文章编号:1671-4598(2022)09-0184-09 DOI:10.16526/j.cnki.11-4762/tp.2022.09.028 中图分类号:TP242.2 文献标识码:A

基于多尺度特征注意 Yolact 网络的堆叠工件分拣算法

徐胜军[1], 李康平[1], 韩九强[1,2], 孟月波[1], 刘光辉[1]
(1. 西安建筑科技大学 信息与控制工程学院, 西安 710055;
2. 西安交通大学 电信学部, 西安 710061)

摘要: 针对非结构化场景中存在的多工件堆叠遮挡等问题, 提出了基于多尺度特征注意 Yolact 网络的堆叠工件识别定位算法; 所提算法首先在 Yolact 网络的掩码模板生成分支中加入多尺度融合与特征注意机制, 提升网络预测堆叠工件掩码的质量, 并设计了基于膨胀编码的目标检测模块, 增强网络对不同尺度堆叠工件的适应能力, 构建了多尺度特征注意 Yolact 网络; 其次, 利用构建的多尺度特征注意 Yolact 网络预测堆叠工件的掩码与边界框, 并对堆叠工件掩码进行最小外接矩形生成, 根据掩码边界框与掩码的最小外接矩形确定目标工件的抓取点与旋转角度; 最后, 基于堆叠工件识别定位算法研发了视觉机器人工件分拣系统; 实验结果表明, 所提模型在边界框回归、掩码预测两项任务上的识别精度均有提升, 机器人工件分拣系统进行堆叠工件分拣作业的成功率达到 97.5%。

关键词: 注意力机制; 膨胀编码; Yolact 网络; 堆叠工件分拣; 视觉机器人

Stacking Workpieces Sorting Algorithm Based on Multi-scale Feature Attention Yolact Network

XU Shengjun[1], LI Kangping[1], HAN Jiuqiang[1,2], MENG Yuebo[1], LIU Guanghui[1]
(1. School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China;
2. Department of telecommunications, Xi'an Jiaotong University, Xi'an 710061, China)

Abstract: Aiming at the problems of multi workpieces stacking occlusion in the unstructured scenes, a stacked workpiece recognition and location algorithm based on multi-scale feature attention Yolact network is proposed. Firstly, the proposed algorithm adds multi-scale fusion and feature attention mechanism to the mask template generation branch of Yolact network to improve the quality of network prediction stacking workpieces mask, and a target detection module based on expansion coding is designed to enhance the adaptability of the network to stacking workpieces with the different scales, and a multi-scale feature attention Yolact network is constructed. Secondly, using the constructed multi-scale feature attention Yolact network to predict the mask and bounding box of the stacked workpieces, the minimum circumscribed rectangle of the stacked workpieces mask is generated, and the grab point and rotation angle of the target workpieces are determined by the mask bounding box and the minimum circumscribed rectangle of the mask. Finally, a visual robot workpieces sorting system is developed based on the stacking workpieces recognition and positioning algorithm. The experimental results show that the recognition accuracy of the proposed model in the two tasks of bounding box regression and mask prediction is improved, and the success rate of stacking workpieces sorting by the robot workpieces sorting system reaches up to 97.5%.

Keywords: attention mechanism; expansion coding; Yolact network; stacking workpieces sorting; vision robot

0 引言

随着国家工业的发展, 仅依靠传统人力对工件进行分拣已无法满足当今工业的生产需求, 以工业机器人代替人力完成工件分拣任务逐渐成为研究的热点。利用工业机器人的自动识别抓取工件是提高自动化生产线效率的关键环节之一。在空调驱动电机的生产过程中, 需要将大量电机

收稿日期:2022-03-13; **修回日期**:2022-04-18。
基金项目:国家自然科学基金(51678470);陕西省自然科学基础研究计划(2020JM472, 2020JM473, 2019JQ760)。
作者简介:徐胜军(1976-),男,陕西西安人,博士,副教授,主要从事图像处理、模式识别、人工智能及自动化等方向的研究。
李康平(1996-),男,陕西西安人,硕士研究生,主要从事图像处理、模式识别、视觉机器人等方向的研究。
韩九强(1951-),男,陕西咸阳人,博士,教授,主要从事智能检测理论及应用、图像信息融合与模式识别方向的研究。
孟月波(1979-),女,陕西西安人,博士,副教授,主要从事机器视觉信息处理与分析、建筑智能化方向的研究。
刘光辉(1976-),男,陕西西安人,博士,副教授,主要从事建筑智能化、机器视觉信息处理与分析方向的研究。
引用格式:徐胜军,李康平,韩九强,等.基于多尺度特征注意 Yolact 网络的堆叠工件分拣算法[J].计算机测量与控制,2022,30(9):184-192,200.

转轴从容器中分拣出来放置于工业流水线上，由车床、铣床进行精加工。然而由于随机摆放的大量电机转轴存在复杂的堆叠遮挡问题，因此这种非结构化场景的堆叠遮挡工件的个体识别难题给基于机器人的自动识别抓取工作带来了很大挑战性。

当前，基于视觉机器人的工件识别分拣方法受到了重点关注。这些研究方法主要分为基于传统视觉方法和基于深度学习方法。基于传统视觉方法主要包括边缘检测方法、特征匹配方法和图像分割方法。伍锡如等[1]基于 Otsu 阈值分割、形态学处理、与边界像素检测等方法实现了对象棋棋子的识别。高赢等[2-3]通过 VisonPro 视觉工具中的 PatMax 特征匹配方法识别发动机瓦盖，实现了发动机瓦盖上料机器人系统。宋海涛等[4]采用 SIFT 特征匹配算法研发了移动机器人物体抓取系统。熊俊涛等[5]利用 K-means 聚类分割和 Hough 圆检测实现了柑橘的识别分割。由于传统机器视觉算法对外界环境变化较为敏感，且算法各阶段所采用的阈值参数均是根据人为经验设定，因此难以准确提取待识别物体的特征，在实际工件分拣机器人系统中不能准确定位抓取非结构化场景的堆叠遮挡工件的个体。

面向图像处理的深度学习网络已被证明在感知问题上具有强泛化能力，因此基于深度学习的机器视觉技术受到了广泛关注。基于深度学习的目标检测算法[6,8-10]是机器人自动识别与抓取物件的常用算法。武星等[11]提出了一种基于轻量化 YOLOv3 网络的苹果检测算法。杨长辉等[12]基于改进 YOLOv3 网络设计了一种柑橘采摘机器人系统，可实现柑橘自动采摘，并对多类障碍物进行识别、避障。朱江等[13]基于 Faster R-CNN 模型实现了曲轴瓦盖上料机器人系统，提升了发动机装配生产线的工作效率。王欣等[14]基于快速 SSD 深度学习算法开发了一种水果识别分拣机器人。杜学丹等[15]利用 Faster R-CNN 目标检测网络得到目标物体的类别与位置，再根据对目标物体分类检测的结果使用基于深度学习的方法学习抓取位置以达到对物体的自动抓取。薛腾等[16]根据抓取过程中物体抵抗外界扰动的能力，提出一种基于触觉信息的抓取质量评估方法，应用此方法搭建了机器人抓取系统实现了对 10 种目标物体的抓取。虽然基于深度学习的目标检测算法对抓取物体与背景有明显差异情况下的简单场景物体识别效果较好，但是由于空调驱动电机的生产过程中大量电机转轴堆叠在工业流水线上，利用目标检测算法在使用样本学习时会学习到过多冗余特征，因此这种算法难以有效识别这种复杂堆叠场景下的电机转轴。

电机转轴抓取属于复杂背景下的堆叠工件抓取问题。实例分割是目标检测与语义分割任务的结合，可以同时得到图片中感兴趣物体的位置、所属类别、与掩码信息。因此像素级的实例分割算法更适用于三维堆叠工件的图像处理。Li 等[17]基于实例感知全卷积网络中位置感知特征图的概念提出 FCIS 实例分割方法。He 等[19]基于 Faster-RCNN[18]边界框识别分支的基础上增加了一个 mask 预测分支，同时使用 ROI Align[19]解决了在候选区域与特征图像进行映射时由于 ROI pooling[18]造成的匹配误差问题，在检测目标对象的同时为每个实例生成掩码。杨攀等[20]使用 Mask-RCNN 算法对不同尺寸的木材端面图像进行了分割，实现了复杂背景下不同尺寸木材的检测、计数功能。赵庶旭[21]基于改进 Mask-RCNN 算法实现了牙齿的识别与分割。王涛[22]基于平衡金字塔思想对 Mask-RCNN 的特征提取网络进行改进，实现了零件回收抓取系统。但由于 Mask-RCNN 属于两阶段实例分割算法，需要 RPN 网络产生建议区域再将建议区域映射至特征图像，根据建议区域与其特征预测实例掩码。建议区域的产生过程与分割操作的不并行会导致模型处理图像速度较慢，无法满足工业生产中的实时性要求。

针对非结构化场景中存在的多个物体堆叠遮挡等问题，提出了基于多尺度特征注意 Yolact 网络的堆叠工件识别定位算法。首先，针对堆叠工件图像分割结果边缘模糊及边界框定位不准问题，提出了多尺度特征注意 Yolact 网络；其次，针对工件位置难以确定问题，利用目标掩码进行最小外接矩形生成，结合目标边界框与目标掩码的最小外接矩形确定了工件的位置信息。最终基于所提算法设计了一种工件分拣机器人系统，并应用于实际空调电机转轴分拣作业场景，通过实验证明了该系统的有效性。

1　Yolact 网络基本原理

Yolact[23]一种基于单阶段目标检测网络的实例分割网络，该网络在单阶段目标检测网络[24-25]上增加了一个掩码模板生成分支，此分支与目标检测分支并行，同时在目标检测分支中增加掩码系数预测分支，最终利用掩码系数与掩码模板产生实例掩码。由于该网络无需等待 RPN[18]生成建议区域后再进行特征映射产生实例掩码，因此 Yolact 的速度远远高于双阶段的实例分割网络，适用于工业场景下的实时检测任务。

Yolact 网络的基本结构如图 1 所示。该网络主要由骨干网络（Backbone）、掩码模板生成分支（Protonet）、预测模块（Prediction module）、聚合分支（assembly）和剪裁模块 5 个部分组成。

骨干网络由 Resnet101[26]与特征金字塔网络（FPN）构成，基于 FPN 得到特征图像 P_5、P_4、P_3，并对特征图像 P_5 进行卷积操作得到特征图像 P_6、P_7。随后将实例分割分为两个并行的子任务，一个子任务将特征图像 P_3 输入 Protonet 生成一系列掩码模板（prototype masks），不同的掩码模板对不同实例的敏感程度不同。另一个子任务在目标检测分支中增加了掩码系数预测分支，在预测目标物体边界框位置与类别的同时，产生掩码模板中表示实例掩码的掩码系数（mask coefficients）。最终，将掩码系数与掩码模板进行线性组合得到实例掩码，再根据预测所得的边界框对图像进行剪裁实现实例分割。

2　多尺度特征注意 Yolact 网络

由于 Yolact 网络采用掩码模板与掩码系数线性组合的方式获得目标物体的分割结果，因此掩码模板的质量好坏

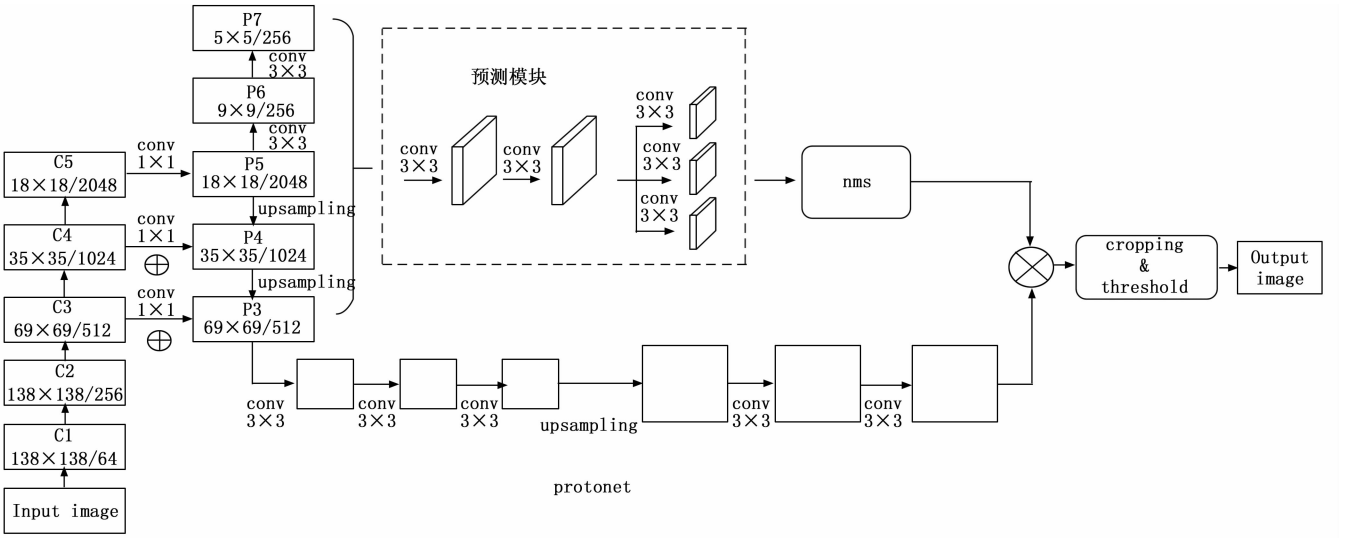

图 1 Yolact 实例分割网络

对目标物体的分割结果有很大的影响。由于本文需要对堆叠工件进行单体识别并进行抓取，在这种堆叠工件识别抓取场景中，最顶层无遮挡的工件与叠加工件的形状、颜色、纹理特征极为相似，故需要深度学习网络提取高级抽象语义特征并识别出最上方可分拣的工件。然而堆叠工件的分割任务不同于常规物体检测，这种任务既需要识别网络能捕获长距离范围内像素点之间的语义关系，同时也需要准确提取工件的空间细节特征。由于标准 Yolact 网络的 Protonet 分支利用 CNN 进行特征提取时仅简单利用卷积层增大其感受野，忽略了不同尺度特征图像之间的特征差异与不同特征对于目标分割精度的重要性。与此同时，由于工件的数量与堆叠程度不同，工件在图像中的尺度大小会发生变化，而 Yolact 算法的预测模块使用特征图像的感受野丰富度较低。因此会造成目标工件边缘分割结果精度低、工件边界定位不准确等问题。为解决此问题，提出了多尺度特征注意 Yolact 实例分割网络，该网络由骨干网络、多尺度特征注意掩码模板生成分支、膨胀编码预测模块、聚合分支与剪裁模块组成。所提网络如图 2 所示。首先利用由 Resnet101 与特征金字塔网络构成的骨干网络对工件图像的特征进行提取。随后将提取到的特征图像 P_3 输入多尺度特征注意掩码模板生成分支获取掩码模板。多尺度特征注意掩码模板生成分支在标准 Yolact 网络的 Protonet 分支中嵌入多尺度融合与注意力机制，可以在聚合多尺度工件图像特征的同时，引导网络学习与目标工件相关特征生成一系列高质量的掩码模板。同时将由骨干网络获得的特征图像 P_3 至 P_7 输送至膨胀编码预测模块预测工件图像中实例的类别、边界框位置、掩码系数。膨胀编码预测模块在标准 Yolact 网络的预测模块中引入膨胀编码机制增强网络对于不同尺度大小目标的适应能力。聚合分支将实例的掩码系数与掩码模板进行线性组合得到实例掩码。剪裁模块根据膨胀编码预测模块预测的边界框对图像进行剪裁完成实例分割。

2.1 多尺度特征注意掩码模板生成分支

提出的多尺度特征注意掩码模板生成分支结构如图 3 所示。该分支利用五层卷积层对特征图像 P_3 进行特征提取，然后使用空洞空间金字塔池化层对由 conva 得到的特征

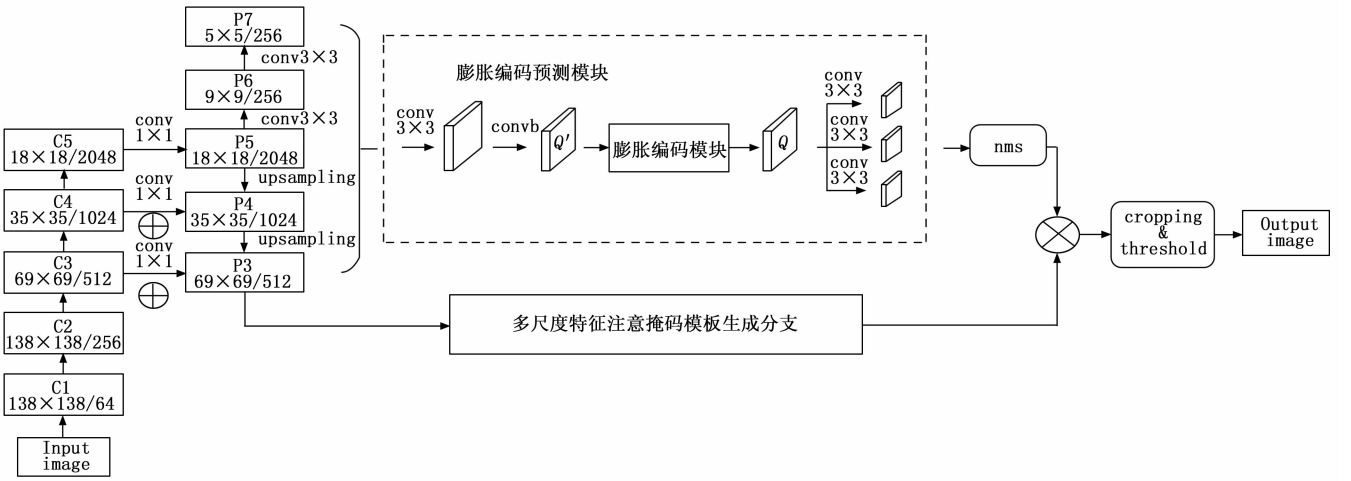

图 2 多尺度特征注意 Yolact 实例分割网络

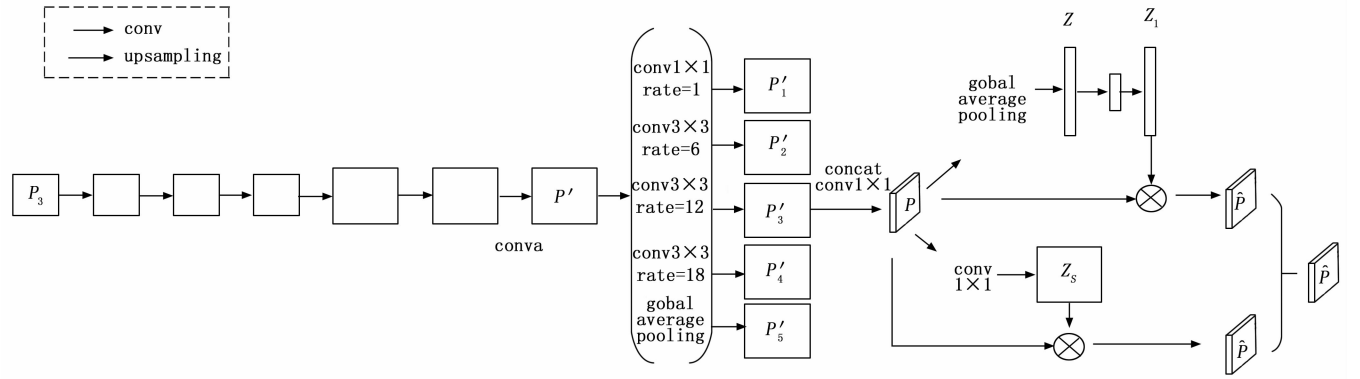

图 3　多尺度特征注意掩码模板生成分支

图像 P' 进行处理，并对得到的特征图像 P'_1、P'_2、P'_3、P'_4、P'_5 通过通道拼接的方式进行融合。为避免多层卷积造成的特征图像分辨率下降问题，应用空洞卷积在不加深网络深度的情况下增大特征图像的感受野，不同大小空洞率的卷积核会产生包含不同尺度特征的特征图像 P'_1 至 P'_5，将不同尺度的特征图像以通道拼接的方式进行融合，可以获得包含低级颜色、纹理特征与高级抽象语义特征的特征图像 P。因此使用此特征图像生成的掩码模板可在对图片中像素点进行精确分类的同时确保图片的空间分辨率。

为了增强所提网络对感兴趣区域细节特征的学习能力，基于注意力机制构成空间金字塔注意力模块即在其后加入通道注意力分支、空间注意力分支两个并行的网络分支。通道注意力分支由全局池化层与两组卷积核尺寸大小为 1×1 的卷积层构成。该分支的输入为特征图像 $P=\{p_1,p_2,\cdots,p_c\}$，其中 $p_i\in\mathbf{R}^{H*W}$ 表示第 i 个通道，H,W 分别表示特征图像的高和宽。该分支使用全局平均池化对特征图像 P 进行压缩操作，即将每个特征通道都压缩成一个实数，从而将感受野扩展到全局范围。特征图像 P 经过全局平均池化后得到一个向量 Z，且 $Z\in\mathbf{R}^{1*1*C}$，则每个位置的值 z_c 为：

$$z_c=\frac{1}{H\times W}\sum_{i}^{H}\sum_{j}^{W}p_c(i,j) \tag{1}$$

式中，c 为特征图像的通道数，H,W 分别为特征图像 P 的高和宽。

将每个特征通道 p_i 压缩后得到的向量 Z 使用一个卷积层降维，然后通过整流线性单元 ReLU 激活，再经过一个卷积层升维，最后利用 sigmoid 激活函数生成长度为 32 的特征注意权重向量 Z_1，整个过程表示如下：

$$Z_1=\delta[w_2(\sigma(w_1Z))] \tag{2}$$

式中，w_1,w_2 分别为两个卷积层的权重，$\sigma(\cdot)$ 表示整流线性单元 ReLU，$\delta(\cdot)$ 表示 sigmoid 激活函数。

通过特征向量 Z_1 对空洞空间金字塔池化层产生的特征图像 P 进行加权，加权特征图像 $\widetilde{P}$ 中不同特征通道 $\{\widetilde{p}_1,\widetilde{p}_2,\cdots\widetilde{p}_c\}$ 按其重要性可以获得不同的权重。与原图中较为重要位置（如空调电机转轴的边缘）有关的特征通道分配较大的权重，反之分配较小的权重。加权的特征图像 $\widetilde{P}$ 计算过程如下所示。

$$\widetilde{P}=F_c(P)=\{z_{11}p_1,z_{12}p_2,\cdots,z_{1c}p_c\} \tag{3}$$

式中，z_{1i} 为第 i 个通道的权重，衡量了第 i 个特征通道 p_i 的特征重要程度。

空间注意力分支由尺寸大小为 1×1、卷积核数量为 1 的卷积层构成。此分支的输入图像为 $P=\{p_{1,1},p_{1,2},\cdots\cdots,p_{i,j},\cdots\cdots,p_{H,W}\}$，$H,W$ 为特征图像的尺寸，i,j 为特征图像的空间位置。首先使用卷积层 $w_3\in\boldsymbol{R}^{1*1*c*1}$ 对特征图像 P 进行空间压缩操作，得到一个特征注意权重张量 Z_s：

$$Z_s=w_3(P) \tag{4}$$

式中，特征张量 Z_s 中的权重表示图像上不同位置对于图像分割任务的贡献度。

然后，利用特征张量 Z_s 对特征图像进行加权，原始图像上实例所在位置的特征可以得到更大的权重。加权后的特征图像 $\overline{P}$ 表示为：

$$\overline{P}=F_s(P)=\{\delta(Z_{s_{(1,1)}})p_{1,1},\cdots,\delta(Z_{s_{(i,j)}})p_{i,j},\cdots,\delta(Z_{s_{(H,W)}})p_{H,W}\} \tag{5}$$

式中，$\delta(\cdot)$ 为 sigmoid 激活函数，$\delta(Z_{s(i,j)})$ 代表特征图 P 中位置 (i,j) 处的特征重要度。

为使得网络重点学习到原始图像中目标工件所处空间位置的特征及与目标工件相关的特征通道。将由通道注意力分支和空间注意力分支分别得到的特征图像 $\widetilde{P},\overline{P}$ 以逐像素点相加的方式进行融合，利用融合后的特征图像 $\hat{P}$ 生成掩码模板，提升网络对于目标工件的分割效果。融合后的特征图像 $\hat{P}$ 表示为：

$$\hat{P}=\widetilde{P}+\overline{P} \tag{6}$$

2.2　膨胀编码

Yolact 网络把 FPN 网络输出的 5 层特征图像 $P_3\sim P_7$ 输入预测网络实现边界框的预测，利用边界框对集成掩码进行剪裁并生成实例掩码。如果边界框定位准确，那么网络会生成高质量的实例掩码；反之生成的实例掩码包含大量“噪声”，对分割结果带来很大干扰。由于特征图像感受野的丰富程度对边界框的定位精度至关重要，为使检测网络对不同尺度的物体均实现准确检测，本文选择感受野较大的特征图像检测大尺度目标，感受野较小的特征图像检测小尺度目标。所提网络在预测模块的 Convb 卷积层后加入膨胀编码模块，膨胀编码模块首先使用 1×1 的卷积将特

征图像的通道数缩小为原来的四分之一，之后使用 3×3 的卷积对特征图像的上下文语义信息进行整合，并连续叠加 4 个相同结构的残差块，残差块的结构如图 4 所示。

每个残差块均由 3 个卷积层构成，卷积核的尺寸大小分别为 1×1、3×3、1×1，其中 3×3 的卷积层使用了空洞卷积，4 个残差块中尺寸大小为 3×3 的卷积层的空洞率分别设置为 2、4、6、8。通过 4 个残差块的叠加将不同感受野大小的特征图像以逐像素点相加的方式进行融合，得到上下文信息足够去检测不同尺度目标的特征图像 Q。由于特征图像 Q 中蕴含的上下文信息较原始网络 Q' 更加丰富，涵盖了在实际分拣作业场景中可能遇到的各种不同大小尺度工件的特征。因此基于此特征图像进行目标边界框回归，可以提升网络对于不同尺度目标的适应调节能力进而提高边界框的预测精度。

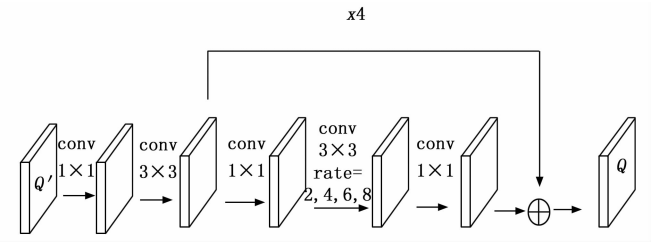

图 4 膨胀编码模块

2.3 损失函数

本文所使用的损失函数由类别损失、边界框位置损失与掩码预测损失三部分组成。总损失函数 $L = \partial_1 L_{class} + \partial_2 L_{bbox} + \partial_3 L_{mask}$ 如下式所示：

$$L = \partial_1 L_{class} + \partial_2 L_{bbox} + \partial_3 L_{mask} \tag{7}$$

式中，$L_{class} = -\sum_{i \in positive}^{N} x_{ij}^{p} \log(c_i^p) - \sum_{i \in negative} \log(c_i^0)$ 为边界框内物体类别损失；$L_{class} = -\sum_{i \in positive}^{N} x_{ij}^{p} \log(c_i^p) - \sum_{i \in negative} \log(c_i^0)$ 为边界框位置损失；$L_{class} = -\sum_{i \in positive}^{N} x_{ij}^{p} \log(c_i^p) - \sum_{i \in negative} \log(c_i^0)$ 为框内物体掩码预测损失；$L_{class} = -\sum_{i \in positive}^{N} x_{ij}^{p} \log(c_i^p) - \sum_{i \in negative} \log(c_i^0)$，$L_{class} = -\sum_{i \in positive}^{N} x_{ij}^{p} \log(c_i^p) - \sum_{i \in negative} \log(c_i^0)$，$L_{class} = -\sum_{i \in positive}^{N} x_{ij}^{p} \log(c_i^p) - \sum_{i \in negative} \log(c_i^0)$ 分别为各项损失的权重，实验中令 $L_{class} = -\sum_{i \in positive}^{N} x_{ij}^{p} \log(c_i^p) - \sum_{i \in negative} \log(c_i^0)$，$L_{class} = -\sum_{i \in positive}^{N} x_{ij}^{p} \log(c_i^p) - \sum_{i \in negative} \log(c_i^0)$，$L_{class} = -\sum_{i \in positive}^{N} x_{ij}^{p} \log(c_i^p) - \sum_{i \in negative} \log(c_i^0)$。

$L_{class} = -\sum_{i \in positive}^{N} x_{ij}^{p} \log(c_i^p) - \sum_{i \in negative} \log(c_i^0)$ 类别损失函数定义为：

$$L_{class} = -\sum_{i \in positive}^{N} x_{ij}^{p} \log(c_i^p) - \sum_{i \in negative} \log(c_i^0) \tag{8}$$

式中，$L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$。当 $L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 时表示第 $L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 个预测框匹配至第 $L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 个的 GroundTruth 边界框，且框内物体预测类别为 $L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$，当 $L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 时表示预测框为背景；$L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 表示第 $L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 个预测框内物体预测类别为 $L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 的概率值，$L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 表示类别为背景；$L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 表示预测框的序号，$L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 表示 GroundTruth 边界框的序号，$L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 表示边界框总数。

$L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q)$ 边界框位置损失函数定义为：

$$L_{bbox} = \sum_{i \in positive}^{N} \sum_{Q \in \{Cx,Cy,w,h\}} x_{ij}^{k} smoothL1(l_i^Q - \hat{g}_j^Q) \tag{9}$$

式中，l_i^Q 表示网络预测值，i 表示预测框序号，j 表示 GroundTruth 边界框序号；$\hat{g}_j^Q$ 表示 GroundTruth。由于网络输出的是预测框对于先验框的偏移量，因此 $\hat{g}_j^Q$ 为 GroundTruth 对于先验框的偏移量，$\hat{g}_j^Q$ 定义为：

$$\hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx})/d_i^{w} \tag{10}$$

$$\hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy})d_i^{h} \tag{11}$$

$$\hat{g}_j^{w} = \log(g_j^{w}/d_i^{w}) \tag{12}$$

$$\hat{g}_j^{h} = \log(g_j^{h}/d_i^{h}) \tag{13}$$

式中，g_j^{cx}，g_j^{cy} 分别表示 GroundTruth 边界框中心点的横、纵坐标。g_j^{w}，g_j^{h} 分别代表 GroundTruth 边界框的宽和高。d_i^{cx}，d_i^{cy} 分别表示先验框中心点的横、纵坐标。d_i^{w}，d_i^{h} 分别代表先验框的宽和高。smoothL1 为损失函数，定义为：

$$smoothL1(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & else \end{cases} \tag{14}$$

式中，输入 x 表示为网络预测值与 GroundTruth 对于先验框偏移量的差值 $l_i^Q - \hat{g}_j^Q$。

L_{mask} 掩码预测损失定义为：

$$L_{mask} = -1/n\sum_{i=1}^{N}[y_i\log x_i + (1-y_i)\log(1-x_i)] \quad (15)$$

式中 i 表示图片中的像素点序号，N 为像素点总数。x_i 代表第 i 个像素点所属类别的预测值，y_i 代表第 i 个像素点所属类别的真实值。

3　基于视觉引导的堆叠工件分拣机器人

堆叠工件分拣机器人以广州数控 RB06-900 六轴工业机器人为主体，在其机械臂上安装了英特尔 D435i 深度相机与 SRT 公司生产的 SFG-FMA4-M5072 柔性夹爪末端执行机构。由深度相机获取容器内杂乱堆放的工件图像，控制器对获取的容器内堆叠工件图像进行分析处理，根据工件识别定位结果引导工业机器人进行工件夹取、搬运等分拣操作。研发的堆叠工件分拣机器人作业系统如图 5 所示。

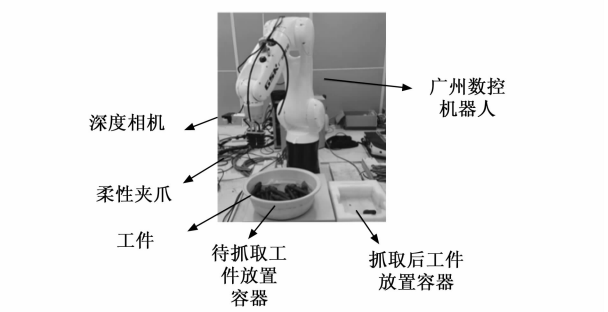

图 5　堆叠工件分拣机器人系统

基于构建的多尺度特征注意 Yolact 网络可实现对堆叠工件中可分拣工件的识别，但要实现对工件的抓取，需要计算工件抓取点在世界坐标系下的坐标与旋转角度，进而确定机器人的抓取位姿。令机器人的抓取位姿 $G=\{x_w,y_w,z_w,a,b,c\}$，其中 x_w，y_w，z_w 为工件抓取点在世界坐标系下的坐标值，a，b，c 为机器人末端柔性夹爪中心点绕世界坐标系的 x，y，z 轴的旋转角度。坐标转换示意图如图 6 所示，图中 $u-o_1-v$ 为像素坐标系，$x-o_2-y$ 为物理坐标系，$x_c-o_3-y_c-z_c$ 为相机坐标系，$x_w-o_4-y_w-z_w$ 为世界坐标系。

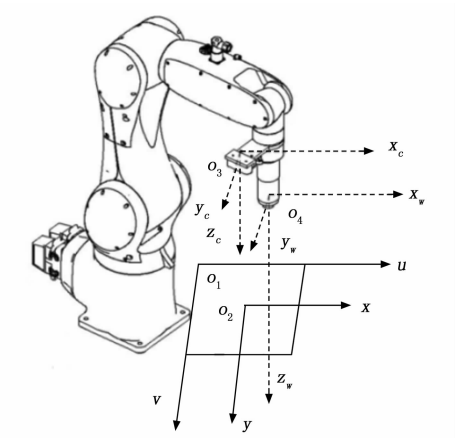

图 6　坐标转换示意图

获取工件抓取点像素坐标与旋转角度方法的具体步骤如下。

步骤 1：令机器人的抓取位姿 $G=\{x_w,y_w,z_w,a,b,c\}$，其中 x_w,y_w,z_w 为工件抓取点在世界坐标系下的坐标值，a，b，c 为机器人末端柔性夹爪中心点绕世界坐标系的 x，y，z 轴的旋转角度。工件抓取点坐标 x_w,y_w,z_w 定义如下：

$$\begin{bmatrix}x_w\\y_w\\z_w\\1\end{bmatrix}=\begin{bmatrix}\boldsymbol{R} & \boldsymbol{T}\\0 & 1\end{bmatrix}\begin{bmatrix}\dfrac{(c_x-u_0)\cdot d\cdot d_x}{f}\\[2mm]\dfrac{(c_y-v_0)\cdot d\cdot d_y}{f}\\[2mm]d\\1\end{bmatrix} \quad (16)$$

式中，$\boldsymbol{R}$、$\boldsymbol{T}$ 为世界坐标系相对于相机坐标系的旋转、平移矩阵。u_0、v_0 为图像中心点的像素坐标，c_x、c_y 为工件抓取点在像素坐标系下的像素坐标，f 为相机焦距，d 为由深度相机获取的工件抓取点与摄像头光心之间的距离，d_x,d_y 为图像中 u，v 方向上一个像素点对应的真实距离。

步骤 2：利用多尺度特征注意 Yolact 实例分割算法对图像进行处理得到待抓取工件的分割结果与边界框 f，选取待抓取工件目标中置信度最高的目标 $\hat{f}_k$；

$$\hat{f}_k=\max_{k\in K}\{conf(f_k=F(\Omega))\mid\forall k\in\{1,2,\cdots,K\}\} \quad (17)$$

式中，f_k 表示预测出的第 k 个目标工件，Ω 为采集到的工件图像；K 表示网络预测出的目标总数，F 表示实例分割网络的推理过程，$conf(f_k)$ 表示第 k 个待抓取工件的置信度。

步骤 3：建立待抓取工件目标 $\hat{f}_k$ 的图像轮廓点集 P_k，使用 Graham 扫描法对点集 P_k 构建凸包点集 $\hat{P}_k$；

步骤 4：计算凸包多边形各边 $\{l_k\mid\forall k\in\{1,2,\cdots,K\}\}$ 与图像横轴的夹角 θ，并将凸包多边形各边 l_k 依次围绕凸包点 $p_i(x_i,y_i)$ 旋转 θ，旋转后的点集 $\widetilde{P}_k$ 表示为：

$$\widetilde{P}_k=\begin{bmatrix}x_1 & y_1 & 1\\ \vdots & \vdots & \vdots\\ x_K & y_K & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ -x_i & -y_i & 1\end{bmatrix}\begin{bmatrix}\cos\theta & \sin\theta & 0\\ -\sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ x_i & y_i & 1\end{bmatrix} \quad (18)$$

步骤 5：依据旋转后的点集 $\widetilde{P}_k$ 获取待抓取工件的最小外接矩形与其对角顶点坐标值 $\{[x_l\quad y_l],[x_r,y_r]\}$，并计算其抓取点像素坐标 $\{c_x,c_y\}=\left\{\dfrac{(x_l+x_r)}{2},\dfrac{(y_l+y_r)}{2}\right\}$。

步骤 6：依据步骤 5 对角顶点坐标值 $\{[x_l,y_l],[x_r,y_r]\}$ 计算待抓取工件最小外接矩形框对角线偏移垂轴夹角 $a_1=\arctan\dfrac{x_r-x_l}{y_l-y_r}$。

步骤 7：依据工件实例分割边界框顶点坐标 $f=\{[m_1,n_1],[m_2,n_2],[m_3,n_3],[m_4,n_4]\}$ 分别计算两条对角线与垂轴夹角 $a_2=\arctan\dfrac{m_1-m_2}{n_2-n_1}$、$a_3=\arctan\dfrac{m_3-m_4}{n_4-n_3}$。

步骤 8：若 a_2 与 a_1 的符号相同，则取 a_2 为工件的旋转角度。若 a_3 与 a_1 的符号相同，则取 a_3 为工件的旋转角度。

即机器人抓取工件旋转角度 φ 定义为：

$$\varphi = \begin{cases} a_2 & a_2 \times a_1 > 0 \\ a_3 & a_3 \times a_1 > 0 \end{cases} \tag{19}$$

步骤 9：根据步骤 5 得到的抓取点像素坐标 $\{c_x, c_y\}$ 和步骤 8 得到的机器人抓取工件旋转角度 φ，输出机器人抓取位姿参数 $G = \{x_w, y_w, z_w, a, b+\varphi, c\}$，机器人执行抓取工件操作。

4 实验分析

4.1 模型训练平台与工件数据集

深度学习平台的配置为 Ubuntu18.04 系统，两块显存为 11G 的 GeForce _ RTX _ 2080 _ TiGPU，python 版本为 3.6.5，pytorch 版本为 1.2.0，Cuda 版本为 10.1。

训练参数配置中批大小 Batchsize 设置为 8，学习率设置为 0.002，训练回合数为 2 000，采用随机梯度下降法（SGD）对网络各层权重进行优化。

工件数据集使用自建的堆叠空调电机转轴数据集，利用深度相机对图像进行采集，设计了随机摆放的不同堆叠程度、不同数量的工件场景，涵盖了在实际工业分拣场景中工件可能出现的所有场景。数据集共采集 800 张图片，部分示例图像场景如图 7 所示。为了提高网络泛化能力，本文采用随机镜像、随机翻转、随机旋转、光度扭曲等数据增强方法对实验数据集进行增强，增强后的数据集共 3 00 张图片，其中复杂堆叠场景下的图片 1 148 张，轻微堆叠场景下的图片 1 184 张，无堆叠场景下的图片 868 张。将全部标注图片转化为 coco 数据集格式，并按照 8∶2 的比例划分为训练集与验证集。利用 labelme 软件对工件数据集进行标注，标注信息可提供深度学习网络训练所需要的工件掩码特征。

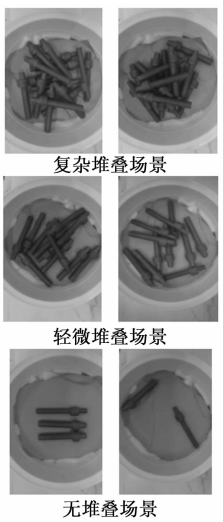

图 7 数据集部分图像

4.2 评价指标

本实验采用实例分割模型通用评价指标 MAP（平均精度均值）来评判算法模型的好坏，MAP 的计算方法如下：

首先根据样本的真实值与预测值将样本划分为 4 种类型：

表 1 样本类型划分

预测值 \ 真实值	负样本	正样本
正样本	TP（True positive）	FP(False positive)
负样本	TN(True negative)	FN(False negative)

然后计算精确度（Precision）与召回率（Recall），P 指精确度（Precision）计算方式如式所示：

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} \tag{20}$$

R 指召回率（Recall）计算方式如式所示：

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{21}$$

以 P 为纵轴，R 为横轴绘制 $P-R$ 曲线，该曲线与坐标轴围成的面积即为单类别的 AP。MAP 即为所有类别 AP 的平均值。

4.3 实验结果对比分析

为了验证本文提出的多尺度特征注意 Yolact 网络的有效性，分别与 Yolact、改进掩码模板生成分支的 Yolact、引入膨胀编码的 Yolact、Mask-RCNN（Resnet50）[19]、Mask-RCNN（Resnet101）[19]、改进 Mask-RCNN[21]、FCIS[17] 共 7 个网络进行对比实验。

定性分析。为验证所提出网络对工件掩码的提取效果，堆叠工件掩码提取效果对比结果如图 8 所示。图 8（a）为标准 Yolact 网络对堆叠工件掩码提取结果，图 8（b）为本文提出的多尺度特征注意 Yolact 网络对堆叠工件掩码提取结果。由图 8 第一行工件图像掩码提取结果可看出标准 Yolact 网络在对工件图片掩码进行预测时存在掩码不完整的情况。而本文构建的多尺度特征注意 Yolact 网络可以完整的将掩码预测出来。由图 8 第二行工件图像掩码提取结果可以看出由标准 Yolact 网络得到的工件掩码的边缘比较模糊，而由本文提出网络得到的工件掩码的边缘细节效果较好，掩码更加贴近工件边缘。由图 8 第三行工件图像掩码提取结果可以看出，由标准 Yolact 网络提取的工件掩码存在误判情况，而由本文提出网络提取的工件掩码较为准确，可以更好地区分出上下方工件。

为验证所提算法在实际分拣作业中采集的不同场景下堆叠工件图像进行识别的效果，分别与 Yolact、Mask-RCNN（Resnet50）、Mask-RCNN（Resnet101）、改进 Mask-RCNN 等 4 个网络进行实验对比，对比结果如图 9 所示。图中，第一列为不同对比场景中相机采集的实际工件场景图，从左向右分别为 Mask-RCNN（Resnet50）、Mask-RCNN（Resnet101）、改进 Mask-RCNN、Yolact 等算法识别结果图。图 9（a）为无堆叠场景下识别对比结果，图 9（b）为轻微堆叠场景下识别对比结果，图 9（c）为复杂堆叠场景下识别对比结果。由图 9（a）识别结果可以看出，虽然所有对比算法均识别出了简单无堆叠场景下的工件图像，但是 Mask-RCNN（Resnet50）、Mask-RCNN（Resnet101）、

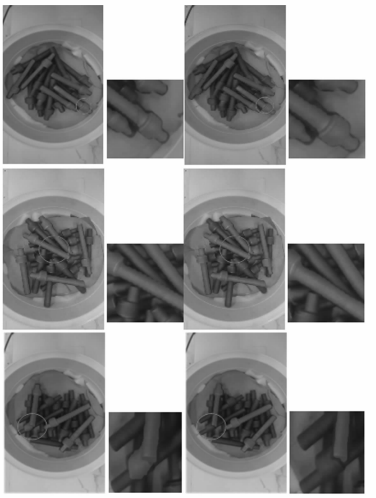

图 8　堆叠工件识别结果对比图

改进 Mask- RCNN 等对比算法识别结果不完整，均出现了识别结果部分缺失的问题，Yolact 算法虽然较完整的识别出了工件图像，但是存在识别结果边界不能完全贴合实际工件的图像，由此导致机器人在抓取过程中出现定位不准确问题。由图 9（b）识别结果可以看出，在轻微堆叠场景下，Mask-RCNN（Resnet50）、Mask-RCNN（Resnet101）、改进 Mask-RCNN 等对比算法识别结果仍然存在识别结果不完整情况，并出现了漏检问题；而 Yolact 算法和本文所提算法均完整识别出了所有可分拣工件图像，但是本文所提算法识别结果边缘更贴合实际工件图像。由图 9（c）识别结果可以看出，对于复杂堆叠场景下工件图像，Mask-RCNN（Resnet50）、Mask-RCNN（Resnet101）、改进 Mask-RCNN 等的识别结果出现了较多的漏检问题，甚至 Mask-RCNN（Resnet101）、改进 Mask- RCNN 的识别结果出现了误检问题，导致下方叠压工件被错误识别；Yolact 算法在复杂堆叠场景下也出现了误识别问题，导致不易抓取工件的误识别。本文所提算法均完整识别出了所有可分拣工件图像，并且识别结果边缘更贴合实际工件图像。因此，由大量实验结果可以看出，在非结构化场景下，本文提出算法较对比算法对工件图像中可分拣工件的识别更为准确，并且在掩码预测的完整度与边缘细节效果上均强于对比算法，不会出现对比算法造成的掩码缺失与边界不贴合工件等问题。

定量分析。为进一步定量分析所提算法性能，分别与 Yolact、改进掩码模板生成分支的 Yolact、引入膨胀编码的 Yolact、Mask-RCNN（Resnet50）、Mask-RCNN（Resnet101）、改进 Mask-RCNN、FCIS 等 7 个网络进行实验对比，对比结果如表 2 所示。由表 2 可知，改进掩码模板生成分支的 Yolact 网络与标准的 Yolact 网络相比其 Mask_MAP 提升了 1.22 个百分点，证明本文提出的多尺度融合与特征注意机制可以提高网络对目标物体的分割精度。引入膨胀编码的 Yolact 网络对比标准的 Yolact 网络其 B_box_MAP 提升了 1.74 个百分点，Mask_MAP 提升了 1.5 个百分点，证明以感受野更加丰富的特征图像进行预测会提高网络完成目标检测和语义分割两大任务的效果。多尺度特征注意 Yolact 网络在 B_box_MAP 与 Mask_MAP 上均达到最大值，证明本文提出的网络相对于标准的 Yolact 网络在边界框回归与掩码预测两个方面均有所提升，可以更好地引导视觉机器人系统进行工件分拣任务。

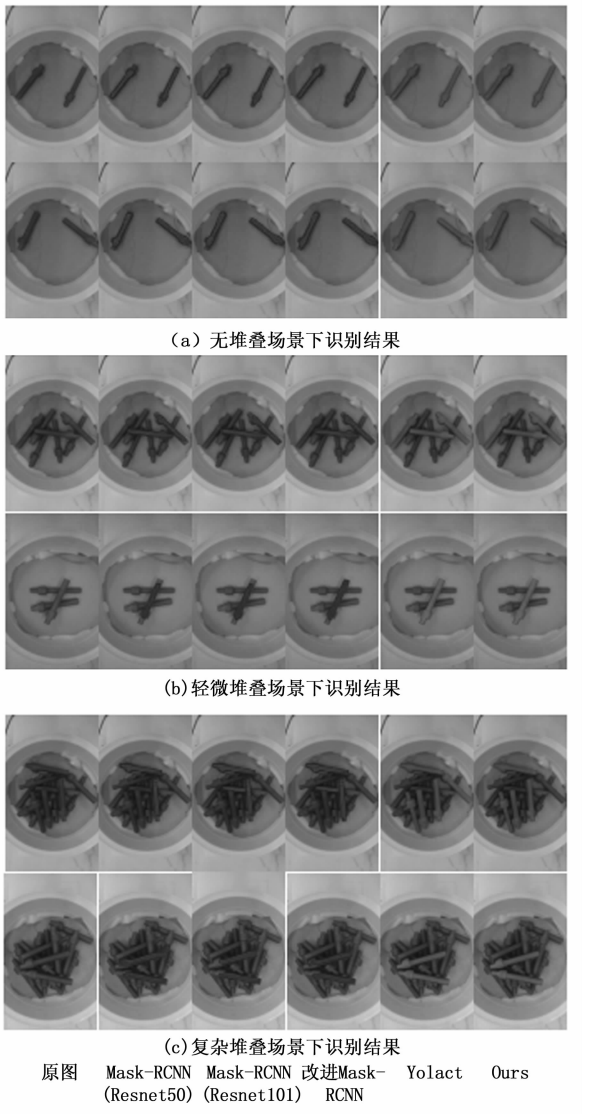

图 9　对比网络识别结果对比图

表 2　平均精度均值对比

网络	B_box_MAP/%	Mask_MAP/%
Yolact	79.65	80.10
Yolact(改进掩码模板生成分支)	80.67	81.32
Yolact(引入膨胀编码)	81.39	81.60
Mask-RCNN(Resnet50)	74.33	71.30
Mask-RCNN(Resnet101)	79.04	78.97
改进 Mask-RCNN(文献[19])	80.40	80.25
FCIS	77.14	76.87
Ours	**81.85**	**81.93**

模型大小与处理单张图片所用时间对比如表 3 所示。由表 3 可以得知本文提出的多尺度特征注意 Yolact 算法模型较标准 Yolact 算法模型在模型大小上增加了 2.1 M，处理单张图片所需时间增加了 0.14 s。此时长在满足实际工件分拣作业需求的可接受范围之内。本文所提出的算法网络在与标准 Yolact 算法处理单张图片所用时间相差无几的情况下有效提升了工件识别的准确率。

表 3 实时性对比

网络	模型大小/M	处理时间/s
Yolact	189.7	1.48
Yolact＋掩码模板生成分支	189.9	1.53
Yolact＋膨胀编码	191.6	1.60
Ours	**191.8**	**1.62**

为进一步验证所提网络的有效性，将所提网络部署在工业机器人系统中，并研发了一种堆叠工件分拣机器人。按照空调电机转轴实际生成场景构建测试环境，每次将 15 个空调电机转轴随机摆放在待抓取区域，由机器人连续抓取直至将所有空调电机转轴全部取出，并利用工件分拣机器人的作业情况分别进行了 900 次抓取对比测试实验。对比实验分别采用基于所提网络、标准 Yolact 算法的视觉工件分拣机器人作业情况进行对比，机器人根据对比算法得到的工件位置信息对工件进行抓取并放置于目标位置。本文以空调电机转轴分拣成功率作为性能评价指标，定义如下：

$$\text{分拣成功率} = \frac{\text{成功抓取次数}}{\text{总抓取次数}} \times 100\% \tag{22}$$

设计的实验分为 A 组与 B 组两组，在 A 组实验中使用标准的 Yolact 算法对工件进行识别，B 组实验使用所提的多尺度特征注意 Yolact 算法进行工件识别。两组实验中工件的位置信息与机器人的抓取位姿均由对应算法获得。实验对比结果如表 4 所示，基于本文所提网络的视觉机器人工件分拣系统完成堆叠工件分拣作业的成功率为 97.5%。与标准 Yolact 算法的视觉机器人工件分拣系统 95.8%的成功率相比提高了 1.7%。识别并输出每个电机转轴的机器人抓取位姿并由机器人进行抓取及放置的平均时间为 4.62 s。因此基于所提算法的视觉机器人工件分拣系统满足实际应用场景需求，能有效提高空调电机转轴自动化生产的智能化水平。

表 4 空调电机转轴抓取成功率对比

测试组别	实验次数	抓取成功次数	成功率%
A 组	900	863	95.8
B 组	900	878	97.5

5 结束语

针对非结构化场景中存在的多工件堆叠遮挡等问题，提出了基于多尺度特征注意 Yolact 网络的堆叠工件图像分割网络。所提网络基于多尺度特征融合器与注意力机制，有效提高了网络对于堆叠工件图像特征的学习能力；基于膨胀编码增强算法对于不同尺度大小目标的适应能力，提升了网络对堆叠工件定位回归的精度，最终研发了一种机器人工件分拣系统，并将该系统应用于实际空调电机转轴分拣任务中，实验结果证明了机器人工件分拣系统具有良好的性能。

参考文献：

[1] 伍锡如，黄国明，孙立宁．基于深度学习的工业分拣机器人快速视觉识别与定位算法 [J]．机器人，2016，38 (6)：711－719.

[2] 高 赢，张 莹，闫 璠，等．发动机瓦盖上料机器人研究 [J]．计算机工程与应用，2016，52 (20)：257－262.

[3] 高 赢．发动机瓦盖上料和检测系统研究 [D]．湘潭：湘潭大学，2016.

[4] 宋海涛，何文浩，原 魁．一种基于 SIFT 特征的机器人环境感知双目立体视觉系统 [J]．控制与决策，2019，34 (7)：1545－1552.

[5] 熊俊涛，邹湘军，彭红星，等．扰动柑橘采摘的实时识别与采摘点确定技术 [J]．农业机械学报，2014，45 (8)：38－43.

[6] CAI Z，VASCONCELOS N. Cascade r-cnn：Delving into high quality object detection [C] //Proceedings of the IEEE conference on computer vision and pattern recognition. 2018：6154－6162.

[7] GIRSHICK R. Fast r-cnn [C] //Proceedings of the IEEE international conference on computer vision. 2015：1440－1448.

[8] 彭红星，黄 博，邵园园，等．自然环境下多类水果采摘目标识别的通用改进 SSD 模型 [J]．农业工程学报，2018，34 (16)：155－162.

[9] TIAN Y，YANG G，WANG Z，et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model [J]. Computers and electronics in agriculture，2019，157：417－426.

[10] 王丹丹，何东健．基于 R-FCN 深度卷积神经网络的机器人疏果前苹果目标的识别 [J]．农业工程学报，2019，35 (3)：156－163.

[11] 武 星，齐泽宇，王龙军，等．基于轻量化 YOLOv3 卷积神经网络的苹果检测方法 [J]．农业机械学报，2020，51 (8)：17－25.

[12] 杨长辉，刘艳平，王 毅，等．自然环境下柑橘采摘机器人识别定位系统研究 [J]．农业机械学报，2019，50 (12)：14－22，72.

[13] 朱 江，杜 瑞，李建奇，等．基于注意力机制的曲轴瓦盖上料机器人视觉定位和检测方法 [J]．仪器仪表学报，2021，42 (5)：140－150.

[14] 王 欣．基于快速 SSD 深度学习算法的机器人抓取系统研究 [D]．武汉：武汉科技大学，2018.

[15] 杜学丹，蔡莹皓，鲁 涛，等．一种基于深度学习的机械臂抓取方法 [J]．机器人，2017，39 (6)：820－828，837.

(下转第 200 页)