

针对 X 光安检场景的多目标违禁品识别算法

曹洋*, 张莉, 孟俊熙, 宋倩, 张乐天

西安工程大学电子信息学院, 陕西 西安 710600

摘要 针对 X 光安检图像中多目标违禁品识别难度高的问题, 提出了一种多目标违禁品识别算法。首先, 综合考虑实际应用需求、网络性能和运行速度, 用残差网络 (ResNet50) 作为骨干网络, 并添加局部强化模块弥补空洞卷积导致的棋盘格现象。然后, 用空洞残差特征增强模块和可变形空洞空间金字塔池化分别处理不同层级特征, 自适应学习违禁品的多尺度特性。最后, 引入注意力机制, 强化对重点通道的学习能力并实现空间维度上的特征聚焦, 加强违禁品区域的细节表征能力。在安检违禁品图像数据集上的测试结果表明, 相比其他对比算法, 本算法可在保证实时性的前提下取得更优的分割精度, 平均交并比为 82.26%, 图像处理速度为 16.21 frame/s。

关键词 机器视觉; X 光图像; 深度学习; 语义分割; 违禁品识别

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP202259.1015009

Multi-Target Prohibited Item Recognition Algorithm for X-Ray Security Scene

Cao Yang*, Zhang Li, Meng Junxi, Song Qian, Zhang Letian

College of Electronics and Information, Xi'an Polytechnic University, Xi'an 710600, Shaanxi, China

Abstract Aiming at the difficulty of identifying multi-target prohibited item in X-ray security images, a multi-target prohibited item recognition algorithm is proposed in this paper. First, considering practical application requirements, network performance and running speed, the residual network (ResNet50) is used as the backbone network, and a local reinforcement module is added to compensate for the checkerboard phenomenon caused by dilate convolution. Then, the features of different levels are processed by the dilated residual feature enhancement module and the transformable dilated space pyramid pooling respectively, and the multi-scale characteristics of prohibited item are adaptively learned. Finally, the attention mechanism is introduced to strengthen the learning ability of key channels and realize the feature focusing in the spatial dimension, so as to strengthen the detailed representation ability of the prohibited item area. The test results on the security inspection prohibited item image dataset show that compared with other comparison algorithms, the algorithm can achieve better segmentation accuracy on the premise of ensuring real-time performance, the mean intersection-over-union is 82.26%, and the image processing speed is 16.21 frame/s.

Key words machine vision; X-ray image; deep learning; semantic segmentation; prohibited item recognition

1 引言

安检工作对维护社会稳定、保障公众生命财产

安全至关重要。目前, X 光安检机下违禁品的检测主要依赖于人工判别, 而 X 光图像背景繁杂、目标多、色彩对比度低且遮挡现象比较严重, 安检人员

收稿日期: 2021-06-23; 修回日期: 2021-07-21; 录用日期: 2021-08-17

基金项目: 国家自然科学基金(51607133)、陕西省教育厅研究项目(10JK510)、西安市科技局产业化项目(CXY1517(4))、安检违禁品检测软件及设备研制设计(2020KJ-391)

通信作者: *827602052@qq.com

长时间处于高压密集的工作状态下极易产生漏检、错检问题。

随着深度学习的发展,人们逐渐倾向用卷积神经网络实现更高效的违禁品检测,如通过目标检测算法^[1]输出矩形框标定目标的位置。Akay 等^[2]通过迁移学习将卷积神经网络应用到 X 光图像的分类与检测任务中;Liu 等^[3]建立了基于深度卷积神经网络的检测器,可对危险品区域进行标注。但这类检测算法不完全适用于 X 光成像下的违禁品识别,实际应用场景中用标准矩形检测框无法准确获取违禁品的轮廓信息,给图像识别带来了一定的困难。通过语义分割算法也可以对违禁品图像中像素点的类别进行预测^[4],得到语义分割图,兼具定位与分类功能,从而更好地提取 X 光下违禁品的本质特征,实现对安检图像中违禁品的精细化分割。Xu 等^[5]创新性地将空间注意力机制加入传统图像分割技术^[6]中,设计了一种安检违禁品识别算法,通过空间位置、细节及类别信息识别违禁品;An 等^[7]将语义分割技术与 X 光违禁品图像识别算法相融合,实现了对安检违禁品单一目标的高精度识别,但识别速度并不理想,难以应用到多目标的实时检测中^[8-9]。

目前暂无公开的安检图像语义分割数据集,且对基于语义分割的安检违禁品识别算法研究较少。全卷积神经网络(FCN)^[10]的出现有效解决了像素级对象的预测及分类问题,但依旧存在部分像素丢失问题,且未考虑图像的上下文信息。反卷积网络(DeconvNet)^[11]通过卷积层镜像处理,构成编码-解码(Encoder-decoder)结构,改善了 FCN 的分割效果。U-Net^[12]与高效神经网络(ENet)^[13]在该结构的基础上进行优化处理,显著提升了目标轮廓的实时分割效果。轻量化网络(LDENet)^[14]用非对称编解码结构改进了残差模块并引入注意力机制,减少网络模型参数的同时增强了特征的表达能力。针对整合不同尺度空间信息以及权衡局部与全局信息的问题,金字塔场景解析网络(PSPNet)^[15]通过加入空间金字塔池化(SPP)模块,更好地利用了深层特征中的语义信息,实现了更高质量的像素级场景解析。DeepLabv2^[16]提出了空洞空间金字塔池化(ASPP)^[17]模块,实现了多尺度目标处理。DeepLabv3+^[18]将编解码结构与 ASPP 模块融合,在多个公开数据集上得到了较好的分割结果,为解决不同尺度目标的分割问题提供了新思路。

结合 X 光安检违禁品识别任务的特点以及实际应用场景的需求,本文提出了一种 X 光安检场景下的多目标违禁品识别算法。首先,为了更好地权衡模型性能与参数量,用残差网络(ResNet50)作为特征提取基准网络,引入空间和通道注意力筛选重要特征,同时抑制不相关区域的特征。然后,在网络末端拼接局部强化(FE)模块,增强像素间的交互关系,有效缓解棋盘格现象。最后,用空洞残差特征增强(ARFE)模块进一步丰富高级特征的局部语义信息与低级特征的空间位置信息,用可变形空洞空间金字塔池化(TASPP)模块提高模型对多尺度、多形态违禁品特征的学习能力。采集多种 X 光安检违禁品图像并进行人工标注,建立多目标 X 光安检违禁品语义分割专用数据集。实验结果表明,本算法对 X 光图像中的安检违禁品具有良好的分割处理能力,且能保持较理想的处理效率。

2 网络模型设计

2.1 算法模型的整体网络设计

本算法整体采用编解码结构,以 ResNet50 作为特征提取基准网络,其结构如图 1 所示。其中,Conv 为卷积模块。在编码阶段,将卷积块注意力模块(CBAM)^[19]集成至基准网络的残差块中,通过通道注意力和空间注意力分别实现通道维度上的加权标定、空间维度的特征聚焦,强化对重点通道的特征学习能力并恢复空间局部细节,捕捉关键性区域,有效提升模型的分割精度。

在基准网络末端加入 FE 模块,将提取到的深层特征输入 TASPP 模块。通过多路并行的可变形空洞卷积和全局平均池化模块,学习多尺度特征并捕获全局上下文信息。其中,空洞卷积的膨胀率(Dilation)分别为 6、12、18。根据目标的尺度变化灵活调整感受野,融合不同感受野的语义信息,增强对姿态多样目标的建模能力,在只增加少量模型参数和计算量的情况下有效提升网络的性能。在解码阶段,对编码阶段输出的高级特征张量进行两次 2 倍上采样和一次 4 倍上采样,获得与输入图像尺寸相同的类别置信图。此外,用 ARFE 模块处理特征提取网络中的低级特征,强化更多细节信息表征。将高级特征分别与筛选得到的低级特征相融合,以减少上采样过程中有效信息的丢失,丰富图像的整体语义信息与细节信息,进一步提升网络的预测精度。

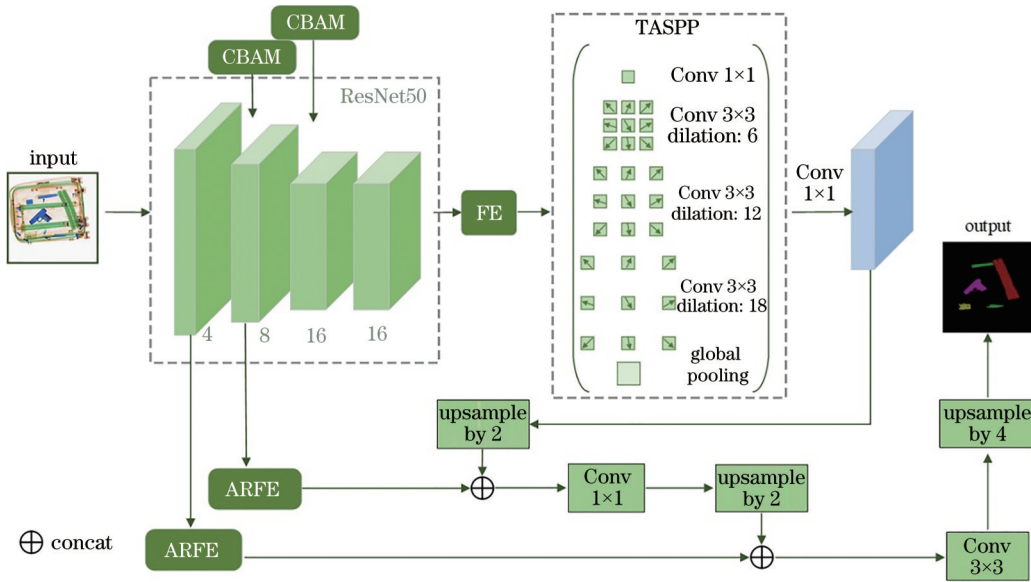


图 1 本算法的网络结构

Fig. 1 Network structure of the algorithm

2.2 注意力机制

CBAM 依次采用通道注意力与空间注意力,如图 2 所示。通过在 ResNet 中集成残差块的 CBAM 自适应地细化违禁品特征,关注感兴趣的区域,抑制非相关特征。

在通道域内,特征 F 经过平均池化与最大池化汇聚空间信息,将处理后的通道平均池化特征 F_{avg}^c 与通道最大池化特征 F_{max}^c 相加再通过 Sigmoid 激活函数 σ 输出。通道注意力特征 $M_c(F)$ 可表示为

$$M_c(F) = \sigma \left\{ W_1 \left[W_0 \left(F_{avg}^c \right) \right] \right\} + W_1 \left[W_0 \left(F_{max}^c \right) \right], \quad (1)$$

式中, W_0 为通道降维处理时全连接层的权重, W_1 为通道升维处理时全连接层的权重。空间域中,先在通道维度上用平均池化和最大池化得到空间平均池化特征 F_{avg}^s 和空间最大池化特征 F_{max}^s ,然后将两个通道信息的特征图串联,并通过 7×7 的卷积操作提取特征,最后通过 Sigmoid 激活函数 σ 得到空间注

意力特征 $M_s(F)$, 可表示为

$$M_s(F) = \sigma \left[f^{7 \times 7} \left(F_{avg}^s; F_{max}^s \right) \right], \quad (2)$$

式中, $f^{7 \times 7}$ 为卷积核尺寸为 7×7 的卷积操作。

特征图的通道中蕴含着丰富的信息,通过对违禁品特征图的通道添加权重,筛选得到每个分支通道上不同重要程度的特征信息,从而在一定程度上降低处理高维信息的复杂度,强化对重要通道的特征学习能力,在通道维度上了解目标对象的类别属性,进一步提升模型的分割精度及泛化能力。相比通道注意力,空间域注意力更针对空间维度上的特征,从而对通道注意力进行补充,强化各通道中像素权重较大的特征^[20],抑制像素权重小的特征,进一步提升网络对目标重点区域的捕获能力。实验结果表明,通过依次引入通道注意力和空间注意力,能在保证较低复杂度的情况下有效提升模型性能。

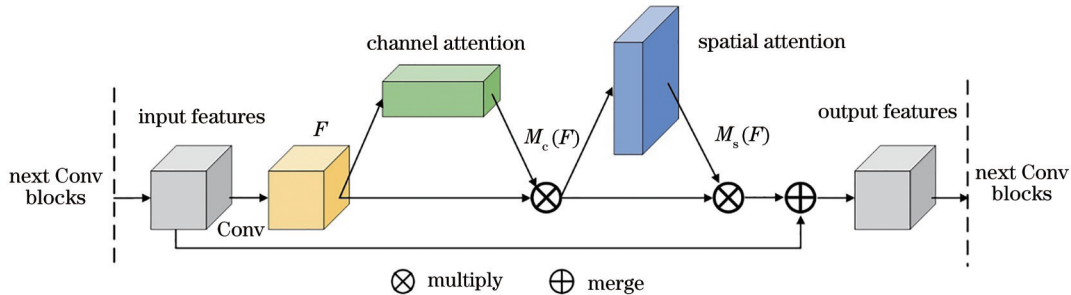


图 2 基于注意力机制的残差块

Fig. 2 Residual block based on attention mechanism

2.3 局部强化模块

采用空洞卷积在不增加计算量的同时扩大感受野,通过对卷积核间隔位置补零的操作,取得有效像素值,即只提取区域中非零值部分的信息。由于空洞卷积无法覆盖到全部的图像特征,会造成局部信息的连续性损失(棋盘格现象),导致模型最终

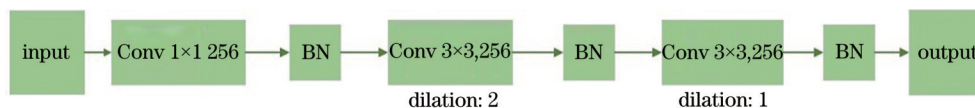


图3 FE模块的结构

Fig. 3 Structure of the FE module

在最后级联的空洞卷积与TASPP模块之间通过加入两个膨胀率较小的空洞卷积,使采样方式更加密集,从而加强了像素相互间的依赖性,获取更多的有效细节信息。为提升特征信息的学习效率、减少计算量,将经过降维处理的图像输入膨胀率为2的空洞卷积单元,可表示为

$$[\mathbf{C}(m) \times \mathbf{k}(n)](m+2n) = \sum_{m+2n} \mathbf{C}(m) \mathbf{k}(n), \quad (3)$$

式中, m 为步长, n 为位置偏移量, $\mathbf{C}(m)$ 为卷积向量, $\mathbf{k}(n)$ 为卷积核, \times 为卷积运算。然后,将空洞卷积单元得到的结果输入膨胀率为1的空洞卷积,可表示为

$$[\mathbf{C}(m) \times \mathbf{k}(n)](m+n) = \sum_{m+n} \mathbf{C}(m) \mathbf{k}(n). \quad (4)$$

此外,在每个卷积层后加入BN层,利用可训练的参数对卷积单元部分批量数据进行BN理。引入BN层可提高模型的泛化能力,加快深度网络的训练速度,简化调参过程,解决网络深度增加导致的梯度弥散问题,大幅降低对初始学习率的依赖。实验结果表明,FE模块能使该层提取的特征相互依赖性更强,能有效提高违禁品图像信息提取的准确度。

2.4 空洞残差的特征增强

由于浅层特征包含更多的全局空间信息和初级阶段较为模糊的轮廓信息,而X光成像存在色彩单调、纹理性缺失等问题。针对此类问题,设计了ARFE模块,其结构如图4所示。将输入的特征图分别经过两路处理,以强化局部细节信息,进而提

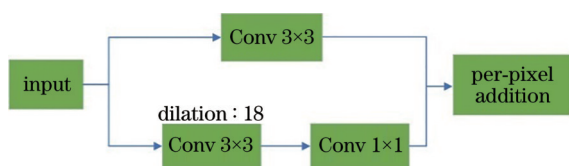


图4 ARFE模块的结构

Fig. 4 Structure of the ARFE module

学习到的特征缺乏相关性。利用并行多扩张率空洞卷积可增强特征信息的利用率,但当膨胀率大幅增加时有效像素点会急剧减少,依旧无法有效解决棋盘格现象,导致网络训练失败,丢失小目标局部特征信息。针对上述问题,设计了特征图的FE模块,其结构如图3所示。其中,BN为批归一化。

取全局空间信息。一路分支通过扩大感受野得到全局轮廓信息,采用膨胀率为18、卷积核尺寸为 3×3 的空洞卷积,同时采用 1×1 的卷积整合通道维度,便于特征融合;另一路分支则保留原始空间特征,通过 3×3 卷积优化局部重点细节信息。最后,将强化的全局位置信息与局部边缘信息以残差形式融合,完成浅层特征的整体空间信息优化。

2.5 可变形空洞空间金字塔池化

为解决多目标违禁品尺度不一的问题,提出了一种改进的空间金字塔池化模块。ASPP模块由 1×1 的卷积层、三个空洞率分别为6、12、18的 3×3 卷积层与一个池化层并行组成,最后对输出特征进行 1×1 卷积以及上采样,使其恢复至原始图像尺寸。该模块能有效获取不同尺度违禁品图像的语义信息,提高网络学习效率。在训练过程中,一般使用传统尺寸为 3×3 的标准卷积核,如图5(a)所示。标准卷积核在不同位置的感受野均相同,无法根据不同尺度、形变的物体自动调节尺度、感受野,对未知变化的适应性差,泛化能力弱。因此,引入可变形卷积核,可以根据实际情况给每个卷积核的参数添加方向向量,调整自身形状以达到更好的表征能力,其结构如图5(b)和图5(c)所示。

针对违禁品这类姿态、尺度多样的目标,在可变形卷积的基础上引入空洞卷积思想,形成可变形空洞卷积,取代ASPP内原有的三路并行空洞卷积,构建TASPP模块。可变形空洞卷积的具体学习过程如图6所示,相比传统卷积结构,可变形空洞卷积网络增加了偏置向量以及一个空洞率指标。即将标准卷积利用传统卷积核通过规则网格 R 在输入特征图 x 上进行采样,通过另一平行标准卷积层获得偏移量,生成的通道维度(偏移层)为 $2N$,然后将偏移量添加到输入层中的固定采样区域。

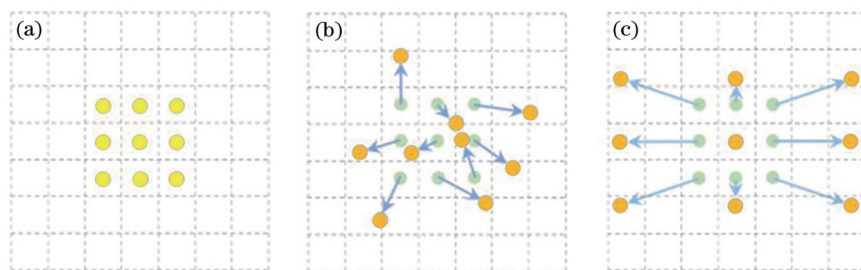


图 5 传统卷积与可变形卷积。(a)标准卷积核;(b)可变形卷积核;(c)可变形卷积核的特殊形式

Fig. 5 Standard convolution and transformable convolution. (a) Standard convolution kernel; (b) transformable convolution kernel; (c) special forms of transformable convolution kernel

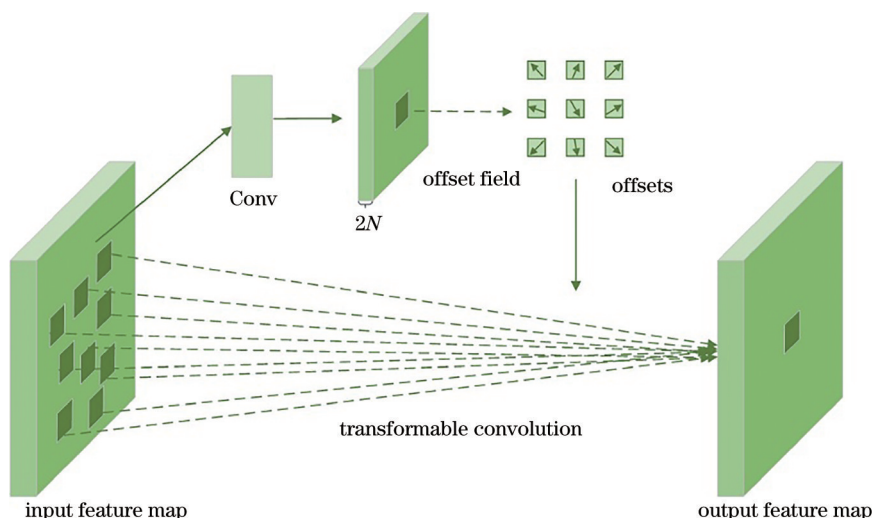


图 6 可变形空洞卷积的学习过程

Fig. 6 Learning process of the transformable dilated convolution

引入偏移量学习后,可变形空洞卷积核的大小与位置可通过图像实际内容自适应调节,但需对新的不连续位置求导。因此,采用双线性插值和后向传播算法进行学习,加入偏移量后输出的特征图可表示为

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta), \quad (5)$$

式中, w 为权重, p_0 为输出特征图 y 中某像素的位置, p_n 为标准采样位置, Δ 为可变形卷积的偏移量。

可变形空洞卷积通过偏移覆盖物体边缘,能很好地满足安检违禁品这类多尺度、多姿态的目标。此外,感受野能随目标物的实际尺度进行自适应调节,只需添加少量模型参数,就能有效提升模型对几何变换的学习能力,获得目标边缘更精准的预测结果。

3 实验结果及分析

3.1 数据集的建立及实验设计

目前暂无公开的 X 光安检违禁品图像语义分

割数据集,因此,参照 PASAL VOC2012 数据集^[21]的标注方式,利用高铁站 FISCAN 神盾 X 射线安检机采集 4653 张 X 光违禁品图像,并用 LabelMe 软件完成像素级数据标注。将实拍的旅客行李安检图像与模拟的单个、多个违禁品图像结合(单个目标违禁品图像与多目标违禁品图像的数量比约为 1:3),创建了一个 X 光下多目标违禁品语义分割数据集(MCDD)。该数据集包含 5 种违禁品类用于模型训练与评估:类别 a 为刀具类(剪刀、水果刀、菜刀、叉子、手工刀、钳子、管制刀具),共计 2307 张图像;类别 b 为枪支类,共计 1560 张图像;类别 c 为违禁液体类(酒精、汽油、白酒等易燃液体),共计 2119 张图像;类别 d 为易爆物(压力罐、充电宝、铅蓄电池、爆竹、打火机等),共计 2035 张图像;类别 e 为危险药品类,共计 1037 张图像。对 X 光安检违禁品图像经镜像、旋转、灰度反转、裁剪、拉伸、缩放等数据增广处理后,共得到 18523 张图像。本研究基于 PyTorch 深度学习框架,实验环境为四块 Nvidia GeForce

GTX1080Ti 显卡。为加速网络收敛,将训练集、测试集和验证集中的图像数量比例设置为 4:3:3。网络训练过程中的学习策略为 Poly 策略,动量因子为 0.9,初始学习率为 0.0001,权重衰减次数为 0.0002,根据计算机配置将 batchsize 设置为 16。

为保证算法的分割效果,选择平均交并比(MIoU)、画面每秒传输帧数(FPS)和运行时间(Time)作为评价指标,以评估算法的分割精度及速度。MIoU 是衡量算法分割精确度的标准度量,可表示为

$$X_{MIoU} = \frac{1}{X_{NUM}} \left[\frac{\sum_{i=1}^P X_{ii}}{T_i + \sum_{j=1}^P (X_{ji} - X_{ii})} \right], \quad (6)$$

式中, X_{NUM} 为全部样本类别的数量, T_i 为第 i 类像素总数, X_{ii} 表示实际类别为 i 、预测类别为 i 的像素总数, X_{ji} 表示实际类别为 i 、预测类别为 j 的像素总数, P 为类别总数。

3.2 模型参数设置

为确定 ARFE 模块分支内空洞卷积的膨胀率,对比设置不同膨胀率时模型的分割性能。在骨干网络浅层加入 ARFE 模块并调整膨胀率大小,采用控制变量法对比分析膨胀率对模型性能的影响,结果如表 1 所示。可以发现,增大膨胀率能在一定程度上提升网络性能,但增加到一定值后继续增加膨胀率会降低网络性能,当膨胀率为 18 时,模型的分割效果最好。

表 1 不同膨胀率时模型的性能

Table 1 Performance of the model with different dilation rate

Group	ARFE	Dilation rate	MIoU / %
1	×	—	80.97
2	√	6	81.65
3	√	12	81.93
4	√	18	82.26
5	√	24	81.11

为测试不同特征提取基准网络对语义分割模型性能的影响,对比了用 4 组不同特征提取网络作为基准网络时模型的性能,结果如表 2 所示。可以发现:用 MobileNetv2 或 ResNet34 作为特征提取基准网络时,模型的运行速度较快,实时性更优,但对于 X 光安检违禁品图像的分割精度较差;用 ResNet101 作为特征提取基准网络时,模型的分割精度略高,但速度较差;而 ResNet50 综合考虑了安检实际需求因素,很好地平衡了分割精度与速度。

表 2 不同特征提取基准网络下模型的分割性能

Table 2 Segmentation performance of models under different feature extraction baseline networks

Model	Baseline network	MIoU / %	FPS	Model size / Mb
Ours 1	MobileNetv2	75.93	28.34	8.97
Ours 2	ResNet34	76.81	18.19	29.31
Ours 3	ResNet101	83.14	9.57	50.33
Ours 4	ResNet50	82.26	16.21	34.46

3.3 与经典语义分割算法的对比

为验证本算法在多目标安检违禁品分割任务中的有效性,将其与语义分割领域内较性能较好的主流算法进行对比。在相同测试环境下,用其他几种算法的公开源码在 MCDD 上完成测试,不同算法的定量对比结果如表 3 所示。可以发现:U-Net 和 FCN-8s 算法的模型结构简单且预测速度较快,但预测过程中丢失了过多的图像空间信息,MIoU 指标较低,难以满足安检应用的需求;引入空洞卷积以及空间金字塔的 PSPNet 和 DeepLabv3 算法的预测准确率较高,但随着网络深度的增加,模型的运算速度有大幅降低,无法保证安检过程的实时性;本算法以损失较少的分割速率为代价,大幅提升了模型的分割精度,预测精度比 DeepLabv3 算法高 2.98 个百分点,FPS 比 U-Net 算法提高了 0.24,很好地平衡了实时性和准确性,能更好地满足工程应用需求。

表 3 不同算法在 MCDD 上的分割性能

Table 3 Segmentation performance of different algorithms on MCDD

Algorithm	Time / ms	MIoU / %	FPS
U-Net	62.62	72.14	15.97
FCN-8 s	64.98	70.85	15.39
PSPNet	118.91	73.03	6.41
DeepLabv3	141.84	79.28	5.05
Ours	61.69	82.26	16.21

不同算法对部分 X 光安检违禁品图像预测的可视化结果如图 7 所示。可以发现:U-Net 算法对于行李内多目标违禁品区域的分割效果不理想,对部分违禁品区域存在严重的错分割问题,将部分金属零件错误预测为刀具类别,且对药品及管制刀具的轮廓分割比较粗糙;DeepLabv3 算法对违禁品小目标的捕获效果较差,违禁品的细节处理不理想,如刀尖部分和枪支扳机处存在一定的漏分割现象,小目标药品区域的分割轮廓比较粗糙;本算法能较

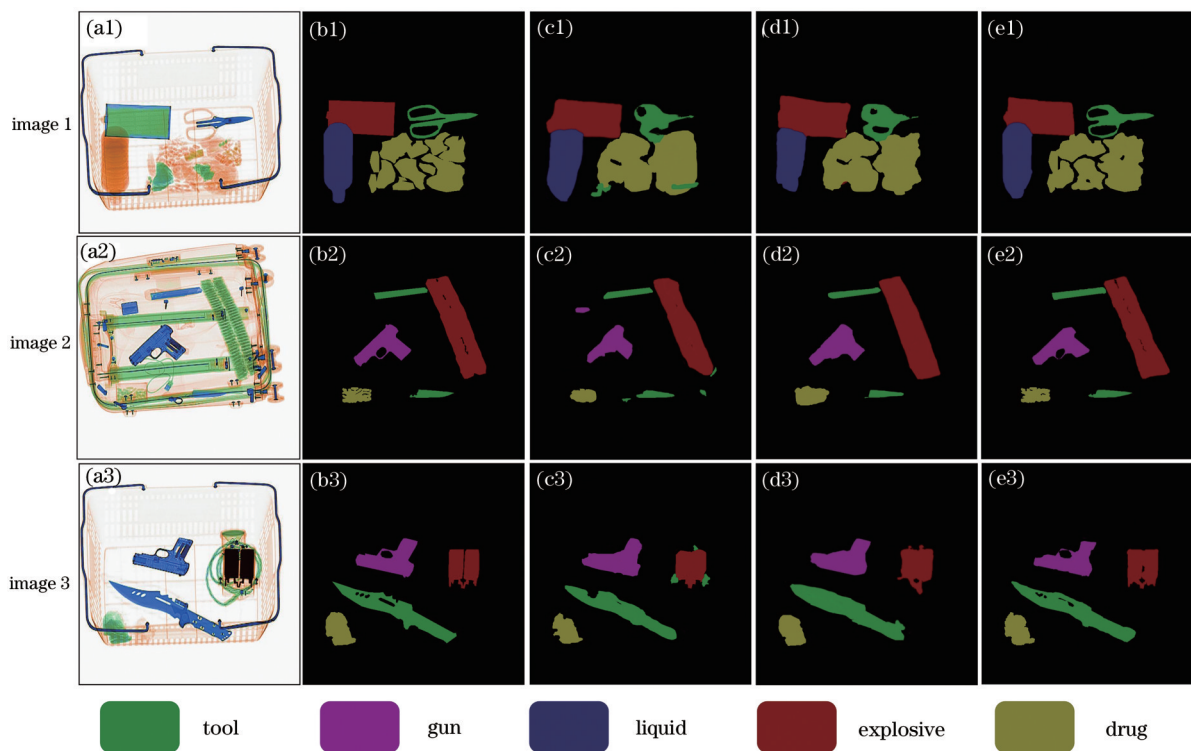


图7 不同算法在MCDD上的可视化结果。(a)原始图像;(b)标签;(c) U-Net; (d) DeepLabv3; (e)本算法
 Fig. 7 Visualization results of different algorithms on MCDD. (a) Original image; (b) label; (c) U-Net; (d) DeepLabv3; (e) our algorithm

好地解决上述问题,具有很好的场景理解能力及局部信息捕获能力,且兼顾了整体空间特征以及局部细节特征,能准确分割出小目标违禁品(如管制刀具细节部分及药品区域),基本拟合了真实的违禁品标签,满足目标的多尺度变化,鲁棒性较强。

3.4 消融实验

为验证本算法中各方案对X光安检违禁品图像分割的有效性,采用控制变量法分析各方案对算法性能的影响,以分割精度与速度为评价指标,对比结果如表4所示。其中,实验 I 未添加任何模块,直接将编码阶段输出的高级特征与特征提取基准网络内的浅层特征融合。对比实验 I、II 可以发现,引入CBAM对提升X光安检违禁品图像分割精度有着积极影响;对比实验 II、III 可以发现,增加ARFE模块加强违禁品局部特征信息后算法的MIoU提升了1.07个百分点,进一步优化了模型的分割效果;对比实验 III、IV 可以发现,使用FE模块后算法的MIoU提升了1.14个百分点,FPS提升了0.17;对比实验 V、VI 可以发现,相比未使用可变形空洞卷积的ASPP模块,TASPP模块能使算法的MIoU提高1.37个百分点,FPS提升了0.08;对比实验 IV、VI 可以发现,引入TASPP模块能在损失少

表4 不同模块对算法性能的影响

Table 4 Effects of different modules on algorithm performance

Group	Algorithm	MIoU / %	FPS
I	-	77.84	17.19
II	CBAM	78.16	17.04
III	CBAM+ARFE	79.23	16.45
IV	CBAM+ARFE+FE	80.37	16.62
V	CBAM+ARFE+FE+ASPP	80.89	16.13
VI	CBAM+ARFE+FE+TASPP	82.26	16.21

部分参数量情况下进一步提升模型的预测精度,最终的MIoU比实验 I 提高了4.42个百分点,FPS为16.21,很好地平衡了分割精度与速度,满足实际工程的应用需求。

3.5 泛化实验

为验证本算法的有效性,将本算法与语义分割领域中主流的几种算法进行对比实验。测试实验硬件环境相同,其他算法模型通过官方公布的源码进行测试,在PASCAL VOC 2012数据集上的实验结果如表5所示。可以发现,相比算法 I、II、III,本算法的MIoU分别提高了19.46、3.41、2.84个百分点,进一步验证了本算法的有效性。

表5 不同算法在PASCAL VOC 2012数据集上的性能
Table 5 Performance of different algorithms on PASCAL VOC 2012 dataset

Group	Algorithm	Backbone network	MIoU / %
I	FCN-8s	VGG-16	61.57
II	PSPNet	ResNet101	77.62
III	DeepLabv3	ResNet101	78.19
IV	ours	ResNet50	81.03

4 结 论

针对安检包裹内物品摆放复杂、多尺度违禁品导致的安检识别难度高问题,提出了一种安检场景下的多目标违禁物品识别算法。首先,引入注意力机制主要关注目标位置,辅助实现对违禁品更精确的定位;然后,用FE模块提升特征间的相互依赖性;最后,用ARFE和TASPP模块强化网络对尺度变化物体的自适应能力,获取更全面的空间信息。在X光多目标违禁品语义分割数据集上的实验结果表明,相比几种经典的语义分割算法,本算法能有效避免棋盘格现象,提高多尺度多目标违禁品的分割能力,在安检违禁品分割任务中的表现较好,具有一定的工程应用价值。

参 考 文 献

- [1] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [2] 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, 13(9): 2203-2215.
- [3] Liu J Y, Leng X X, Liu Y. Deep convolutional neural network based object detector for X-ray baggage security imagery[C]//2019 IEEE 31st International Conference on Tools with Artificial Intelligence, November 4-6, 2019, Portland, OR, USA. New York: IEEE Press, 2019: 1757-1761.
- [4] Miao C J, Xie L X, Wan F, et al. SIXray: a large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 2114-2123.
- [5] Xu M S, Zhang H G, Yang J F. Prohibited item detection in airport X-ray security images via attention mechanism based CNN[M]//Lai J H, Liu C L, Chen X L, et al. Pattern recognition and computer vision. Lecture notes in computer science. Cham: Springer, 2018, 11257: 429-439.
- [6] Gómez D, Yáñez J, Guada C, et al. Fuzzy image segmentation based upon hierarchical clustering[J]. Knowledge-Based Systems, 2015, 87: 26-37.
- [7] An J Y, Zhang H G, Zhu Y, et al. Semantic segmentation for prohibited items in baggage inspection[M]//Cui Z, Pan J S, Zhang S S, et al. Intelligence science and big data engineering. visual data engineering. Lecture notes in computer science. Cham: Springer, 2019, 11935: 495-505.
- [8] 鲍海龙, 万敏, 刘忠祥, 等. 基于区域自我注意力的实时语义分割网络[J]. 激光与光电子学进展, 2021, 58(8): 0810018.
Bao H L, Wan M, Liu Z X, et al. Real-time semantic segmentation network based on regional self-attention[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810018.
- [9] 程晓悦, 赵龙章, 胡穹, 等. 基于膨胀卷积平滑及轻型上采样的实时语义分割[J]. 激光与光电子学进展, 2020, 57(2): 021017.
Cheng X Y, Zhao L Z, Hu Q, et al. Real-time semantic segmentation based on dilated convolution smoothing and lightweight up-sampling[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021017.
- [10] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [11] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1520-1528.
- [12] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [13] Paszke A, Chaurasia A, Kim S, et al. ENet: a deep

- neural network architecture for real-time semantic segmentation[EB/OL]. (2016-06-07) [2021-06-04]. <https://arxiv.org/abs/1606.02147>.
- [14] Wang Y, Zhou Q, Liu J, et al. Lednet: a lightweight encoder-decoder network for real-time semantic segmentation[C]//2019 IEEE International Conference on Image Processing, September 22-25, 2019, Taipei, Taiwan, China. New York: IEEE Press, 2019: 1860-1864.
- [15] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.
- [16] 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.
- [17] Zhang R, Tang S, Lin M, et al. Global-residual and local-boundary refinement networks for rectifying scene parsing predictions[C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, August 19-26, 2017, Melbourne, Australia. California: International Joint Conferences on Artificial Intelligence Organization, 2017: 3427-3433.
- [18] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 801-818.
- [19] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [20] 郭列, 张团善, 孙威振, 等. 融合空间注意力机制的图像语义描述算法[J]. 激光与光电子学进展, 2021, 58(12): 1210030.
- Guo L, Zhang T S, Sun W Z, et al. Image semantic description algorithm with integrated spatial attention mechanism[J]. Laser & Optoelectronics Progress, 2021, 58(12): 1210030.
- [21] Everingham M, Eslami S M A, van Gool L, et al. The pascal visual object classes challenge: a retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.