裁剪区域定制不同的生长准则, 值得注意的是, 在裁剪“技术要求区域”时, 往往使用文字识别方法进行种子的定位, 该方法效率过低。并且该方法对于没有固定特征的区域, 例如“其余粗糙度区域”, 不能自动选取合适的种子点, 导致裁剪失败。

4. 基于改进 YOLOv8 的裁剪方法

YOLO 目标检测算法是一种高效的深度学习目标检测模型, 由 Joseph Redmon 等人于 2016 年提出。不同于传统的两阶段目标检测方法(如 R-CNN 系列), YOLO 算法仅使用一次前向传播, 就能同时预测图像中目标的类别和位置。这种独特的设计使其能够以极高的速度实时处理图像或视频数据, 并显著提高了检测效率。YOLOv8 是由 Ultralytics 于 2023 年推出的一种先进的实时目标检测算法, 它延续了 YOLO 系列高效、轻量化的特点, 并在准确性、易用性和灵活性方面取得了进一步的提升。

(1) 数据采集与标注

初始收集了 100 张机械零件图纸, 在经过数据增强之后, 共有 1596 张图像。由于使用的是 YOLO 系列的目标检测模型, 所以在使用 labeling 工具对图像进行标注之前, 需要制作类别标签的 txt 文件, 将类别标签提前录入, 方便后续的标注, 每行一个标签, 一共三个类别标签, 对应需要检测的区域种类数量。然后使用 labeling 对每张图像进行标注, 分配类别标签, 如图 5 所示。在标注界面, 每个矩形框将表示不同的区域, 并分别被赋予了不同的类别标签, 标签“0”表示的是标题栏区域, 标签“1”表示的是技术要求区域, 标签“2”为其余粗糙度区域。

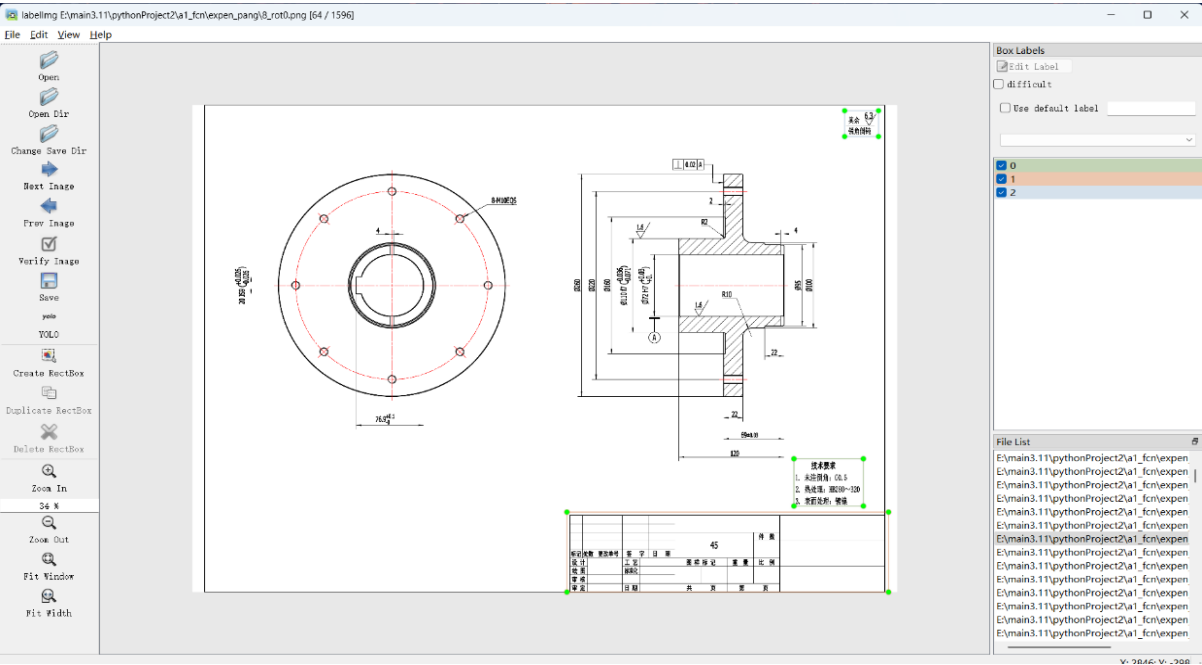


Figure 5. Data annotation diagram
图 5. 数据标注示意图

在全部的图像都标注完成之后, labeling 工具会自动生成与图像对应的 txt 文件, 其包含类别标签以及在标签区域在图像中的位置。其中, 每一行的第一个值是类别标签, 第二个和第三个值分别为矩形框中心的 x 和 y 坐标, 第四个和第五个值分别为矩形框的宽度和高度。所有坐标和尺寸都是归一化值, 即相对于图像宽度和高度的比例, 范围在 0 到 1 之间。

(2) CBAM 注意力机制

卷积块注意力模块(Convolutional Block Attention Module, CBAM)是一种用于卷积神经网络的注意力机制, 旨在通过关注图像中的重要特征和区域来提升模型的性能。CBAM 的结构主要包含两个核心部分: 通道注意力和空间注意力, 如图 6 所示。

对于通道注意力, 它的输入是一个三维特征图, 记为 F_i , 其形状为 $H \times W \times C$ (高度 $\times$ 宽度 $\times$ 通道数), 然后对 F_i 进行两种池化操作: 全局最大池化和全局平均池化。全局最大池化能够提取每个通道的最大值, 生成一个 $1 \times 1 \times C$ 的向量, 捕捉通道中最显著的信息。全局平均池化计算每个通道的平均值, 生成另一个 $1 \times 1 \times C$ 的向量, 反映通道的整体统计信息。这两个向量分别描述了特征图在通道维度上的不同特性。然后将这两个向量输入一个共享的全连接。随后, 将结果相加并通过 sigmoid 激活函数归一化, 生成通道注意力权重, 记为 M_c , 形状为 $1 \times 1 \times C$ 。 M_c 的每个值在 0 到 1 的范围内, 表示对应通道的重

要性。将通道注意力权重 M_c 与输入特征图 F_1 进行逐元素相乘，调整特征图中每个通道的权重。输出结果是一个加权后的特征图，增强了重要通道的比重，抑制了次要通道的影响。

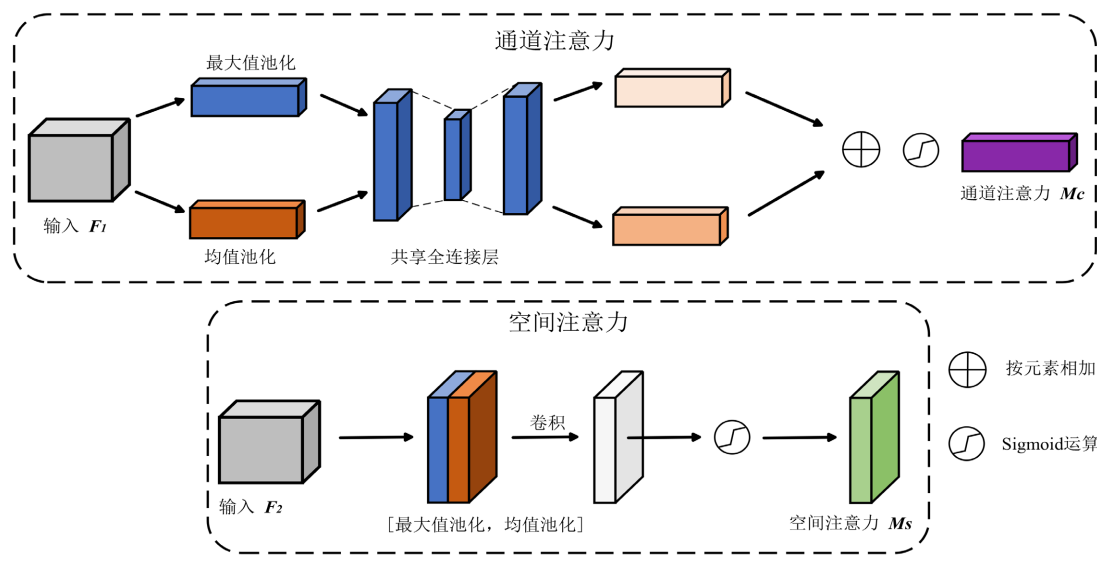


Figure 6. Schematic diagram of CBAM attention mechanism
图 6. CBAM 注意力机制示意图

对于空间注意力，输入是经过通道注意力调整后的特征图，仍然是 $H \times W \times C$ 的形状，记为 F_2 。对输入特征图在通道维度进行通道平均池化和通道最大池化操作。通道平均池化能够生成一个 $H \times W \times 1$ 的二维图，表示每个位置的平均特征强度。通道最大池化会生成另一个 $H \times W \times 1$ 的二维图，表示每个位置的最大特征强度。这两个二维图被拼接，形成一个 $H \times W \times 2$ 的特征图。然后对拼接后的特征图应用一个卷积操作，生成一个 $H \times W \times 1$ 的空间注意力图，记为 M_s 。之后，通过 Sigmoid 激活函数将 M_s 的值归一化到 $[0, 1]$ ，代表每个空间位置的重要性。将空间注意力权重 M_s 与输入特征图进行逐元素相乘，增强关键区域的特征，抑制非重要区域的影响。输出结果是一个二次加权后的特征图。

CBAM 通过关注重要特征和区域，显著提高 CNN 在图像分类、目标检测等任务中的准确性，并且计算开销和参数量小(仅涉及池化、全连接层和少量卷积)，易于集成到现有网络中。而且还具有一定的自适应性，注意力权重可以通过训练自动学习，适应不同任务和数据集。

(3) 模型训练

总共 1596 张图像，其中 1120 张是训练集，320 张为验证集，156 张为测试集。使用 Ultralytics YOLOv8 库来训练目标检测模型，然后创建了一个 YOLO 模型实例，如图 7 所示，并加载了预训练权重文件“yolov8n.pt”，之后调用模型的 train 方法，在指定的数据集上训练目标检测模型。模型会训练 100 次，并将所有输入图像调整为 512×512 像素。在训练完成之后会生成一个模型权重文件，可以直接用于推理、微调或部署。

```
from ultralytics import YOLO
model = YOLO("yolov8n.pt")
model.train(data="data.yaml", epochs=100, imgsz=512)
```

Figure 7. Model training parameter settings
图 7. 模型训练参数设置

(4) 结果与分析

在引入 CBAM 注意力机制后, 平均置信度上升了 9.6%, 模型大小只增加了 1.2%, 具体对比信息如表 1 所示。模型在测试集上的表现如图 8 所示, 平均单张图像的检测时间不超过 0.4 s, 成功检测到了三个目标区域。其中, “0” 表示的是标题栏区域, “1.00” 表示模型对该区域是标题栏区域的置信程度为 100%, 对正确区域预测的置信度越高代表该模型的目标检测性能越好。

Table 1. Comparison of the two models
表 1. 两种模型比较

预测区域	YOLOv8	YOLOv8 + CBAM	比较
mAP@0.5	0.927	0.955	上升 2.8%
mAP@0.5:0.95	0.806	0.861	上升 5.5%
Recall	0.842	0.918	上升 7.6%
F1-score	0.825	0.887	上升 6.2%

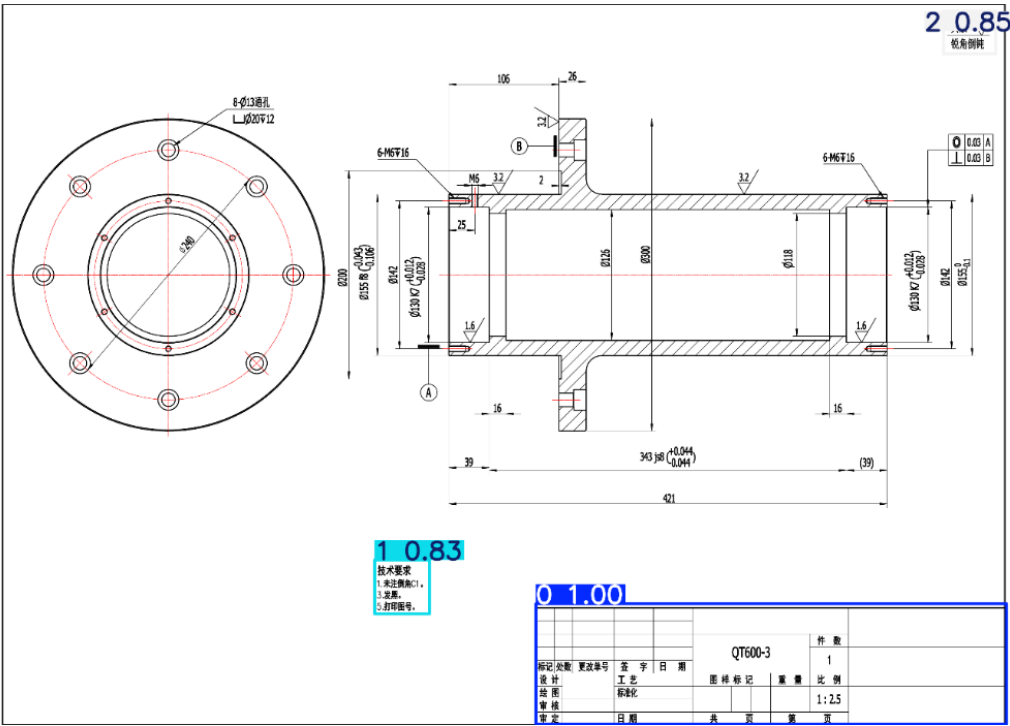


Figure 8. Test set detection effect
图 8. 测试集检测效果

5. 基于轮廓检测的图元裁剪方法

本文先使用 OpenCV 库的 cv2.findContours 方法对图像进行轮廓检测, 并将图像的每个黑色区域的最外层轮廓标记为绿色, 效果如图 9 所示。

经过局部放大, 如图 10 所示, 发现本应该被视为一个整体的小部分轮廓被单独标记了出来。

鉴于这些小面积轮廓主要是视图中的标注信息, 与标注线部分接触或者不完全接触, 所以不能直接通过检测出来的轮廓去分割图元区域, 否则会使得裁剪出来的视图信息不完整。

所以在轮廓检测之后,本文考虑了轮廓合并的方法,引入最近大、小面积块绑定思想,流程如图 11 所示。它的具体逻辑如下:

- 1) 合并小面积轮廓: 将距离相近的(距离不超过 10 个像素)小面积轮廓视为一个整体小面积轮廓;
- 2) 绑定大面积区域: 以每个小轮廓区域的中心点为起点, 分别寻找距离最近的大面积轮廓区域。为了便于观察, 绘制一条线将小面积轮廓和与之匹配的大面积轮廓连接起来; 将大面积区域填充为红色, 如图 12 所示, 并绘制蓝色线连接两种轮廓, 如图 13 所示。

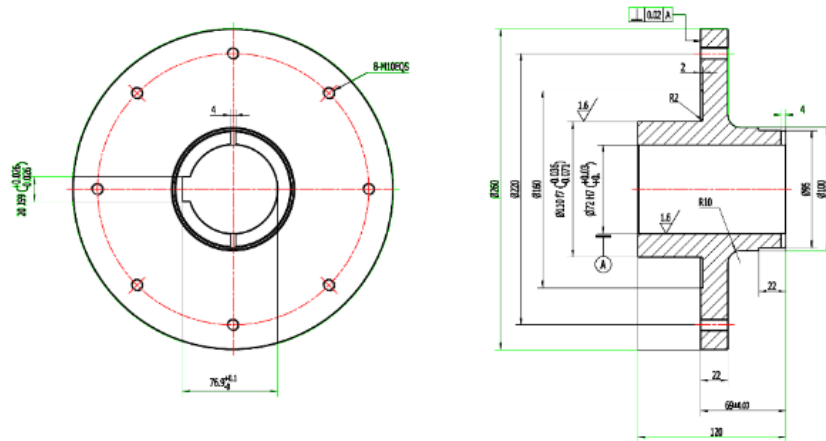


Figure 9. Green outline illustration

图 9. 绿色轮廓示意



Figure 10. Small area contour

图 10. 小面积轮廓

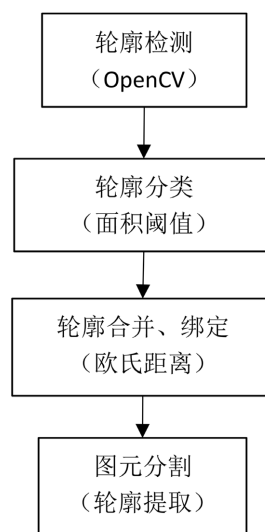


Figure 11. Graph segmentation process

图 11. 图元分割流程

对逻辑的解释如下：

- 1) 大、小面积轮廓的分类：计算每个轮廓的面积，并记录面积最大轮廓与每个轮廓面积的比值，将比值大于 20 的面积轮廓认定为小面积轮廓，其余的为大面积轮廓。
- 2) 轮廓合并：在图 13 中，一个标注被检测成了若干个距离相近的轮廓。优先处理距离最近的小轮廓，并不断重复合并步骤，直至合并后的轮廓周围不存在距离小于 10 像素的小面积轮廓。
- 3) 轮廓绑定：将距离最近的大、小面积区域绑定，合并轮廓。计算合并之后的小面积轮廓的中心点，以该点为圆心逐像素地在周围寻找距离最近的红色像素点(大面积区域)，并将小面积区域与红色像素点所属的大面积区域形成绑定关系。

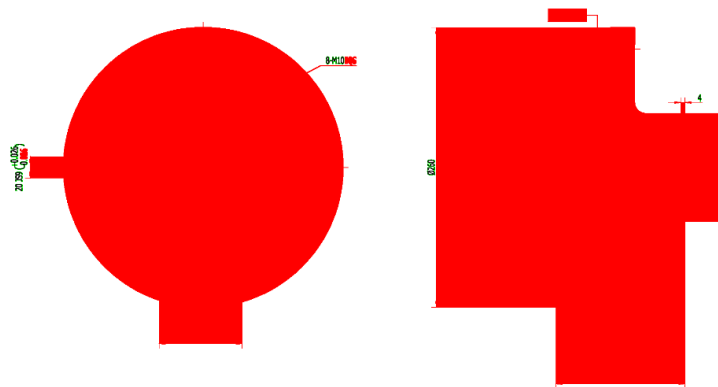


Figure 12. Large and small area contours
图 12. 大、小面积轮廓

连接图局部放大之后，可以清楚地看到蓝色线将大、小面积轮廓连接起来。

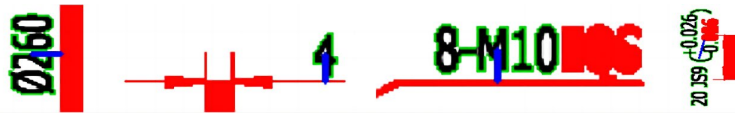


Figure 13. Local amplification binding relationship
图 13. 局部放大绑定关系

这样便可以将小面积区域精准地分配到本应归属的视图中，然后将其完全填充为红色，将蓝色连线也修改为红色，如图 14 所示。

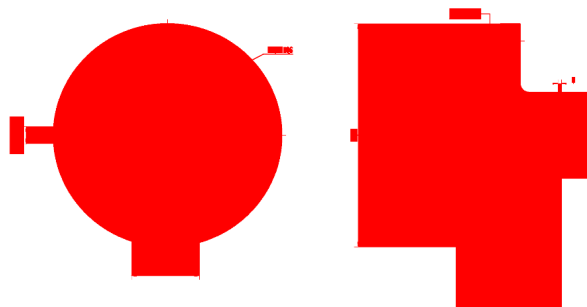


Figure 14. Binding results
图 14. 绑定结果

然后再通过 `cv2.findContours` 方法提取红色区域的边缘, 以该边缘作为分割线, 可将图元区域分割为两个视图。至此, 图纸所有的区域都已经裁剪和分割完成, 效果如图 15 所示。

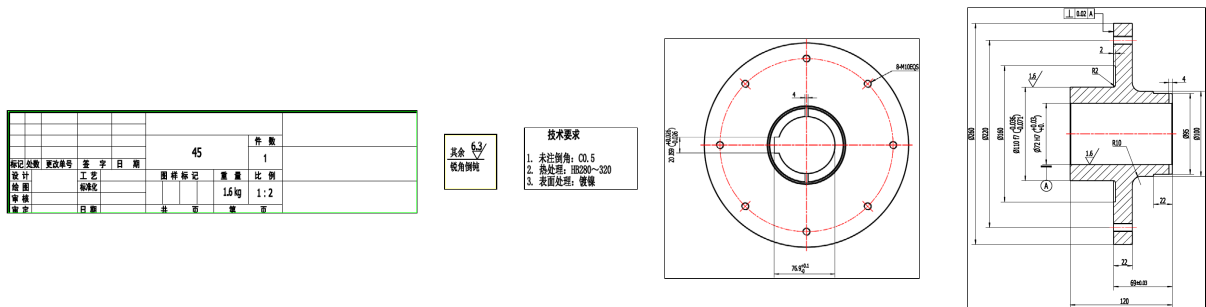


Figure 15. Segmentation completed
图 15. 分割完成

基于 OpenCV 的轮廓检测方法是最常用的轮廓检测方法之一, 尤其是在黑色为前景, 白色为背景的机械零件图纸中, 效果更加突出, 不仅检测精度可达到像素级, 而且在实际应用场景中可操作性强, 可进行面积的计算, 轮廓的连接等操作, 使得裁剪效率得到了提升。

6. 结论

1) 本文使用的改进 YOLOv8 目标检测模型对机械零件图纸进行裁剪, 比传统的区域生长方法效率更高, 且裁剪精度也足以满足后续的信息提取需要。该模型比原始的 YOLOv8 模型, 虽然模型大小增加了 1.2%, 但是平均置信度上升了 9.6%, $mAP@0.5$ 上升了 2.8%, $mAP@0.5:0.95$ 提升了 5.5%, 召回率提升了 7.6%。

2) 本文使用轮廓检测方法对图元区域进行裁剪, 引入大小面积块绑定的方法, 精确的将标注连同其所属区域裁剪, 相较于原始方法裁剪精度更高。

参考文献

- [1] 汪昊, 刘向阳. 基于区域中心的交互式图像前景提取方法[J]. 计算机技术与发展, 2020, 30(2): 12-16.
- [2] 于灏, 杨建鸣, 王小刚. 基于改进果蝇算法的工程图纸分割方法研究[J]. 计算机技术与发展, 2018, 28(10): 124-128.
- [3] Brar, K.K., Goyal, B., Dogra, A., Mustafa, M.A., Majumdar, R., Alkhayyat, A., *et al.* (2025) Image Segmentation Review: Theoretical Background and Recent Advances. *Information Fusion*, **114**, Article ID: 102608. <https://doi.org/10.1016/j.inffus.2024.102608>
- [4] 黄鹏, 郑洪, 梁超. 图像分割方法综述[J]. 武汉大学学报(理学版), 2020, 66(6): 519-531.
- [5] Zhang, W., Joseph, J., Yin, Y., Xie, L., Furuhashi, T., Yamakawa, S., *et al.* (2023) Component Segmentation of Engineering Drawings Using Graph Convolutional Networks. *Computers in Industry*, **147**, Article ID: 103885. <https://doi.org/10.1016/j.compind.2023.103885>
- [6] Ghosh, S., Das, N., Das, I. and Maulik, U. (2019) Understanding Deep Learning Techniques for Image Segmentation. *ACM Computing Surveys*, **52**, 1-35. <https://doi.org/10.1145/3329784>
- [7] Buric, M., Grozdanic, S. and Ivacic-Kos, M. (2024) Diagnosis of Ophthalmologic Diseases in Canines Based on Images Using Neural Networks for Image Segmentation. *Heliyon*, **10**, e38287. <https://doi.org/10.1016/j.heliyon.2024.e38287>
- [8] Yong, S.Y., O'Grady, J., Gregory, R. and Lynton, D. (2024) Regional-Scale Image Segmentation of Sandy Beaches in Southeastern Australia. *Remote Sensing*, **16**, Article 3534. <https://doi.org/10.3390/rs16183534>
- [9] Juhong, A., Li, B., Liu, Y., Yang, C., Yao, C., Agnew, D.W., *et al.* (2024) Multihead Attention U-Net for Magnetic Particle Imaging-Computed Tomography Image Segmentation. *Advanced Intelligent Systems*, **6**, Article ID: 2400007. <https://doi.org/10.1002/aisy.202400007>
- [10] 郭宏志, 吕征南, 张志成. 基于改进蜜蜂觅食算法的多阈值图像分割[J]. 测控技术, 2024, 43(9): 28-34.
- [11] 马宇超. 面向信息化管理的机械图纸信息智能提取和识别技术研究[D]: [硕士学位论文]. 常州: 常州大学, 2023.