·农林智能装备与技术 Intelligent Equipment and Technology for Agriculture and Forestry·

基于改进YOLOv11n-seg的木椅腿实例分割方法

徐冀策，王海艳，王宁，魏泽辉，吴哲*
（东北林业大学 机电工程学院，哈尔滨 150040）

摘　要：准确识别木椅腿并定位抓取位置是实现机器人高效上料作业的关键，然而现有的基于实例分割图像处理算法难以兼具精度与速度，且模型参数量庞大，不利于工业场景实际部署，在复杂场景下还容易出现漏检和误检情况。针对上述问题，提出一种基于改进的YOLOv11n-seg模型，通过引入轻量化模块并结合新的损失函数，使其兼具精度与效率，优化部署性能。具体改进方法为，首先引入RepViT作为骨干网络，增强特征提取能力并降低计算复杂度；其次在颈部网络中使用轻量化动态上采样器DySample，提升特征融合效率；最后使用Shape-IoU替换传统CIoU Loss损失函数，优化检测框的回归精度。消融试验和对比试验结果表明，改进后的YOLOv11n-seg算法较原算法相比在推理速度和分割精度上均有提升，其中，模型整体参数量下降11.31%，浮点运算量下降9.62%，分割平均精度均值mAP0.5∶0.95增长4.61个百分点，验证了改进算法的有效性，为实现下一步机器人精准定位并抓取木椅腿提供技术支持。

关键词：木椅腿加工；实例分割；YOLO；轻量化；深度学习
中图分类号：S777　　**文献标识码：**A　　**DOI：**10.7525/j.issn.1006-8023.2026.05.013

Instance Segmentation Method for Wooden Chair Legs Based on Improved YOLOv11n-seg

XU Jice, WANG Haiyan, WANG Ning, WEI Zehui, WU Zhe*
(College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China)

Abstract: Accurate identification of wooden chair legs and localization of grasping positions are crucial for achieving efficient robotic loading operations. However, existing instance segmentation-based image processing algorithms often struggle to balance accuracy and speed, and their large number of parameters hinders practical deployment in industrial scenarios. Furthermore, these algorithms are prone to missed detections and false detections in complex environments. To address the aforementioned issues, this paper proposes an improved YOLOv11n-seg model. By incorporating lightweight modules and a novel loss function, the model achieves a favorable balance between accuracy and efficiency, thereby optimizing its deployment performance. The specific improvements are as follows: firstly, RepViT is introduced as the backbone network to enhance feature extraction capability while reducing computational complexity. Secondly, the lightweight dynamic upsampler DySample is adopted in the neck network to improve feature fusion efficiency. Finally, Shape-IoU is used to replace the traditional CIoU Loss loss function, optimizing the regression accuracy of bounding boxes. Results of ablation experiments and comparative experiments demonstrate that the improved YOLOv11n-seg algorithm outperforms the original model in both inference speed and segmentation accuracy. Specifically, the total number of model parameters is reduced by 11.31%, the floating-point operations (FLOPs) are decreased by 9.62%, and the mean average precision for segmentation (mAP@0.5:0.95) is increased by 4.61 percentage points. These findings validate the effectiveness of the improved algorithm and provide technical support for the subsequent precise positioning and grasping of wooden chair legs by robots.

Keywords: Wooden chair leg machining; instance segmentation; YOLO; lightweight; deep learning

收稿日期：2025-10-13
基金项目：国家重点研发计划项目(2024YFD2200700)。
第一作者简介：徐冀策，硕士研究生。研究方向为木工机械装备设计。E-mail:1219237200@qq.com
***通信作者：**吴哲，博士，教授。主要从事木工机械设计。E-mail:wuzhe@nefu.edu.cn
引文格式：徐冀策，王海艳，王宁，等. 基于改进YOLOv11n-seg的木椅腿实例分割方法[J]. 森林工程，2026，42(5)：1022-1032.
Xu Jice, Wang Haiyan, Wang Ning, et al. Instance Segmentation method for wooden chair legs based on improved YOLOv11n-seg[J]. Forest Engineering, 2026, 42(5): 1022-1032.

0　引言

在异形弯曲木椅腿的生产过程中，使用木材加工中心进行车、铣、雕刻等多工序复合加工，同时配合机器人完成上下料作业，可极大提高加工效率[1]。近年来，随着机器视觉、深度学习和机器人控制技术的快速发展，机器人上下料技术日益成熟，实现了零件坐标的实时获取，进而求解出机器人抓取位姿。为保障机器人抓取精度，本研究采用单目 2D 相机结合激光测距模块的视觉方案，见图 1，以获取产线上木椅腿初始图像信息，通过实例分割方法对木椅腿进行识别与定位。然而，实际车间生产环境中木椅腿存在堆叠、遮挡、光照不均等问题，传统图像处理算法易受上述情况干扰，鲁棒性较差，难以满足工业场景生产的需求。自 2006 年 Hinton[2] 首次提出神经网络的概念以来，基于深度学习技术的目标检测与实例分割技术获得前所未有的发展与应用，为上述问题提供了新的解决思路。

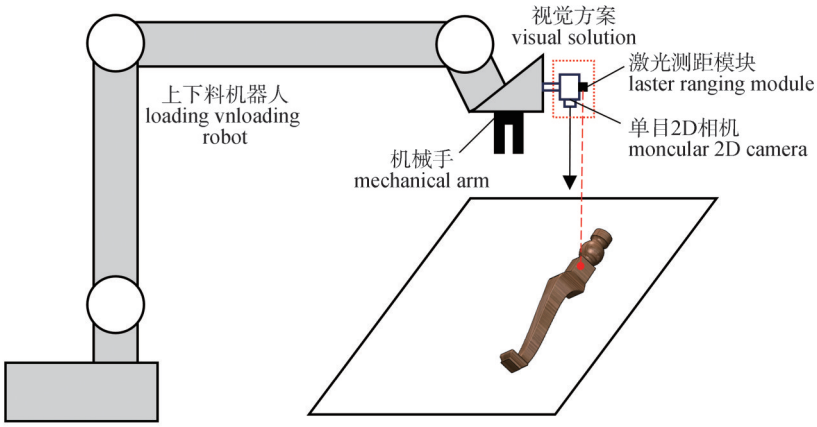


图 1　上下料机器人视觉方案示意图

Fig. 1　Schematic of the vision scheme for loading/unloading robot

基于深度学习的智能分割算法一方面包括以 U-Net[3]、DeepLab 等[4]为代表的语义分割算法，另一方面是基于目标检测框架的实例分割算法。Yu 等[5]提出一种机器人抓取检测网络 SA-U2GNet 结合 U2 网络和随机注意力网络，提高机器人在复杂环境下抓取的精度；白强[6]采用改进的 DeepLabV3+模型并结合三角形抓取策略进一步提高了机器人抓取精度与可靠性。语义分割算法无法分割密集物体中的独立目标，且容易出现前景目标分割不完整的缺陷。基于目标检测框架的实例分割算法不仅可以定位目标的边界框，还能实现像素级精确分割目标，具体可以分为双阶段实例分割算法和单阶段实例分割算法两类。双阶段算法以 R-CNN 系列为代表，藏春华等[7]提出一种基于 Mask R-CNN 实例分割的机械零件识别方法，通过加入 PointRend 模块提升原始模型的分割精度。双阶段算法有较高的精度，但算法流程复杂，在实时性、部署成本及检测效率方面存在一定局限。以 YOLO 系列为代表的单阶段实例分割算法直接从输入图像中提取特征并生成掩码，省略了区域建议网络（regional proposal network，RPN）生成步骤，在时效性与检测效率上优于双阶段实例分割算法，受到了更多研究者的青睐。翟维枫等[8]使用 YOLOv8s-seg 对 RGB 图像中的抓取目标进行分割，结合深度图像生成点云，借助 GPD 检测点云得到候选抓取位姿。李乐[9]基于深度可分离卷积和 Ghost 瓶颈结构对 YOLOv4 目标检测算法进行轻量化改进，提升了实时检测速度，但检测精度有所损失。区树晖[10]在 YOLOv8-seg 中引入 C2f-SCConv 与 ADown 模块，构建轻量化实例分割网络，以降低精度换取更快的分割速度。

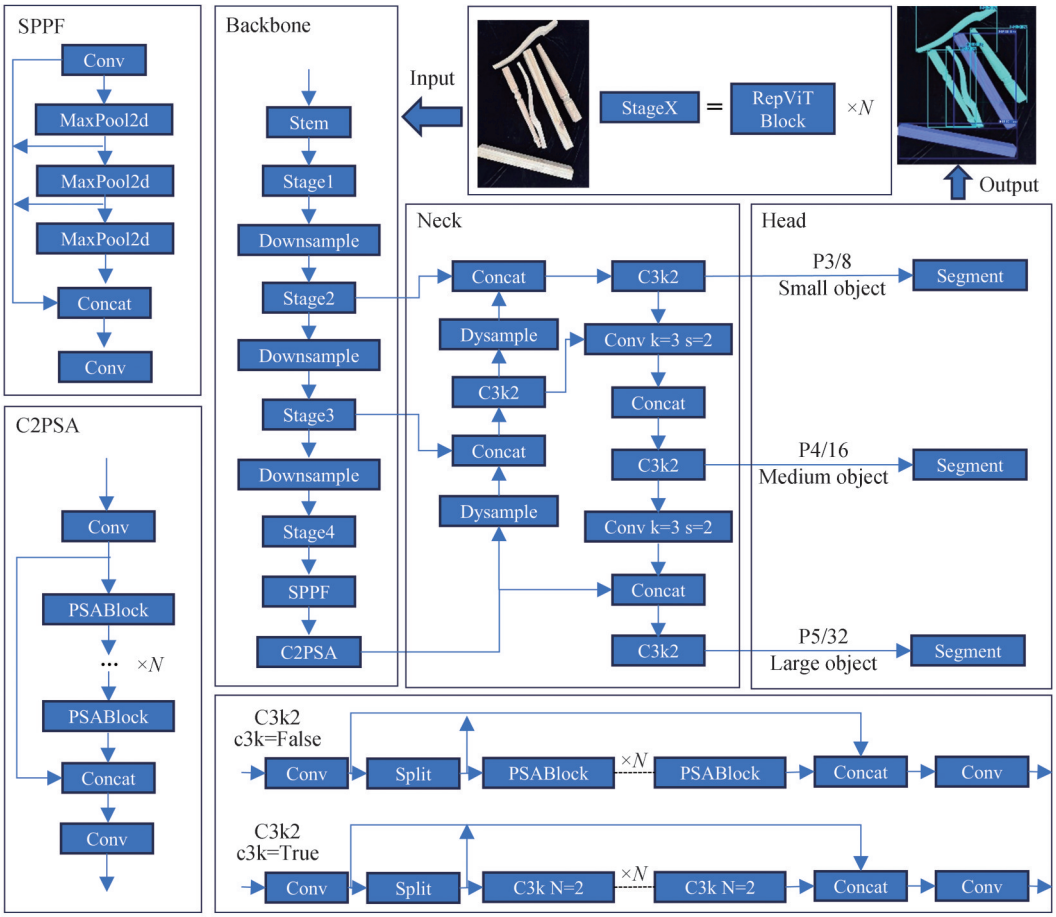
YOLOv11n-seg 作为该系列最新的版本，在分割精度和效率上有显著优势[11]，然而将 YOLOv11n-seg 算法直接用于输送带上木椅腿的实例分割难以兼具精度与速度，且计算参数量大，不利于终端设备的部署和应用[12]。针对上述问题，本研究对 YOLOv11n-seg 算法网络进行优化改进，针对主干网络、颈部网络进行轻量化改进，引入 Shape-IoU 边界框损失函数，提升边界框回归的精度和鲁棒性。最后经消融试验和对比试验，证明本研究改进后的 YOLOv11n-seg 有效性。

1　改进 YOLOv11n-seg 网络结构设计

YOLOv11 的整体结构包含 4 个部分，Input、Backbone、Neck 和 Head，分别对应读取图像、特征提取、特征融合以及检测分割任务。YOLOv11 的主干网络由卷积批量归一化和 SiLU 激活函数（CBS）、C3k2、快速

空间金字塔池化(SPPF)和C2PSA组成。CBS模块是基础卷积单元,负责特征初步提取与维度变换。C3k2取代YOLOv8的C2f,采用多个3×3小卷积核进行分阶段特征提取。SPPF进行快速池化与特征拼接,实现高效多尺度特征金字塔融合,提升小目标检测性能。最后C2PSA结合部分空间注意力(partial spatial attention,PSA)机制,通过多分支特征提取与加权融合,优化了在复杂场景下对小目标及被遮挡目标的检测精度。颈部网络通过路径聚合网格-特征金字塔网络(path aggregation network-feature pyramid network,PAN-FPN)结构融合深浅层特征,增强了目标定位能力。检测头通过解耦设计和DWConv操作,减少了模型参数和计算开销[13]。

YOLO系列有较好的可拓展性,本研究以YOLOv11n-seg作为基线模型,提出一种改进的YOLOv11n-seg模型,期望能在提高实例分割精度的同时降低模型复杂度与参数计算量,进而提升机器人上料时对抓取目标位姿估计的精度,并满足在资源有限的平台上部署和应用的需求。图2为改进的YOLOv11n-seg结构图,采用RepViT作为主干网络,利用多分支结构强化局部表征能力,增强小目标特征的提取能力,使木椅腿边缘细节在网络传播过程中能够更完整地保留。Neck层引入轻量化动态上采样器DySample,减少上采样过程中的信息损失,提升特征融合效率。最后引入Shape-IoU损失函数,优化检测框的回归精度,增强目标定位,提升木椅腿分割精度。



Conv为卷积模块;MaxPool2d为二维最大池化层;Concat为特征融合模块;PSABlock为位置敏感注意力块;Downsample为下采样模块;SPPF为空间金字塔池化模块;C2PSA为跨阶段注意力模块;C3k2为残差特征提取模块;Segment为分割头。

Conv denotes the convolution layer; MaxPool2d denotes the 2D maximum pooling layer; Concat denotes the feature concatenation layer; PSABlock denotes the Position-Sensitive Attention Block; Downsample denotes the downsampling module; SPPF denotes the spatial pyramid pooling fast module; C2PSA denotes a crossstagepartial attention module; C3k2 denotes the residual feature extraction block; Segment denotes the segmentation head.

图2 改进YOLOv11n-seg结构

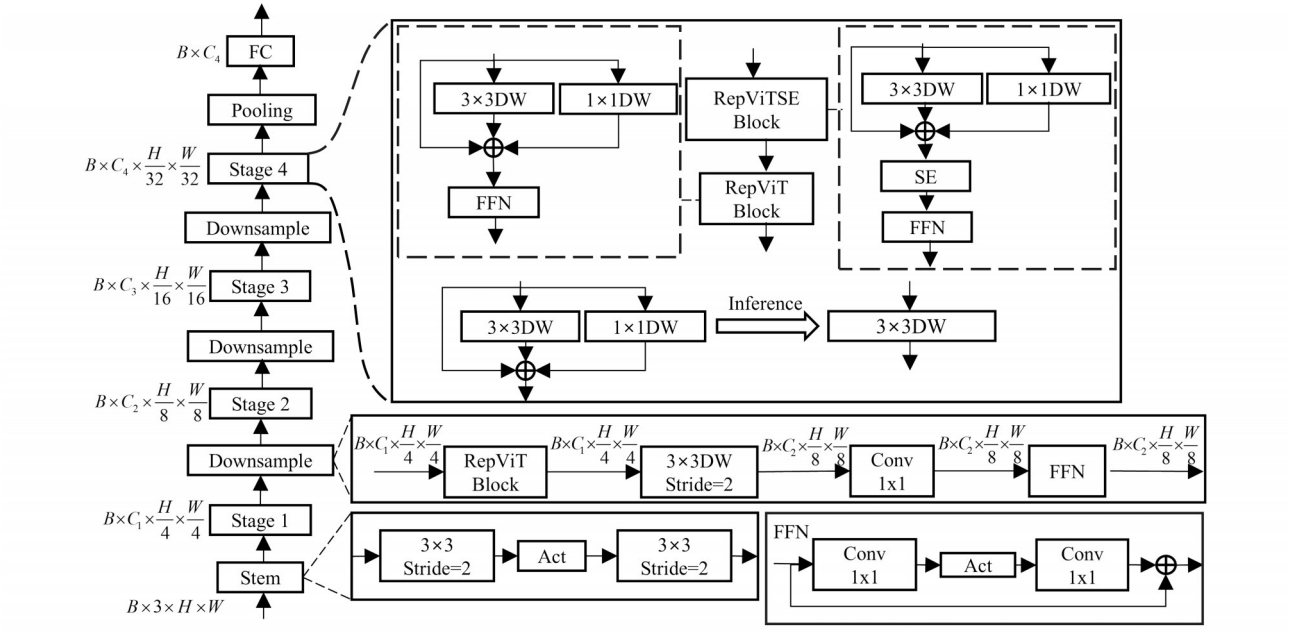
Fig. 2 Architecture of the improved YOLOv11n-seg algorithm

1.1 RepViT主干网络

RepViT由Fu[14]提出,旨在为低算力资源的移动设备设计兼具高性能与低延迟的轻量级模型。该团队受轻量级视觉转换器(vision transformers,ViTs)的高效

图像特征处理过程启发，从 ViTs 视角对轻量级 CNNs 的块结构、宏观架构、微观组件进行系统优化，逐步融合 ViTs 的高效架构，并结合结构重参数化技术消除额外推理开销[15]。相较于传统的 ViTs 模型，RepViT 拥有更加紧凑的网络结构。

RepViT 的结构主要由 5 个部分组成：特征提取 Stem、分阶段特征提取层 Stage、下采样模块 Downsample、池化层 Pooling 以及全连接层 FC，见图 3。Stem 作为主干网络起始层，对图像进行两次堆叠的 3×3 卷积（stride=2）下采样初步处理，小卷积核的堆叠操作可以增强对曲线木椅腿边缘弧度的特征初步捕捉能力，缓解计算瓶颈。随后，图像经分阶段特征提取层 Stage 和下采样模块 Downsample 交替处理，4 个特征提取阶段对应的图像分辨率依次下降为 $\frac{H}{4}\times\frac{W}{4}$、$\frac{H}{8}\times\frac{W}{8}$、$\frac{H}{16}\times\frac{W}{16}$、$\frac{H}{32}\times\frac{W}{32}$，其中 H、W 为输入图像的高、宽，C_i 代表第 i 阶段的通道维度，B 表示训练时的批次处理大小。每个 Stage 阶段包含若干个重复的特征提取块 RepViT Block（图 2）。该模块通过结构重参数化技术分离令牌混合器（token mixer）与通道混合器（channel mixer），用深度卷积捕捉木椅腿粗细、弯曲空间局部特征，与前馈网络（feed-forward neural network，FFN）结合实现跨通道非线性变换。压缩和激励网络层（squeeze-and-excitation network，SE）是一种卷积通道注意力模块，可以增强网络特征学习能力，自适应强化曲线椅子腿轮廓特征对应的通道权重，抑制无关背景信息干扰。RepViT 使用更深的下采样层，额外在 1×1 卷积的输入和输出端增加 RepViT Block 与 FFN 模块，增加网络深度，减少信息损失。综上所述，RepViT 通过层次化结构处理和多阶段特征融合设计，融合传统卷积神经网络的局部感知受野优势和 ViTs 的全局表征能力，在模型轻量化的前提下，实现对木椅腿的高精度实例分割。



Act 为激活函数；Stem 为主干初始层；Stage1~4 为 RepViT 的四个阶段；Stride 为卷积步长；FFN 为前馈网络；Pooling 为池化层；DW 为深度卷积；FC 为全连接层。

Act denotes the activation function; Stem denotes the initial backbone layer; Stage 1-4 denote the four stages of RepViT; Stride denotes the convolution stride; FFN denotes the feed-forward network; Pooling denotes the pooling layer; DW denotes the depthwise convolution; FC denotes the fully connected layer.

图 3　RepViT 结构

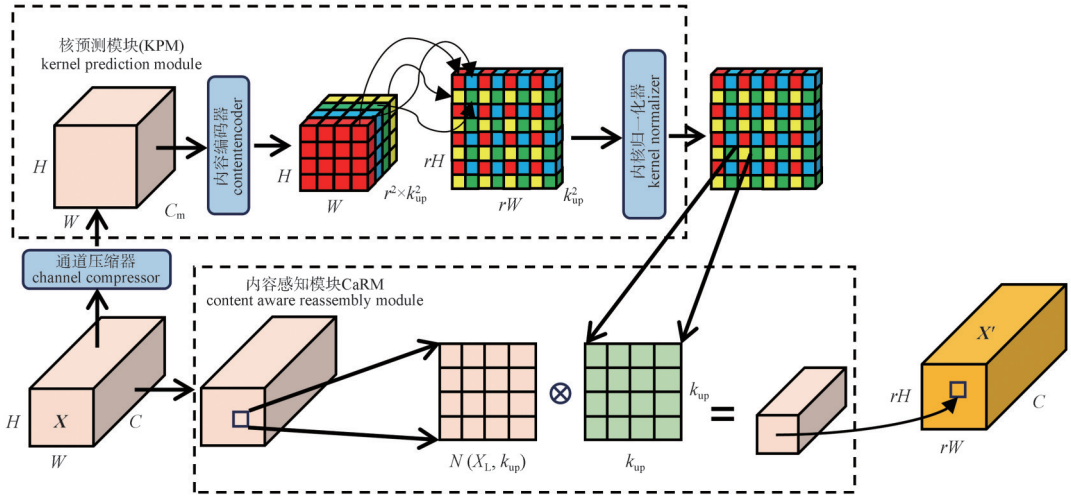
Fig. 3　Architecture of the RepViT

1.2　轻量化动态上采样器 DySample

YOLOv11 的原始上采样模块采用最近邻操作符，该模块作为一种静态插值方法，只能机械地进行像素值的复制与缩放，在处理低分辨率图像时容易引入噪声，丢失细节特征。在木椅腿实例分割场景下，尤其是处理低分辨率的特征图时，原有方法提取木椅腿外形

曲线等细小边缘轮廓表现不佳，上采样过程中会破坏特征图轮廓完整性，导致特征边界出现锯齿状的噪声和模糊，从而影响模型最终分割精度。为增强模型对外轮廓特征信息的表达能力，本研究使用 DySample 动态上采样器替换原有的近邻插值模块。

DySample[16]通过可学习偏移量动态调整采样点位置，使上采样过程更贴合目标形态特征，同时提升低分辨率图像细节恢复能力。相比基于动态卷积的上采样方法感知特征重组（content-aware reassembly of features，CARAFE）[17]，DySample 结构更加轻量化，分组策略的引入，降低了计算复杂度和显存占用。图 4 为 CARAFE 上采样结构，CARAFE 通过上采样核预测模块（kernel prediction module，KPM）和特征重组模块 CaRM 利用底层内容信息预测重组核，并在预定义的区域内重组特征，实现特征增强。CARAFE 内部的通道压缩器、内容编码器、内核归一化器等组件所包含的参数量和计算成本高于 DySample 卷积，导致前向传播速度慢、反向传播的梯度计算量大。



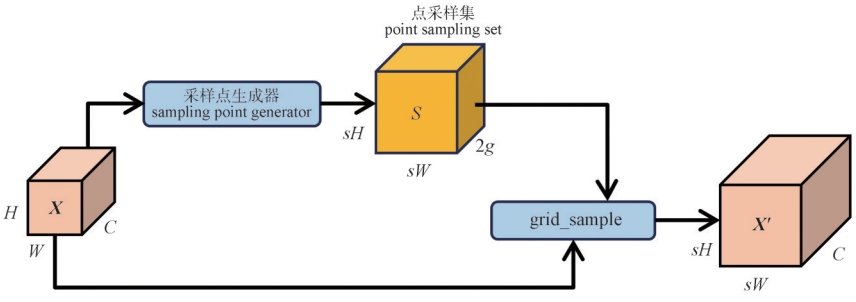
C 为特征图通道数；C_m 为通道压缩后的通道数；r 为上采样的缩放因子；k_{up} 为上采样核大小；rH、rW 为上采样后特征图高度和宽度

C denotes the channel number of the input feature map; C_m denotes the channel number after channel compression; r denotes the upsampling scaling factor; k_{up} denotes the size of the upsampling kernel; rH and rW denote the height and width of the feature map after upsampling.

图 4 CARAFE 上采样结构

Fig. 4 CARAFE up-sampler architecture

DySample 采样过程见图 5。给定一个大小为 $C \times H \times W$ 的输入特征图 $\boldsymbol{X}$，由采样点生成器创建采样集 S，大小为 $2g \times sH \times sW$，其中 2 表示坐标维度，每个采样点需要 x、y 2 个坐标值进行定位，g 表示通道分组数，s 表示上采样比例因子。随后通过内置函数 grid-sample[18]使用采样集 S 中的坐标位置，对特征图 $\boldsymbol{X}$ 按照双线性插值方法重新采样得到大小为 $C \times sH \times sW$ 的高分辨率特征图 $\boldsymbol{X}'$。



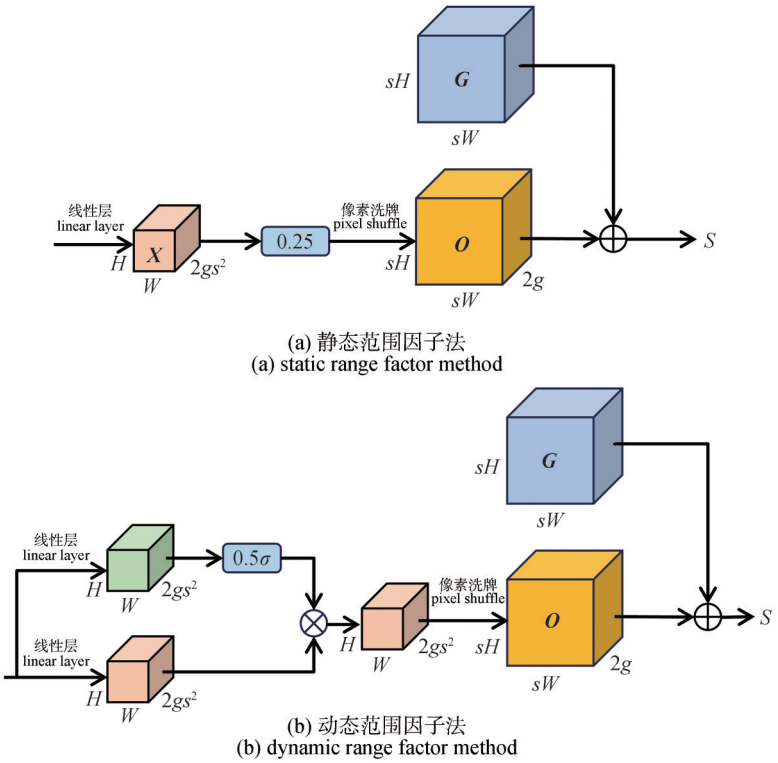
$2g$ 为坐标通道；sH、sW 为上采样后特征图的高度和宽度。

$2g$ denotes the coordinate channel; sH and sW denote the height and width of the feature map after upsampling.

图 5 DySample 采样过程

Fig. 5 Sampling process of DySample

生成采样点的 2 种方法包括静态范围因子法和动态范围因子法，见图 6。静态范围因子法通过一个线性层对输入特征进行变换，并利用像素洗牌操作 pixel shuffle 将通道信息转换为空间偏移量，通过预设的、固定的范围因子缩放偏移量大小，进而生成初始偏移量 $\boldsymbol{O}$。初始偏移量 $\boldsymbol{O}$ 与动态生成的原始采样网格 $\boldsymbol{G}$ 相加得到采样点集 S。原始采样网格 $\boldsymbol{G}$ 是根据输入特征图的分辨率和上采样倍率动态生成的，用于标定上采样后目标像素点在原始低分辨率特征图上的默认位置。静态范围因子法的局限在于可能无法适应所有区域的特征分布。



$\boldsymbol{X}$ 为线性层输出中间特征张量；$\boldsymbol{G}$ 为基础网格坐标张量；$\boldsymbol{O}$ 为像素洗牌后的偏移张量；S 为 $\boldsymbol{G}$ 与 $\boldsymbol{O}$ 相加得到的最终点采样坐标集；sH、sW 为上采样后特征图高、宽；g 为单网格采样点数，s 为上采样缩放因子；σ 为自适应标准差因子。

$\boldsymbol{X}$ denotes the intermediate feature tensor output from the linear layer; $\boldsymbol{G}$ denotes the base grid coordinate tensor; $\boldsymbol{O}$ denotes the offset tensor after pixel shuffle; S denotes the final point sampling coordinate set obtained by adding $\boldsymbol{G}$ and $\boldsymbol{O}$; sH and sW denote the height and width of the feature map after upsampling; g denotes the number of sampling points per grid; s denotes the upsampling scaling factor; σ denotes the adaptive standard deviation factor.

图 6　采样点生成器结构

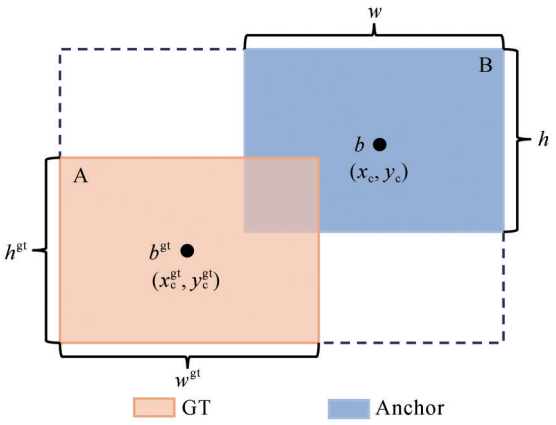
Fig. 6　Structure of sampling point generator

为克服上述方法存在的局限，避免产生采样点过度重叠导致的边界伪影现象，动态范围因子法是一种特殊调控机制。在静态范围因子法的基础上，引入一个动态范围因子，为确保该因子数值稳定，使用 sigmoid[19] 激活函数将动态因子限制在合理范围内，使用动态范围因子调控静态范围因子，生成经校准的偏移量 $\boldsymbol{O}'$，确保采样点在理论边界内，减少预测误差。综上所述，DySample 动态上采样器能保留更多高频细节，具备更大的感受野，在减少计算开销的同时，提升模型的准确性和整体性能。

1.3　Shape-IoU 边界框损失函数

YOLOv11n-seg 采用 CIoU Loss 计算边界框的回归损失，CIoU Loss 引入了预测边界框与真实边界框的中心点距离这一概念，通过在 IoU 损失函数的基础上添加最小中心点惩罚项实现，同时考虑预测边界框宽高比与真实边界框宽高比的差异，进一步提升检测精度。然而其对宽高比的定义较为模糊，并且忽略了边界框自身的形状和尺寸会对边界框回归产生影响，无法进行动态调整。

木椅腿随输送带移动，检测时目标角度不断变化，为解决传统损失函数 IoU、CIoU 等在定位和形状拟合方面存在的问题，提高对产线上的木椅腿边界框回归预测准确性，本研究采用 Shape-IoU[20] 损失函数替换 CIoU 损失函数。Shape-IoU 更容易捕获目标物体边界形状的贴合程度和真实框之间的匹配程度，引入边界框几何特征作为动态权重因子，解决 CIoU 的局限性，更适用于不规则目标区域。Shape-IoU 损失函数见图 7。



A 为真实标注框，w^{gt}、h^{gt} 为真实框宽、高，x_c^{gt}，y_c^{gt} 为真实框中心坐标；B 为锚框；w、h 为锚框宽、高，x_c，y_c 为锚框中心坐标；GT 为真实标注框，Anchor 为先验锚框。

A denotes the ground truth box; w^{gt} and h^{gt} denote the width and height of the ground truth box; x_c^{gt} and y_c^{gt} denote the center coordinates of the ground truth box; B denotes the anchor box; w and h denote the width and height of the anchor box; x_c and y_c denote the center coordinates of the anchor box; GT stands for ground truth; Anchor denotes the predefined anchor box.

图 7 Shape-IoU 结构

Fig. 7 Architecture of Shape-IoU

Shape-IoU 损失函数[21-22]如下列出。

预测框与真实框重叠度交并比 IoU（式中记为 I_{oU}），公式为

$$I_{oU} = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}。 \tag{1}$$

式中：B 为预测框；B^{gt} 为真实框。

距离损失计算式（式中记为 $L_{distance}$），表示预测框与真实框中心距离差，公式为

$$L_{distance} = a_{hh} \times (x_c - x_c^{gt})/c^2 + a_{ww} \times (y_c - y_c^{gt})/c^2。 \tag{2}$$

式中：c 为最小包围框的对角线长度；权重系数 a_{ww} 和 a_{hh} 的公式为

$$\begin{cases} a_{hh} = \dfrac{2 \times (h^{gt})^{s_{cale}}}{(w^{gt})^{s_{cale}} + (h^{gt})^{s_{cale}}} \\ a_{ww} = \dfrac{2 \times (w^{gt})^{s_{cale}}}{(w^{gt})^{s_{cale}} + (h^{gt})^{s_{cale}}} \end{cases}。 \tag{3}$$

式中：s_{cale} 是尺度因子，与数据集中目标尺度有关；a_{ww} 和 a_{hh} 分别为水平方向和竖直方向的权重系数，取值与 GT 框的形状有关；h^{gt} 为真实框高；w^{gt} 为真实框宽度。

外形损失（Ω^{shape}）、预测框与真实框尺寸差异公式为

$$\Omega^{shape} = \sum_{t \in (w,h)} (1 - e^{-\omega_t})^{\theta}。 \tag{4}$$

式中：θ 为形状惩罚敏感度超参数，通常设置为 4；ω_t 为尺寸权重系数，下标 $t \in (w,h)$，分别代表目标框的高和宽；ω_h 和 ω_w 为高度、宽度方向的形状损失权重，其计算式为

$$\begin{cases} \omega_h = a_{ww} \times \dfrac{|h - h^{gt}|}{\max(h, h^{gt})} \\ \omega_w = a_{hh} \times \dfrac{|w - w^{gt}|}{\max(w, w^{gt})} \end{cases}。 \tag{5}$$

式中：w^{gt} 和 h^{gt} 分别为真实框的高度和宽度。边界框回归损失公式为

$$L_{shape-IoU} = 1 - I_{oU} + L_{distance} + 0.5 \times \Omega^{shape}。 \tag{6}$$

2 试验与分析

试验平台操作系统采用 Windows 11，基于 Python 3.10、PyTorch 2.7.1 及 CUDA 12.8 构建开发环境，GPU 使用 RTX 5060 作为算力资源。训练参数设置为：初始学习率为 0.01，动态因子和权重衰减分别为 0.937 和 0.000 5，训练批次设置为 16，选择随机梯度下降（stochastic gradient descent，SGD）优化器，训练轮次设置为 300 轮。

2.1 试验数据

试验基于自主采集的木椅腿毛坯件及初加工椅腿数据集进行，使用的图像采集设备为大恒图像水星二代工业相机 MER2-202-60GC，搭配的镜头为 ZX-SF0620C 工业镜头，共采集照片 330 余张，包含矩形毛坯件以及 4 种不同类型的初加工木椅腿。

为增强模型的鲁棒性和泛化性，对原始采集的照片进行扩充处理，图 8 为数据集建立过程，包括原始采集照片、图片扩充以及标注界面。图片扩充步骤包含几何变换（随机翻转、随机裁剪、随机缩放等）、色彩光照变换、噪声模糊增强等，得到近 2 000 张图像，接着使用 labelme 软件对图形进行标注，构建数据集，最后将上述数据集按照 7∶1∶2 的比例划分为训练集、验证集和测试集。

2.2 评价指标

试验选用分割精确率 Precision（Mask，P）、分割召回率 Recall（Mask，R）、分割平均精度均值 mAP0.5∶0.95（Mask）、参数量 Params、计算复杂度 GFLOPs 和帧率 FPS 作为指标进行评估[23]。精确率 P 表示模型预测为正类样本中真正属于正类的比例；召回率 R 表示在所有实际为正类的样本中被模型正确预测为正类的比例。分割平均精度均值 mAP 是对所有检测类别精度的评估指标，计算方式为各类别平均精度 AP 的均值，mAP0.5 表示 IoU 阈值为 50% 时的 mAP，mAP0.5∶0.95 是一个更严谨的评价指标，表示 IoU 阈值从 50% 到



图 8　数据集建立过程

Fig. 8　Dataset establishment process

95% 范围内的平均 mAP,指标值越高,表明模型在不同类别目标上的检测效果越好。参数量 Params 和计算复杂度 GFLOPs 表征模型的大小及计算复杂度,帧率 FPS 用来评估模型推理速度和实时性能。

利用训练集与验证集进行 300 次迭代训练及验证后,使用训练所得的最优权重在测试集上进行测试得到了木椅腿的检测与分割结果,见图 9,可以直观地看出改进的 YOLOv11n-seg 模型对于不同方向、大小、形状的木椅腿能做到精准检测与分割,表现出色。训练过程中,改进的 YOLOv11n-seg 相关评估指标 Precision (Mask)、Recall(Mask)、mAP0. 5:0. 95(Mask)以及分割损失(Seg Loss)、分类损失(Cls Loss)和蒸馏特征损失(Dfl Loss)随着迭代训练次数变化的曲线见图 10,ml、yt 分别指木料与椅腿,在迭代接近 100 次时,精确率和召回率曲线趋向平缓,达到了近 90%;平均分割精度均值在迭代接近 200 次时达到 83% 左右,表明该模型的检测收敛效果优秀,精度边界拟合能力较强。分割损失(Seg Loss)、分类损失(Cls Loss)和蒸馏特征损失(Dfl Loss)函数训练曲线呈现下降趋势,训练迭代 150 次以后模型趋于稳定,达到了整体最优解。

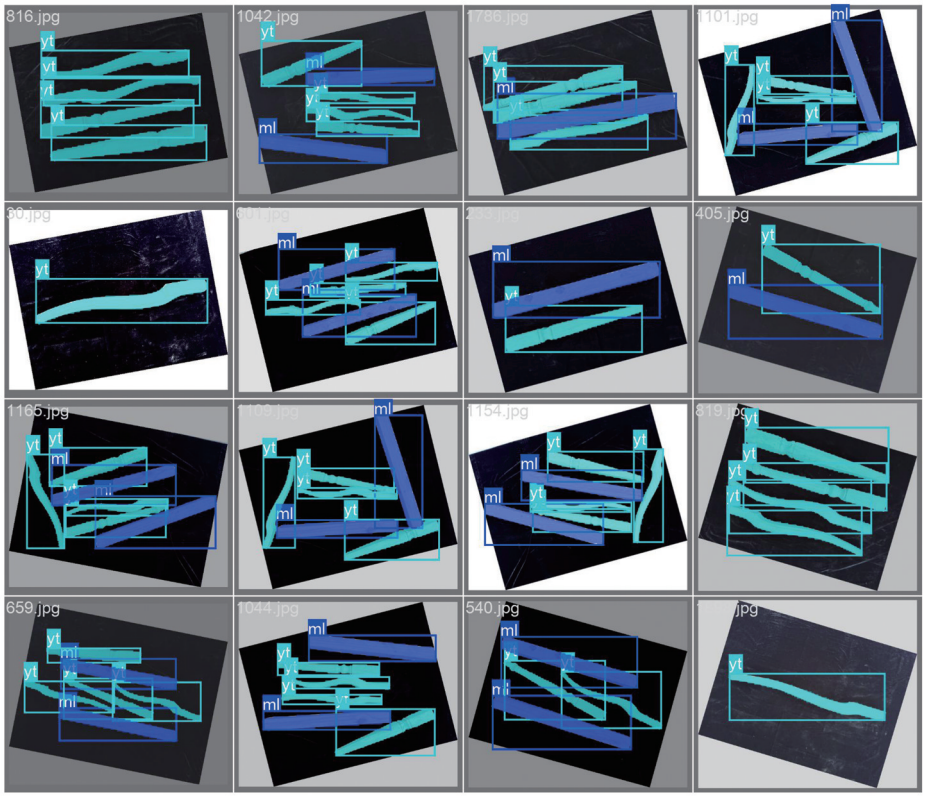


图 9　不同类型木椅腿检测与分割结果

Fig. 9　Detection and segmentation results of different types of wooden chair legs

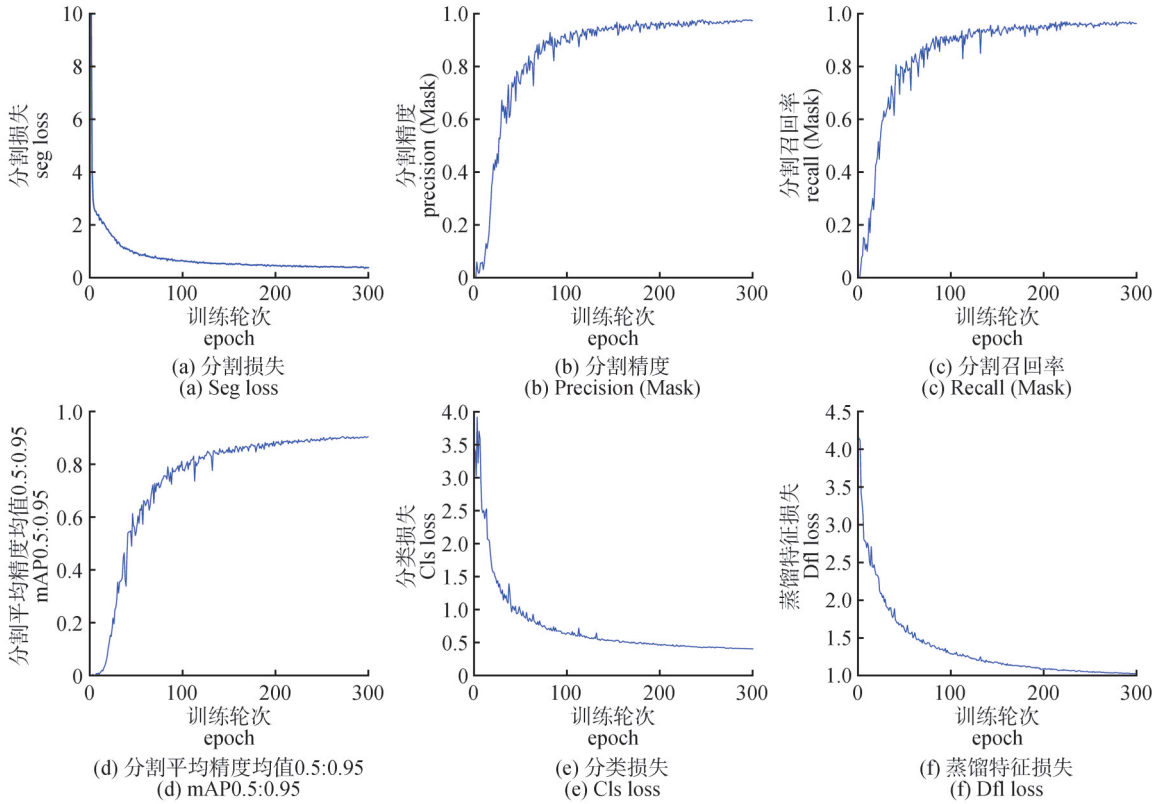


图 10 相关指标的变化曲线

Fig. 10 Variation curves of the relevant metrics

2.3 消融试验

为验证改进模块的有效性，进行消融试验。表 1 为消融试验结果，左侧模型种类依次为基线模型 YOLOv11n-seg、YOLOv11n-seg+RepViT（模型 1）、YOLOv11n-seg+RepViT+DySample（模型 2）、YOLOv11n-seg+RepViT +CARAFE（模型 3）、YOLOv11n-seg+RepViT+DySample +Shape-IoU（模型 4）。

表 1 消融试验结果分析

Tab. 1 Analysis of ablation experiment results

模型 model	分割精度 P/% precision (Mask, P)	分割召回率 R/% recall (Mask, R)	分割平均精度均值/% mAP0.5:0.95 (Mask)	参数量/10^6 params	计算复杂度/G GFLOPs	帧率/(帧·s^{-1}) FPS
基线模型 baseline model	95.08	93.86	83.77	2.83	10.4	197
模型 1 model 1	95.31	94.23	84.06	2.37	9.4	192
模型 2 model 2	96.83	96.53	86.89	2.51	10.4	189
模型 3 model 3	97.17	96.34	87.21	2.98	10.6	143
模型 4 model 4	97.23	97.05	88.38	2.51	9.4	190

主干网络替换为 RepViT 后，分割精度 P 略有下降，分割召回率 R、分割平均精度均值有所提升，整体分割精度变化较小，参数量与计算量降低明显，表明 RepViT 在不影响模型精度的情况下，降低了模型体积，缩减了计算量，整体符合轻量化目标。继续添加轻量化动态上采样器 DySample，分割精度 P、分割召回率 R 以及平均精度均值分别提升 1.52、2.30、2.83 个百分点。对比内容感知特征重组 CARAFE 上采样算子，尽管平均精度均值比 CARAFE 略低 0.4%，但 CARAFE 因复杂的特征重组机制增加了计算量，推理速度下降，DySample 通过动态自适应采样等轻量化机制设计，提升计算效率，综合性能优良。引入 Shape-IoU 损失函数，分割精度 P 增长 0.40%，分割召回率 R 增长 0.52%，分割平均精度均值 mAP0.5:0.95 增长 1.49%。

Shape-IoU对目标形状更敏感,通过增强边界框与目标形状之间的重合度,提高检测精度。

综合分析可得,经改进的YOLOv11n-seg相较基线模型各方面均有出色表现,其中分割精度P增长2.15个百分点,分割召回率R增长3.19%,分割平均精度均值mAP0.5:0.95增长4.61%,模型整体参数量下降11.31%,浮点运算量下降9.62%,验证了改进算法的有效性。

2.4　对比试验

选取多个主流实例分割模型与改进的YOLOv11n-seg模型(ours)进行对比试验,包括单阶段实例分割模型YOLOv5n-seg、YOLOv8n-seg、YOLOv8s-seg、YOLO v11s-seg以及双阶段实例分割模型Mask R-CNN,试验结果汇总见表2。

表2　对比试验结果分析

Tab. 2　Analysis of comparative experimental results

模型 model	分割精度P/% precision (Mask, P)	分割召回率R/% recall(Mask, R)	分割平均精度均值/% mAP0.5:0.95 (Mask)	参数量/10^6 params	计算复杂度/G GFLOPs	帧率/s FPS
Mask-RCNN	96.87	97.53	91.02	44.1	22.9	12
YOLOv5n-seg	92.08	90.56	80.56	2.76	11.1	177
YOLOv8n-seg	94.17	91.60	81.71	3.26	12.1	188
YOLOv8s-seg	97.37	97.62	93.50	11.78	42.7	139
YOLOv11s-seg	98.53	97.97	93.87	10.12	35.6	145
ours	97.23	97.05	88.38	2.51	9.4	190

Mask R-CNN在Faster R-CNN现有的目标检测框架基础上增添一个分支,使其具备分割功能。该模型分割精度很高,但参数量大,计算复杂度高,并且对硬件资源要求严格,推理速度相对较慢,难以满足实时性应用的需求。YOLO系列模型中,s系列实例分割模型相较n系列有更大的参数量,分割精度也有明显提升。由表2可以看出,本研究提出的改进模型性能优于YOLOv5n-seg与YOLOv8n-seg,分割精度接近Mask R-CNN与YOLOv8s-seg。与YOLOv11s-seg相比较,本模型在分割精度P和召回率R上略低约1%~2%,在分割平均精度均值mAP0.5:0.95评估指标上约存在5个百分点的差距。值得注意的是,本研究提出的改进模型在参数量与计算量方面远低于上述实例分割模型,推理速度也有明显优势,在分割精度和速度之间取得了更好的平衡,利于部署在边缘设备等算力资源较低的平台。

3　结论

本研究提出一种基于改进YOLOv11n-seg的木椅腿实例分割轻量化模型。针对基线模型算法精度低、计算复杂度高以及模型参数量大等问题进行优化。首先替换主干网络为参数量更小的RepViT网络;其次选用轻量化动态上采样器DySample;最后引入Shape-IoU边界框损失函数。在压缩模型参数量的同时保障了分割精度,有效提升了模型的实用性与部署能力。主要结论如下。

1)改进后的YOLOv11n-seg模型能保持较高的精度的同时减小了模型复杂度,参数量下降11.31%,浮点运算量下降9.62%,分割精度P、分割召回率R分别增长2.15%、3.19%,分割平均精度均值mAP0.5:0.95增长4.61%,验证了改进算法的有效性。

2)通过试验对比轻量化动态上采样器DySample与CARAFE上采样算子,得出DySample上采样器在分割精度上虽较CARAFE略低,但计算复杂度与推理速度方面优于CARAFE,因此选用轻量化动态上采样器DySample作为改进模块替换原始上采样模块。

3)与主流的实例分割模型进行对比试验,得出本研究改进的模型在推理速度和模型大小等方面有一定优势,但在分割精度方面仍有进一步提升空间,未来将进一步优化这些性能指标。

参　考　文　献

[1] 李亚韫. 基于工业机器人的CNC上下料加工组合生产线设计[J]. 机械工程与自动化,2025,54(2):156-158.
Li Yayun. Design of a combined production line for CNC loading and unloading based on industrial robots[J]. Mechanical Engineering & Automation,2025,54(2):156-158.

[2] Hinton G E,Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786):504-507.

[3] Gowramma B H,Poornima B,Jadekar R B. Enhancing silkworm breeding through advanced image segmentation: a comparative study of U-net, mask R-CNN, and FCN[C]// 2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering, (ICICEC). October 24-25, 2024, Davangere, India. IEEE,

2024:1-6.

[4] Chen L-C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[5] Yu Q S, Xu X R, Liu Y Z, et al. Robot plane grasping pose detection based on U^2-net [C]//2023 IEEE International Conference on Robotics and Biomimetics (ROBIO). December 4-9, 2023. Koh Samui, Thailand. IEEE, 2023: 1-6.

[6] 白强. 机器视觉中语义分割及目标抓取的深度学习技术研究[D]. 贵阳:贵州大学, 2022.
Bai Qiang. Research on deep learning techniques for semantic segmentation and target grasping in machine vision [D]. Guiyang: Guizhou University, 2022.

[7] 臧春华, 周介祺, 刘桂雄. 基于Mask R-CNN实例分割的机械零件识别方法研究[J]. 电子测量技术, 2021, 44(22): 32-36.
Zang Chunhua, Zhou Jieqi, Liu Guixiong. Research on recognition method of mechanical parts based on Mask R-CNN instance segmentation [J]. Electronic Measurement Technology, 2021, 44(22): 32-36.

[8] 翟维枫, 陆文涛, 薛同来. 基于实例分割的机械臂抓取位姿检测方法[J]. 仪表技术与传感器, 2024(10): 89-94, 116.
Zhai Weifeng, Lu Wentao, Xue Tonglai. Robotic arm gripping position detection method based on instance segmentation [J]. Instrument Technique and Sensor, 2024(10): 89-94, 116.

[9] 李乐. 面向非结构化任务场景的机器人目标识别及定位抓取研究[D]. 杭州:浙江大学, 2023.
Li Le. Research on robot object recognition and positioning grasp for unstructured task scenarios [D]. Hangzhou: Zhejiang University, 2023.

[10] 区树晖. 基于合成数据和深度学习的管件无序抓取研究[D]. 广州:华南理工大学, 2024.
Qu Shuhui. Research on unordered pipe component grasping based on synthetic data and deep learning [D]. Guangzhou: South China University of Technology, 2024.

[11] Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements [J]. arXiv. 2024. 10. 48550/arXiv. 2410. 17725

[12] 王建超, 李玮, 谌海龙, 等. 基于改进 YOLOv11n 普洱龙珠茶外观质量检测分级算法[J/OL]. 森林工程, 2025: 1-21[2025-09-28].
Wang Jianchao, Li Wei, Ti Hailong, et al. Appearance quality detection and grading algorithm for Pu'er dragon ball tea based on improved YOLOv11n [J/OL]. Forest Engineering, 2025: 1-21[2025-09-28].

[13] 李彬, 李生林. 改进YOLOv11n的无人机小目标检测算法[J]. 计算机工程与应用, 2025, 61(7): 96-104.
Li Bin, Li Shenglin. Improved YOLOv11n small object detection algorithm in UAV view [J]. Computer Engineering and Applications, 2025, 61(7): 96-104.

[14] Fu Z. Vision transformer: vit and its derivatives [J]. arXiv. 2022. 10. 48550/arXiv. 2205. 11239

[15] Wang A, Chen H, Lin Z J, et al. Rep ViT: revisiting mobile CNN from ViT perspective [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 15909-15920.

[16] Liu W Z, Lu H, Fu H T, et al. Learning to upsample by learning to sample [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). October 1-6, 2023. Paris, France. IEEE, 2023: 6004-6014.

[17] Wang J Q, Chen K, Xu R, et al. CARAFE: content-aware ReAssembly of features [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). October 27-November 2, 2019. Seoul, Korea. IEEE, 2019: 3007-3016.

[18] Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library [J]. arXiv. 2019. 10. 48550/arXiv. 1912. 01703

[19] Huybrechs D, Trefethen L N. Sigmoid functions and multiscale resolution of singularities [J]. arXiv. 2023. 10. 48550/arXiv. 2303. 01967

[20] Zhang H, Zhang S. Shape-IoU: more accurate metric considering bounding box shape and scale [J]. arXiv. 2023. 10. 48550/arXiv. 2312. 17663

[21] 邢远秀, 刘颛玮, 邢玉峰, 等. BDD-DETR:高效感知小目标的锂电池表面缺陷检测[J]. 储能科学与技术, 2025, 14(1): 370-379.
Xing Yuanxiu, Liu Zhuanwei, Xing Yufeng, et al. BDD-DETR: an efficient algorithm for detecting small surface defects on lithium batteries [J]. Energy Storage Science and Technology, 2025, 14(1): 370-379.

[22] 陈鹏宇, 王烈, 梁钰墁, 等. AFL-YOLO:基于YOLOv8改进的小目标检测算法[J/OL]. 电讯技术, 2025: 1-15[2025-01-21].
Chen Pengyu, Wang Lie, Liang Yuman, et al. AFL-YOLO: an improved small object detection algorithm based on YOLOv8 [J/OL]. Telecommunication Engineering, 2025: 1-15[2025-01-21].

[23] 杨鸿丹, 付贵, 邵慧超, 等. 融合多尺度层级特征的航拍小目标检测[J]. 计算机工程与应用, 2025, 61(9): 230-241.
Yang Hongdan, Fu Gui, Shao Huichao, et al. Small object detection in aerial imagery using multi-scale hiearchical feature fusion based approach [J]. Computer Engineering and Applications, 2025, 61(9): 230-241.