第 60 卷 第 22 期/2023 年 11 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于语义对齐与图节点交互的实例分割算法

张敏，邓洋洋，李亚军，张苗辉*

河南大学人工智能学院，河南 郑州 450046

摘要 针对主流单阶段实例分割算法因冗余语义信息造成实例掩码缺失和泄漏的问题，提出一个基于语义对齐和图节点交互的实例分割算法。在全局掩码生成阶段，设计一个语义对齐模块，通过全局映射和高斯映射评估语义信息对全局和局部语义完整性的影响，从而对冗余语义信息进行抑制。此外，在实例掩码组装阶段，设计一个图节点交互模块。该模块通过对特征图进行图结构数据变换和图节点信息交互，提取拓扑图的空间特征，补充了掩码组装信息，进一步提高了实例掩码的准确度。实验结果表明，所提算法在MS COCO数据集上实现了38.3%的平均精度均值(mAP)，与其他先进算法相比，有很强的竞争力。

关键词 图像处理；实例分割；语义对齐；图节点交互；MS COCO数据集

中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/LOP231402

Instance Segmentation Algorithm Based on Semantic Alignment and Graph Node Interaction

Zhang Min, Deng Yangyang, Li Yajun, Zhang Miaohui*

School of Artificial Intelligence, Henan University, Zhengzhou 450046, *Henan, China*

Abstract In order to address the issues of missing and leaking instance masks caused by redundant semantic information in mainstream single-stage instance segmentation algorithms, this paper proposes an instance segmentation algorithm based on semantic alignment and graph node interaction. In the global mask generation stage, a semantic alignment module was designed to evaluate the influence of semantic information on global and local semantic integrity through global mapping and Gaussian mapping, thereby suppressing redundant semantic information. In addition, a graph node interaction module was designed in the instance mask assembly stage that extracts spatial features of the topological graph by transforming the feature map into graph-structured data and interacting with graph node information, supplementing the mask assembly information and further improving the accuracy of the instance masks. The experimental results demonstrate that the proposed algorithm achieves a mean average accuracy (mAP) of 38.3% on the MS COCO dataset, exhibiting strong competitiveness against other state-of-the-art algorithms.

Key words image processing; instance segmentation; semantic alignment; graph node interaction; MS COCO dataset

1 引 言

实例分割任务[1-2]是目标检测任务和语义分割任务的进阶任务。实例分割任务预测实例掩码时，既能够像目标检测任务[3-6]一样对图像中的同类别或者不同类别的实例目标进行分类和定位，也能够像语义分割任务[7-9]一样给出像素级别而不是锚框级别的实例区分。目前，实例分割任务广泛地应用于各个方向，例如农业果实定位、医学病灶分割(肿瘤和息肉)和汽车自动驾驶等方向。

随着实例分割任务的应用场景的扩大，单阶段实例分割不断受到研究工作者们的关注，并且涌现了大量优秀的工作。单阶段实例分割算法可以按照实例掩码的表征方式分为：基于轮廓表征的分割算法和基于像素表征的分割算法。基于轮廓表征的分割算法将实例分割任务视为实例目标的轮廓回归任务。PolarMask[10]通过使用与预测锚框顶点的FCOS[11]一样的方式，直接预测实例的轮廓顶点，实现对实例掩码的预测。Deep Snake[12]通过回归预定义的八边形初始轮廓，对初始轮廓进行迭代变形，提高了实例掩码的准

收稿日期：2023-05-30；**修回日期**：2023-06-11；**录用日期**：2023-06-27；**网络首发日期**：2023-07-07

通信作者：*zhmh@henu.edu.cn

确性。E2EC[13]具有可学习的初始化轮廓和初始化架构，进一步增强了算法的灵活度。虽然一系列的基于轮廓的算法取得了很大的进步，但是受限于回归和优化轮廓顶点数量，这种算法还是难以刻画出轮廓形状复杂与多连通区域形状的实例掩码，算法精度不高。

目前，主流的单阶段实例分割算法是基于像素表征的分割算法。YOLACT[14]将实例分割分为两个并行任务，通过并行地预测掩码系数、锚框和原型掩码然后组装的方式实现了实例掩码的分割。CondInst[15]在组装中消除了锚框的影响，并且将原型掩码的系数与原型掩码简单的线性组合替代为了利用动态卷积进行非线性变换的方式，提高了掩码的精度。BlendMask[16]和SipMask V2[17]通过在组装阶段引入实例掩码的空间信息，使分割的掩码更加具有细粒度。然而，这些研究都忽略了全局掩码生成阶段的冗余语义信息的不利影响。全局掩码生成阶段的冗余语义信息会造成生成的全局掩码不清晰，最终实例掩码会出现缺失和泄漏等问题。同时，从上述工作中发现仅仅通过通道信息实现实例掩码的组装是不足的，通过引入额外的信息（如空间信息）可以帮助算法生成更具细粒度的实例掩码。

针对主流的单阶段实例分割算法中全局掩码生成阶段存在冗余语义信息干扰造成实例掩码缺失和泄漏的问题，本文受自然语言翻译领域启发[18-19]，设计了一个语义对齐模块（SAM）。该模块通过全局映射组合局部高斯映射的方式获取语义信息对全局语义完整性和局部语义完整性的贡献，进而调整特征图语义分布，增加有效信息的区分度，抑制冗余语义信息，提高了实例掩码的分割精度。针对主流的单阶段实例分割算法中掩码组装可依据的信息不足的问题，本文设计了一个图节点交互模块（GNIM）。该模块利用先通过数据结构转换，再使用图卷积[20]进行节点信息交互的方式，从图结构数据中提取拓扑图的空间特征对掩码组装进行信息补充，进一步提高了实例掩码的生成质量。

2　方法内容

首先介绍所提算法的总体架构，然后分别具体介绍语义对齐模块和图节点交互模块的实现方式。

2.1　算法框架

如图1所示，所提算法分为实例特征编码器与实例特征解码器两个部分，来实现原始图像到实例分割图像的映射。实例特征编码器由特征提取网络和多尺度特征融合网络构成。实例特征解码器由预测分支网络、全局掩码生成分支网络和组装网络构成。

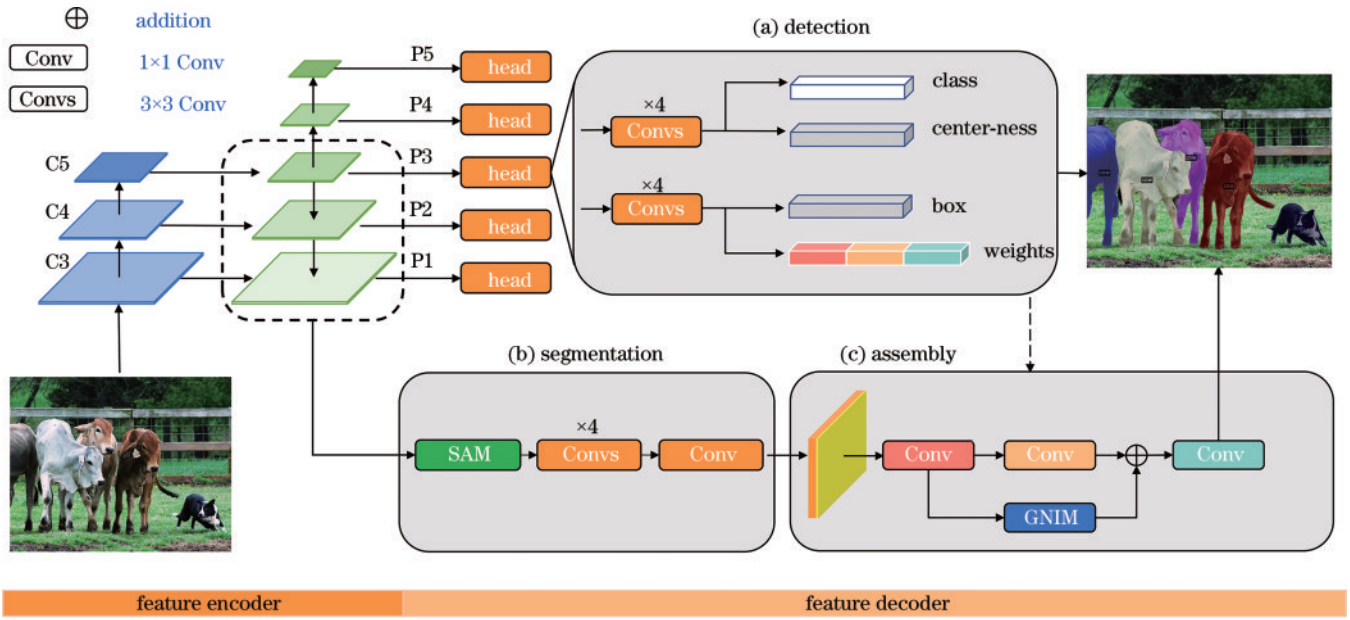

图1　所提算法的框架

Fig. 1　Framework of the proposed algorithm

实例分割任务是一个具有挑战性的任务，需要有高质量的图像特征。综合考虑特征的提取速度和提取特征的丰富性，特征编码器引入骨干网络ResNet[21]，从图片中提取高、中、低层次的语义特征的同时引入特征金字塔网络（FPN）[22]，增强特征的表达，提高算法处理不同尺度实例的性能。具体地，对于初始的图片 $\boldsymbol{I}\in\mathbb{R}^{H\times W\times 3}$，为了便于算法模型的训练优化和提高算法的推理速度，首先将基本图像的大小缩放至550×550，得到 $\boldsymbol{I}_0\in\mathbb{R}^{550\times 550\times 3}$。图片经骨干网络ResNet的提取，得到C1到C5五层语义特征图，从低级语义到高级语义排列。遵循RetinaNet[23]的设计原则，舍弃C1和C2低级语义特征图，将C3至C5送入FPN进行多尺度特征融合，得到{P1，P2，P3，P4，P5}五层特征图，获得编码特征，特征图的大小分别为原图的1/8、1/16、1/32、1/64、1/128，其中P4和P5是由P3通过卷积直接得到的。

特征解码框架基本遵循了 YOLACT 的特征解码方式，由一条预测分支网络、一条语义分支网络和组装网络构成。经主干网络提取的特征图并行地由预测分支和语义分支预测实例掩码的类别、动态卷积权重和全局掩码，再通过组装网络结合两条支路的信息将图像中的实例分割出来。

预测分支网络如图 1(a)所示，本文的预测分支基本遵循了 FCOS 的预测方式。预测分支包含 4 个预测子分支：类别(class)预测、中心度(center-ness)预测、锚框(box)预测和卷积核权重(weights)预测。每个子分支都会先通过 4 个 3×3 卷积、批归一化和激活的组合模块，然后再通过 1 个卷积核大小为 3、步距为 1 的卷积层，得到最终的预测结果。下一步进行卷积核权重预测。为了向条件卷积提供区分不同实例所需的权重，在锚框预测分支上并行地添加了权重预测头，权重预测分支与锚框预测分支共享参数，只在最后一个预测卷积不共享参数。权重参数维度由条件卷积的权重设置和输入输出通道决定，具体计算公式为

$$D_{\mathrm{dim}}=(10\times 8)w_1+8b_1+(8\times 8)w_2+8b_2+8w_3, \tag{1}$$

式中：D_{dim} 表示权重预测分支的输出维度；w 表示卷积权重；b 表示卷积偏置。

全局掩码分支网络如图 1(b)所示。选取 P1、P2、P3 作为输入特征图，使用 1×1 的卷积将特征图的通道都统一到 128，对 P2 和 P3 进行双线性插值上采样操作，使 P2、P3 的大小和 P1 对齐并且进行相加特征融合。然后将融合特征输入语义对齐模块，得到精炼过的特征图，最后使用 4 个 3×3 卷积组成的全卷积网络对特征图进行解码，得到全局掩码。

组装网络如图 1(c)所示。使用全局掩码、条件卷积和图节点交互模块结合的方式来生成实例掩码，条件卷积的卷积核参数由卷积核参数分支给出。具体操作如下：用 1×1 的卷积对 128 维的全局掩码进行降维，以便后续条件卷积的筛选组合，过多的通道信息会增加网络的优化难度，并且会提高网络的运算量。将实例目标中心点为原点的相对位置坐标图附加到降维后的特征图上，对特征图进行空间信息的编码，结合第一个 1×1 的条件卷积将粗糙的实例掩码特征图解码出来。之后对粗糙的实例掩码通过并行的网络进行掩码信息筛选，一条分支是条件卷积，另一条是节点交互模块，再进行相加操作融合获得综合考虑语义信息和图拓扑信息的精炼实例掩码特征图，最后通过一个条件卷积整合特征，进行 Sigmoid 激活操作，获得实例掩码。

2.2　语义对齐模块

语义对齐模块的详细结构如图 2 所示。由 P1、P2、P3 特征融合的特征图 $\boldsymbol{F}\in\mathbb{R}^{C\times H\times W}$ 经过语义对齐模块的计算，得到了一个 3D 的通道信息分布重构矩阵 $D(\boldsymbol{F})\in\mathbb{R}^{C\times H\times W}$。通道信息重构后的特征图 $\boldsymbol{F}'$ 表示为

$$\boldsymbol{F}'=\boldsymbol{F}+\boldsymbol{F}\otimes D(\boldsymbol{F}), \tag{2}$$

式中：$\otimes$ 表示逐元素乘法。引入残差机制来平滑重构通道特征，方便网络进行梯度优化。为了使语义对齐模块能够兼顾全局语义信息的完整性和局部语义信息的完整性，由全局通道信息分布重构矩阵 $D_{\mathrm{g}}(\boldsymbol{F})\in\mathbb{R}^{C\times 1\times 1}$ 和局部通道信息分布重构矩阵 $D_{\mathrm{l}}(\boldsymbol{F})\in\mathbb{R}^{C\times 1\times 1}$ 组成通道信息分布重构矩阵 $D(\boldsymbol{F})$，具体的计算公式为

$$D(\boldsymbol{F})=\delta\left[D_{\mathrm{g}}(\boldsymbol{F})+D_{\mathrm{l}}(\boldsymbol{F})\right], \tag{3}$$

式中：δ 为 Softmax 激活函数。经过激活函数之后，$D(\boldsymbol{F})$ 的维度由 $\mathbb{R}^{C\times 1\times 1}$ 扩展到 $\mathbb{R}^{C\times H\times W}$。

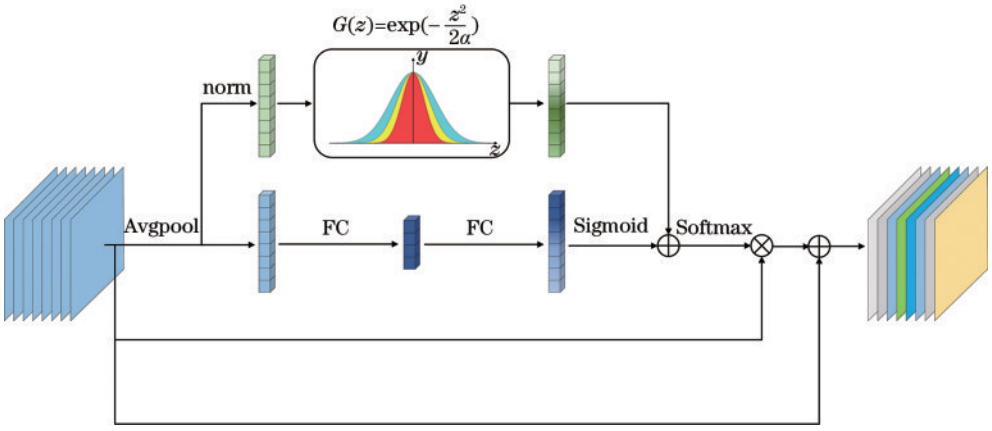

图 2　语义对齐模块

Fig. 2　Semantic alignment module

语义信息压缩。由于二维图像的语义表征形式复杂，庞大的参数量不利于神经网络对输入特征图的语义特征进行学习和调整优化。为了解决这个问题，使用全局平均池化对输入特征图的语义信息进行压缩表征，压缩后的语义特征为 $\boldsymbol{F}_{\mathrm{sq}}\in\mathbb{R}^{C\times 1\times 1}$，表达式为

$$\boldsymbol{F}_{\mathrm{sq}}=\mathrm{Avgpool}(\boldsymbol{F})=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}F(i,j), \tag{4}$$

式中：$F(i,j)$ 表示特征图像素点 (i,j) 上的值。

全局通道信息分布重构。采用与 SENet[24] 类似的机制，为每层压缩后的语义特征引入权重 w，通过梯度

更新建立全局通道信息分布重构矩阵与语义信息之间的映射关系，具体公式为

$$D_g(\boldsymbol{F})=\delta\left\{\Phi\left\{\mathrm{ReLU}\left\{\mathrm{BN}\left[\Psi(\boldsymbol{F}_{sq})\right]\right\}\right\}\right\},\tag{5}$$

式中：$\Phi(\cdot)$为1×1卷积（输入通道为64，输出通道为128，偏置项为False）；$\Psi(\cdot)$为1×1卷积（输入通道为64，输出通道为128，偏置项为False）；BN为批归一化层；ReLU为激活函数。

局部通道信息分布重构。受自然语言处理（NLP）领域的启发，即局部语义信息的完整性分布符合高斯分布，所以使用高斯函数作为映射函数来构建输入的语义信息与局部通道信息分布重构矩阵之间的关系。具体来说，假设构建高斯分布函数的函数方程为

$$G(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],\tag{6}$$

式中：μ为总体均值；σ^2为总体方差。因为输入特征图包含的语义信息不同，为了增强映射关系的鲁棒性，对输入的特征信息进行了归一化处理。定义了一个新的变量来适应输入特征图的变化：

$$z=\mathrm{norm}(x)=x-\mu/\sigma,\tag{7}$$

将式（7）代入式（6），可以得到简化后的高斯分布函数：

$$G(z)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right),\tag{8}$$

为了使映射关系更加灵活，引入变量$\alpha\in[1,4]$增加映射的灵活性，同时对高斯分布函数的幅值进行放缩，变换后的高斯分布函数可以表示为

$$G(z)=\exp\left(-\frac{z^2}{2\alpha}\right),\tag{9}$$

综上所述，局部通道信息分布重构矩阵与语义信息之间的映射关系为

$$D_l(\boldsymbol{F})=\exp\left\{-\frac{\mathrm{norm}\left[\mathrm{Avgpool}(\boldsymbol{F})\right]^2}{2\alpha}\right\}。\tag{10}$$

2.3　图节点交互模块

如图3所示，图节点交互模块将分为图结构转换和图节点交互两个部分来实现。

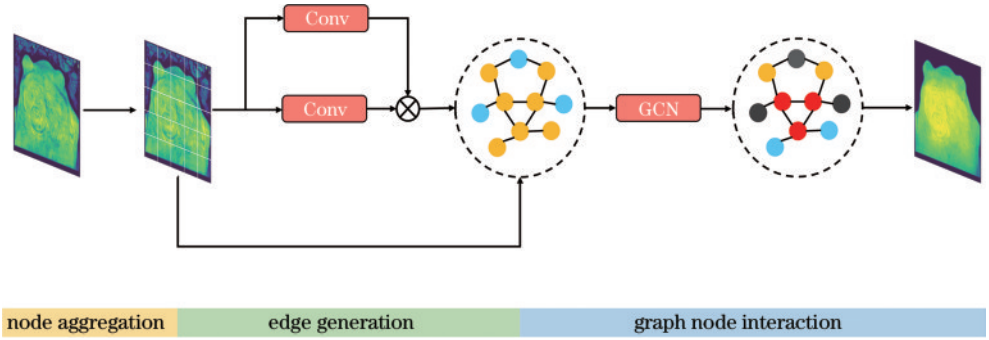

图3　图节点交互模块

Fig. 3　Graph node interaction module

图结构转换。要实现图节点信息的交互，首先要将图像模型构建为图结构的模型，图模型通常表示为$g=\{N,V\}$，N表示图中节点的点集$\{n_1,n_2,\cdots,n_k\}$，V表示图中节点之间的边集$\{v_1,v_2,\cdots,v_k\}$。图节点的点集表示为

$$\boldsymbol{X}_{node}=\mathrm{downsample}(\boldsymbol{X}),\tag{11}$$

式中：$\boldsymbol{X}$为输出图像特征；$\boldsymbol{X}_{node}$表示节点特征；downsample(•)为8倍下采样。

其次，构建图节点的边集V，使用节点的邻接矩阵$\boldsymbol{A}$来描述边集V，邻接矩阵$\boldsymbol{A}$由节点与节点间的相似性决定。具体而言，对图节点的特征$\boldsymbol{X}_{node}$进行特征映射，生成新的节点特征向量$\boldsymbol{X}'_{node}$，再使用相似性度量函数计算节点与节点之间的边的关系，选用点积作为相似性度量函数，计算表达式为

$$\begin{cases}A_{i,j}=f(\boldsymbol{x}_i,\boldsymbol{x}_j)=\theta(\boldsymbol{x}_i)^{\mathrm{T}}\cdot\varphi(\boldsymbol{x}_j)\\ \theta(\boldsymbol{x})=\mathrm{ReLU}\left\{\mathrm{BN}\left[\mathrm{Conv}(\boldsymbol{x})\right]\right\}\\ \varphi(\boldsymbol{x})=\mathrm{ReLU}\left\{\mathrm{BN}\left[\mathrm{Conv}(\boldsymbol{x}\right]\right\}\end{cases}。\tag{12}$$

图节点交互。完成图模型转换之后，使用图卷积即可使图像中的节点得到信息交互，具体计算公式为

$$\boldsymbol{X}_f=\mathrm{upsample}\left[\mathrm{GCN}(\boldsymbol{X}_{node})\right]=\mathrm{upsample}(\boldsymbol{A}\boldsymbol{X}_{node}\boldsymbol{W}),\tag{13}$$

式中：$\boldsymbol{X}_f$为拓扑图的空间特征；upsample(•)为双线性插值操作；GCN(•)为图卷积操作。

3　实验结果与分析

3.1　实验设置

在实例分割领域，开源数据集MS COCO[25]由于具有物体类别丰富、数据量适中、能全面地反映相关现实场景的优点，常用来验证新算法的有效性。该数据集覆盖了绝大多数常见事物，如人、动物、交通工具、生活用品等80类物体。MS COCO2017数据集由训练集、验证集和测试集3个部分组成，它们的数量分别为118287、5000和40670。其中，测试集COCO2017 test被分成两个大致相同大小的split约为20000张的图像：test-dev和test-challenge。通常在test-dev中报告

论文的结果，test-challenge用于每年的COCO挑战。

在实例分割方向，研究者们通常使用平均精度均值（mAP）来对实例分割算法生成的掩码质量进行评估，本文使用AP来评估。除此以外，AP_{50}和AP_{75}分别表示在相应阈值下的分割精度，AP_S、AP_M、AP_L分别表示小、中、大物体的分割精度。其中，目标对象面积小于32×32像素的为小目标，面积大于96×96像素的归为大目标，介于两者之间的为中等目标。

使用NVIDIA RTX A5000作为训练和验证所提算法的硬件基础。在使用MS COCO数据集对所提算法进行训练时，采用随机梯度下降（SGD）算法优化算法模型，其中权重衰减（weight decay）参数设置为0.0001，动量（momentum）参数设置为0.9。初始的学习率设置为0.01，并分别在第21万次和第25万次迭代时除以10，总的迭代次数为27万次。每次迭代过程使用的训练批次（batch）大小为16，主干网络中的参数通过在ImageNet[19]上预训练算法模型得到权重初始化。实验采用图片翻转的操作对输入图像进行数据增强。其中，随机地从[640，800]区间内对图像的短边长度进行采样，对长边进行等比例缩放；当缩放后的长边超过1333时，对其进行截断，以控制最大输入尺寸。在测试期间，不会有数据增强的操作，并且会将输入图像的短边严格限制在800的大小，每一个批次都只会有一张图片通过。所提算法的训练是在4张A5000上实现的，验证和测试是在单张A5000上实现的。

3.2 实验结果

对所提算法与当前先进的算法进行比较，如表1所示，实验结果均是使用COCO test-dev作为测试集得出的。在以ResNet-50为主干的情况下，与Mask R-CNN相比，所提算法的AP高0.8个百分点，实现2%的增益；与算法BlendMask、SipMask V2、CondInst和SparseInst相比，所提算法的AP提高了3.5%、15.1%、1.3%和6.1%。结果表明，所提算法与当前的先进算法相比有相当强的竞争力。值得注意的是，所提算法在小目标分割上的效果比较好，与YOLACT++相比，所提算法的AP_S提高了75.2%。

表1 算法对比实验

Table 1 Algorithm comparison experiment

unit: %

Method	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Mask R-CNN[26]	R-50-FPN	37.5	59.3	40.2	21.1	39.6	48.3
YOLACT++[27]	R-50-FPN	34.1	53.3	36.2	11.7	36.1	53.6
BlendMask[16]	R-50-FPN	37.0	58.9	39.7	17.3	39.4	52.5
SipMask V2[17]	R-50-FPN	33.1	52.5	35.2	12.3	35.5	50.7
CondInst[15]	R-50-FPN	37.7	58.9	40.3	20.4	40.2	48.9
SparseInst[28]	R-50-d	36.1	57.0	38.2	15.0	37.7	53.1
OrienMask[29]	D-53-FPN	34.8	56.7	36.4	16.0	38.2	47.8
SAGinst(ours)	R-50-FPN	38.3	59.7	41.1	20.5	40.6	51.7

图4是所提算法在COCO val数据集上的可视化结果，实验结果表明所提算法能够较好地区分同类实例和不同类实例，如图4中第1行第2列的“大象”和第2行第4列的“小孩与皮球”。此外，图4的可视化结果也证明了所提算法对小目标分割的优越性能，如第3行第2列“飞机”图片中的“人”也能被分割出来。

3.3 消融实验

3.3.1 模块有效性消融实验

评估语义对齐模块和图节点交互模块对基准算法的贡献，结果如表2所示。语义对齐模块对输入特征图中全局语义完整性和局部语义完整性的贡献进行筛选，通过抑制冗余特征、保留有效特征，提高了全局掩码的解码正确率，降低了解码难度。相比基准算法，添加语义对齐模块（SAM）后的算法的AP提高了0.3个百分点，验证了语义对齐模块的有效性，此外，添加图节点交互模块（GNIM）后的算法的AP提高了0.2个百分点，验证了图节点交互模块的有效性。最后，在两个模块的共同作用下，基准算法的各个指标都有了提高，最终模型的AP提高了0.5个百分点，获得了1.4%的绝对增益，验证了两个模块联合作用的有效性。

3.3.2 语义对齐模块分支消融

通过对语义对齐模块的映射分支进行控制，对比分析了语义对齐模块的两个映射分支对全局掩码所造成的影响，如表3所示。实验结果表明：仅仅使用全局映射对语义特征图中全局语义完整性的贡献进行特征筛选，对全局掩码的生成是有积极影响的，AP值比基准算法提高了0.5%；仅使用局部映射对语义特征图中局部语义完整性的贡献进行特征筛选，对全局掩码的生成是有一定消极影响的，特别是在AP_{50}和AP_L的表现上。然而，若将局部映射作为全局映射的补充，综合考虑特征图对全局完整性和局部完整性的影响，对特征进行筛选将会得到更加积极的影响，AP值比基准算法高了0.8%。

3.3.3 图节点交互模块节点聚合度消融

对图节点的聚合度进行改变，对比分析了图节点聚合度的改变对图节点交互模块性能的影响，如表4

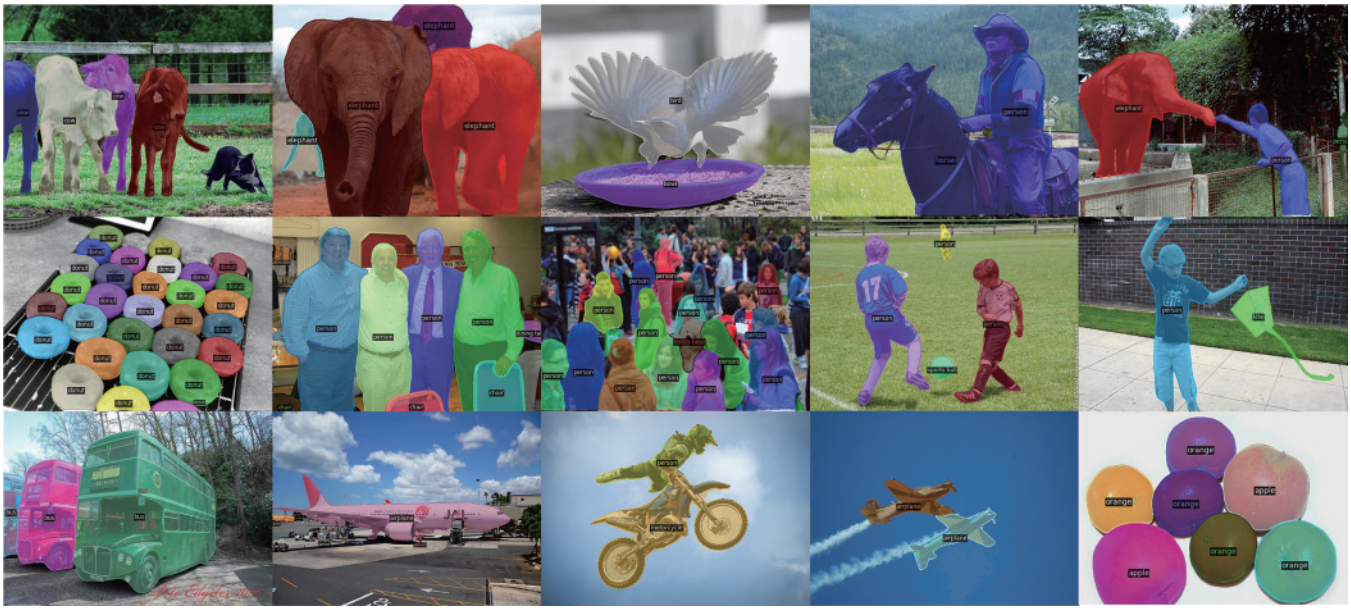

图 4　算法可视化结果

Fig. 4　Algorithm visualization result

表 2　模块有效性消融实验

Table 2　Module validity ablation experiment　unit: %

Baseline	SAM	GNIM	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
✓			36.2	57.3	38.5	17.0	39.6	51.4
✓	✓		36.5	57.6	38.9	17.5	40.0	51.6
✓		✓	36.4	57.3	38.7	17.5	40.1	51.4
✓	✓	✓	36.7	57.7	39.2	17.9	40.5	52.4

表 3　语义对齐模块分支消融实验

Table 3　Semantic alignment module branch ablation experiment　unit: %

Baseline	Global	Local	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
✓			36.2	57.3	38.5	17.0	39.6	51.4
✓	✓		36.4	57.2	38.8	17.4	39.8	51.6
✓		✓	36.1	56.9	38.5	17.4	39.6	51.2
✓	✓	✓	36.5	57.6	38.9	17.5	40.0	51.6

表 4　图节点交互模块节点聚合度消融实验

Table 4　Node aggregation ratio ablation experiment of graph node interaction module　unit: %

Ratio	FLOPs/10^6	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
4×	880	36.4	57.3	38.7	17.3	40.0	51.6
8×	101	36.4	57.3	38.7	17.5	40.1	51.4
16×	46	36.3	57.0	38.7	17.6	40.0	51.3

所示。实验结果表明，当节点聚合度从 4 倍下采样（4×）变化至 16 倍下采样（16×）时，模块的计算量（FLOPs）得到了显著的降低，但是模块性能（AP）有所下降。综合考虑模块的计算复杂度和模块的性能表现，选用 8 倍下采样对图节点特征进行聚合。

3.3.4　图节点交互模块深度消融

对图节点交互模块的深度在掩码组装阶段造成的影响进行了评估，如表 5 所示。实验改变了图节点交互模块的深度，使图节点交互模块的深度从 1 增加到 3。实验结果表明，随着图节点交互深度的增加，虽然 AP_S 和 AP_L 等个别指标有所提升，但是总的掩码精度 AP 在不断下降。图节点交互模块深度从 1 增加到 3 时，会给算法带来 0.8% 的精度下降。

表 5　图节点交互模块深度消融实验

Table 5　Depth ablation experiment of graph node interaction module　unit: %

Depth	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
1	36.4	57.3	38.7	17.5	40.1	51.4
2	36.3	57.1	38.5	17.6	39.8	51.5
3	36.1	57.1	38.5	17.0	39.7	51.7

4　结　　论

提出了一种基于语义对齐和图节点交互的实例分割算法，旨在抑制主流单阶段算法中存在的冗余语义信息，改进实例掩码缺失和泄漏的问题。设计了一个语义对齐模块，通过评估语义信息对全局和局部语义完整性的影响，对特征进行筛选，抑制冗余语义特征。为了进一步提高实例掩码的准确性，还设计了一个图节点交互模块。该模块在实例掩码组装阶段引入了拓扑图的空间特征信息，对掩码的组装信息进行了丰富，提高了掩码的分割质量。所提算法在 MS COCO 数据集上的定量和定性结果表明，所提算法具有有效性和先进性，在解决实例掩码缺失和泄漏问题方面具有潜力。

参 考 文 献

[1] 张绪义, 曹家乐. 基于轮廓点掩模细化的单阶段实例分割网络[J]. 光学学报, 2020, 40(21): 2115001.
Zhang X Y, Cao J L. Contour-point refined mask prediction for single-stage instance segmentation[J]. Acta Optica Sinica, 2020, 40(21): 2115001.

[2] 周苏, 吴迪, 金杰. 基于卷积神经网络的车道线实例分割算法[J]. 激光与光电子学进展, 2021, 58(8): 0815007.
Zhou S, Wu D, Jin J. Lane instance segmentation algorithm based on convolutional neural network[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0815007.

[3] 王友伟, 郭颖, 邵香迎. 基于改进级联算法的遥感图像目标检测[J]. 光学学报, 2022, 42(24): 2428004.
Wang Y W, Guo Y, Shao X Y. Target detection in remote sensing images based on improved cascade algorithm[J]. Acta Optica Sinica, 2022, 42(24): 2428004.

[4] 吴萌萌, 张泽斌, 宋尧哲, 等. 基于自适应特征增强的小目标检测网络[J]. 激光与光电子学进展, 2023, 60(6): 0610004.
Wu M M, Zhang Z B, Song Y Z, et al. Small-target detection network based on adaptive feature enhancement [J]. Laser & Optoelectronics Progress, 2023, 60(6): 0610004.

[5] 周迎峰, 张荣芬, 刘宇红, 等. 基于RetinaNet的海洋鱼类检测算法[J]. 激光与光电子学进展, 2023, 60(10): 1010014.
Zhou Y F, Zhang R F, Liu Y H, et al. Marine fish detection algorithm based on RetinaNet[J]. Laser & Optoelectronics Progress, 2023, 60(10): 1010014.

[6] Lin J P, Haberstroh F, Karsch S, et al. Applications of object detection networks in high-power laser systems and experiments[J]. High Power Laser Science and Engineering, 2023, 11(1): e7.

[7] 易清明, 张文婷, 石敏, 等. 多尺度特征融合的道路场景语义分割[J]. 激光与光电子学进展, 2023, 60(12): 1210006.
Yi Q M, Zhang W T, Shi M, et al. Semantic segmentation for road scene based on multiscale feature fusion[J]. Laser & Optoelectronics Progress, 2023, 60 (12): 1210006.

[8] 陈兵, 贺晟, 刘坚, 等. 基于轻量化DeepLab v3+网络的焊缝结构光图像分割[J]. 中国激光, 2023, 50(8): 0802105.
Chen B, He S, Liu J, et al. Weld structured light image segmentation based on lightweight DeepLab v3+ network [J]. Chinese Journal of Lasers, 2023, 50(8): 0802105.

[9] 单成响, 李锵, 关欣. 多视图卷积轻量级脑肿瘤分割算法[J]. 激光与光电子学进展, 2023, 60(10): 1010018.
Shan C X, Li Q, Guan X. Multi-view convolution lightweight brain tumor segmentation algorithm[J]. Laser & Optoelectronics Progress, 2023, 60(10): 1010018.

[10] Xie E Z, Sun P Z, Song X G, et al. PolarMask: single shot instance segmentation with polar representation[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 12190-12199.

[11] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2020: 9626-9635.

[12] Peng S D, Jiang W, Pi H J, et al. Deep snake for real-time instance segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 8530-8539.

[13] Zhang T, Wei S Q, Ji S P. E2EC: an end-to-end contour-based method for high-quality high-speed instance segmentation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 4433-4442.

[14] Bolya D, Zhou C, Xiao F Y, et al. YOLACT: real-time instance segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2020: 9156-9165.

[15] Tian Z, Zhang B W, Chen H, et al. Instance and panoptic segmentation using conditional convolutions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 669-680.

[16] Chen H, Sun K Y, Tian Z, et al. BlendMask: top-down meets bottom-up for instance segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 8570-8578.

[17] Cao J L, Pang Y W, Anwer R M, et al. SipMaskv2: enhanced fast image and video instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(3): 3798-3812.

[18] Yang B S, Tu Z P, Wong D F, et al. Modeling localness for self-attention networks[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, October 31-November 4, 2018, Brussels, Belgium. Stroudsburg: Association for Computational Linguistics, 2018: 4449-4458.

[19] Galassi A, Lippi M, Torroni P. Attention in natural language processing[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(10): 4291-4308.

[20] Ye Y, Ji S H. Sparse graph attention networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(1): 905-916.

[21] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[22] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.

[23] Lin T Y, Goyal P, Girshick R, et al. Focal loss for

dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.

[24] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.

[25] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8693: 740-755.

[26] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.

[27] Bolya D, Zhou C, Xiao F Y, et al. YOLACT++ better real-time instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(2): 1108-1121.

[28] Cheng T H, Wang X G, Chen S Y, et al. Sparse instance activation for real-time instance segmentation [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 4423-4432.

[29] Du W T, Xiang Z Y, Chen S Y, et al. Real-time instance segmentation with discriminative orientation maps[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 7294-7303.