

DOI: 10.13382/j.jemi.B2306833

面向安检 X 光图像的违禁物品语义分割与识别研究^{*}

李广睿¹ 刘 琼^{1,2} 张熠卿¹ 张馨瑶¹ 黄景煦³ 傅 健^{2,4,5}
(1. 北京信息科技大学自动化学院 北京 100192; 2. 北京航空航天大学江西研究院 南昌 330200;
3. 西安电子科技大学人工智能学院 西安 710071; 4. 北京航空航天大学机械工程及自动化学院 北京 100083;
5. 北京航空航天大学宁波创新研究院 宁波 315800)

摘 要:针对安检 X 光图像中违禁物品大小不一、物品摆放随意且存在重叠遮挡的技术难题,提出了一种改进的 HRNet 多尺度特征融合网络模型,实现图像中违禁物品的自动分割与识别。在编码阶段,利用 HRNet 网络中的多分辨率并行网络架构,提取多尺度特征,解决安检 X 光图像违禁物品尺度多样化的问题。在解码阶段,提出一种多层级特征聚合模块,采用数据相关上采样方法减少信息丢失,并聚合编码阶段提取的特征,以对物品进行更完整表征。在网络整体架构中,嵌入基于注意力机制的去遮挡模块加强模型的边缘感知能力,缓解安检 X 光图像中物品重叠遮挡严重的问题,提高模型的分割识别精度。通过在 PIDray 安检图像公开数据集进行实验,结果表明,在 Easy、Hard、Hidden 3 个验证子集上分别取得了 73.15%、69.47%、58.33% 的平均交并比,相比原始 HRNet 模型,分别提升了 0.49%、1.17%、5.69%,总体平均交并比提升约 2.45%。

关键词: 安检 X 光图像;语义分割;违禁品识别;深度学习

中图分类号: TP391;TN249 **文献标识码:** A **国家标准学科分类代码:** 520.2040

Semantic segmentation and recognition of contraband for security X-ray images

Li Guangrui¹ Liu Qiong^{1,2} Zhang Yiqing¹ Zhang Xinyao¹ Huang Jingxu³ Fu Jian^{2,4,5}
(1. School of Automation, Beijing Information Science and Technology University, Beijing 100192, China; 2. Jiangxi Research Institute of Beihang University, Nanchang 330200, China; 3. School of Artificial Intelligence, Xidian University, Xi'an 710071, China; 4. School of Mechanical Engineering and Automation, Beihang University, Beijing 100083, China; 5. Ningbo Innovation Research Institute of Beihang University, Ningbo 315800, China)

Abstract: In response to the technical challenges posed by the varying sizes, haphazard arrangement, and overlapping occlusion of prohibited items in security X-ray images, we propose an enhanced HRNet-based multi-scale feature fusion network model. This model aims to achieve automatic segmentation and recognition of prohibited items in images. In the encoding stage, we leverage the multi-resolution parallel network architecture of HRNet to extract multi-scale features, addressing the diverse scale of prohibited items in security X-ray images. In the decoding stage, a multi-level feature aggregation module is introduced that uses data-dependent upsampling instead of bilinear interpolation. upsampling to reduce information loss during aggregation, thus ensuring a more comprehensive representation of the features of the features extracted in the coding stage for a more complete characterisation of objects. In the overall architecture of the network, a de-obscuration module based on the attention mechanism is embedded to strengthen the edge-awareness ability of the model, alleviate the problem of serious overlapping occlusion of items in security X-ray images, and improve the segmentation and recognition accuracy of the model. By experimenting on the public dataset of PIDray security check images, the results show that the average intersection ratio of 73.15%, 69.47%, and 58.33% are achieved in the three validation subsets of Easy, Hard, and Hidden, respectively, which are 0.49%, 1.17%, and 5.69%, respectively, and the overall average intersection ratio is improved

收稿日期: 2023-08-18 Received Date: 2023-08-18

^{*} 基金项目: 国家自然科学基金(62302051)、国家自然科学基金面上项目(51975026)、大科学装置联合基金(U1932111)、国家重点研发计划项目(2022YFF0607400)资助

by about 2.45%.

Keywords: security X-ray images; semantic segmentation; contraband identification; deep learning

0 引言

随着交通运输业、快递物流业的蓬勃发展,极大地促进和推动了国内经济总量的快速增长。然而,对运输过程中安全隐患的控制和把握也提出了更高的要求。当前,安检工作主要依靠肉眼识别和判读进行决策,需要工作人员在长时间内保持高度集中的精神状态,极易产生人为误检和漏检的情况^[1]。因此,用先进的人工智能技术和方法辅助安检人员进行违禁物品识别,将具有重要的研究意义。

X光图像比普通图像的信息较少^[2],且安检X光图像物品尺寸多样化、物品位置摆放随意且存在重叠遮挡。近年来,随着机器视觉技术和深度学习方法在自然目标检测、识别等垂直领域取得突破性进展^[3],面向安检X光图像的违禁物品自动识别方法也从基于传统手工特征和传统分类器相结合的方式转换到基于深度学习的端到端方法。

Turcsany等^[4]采用词袋模型与SVM分类器相结合,实现了对X光图像中的枪械的分类识别。而后,Bastan等^[5]利用双能成像技术获取更多的材料信息和多视角成像技术获取更多的空间信息,进一步提高了违禁品检测的精度。Mery等^[6]通过在多个视图之间匹配信息来过滤仅由单视图检测产生的误报,提升了复杂图像上违禁物品的识别精度。。

然而相比之下,基于深度学习的算法能够自动学习和提取安检图像中的多层语义信息,能够更加完整地表达图像内容,提升识别准确率。Akay等^[7]最先探索了Faster R-CNN^[8]、YOLOv2^[9]等深度学习检测网络在安检X光图像中的适用性,发现它们在性能上超越了传统的基于滑动窗口的检测方法。Subramani等^[10]通过迁移学习,验证目标检测网络SSD^[11]和RetinaNet^[12]在检测刀具、剪刀、扳手和钳4类违禁物品上的性能。此外,针对不同厂商和设备之间图像域存在差异的问题,何彦辉等^[13]提出了基于上下文的透射率自适应域对齐方法,提高了当测试域数据与训练域数据存在区别时的算法性能。林春晖^[14]为了提升对重叠违禁物品的检测效果,对非极大值抑制方法进行了改进,通过比较宽高比例和面积差异来更好的筛选边框,以减少漏检情况。然而,以上这些检测算法仅能对安检X光图像违禁物品进行初步分类和粗略定位,无法获取违禁物品的形状信息。2014年,全卷积神经网络^[15](fully convolutional networks, FCN)的提出,开启了基于深度学习的语义分割方法研

究。其后,语义分割模型,如U-Net^[16]、SegNet^[17]、PSPNet^[18]以及DeepLab系列^[19-22]等高性能算法逐渐应用到各个领域。在安检违禁物品识别领域,语义分割算法无疑可以帮助人们获取更多的有用信息,因此,一些学者和专家做出了探索和研究。Xu等^[23]针对安检违禁物品识别问题,将传统图像分割技术与注意力机制相结合,提出了一种违禁物品识别算法,实现对违禁物品类别、位置以及形状3种信息的提取。但是该方法缺乏对真实世界的数据集的验证,且违禁物品种类较少,不能充分验证模型性能。An等^[24]首次将基于深度学习的语义分割技术引入安检违禁品识别领域,在DeepLabv3+网络的基础上引入注意力机制,提出一种高精度的分割模型用于提取更丰富的X光图像特征,成为了将语义分割应用于检任务的开拓者。然而,该方法没有同时结合通道注意力和空间注意力模型性能提升有限较为有限。邓观林等^[25]使用金字塔池化层加强特征表征,利用级联卷积和注意力机制加强特征融合,并引入条件随机场进行二阶段处理,进一步提升了分割精度。然而,第一阶段的分割加上第二阶段的后处理使得模型效率较低。总体而言,目前基分割技术的安检违禁品识别研究还是相对较少,能够对单目标,多目标,以及遮挡情况统筹研究的学者更是缺乏,因此,本文基于包含47 677张图像的来自真实世界的安检数据集展开研究,在改进并联的高分辨率网络的基础上提出一种新的适用于安检违禁物品识别的语义分割算法。高分辨率网络(high-resolution network, HRNet)^[26]通过保持图像的高分辨率,同时下采样得到并行分支以提取不同分辨率的特征,从而捕捉不同尺度的细节和上下文信息,实现了高准确率的图像语义分割。然而,在解码器部分,HRNet仅使用简单的双线性插值上采样与高分辨率特征图进行融合,未能充分融合多尺度特征信息,难以达到较高的分割精度。针对以上问题,本文在HRNet^[27]语义分割算法的基础上对其优化改进,使其更加适用于安检X光图像语义分割任务。

本文的主要贡献总结如下:

1) 针对安检X光图像存在小目标违禁物品和背景复杂的难题,提出了多层级特征聚合模块,采用逐级数据相关上采样方法,增强多尺度特征的聚合效果和细节保持能力,提升违禁物品分割精度。

2) 嵌入一种改进的基于注意力机制的去遮挡模块,增强网络的边缘信息感知能力,通过边界信息得到物品完整轮廓,解决安检X光图像中物品重叠遮挡难题。

3) 在PIDray安检X光公开数据集进行验证,相比于

HRNet 算法等其他先进算法,本文算法在 Easy、Hard、Hidden 3 个不同难度的验证子集上均取得了最佳效果,证明了本文方法的优越性。

1 本文算法

本文算法旨在对 HRNet 语义分割算法进行优化改进。首先,通过引入基于注意力机制的去遮挡模块,将图

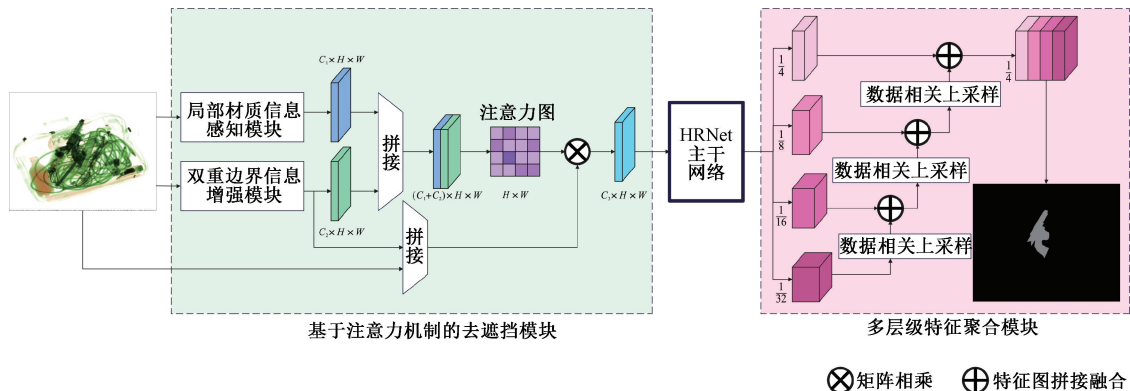


图1 本文算法的网络结构示意图

Fig. 1 The architecture of our semantic segmentation method

1.1 基于注意力机制的去遮挡模块

为了解决安检 X 光图像中存在边界不清晰、重叠遮挡的问题,本文引入了一个基于注意力机制的去遮挡模块,该模块由双重边界信息增强模块、局部材质信息感知模块和注意力图生成模块 3 个子模块构成。具体而言,双重边界信息增强模块和局部材质信息感知模块分别从两个通路提取违禁物品的边界信息和材料信息。然后通过两个分支进行融合,生成一个注意力图,有效捕捉图像中的关键区域,得到改进后的特征图。

1) 双重边界信息增强模块

针对安检 X 光图像所面临的重叠遮挡和物品边界模糊的挑战,构建双重边界信息增强模块。首先从上下左右 4 个方向强化边界信息,并将其聚合成一个完整形状^[28]。其中,提取左边界信息的计算方法,如式(1)所示,其他方向同理。然后,使用 Sobel 算子分别计算水平和垂直方向上的边缘图像^[29]。经过以上步骤,得到边界信息增强图像。

$$F_{ijc}^b = \max F_{ije}^a, j = n, (n+1), \dots, W \quad (1)$$

其中, $F^a \in R^{H \times W \times C}$, $F^b \in R^{H \times W \times C}$, F^a 、 F^b 分别表示边界强化处理前和处理后的特征图。下标 c 表示特征图的第 c 个通道, (i, j) 表示第 c 个通道中特征图 F_b 中横纵坐标分别为 i, j 的位置。 n 为任意自然数且 $n \leq W$ 。

2) 局部材质信息感知模块

材质信息在 X 光安检图像中主要通过颜色和纹理进

像输入该模块中,加强边界感知能力,从而降低遮挡效果的影响。其次,经过去遮挡模块处理后的图像被传入 HRNet 的特征提取主干网络中。在该网络中,利用不同分辨率,并行地提取多尺度特征,并通过不同分辨率之间的信息交互进行一定程度的特征融合。最后,通过将特征图输入到多层级特征聚合模块中进行逐级数据相关上采样,获得最终的语义分割结果。算法网络结构如图 1 所示。

行表述。为了充分利用材质信息来判断违禁物品的种类,本文采用了特征区域信息聚合的方法感知局部材质信息。通过平均池化操作聚合各个区域的局部信息,并生成不同尺度的区域聚合特征图。考虑到本文数据集中图片大小的差异,采取了 8、16 和 24 倍率的平均池化操作,以更好地满足不同局部区域大小范围内信息的聚合,解决安检 X 光图像中违禁物品尺寸大小不一的难题。局部材质信息感知模块的网络结构图如图 2 所示。

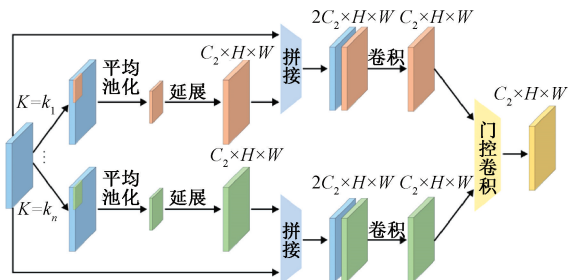


图2 局部材质信息感知模块网络结构示意图

Fig. 2 The architecture of the local material information sensing module

3) 注意力图生成模块

注意力图生成模块主要针对双重边界信息增强模块和局部材质信息感知模块所输出的特征图进行拼接,以融合两个模块的特征信息。随后采用 1×1 的卷积核对接接后的特征图在通道维度上进行压缩,生成用于计算注

意力分数的特征图。计算公式如式(2)所示。

$$F_3 = W_l(F_1 \parallel F_2) + b_l \quad (2)$$

其中, $F_1 \in R^{C_1 \times H \times W}$, $F_2 \in R^{C_2 \times H \times W}$ 分别表示双重边界信息增强模块输出的特征图和局部材质信息感知模块输出的特征图。“ $\parallel$ ”表示拼接操作, W_l 和 b_l 是卷积层的参数。最后, $F_3 \in R^{(C_1+C_2) \times H \times W}$ 输入 Sigmoid 函数生成注意力分数图 S , 公式如下:

$$S = \sigma(F_3) = \frac{1}{1 + e^{-F_3}} \quad (3)$$

其中, $S \in R^{H \times W}$ 。最后, 本文计算注意力分数图 S 和连接图像 P 的内积, 得到最终的注意力图 $F_4 \in R^{C_3 \times H \times W}$, 计算公式如下:

$$F_4 = \sum_{i=1}^{H \times W} S_i \cdot P_i \quad (4)$$

其中, P 为原始输入图像与边界增强图像进行通道拼接后的图像。

1.2 多尺度高分辨率并行特征提取模块

为了提高安检的智能化程度, 计算机辅助判别结果应包含违禁品准确的位置信息和类别信息。然而, 当前的语义分割方法通常采用线性的编码-解码结构, 通过下

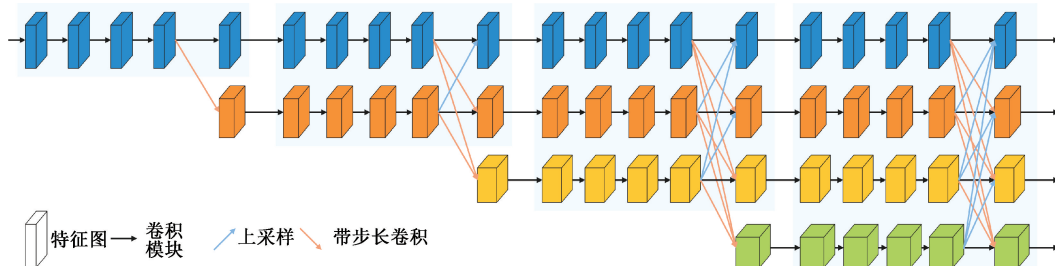


图 3 HRNet 主干网络结构

Fig. 3 The architecture of HRNet backbone network

1.3 多层次特征聚合模块

在解码阶段, 将主干网络提取到的不同尺度的特征进行聚合和分辨率的恢复对像素级的语义分割是至关重要的, 原始的 HRNet 首先将低分辨率特征图通过双线性插值上采样到最高分辨率, 然后进行拼接融合。然而, 这种方法没有考虑图像像素之间的上下文依赖关系, 利用简单的双线性插值上采样来恢复更高分辨率的预测图易导致局部细节信息的丢失, 很难得到精确的语义分割结果^[30]。

为此, 提出了一个多层次特征聚合模块, 通过逐级的数据相关上采样 (data-dependent upsampling, DUpsampling)^[31], 充分利用标签空间中的空间相关性优势, 提高语义重构的能力, 从而聚合多级特征以增强最终特征图的细节保持能力, 解决安检 X 光图像中背景复杂

采样和上采样操作来实现尺度的转换, 其缺点是信息丢失严重, 无法满足高精度的安检任务需求。本文选择 HRNet 作为主干网络, 其中高分辨率分支具有更多的细节表示能力, 擅长捕捉图像的细微变化和细节特征, 能够学习到图像中的纹理、位置等细粒度特征; 低分辨率分支更注重全局语义信息, 具有更大的感受野和上下文信息感知能力, 能够学习到图像的整体形状和全局结构等高层次的语义特征, 更适理解目标物体周围的环境和语义背景。不同分辨率分支学习到的信息在尺度上也存在差异。高分辨率分支更适合处理小尺度的目标物体和细节信息, 而低分辨率分支更擅长处理大尺度的目标物体和整体结构。通过多分辨率并行的网络结构, 利用不同尺度下的特征信息减少特征图在尺度转换过程中的信息损失。且在不同分辨率之间进行了信息交互, 高分辨率到低分辨率的转换使用了 3×3 的步长卷积操作, 而低分辨率到高分辨率的转换则采用了双线性插值的方法, 从而提升不同尺度信息之间的传递交流能力。因此, 本文选择 HRNet 作为特征提取主干网络, 来解决安检 X 光图像中的背景杂乱、尺度多变的难题, 以提高分割准确率。HRNet 的主干网络结构如图 3 所示。

和小目标违禁物品对精确识别所带来的挑战。该模块的结构如图 1 中右侧部分多层次特征聚合模块所示。在逐级上采样过程中, 从最低分辨率开始逐步上采样特征图, 然后将得到的特征图与相应分辨率分支的特征图进行拼接融合, 并继续进行上采样操作, 直到与最高分辨率特征图融合。以此来增加不同分辨率特征之间的交互和融合, 保留不同层级的多尺度特征信息, 提升安检 X 光图像的分割精度。

2 实验及结果分析

本文实验中采用 Ubuntu 操作系统和 NVIDIA GeForce 2080Ti GPU 进行, 并基于 PyTorch 深度学习框架进行神经网络构建, 网络均使用 Adam (adaptive moment estimation) 优化器, 学习率初始化为 0.0005, 采用余弦函

数调整学习率的变化,最低学习率为 0.000 005,训练 100 个 epoch。为了提高网络的泛化性能,先采用比例缩放、随机旋转等方法进行数据增强。为了提高训练效率,将图像统一裁剪为 480×480 大小,不同难度级别测试集示例如图 4 所示。

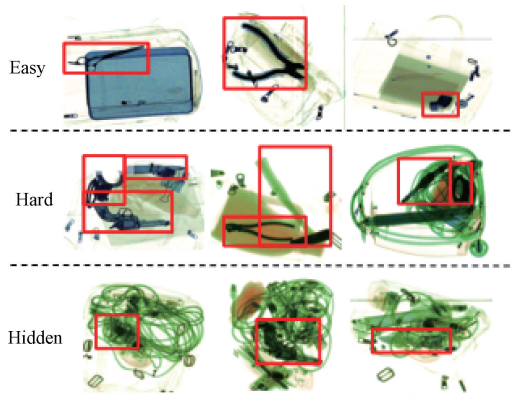


图4 不同难度级别测试集示例

Fig.4 Examples of test sets of three difficulty levels

2.1 数据集简介

本文实验所选择的 PIDray^[32] 安检 X 光图像语义分割数据集来源于机场、地铁站和火车站等不同的真实场景,且来自于 3 台不同制造商所生产的安检设备,具有广泛的覆盖范围和多样性。该数据集包含了常见的 12 个违禁物品,共有 47 677 张图像。训练集包含 29 457 张图像(约 60%),测试集包含 18 220 张图像(约 40%)。每张图像均包含违禁物品,且进行了像素级注释。为了更好地评估算法在不同场景和复杂度下的性能表现,测试集被细分为容易(Easy)、难(Hard)和隐藏(Hidden)3 个图像子集。Easy 图像子集表示测试集中的图像仅包含一个违禁物品,Hard 图像子集表示测试集中的图像包含多个违禁物品,而 Hidden 图像子集表示测试集中的图像中的违禁物品存在严重的重叠遮挡。详细信息如表 1 所示。

表1 PIDray 数据集信息

Table 1 The information of the PIDray dataset

模式	训练	测试		
		简单(Easy)	难(Hard)	隐藏(Hidden)
数量	29 457	9 482	3 733	5 005
总计		47 677		

2.2 评价指标

算法的准确性方面,参考数据集要求,使用平均交并比,即(mean Itersection over union, mIoU)以及平均像素精度(mean pixel accuracy, mPA)作为评价指标进行衡量,计算公式如下:

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \tag{5}$$

$$mPA = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \tag{6}$$

其中, p_{ij} 表示真实值*i*被预测为*j*的数量, $k+1$ 是总类别的数量(包含背景), p_{ii} 是被模型预测为正类的正样本(true positive, TP), p_{ji} 和 p_{ij} 分别表示为被模型预测为正类的负样本(false positive, FP)和被模型预测为负类的正样本(false negative, FN)。

2.3 定量实验分析

针对 PIDray 数据集,本文算法同当前先进的语义分割算法进行了比较,定量实验结果如表 2 所示。从实验结果可知,本文算法在安检 X 光图像语义分割任务上的性能达到了目前领先的水平,尤其是在 Hidden 子集的验证结果上领先更为突出,相比较 HRNet 等其他取得 SOTA 的高精度语义分割算法准确性得到了一定提升。具体地,本文算法在 PIDray 数据集的 3 个测试子集上分别取得了 73.15%、69.47%、58.33%的 mIoU 和 81.43%、80.04%、66.10%的 mPA。表明本文算法优于当前主流语义分割算法。

表2 PIDray 数据集定量实验结果

Table 2 The quantitative results of PIDray dataset

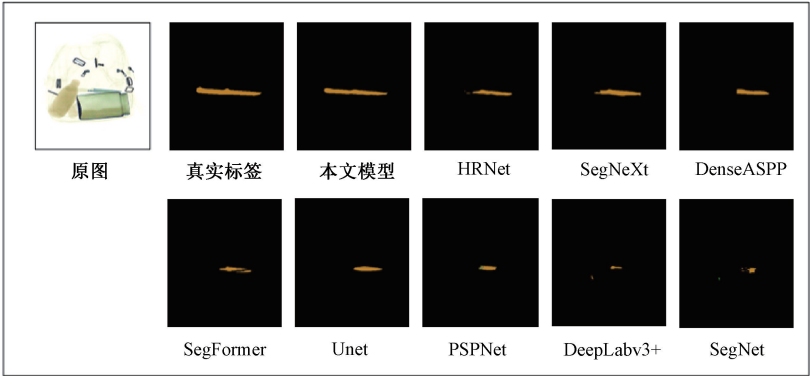
(%)

方法	简单(Easy)		难(Hard)		隐藏(Hidden)	
	mIoU	mPA	mIoU	mPA	mIoU	mPA
PSPNet ^[18]	63.18	72.23	61.50	72.98	40.92	48.03
SegNet ^[17]	64.06	73.68	61.79	73.56	41.90	49.09
UNet ^[16]	66.54	75.44	66.59	77.97	47.01	53.23
DenseASPP ^[33]	67.62	77.07	64.86	75.95	48.49	56.39
DeepLabV3+ ^[22]	70.67	78.62	67.33	77.96	47.32	56.23
SegFormer ^[34]	66.60	74.49	62.89	74.84	46.53	55.14
SegNeXt ^[35]	67.90	75.84	64.91	76.46	43.18	51.88
HRNet ^[27]	72.66	80.77	68.30	79.75	52.64	60.25
本文	73.15	81.43	69.47	80.04	58.33	66.10

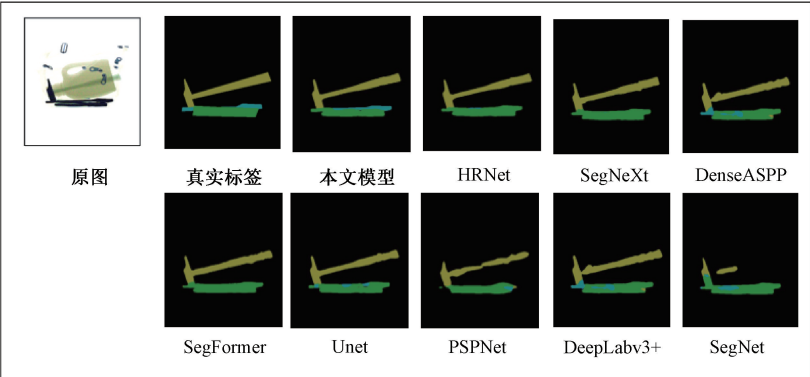
2.4 定性实验分析

基于所提出的图像语义分割算法,本文着重针对安检 X 光图像领域较为复杂的小目标遮挡问题,如图 5(a)所示,其中违禁物品像素点仅占比 3%,多类违禁物品重叠遮挡问题,如图 5(b)所示,以及背景杂乱覆盖遮挡问题,如图 5(c)所示。在 PIDray 数据集上对其进行了定性实验验证,并进行可视化分析,实验结果如图 5 所示。

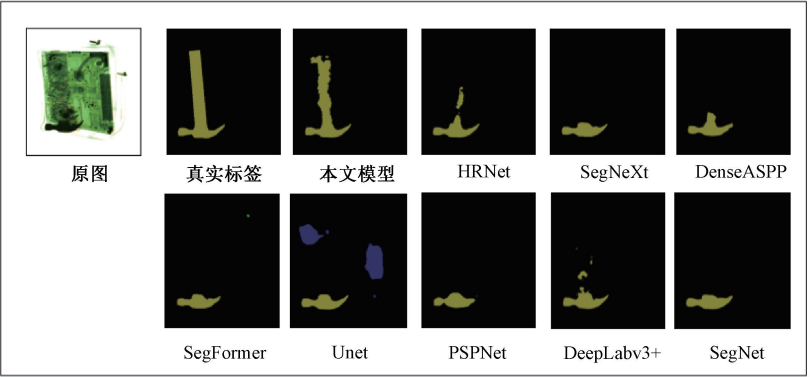
由图 5(a)可以看出,本文算法在处理遮挡的小目标违禁物品图像时,接近真实标签的分割效果,相比于其他先进算法有着显著的优势。包括 HRNet 在内的其他算法都可以识别出刀头,但是 DeepLabv3+、SegNet 等算法对小目标物品的分割识别不太友好,只能识别出少量像



(a) 小目标物品遮挡分割结果
(a) Small objects block segmentation results



(b) 多类违禁物品重叠遮挡分割结果
(b) Multi-class prohibited items overlapping occlusion segmentation results



(c) 背景覆盖遮挡分割结果
(c) The background covers prohibited items segmentation results



图 5 不同模型的分割结果

Fig. 5 Qualitative experimental results

素点。虽然 HRNet、SegFormer 等由于自身算法特性,在未被遮挡且颜色信息较多的刀头部分分割的较为精确,但是对与颜色信息较少且被遮挡的刀柄部分却未能分割识别。充分验证了本文利用高分辨率进行特征提取对位置信息的敏感性以及应用双重边界信息增强的去遮挡模块的有效性,可以比较完整地分割出违禁物品的轮廓。

由图 5(b)可以看出,在对大尺寸违禁物品分割识别和对被成像颜色更浅的物品遮挡的违禁物品分割识别时,各种算法在轮廓上均取得了不错的效果,但是由于扳手和钳子两类违禁物品存在重叠遮挡且材质基本相同,其他模型都存在像素点误分类的现象,而本文算法通过在高分辨率网络架构的基础上进一步聚合不同尺度的特

征信息,有效减少了 HRNet 算法出现的像素点分类错误现象,充分体现了本文设计的多层级特征聚合模块能够提升分割精度。

由图 5(c)可以看出,在对背景杂乱深度遮挡的违禁物品分割识别时,本文算法的分割精度受到影响,违禁品的分割形状开始出现扭曲且不够完整,但是仍然能够识别出绝大部分的像素点的类别信息,和其他算法比较,更为接近真实标签。而未改进的 HRNet 和 DeepLabv3+却只能识别出少量的被遮挡像素点,其他算法几乎只能识别出未被遮挡的锤头,而 Unet 过于注重细节,还出现了把背景像素误分类的现象。总之,本文算法在如此严重的复杂背景遮挡情况下,仍然取得了不错的分割效果,再次验证了本文算法去遮挡效果的优越性,而这种遮挡级别的安检 X 光图像,连人类的肉眼也是很难识别出完整的违禁物品图像的。然而在更严重的遮挡,如更深色的背景物品完全覆盖住更浅色的违禁物品,本文算法尚不能达到较好的效果,主要是因为此种遮挡程度完全掩盖了违禁物品本身的颜色信息,以至于算法无法提取到足够的有用信息,后续会从去遮挡的角度继续深入探究以取得更好的性能。

2.5 消融实验

为验证本算法中各模块对安检 X 光图像违禁物分割的有效性,采用控制变量法分析各模块对算法性能的影响,本环节共设计了 4 组消融实验。

1) 实验 1:未添加任何模块,直接将原 HRNet 网络用于安检 X 光违禁物品图像分割,为了验证原模型的分割效果。

2) 实验 2:在原 HRNet 模型的基础上,只添加基于注意力机制的去遮挡模块,为了验证引入去遮挡模块后,可以综合提升模型的整体分割精度。

3) 实验 3:在原 HRNet 模型的基础上,只添加多层级特征聚合模块,验证该模块确实相较于原方法有着有优秀的特征聚合能力并减少信息损失。

4) 实验 4:同时添加上述两个模块,验证本文整体模型相对于单一模块有着更优秀的语义分割性能和更高的精度。

实验结果如表 3 所示(注:为更便于读者观察,实验结果采取 Easy,Hard 和 Hidden 3 种难度下的平均结果)。对比表 3 中实验 1 和实验 2 的结果可知,增加去遮挡模块后,使得 mIoU 提升了 1.53%,mPA 提升了 1.47%。两项指标均有着不错的提升效果。通过对比实验 1 和实验 3 可知,采用本文设计的多层级特征聚合模块,分别取得了 65.49%的 mIoU 和 74.28%的 mPA,较原模型均有一定程度的提升。通过对比实验 2、实验 3 和实验 4 可知,同时应用两个模块的本文算法在实验 2 和实验 3 的基础上精度得到了进一步提升,充分验证了两种模块的适配性。

表 3 各组成模块消融实验结果					
Table 3 The results of ablation study					
	HRNet	基于注意力机制 多层级特征		mIoU/%	mPA/%
		去遮挡模块	聚合模块		
实验 1	✓			64.53	73.59
实验 2	✓	✓		66.06	75.06
实验 3	✓		✓	65.49	74.28
实验 4	✓	✓	✓	66.98	75.56

此外,为了验证去遮挡模块对于安检中重叠遮挡违禁物品图像分割的有效性,以及本文改进后的去遮挡模块相对于原去遮挡模块 DOAM 效果的提升,本文单独针对 Hidden 测试子集设计了 3 组消融实验。

1) 实验 5:未添加任何模块,直接将原 HRNet 网络在 Hidden 子集上进行实验。为了验证原模型对于物体遮挡情况下的分割效果。

2) 实验 6:在原 HRNet 模型的基础上,添加原去遮挡模块(DOAM),为了验证引入去遮挡模块可以有效减弱遮挡物体对分割效率的干扰。

3) 实验 7:在 DOAM 的基础上,再次加强边界信息处理,构成双重边界信息增强模块,将此改进添加到 DOAM 模块中,为了验证双重边界信息增强模块可以进一步提升去遮挡效果。

实验结果如表 4 所示,对比实验 5 和实验 6 可知,对于每张图像均存在重叠、遮挡现象的 Hidden 子集,在引入 DOAM 后,模型的精度都得到了较大程度的提升,达到了 55.02%的 mIoU 和 63.16%的 mPA。对比实验 6 和实验 7 可知,改进后模块对重叠遮挡图像的分割精度有了进一步的提升,达到了 57.19%的 mIoU 和 64.94%的 mPA,相比于原 DOAM 分别提升了 2.17%和 1.78%,充分验证了本文对原去遮挡模块改进的有效性。

表 4 针对隐藏子集下去遮挡模块消融实验结果					
Table 4 The ablation study of our modified de-occlusion module for hidden dataset					
	HRNet	DOAM	Ours	Hidden	
				mIoU/%	mPA/%
实验 5	✓			52.64	60.25
实验 6	✓	✓		55.02	63.16
实验 7	✓		✓	57.19	64.94

3 结 论

针对 X 光安检图像语义分割任务的实际应用需求,本文对 HRNet 算法进行改进,提出了一种新的网络架构和优化方案。通过引入去遮挡模块和构建多层特征聚合模块,有效地解决了遮挡问题并提升了多尺度特征融合效果。对于提高安检图像分割的准确性和可靠性具有重要意义,并具备潜在的实际应用价值。实验结果表明,相

比较 HRNet 等高精度图像语义分割算法,本文算法在安检 X 光违禁物品图像语义分割任务中得到了更优越的性能。然而,对于违禁品像素点占比图像总像素点数目不足 1% 时的遮挡现象,本文算法还是难以准确识别,无法达到预期效果。从数据样本增强和关联学习角度对更小目标违禁物品识别进行研究,将是下一步的重点研究工作。与此同时,在保证分割识别精度的前提下,对模型进行轻量化研究和设计,满足计算机辅助安检的实际应用部署需求也将是本文的后续重点研究内容。

参考文献

- [1] 邢琛聪,吕淑静,吕岳. 基于小样本学习的 X 光图像违禁物品识别[J]. 电子测量与仪器学报, 2021, 35(3): 204-210.
XING CH C, LYU SH J, LYU Y. Recognition of prohibited items in X-ray images based on small sample learning[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(3): 204-210.
- [2] 司垒,谭超,朱嘉皓,等. 基于 X 射线图像和激光点云的煤矸识别方法[J]. 仪器仪表学报, 2022, 43(9): 193-205.
SI L, TAN CH, ZHU J H, et al. Recognition method of coal gangue based on X-ray image and laser point cloud[J]. Chinese Journal of Scientific Instrument, 2022, 43(9): 193-205.
- [3] 张平,佟昆宏,王学珍. 基于改进 U-net 网络的液压管路分割方法[J]. 电子测量与仪器学报, 2023, 37(1): 123-129.
ZHANG P, TONG K H, WANG X ZH. Hydraulic line segmentation method based on improved U-net network[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(1): 123-129.
- [4] TURCSANY D, MOUTON A, BRECKON T P, et al. Improving feature-based object recognition for X-ray baggage security screening using primed visual words[C]. 2013 IEEE International Conference on Industrial Technology (ICIT). IEEE, 2013: 1140-1145.
- [5] BASTAN M, BYEON W, BREUEL T M, et al. Object recognition in multi-view dual energy X-ray images[C]. BMVC, 2013, 1(2): 11.
- [6] MERY D, RIFFO V, ZUCCAR I, et al. Object recognition in X-ray testing using an efficient search algorithm in multiple views[J]. Insight-Non-Destructive Testing and Condition Monitoring, 2017, 59(2): 85-92.
- [7] AKCAY S, KUNDEGORSKI M E, WILLCOCKS C G, et al. Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery[J]. IEEE Transactions on Information Forensics and Security, 2018: 2203-2215.
- [8] REN S Q, HE K M, ROSS B, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2015): 1137-1149.
- [9] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [10] SUBARMANI M, RAJADUARI K, CHOUDHURY S D, et al. Evaluating one stage detector architecture of convolutional neural network for threat object detection using X-ray baggage security imaging[J]. Revue d'Intelligence Artificielle, 2020, 34(4): 495-500.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
- [12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [13] 何彦辉,吴桂兴,吴志强. 基于域适应的 X 光图像的目标检测[J]. 计算机科学, 2021, 48(1): 175-181.
HE Y H, WU G X, WU ZH Q. Object detection of X-ray images based on domain adaptation[J]. Computer Science, 2021, 48(1): 175-181.
- [14] 林春晖. 基于深度学习的行李安检机危险品自动检测[D]. 南京: 东南大学, 2020.
LIN CH H. Automatic detection of dangerous goods in baggage security machine based on deep learning[D]. Nanjing: Southeast University, 2020.
- [15] LONG J, SHELHAMER E, DARRELL T, et al. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [16] RONNEBERGER O, FISCHER P, BROX T, et al. U-net: Convolutional networks for biomedical image segmentation[C]. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.
- [17] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[C]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(12): 2481-2495.
- [18] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:

- 2881-2890.
- [19] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs [J]. ArXiv Preprint, 2014, ArXiv:1412.7062.
- [20] 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, 2017, 40 (4): 834-848.
- [21] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [J]. ArXiv Preprint, 2017, ArXiv:1706.05587.
- [22] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 801-818.
- [23] XU M, ZHANG H, YANG J, et al. Prohibited item detection in airport X-ray security images via attention mechanism based CNN [C]. Chinese Conference on Pattern Recognition, 2018: 429-439.
- [24] AN J, ZHANG H, ZHU Y, et al. Semantic segmentation for prohibited items in baggage inspection [C]. International Conference on Intelligent Science and Big Data Engineering, 2019: 495-505.
- [25] 邓观林. 基于深度学习的管制物品图像语义分割算法研究[D]. 赣州:江西理工大学, 2021.
- DENG G L. Research on semantic segmentation algorithm of controlled items image based on deep learning [D]. Ganzhou: Jiangxi University of Science and Technology, 2021.
- [26] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5693-5703.
- [27] SUN K, ZHAO Y, JIANG B, et al. High-resolution representations for labeling pixels and regions [J]. ArXiv Preprint, 2019, ArXiv:1904.04514.
- [28] WEI Y, TAO R, WU Z, et al. Occluded prohibited items detection: An x-ray security inspection benchmark and de-occlusion attention module [C]. Proceedings of the 28th ACM International Conference on Multimedia, 2020: 138-146.
- [29] TAO R, WEI Y, JIANG X, et al. Towards real-world X-ray security inspection: A high-quality benchmark and lateral inhibition module for prohibited items detection [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10923-10932.
- [30] XIAO B, WU H, WEI Y. Simple baselines for human pose estimation and tracking [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 466-481.
- [31] TIAN Z, HE T, SHEN C, et al. Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3126-3135.
- [32] WANG B, ZHANG L, WEN L, et al. Towards real-world prohibited item detection: A large-scale X-ray benchmark [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 5412-5421.
- [33] YANG M, YU K, ZHANG C, et al. Denseaspp for semantic segmentation in street scenes [C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 3684-3692.
- [34] XIE E, WANG W, YU Z, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers [J]. Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.
- [35] GUO M H, LU C Z, HOU Q, et al. Segnext: Rethinking convolutional attention design for semantic segmentation [J]. Advances in Neural Information Processing Systems, 2022, 35: 1140-1156.

作者简介



李广睿, 2021 年于东北农业大学获得学士学位, 现为北京信息科技大学硕士研究生, 主要研究方向为安检 X 光图像违禁物品的语义分割。

E-mail: lgr520310@126.com

Li Guangrui received his B. Sc. degree from Northeast Agricultural University in 2021. Now he is a M. Sc. candidate in Beijing Information Science and Technology University. His main research interest includes semantic segmentation of prohibited items in security X-ray image.



刘琼, 2006 年于国防科技大学获得学士学位, 2012 年于北京航空航天大学获得博士学位, 现为北京信息科技大学副教授, 主要研究方向为模式识别、计算机视觉。

E-mail: liuqionglq@126.com

Liu Qiong received her B. Sc. degree from National University of Defense Technology in 2006 and Ph. D. degree from Beihang University in 2012, respectively. Now she is an associate professor in Beijing Information Science and Technology University. Her main research interests include pattern recognition and computer vision.