第 59 卷 第 4 期/2022 年 2 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

基于金字塔卷积和带状池化的X光目标检测

乔靖乾，张良*

中国民航大学电子信息与自动化学院，天津 300300

摘要 安检X光图像违禁品尺度多变、姿态各异，为自动识别带来很大的困难。针对该问题，提出了一种基于金字塔卷积和带状池化的X光目标检测算法。首先，以一阶段无锚框目标检测框架CenterNet为基础，引入金字塔卷积，提出金字塔沙漏网络，丰富Hourglass-104特征提取网络的感受野，增强多尺度特征提取能力。其次，带状池化的引入能够捕捉图像上下文全局信息，防止无关区域的信息干扰，兼顾局部细节信息。最后，在训练过程中将预测目标尺度分支的训练损失替换为交并比(IoU)损失函数，进一步提升尺度预测分支的性能。消融实验结果表明，改进后网络的平均精度(mAP50)由86.6%提升为88.3%，准确率有显著提升。

关键词 图像处理；X光图像目标检测；深度学习；金字塔卷积；带状池化；交并比损失函数

中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP202259.0410017

X-Ray Object Detection Based on Pyramid Convolution and Strip Pooling

Qiao Jingqian, Zhang Liang*

College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, *China*

Abstract The scale of contraband in security X-ray image is changeable and its posture is different, which brings great difficulties to automatic identification. To address this problem, an X-ray target detection algorithm based on pyramid convolution and strip pooling is proposed. First, pyramid convolution is introduced based on CenterNet, a one-stage anchor free frame target detection framework. Then, a pyramid hourglass network is proposed to enrich the receptive field of the hourglass-104 feature extraction network and enhance the ability of multi-scale feature extraction. Second, the introduction of strip pooling can capture the global information of the image context. It can also prevent information interference of irrelevant areas and consider local detail information. Finally, to enhance the performance of the scale prediction branch in the training process, the training loss of the prediction target scale branch is replaced by the intersection over union (IoU) loss function. The ablation experiment results show that the average accuracy (mAP50) of the enhanced network is improved from 86.6% to 88.3%, and the accuracy is significantly improved.

Key words image processing; X-ray image object detection; deep learning; pyramid convolution; strip pooling; intersection over union loss function

1 引 言

目前，通过X光安检机检查旅客行李和货运物品是否含有违禁品，是维护运输安全、保障公共场所免受恐怖主义威胁的重要手段。深度学习的快速发展使计算机自动、快速、高效进行图像判别成

收稿日期：2021-02-05；**修回日期**：2021-03-31；**录用日期**：2021-04-07

基金项目：国家自然科学基金(61179045)

通信作者：*l-zhang@cauc.edu.cn

为可能。同时，X光图像自动检测存在许多挑战。违禁物品的种类繁多且形态、体积差异较大，同一类别也存在不同的尺度，这也使得X光图像中违禁品的尺度、纵横比多种多样。此外，旅客行李包裹中物品的摆放方式是任意的，物体姿态也各异[1-2]。

无锚框的CenterNet，相较于其他算法存在许多优势。CenterNet将物体的定位任务简化为物体关键点预测问题，省略先验框的计算过程；通过预测物体中心点和尺寸对物体边界框进行建模，舍弃锚框（Anchor）[3]的概念，无需设定大量的超参数，模型在数据集之间的迁移也更容易。在预测过程中，CenterNet不需要为前景、背景分类设定阈值；在后处理阶段中，不需要采用非极大值抑制算法（NMS）[4]筛选预测结果。精简的推理过程有利于精度与速度方面达到更好的平衡。

本文在CenterNet目标检测框架基础上，引入金字塔卷积[5]，提出了金字塔沙漏网络；在网络模型中加入带状池化[6]，设计了带状池化检测网络。将原网络预测物体宽高的Smooth L1回归损失替换成交并比（IoU）损失[7]。通过设计金字塔残差卷积块，搭建特征提取网络金字塔沙漏网络（Py-Hourglass-104），所提网络在检测过程中具有更丰富的感受野，可以有效增强网络对多尺度特征的提取能力。基于带状池化模块的检测网络能够沿着水平和垂直空间维度，有效地捕获复杂场景中的长距离上下文信息，同时抑制其他无关区域干扰预测结果。在使用Smooth L1损失的网络训练过程中，宽度和高度两个参数独立学习，缺乏相关性；IoU损失函数可以建立起两个参数学习过程之间的联系。

2 基本原理

2.1 相关工作

2.1.1 Anchor-based目标检测算法

Anchor-based算法将检测任务简化为锚框切割图像后，对产生的大量子特征图的分类任务。通过预先设定的超参数，在图像上逐像素点滑动，按照锚框截取特征图，再对截取后的特征图进行分类。根据是否对子特征图进行筛选操作，Anchor-based算法又分为一阶段和二阶段。

目前，大多数R-CNN检测网络，均是在Faster R-CNN[8]的基础上改进的。通过引入区域推荐网络（RPN），将框的分类分为两个阶段，先对每个锚框进行二分类，判断框中内容是否包含物体，之后根据分类结果，进行感兴趣区域池化，并将池化结果送入检测器，得出精确的分类、位置结果。一阶段算法，省略区域推荐策略，在特征图上均匀地设置稀疏的锚框直接回归出物体类别和位置，能够满足实时检测的需求，例如YOLO-v3[9]、SSD[10]及RetinaNet[11]。YOLO-v3基于图像的全局信息进行预测，对608×608尺寸的图像速度可达到20 frame/s，在COCO数据集上平均精度（mAP）达到57.9%。SSD使用多尺度特征图进行预测，以适应物体的多尺度特性，与Faster-R-CNN精度相当。但因为采样方式，一阶段算法存在着先验框正负样本不平衡的问题。

2.1.2 Anchor-free目标检测算法

先验框大小、纵横比及数量设定会影响算法运行速度及预测效果，不同任务需要设定不同参数。使用同一特征图解决多任务预测问题会带来特征冲突，锚框与特征不对齐的问题。分类任务要求特征具有平移与尺度不变性，定位任务要求模型根据位置、尺度的不同产生不同响应。Anchor-free目标检测算法避免上述的问题。Law等[12]提出的CornerNet将目标边界框视为一对关键点的组合，从而预测出表示角点位置的两组高斯热图。Zhou等[13]认为CornerNet所预测的角点通常不在物体内部，这为特征提取和检测带来了困难，因此有了ExtremeNet，ExtremeNet预测目标边界的上、下、左、右4边界点。Zhou等[14]又提出用目标中心点表示物体的CenterNet，该网络无需关键点匹配便可产生边界框，简单、高效。

2.2 CenterNet网络结构及算法

预测阶段，CenterNet旨在根据输入X光图像$\hat{\boldsymbol{Y}}\in[0,1]^{(W\times H\times C)}$，生成一组关键点高斯热图$\hat{\boldsymbol{Y}}\in[0,1]^{\left(\frac{W}{S}\times\frac{H}{S}\times C\right)}$，可视化结果如图1所示，其中$S$是整体网络为提取语义信息、减小计算量所采用的下采样率，通常设$S=4$，C为需检测的违禁品的种类数。热图中，预测点$\hat{Y}_{x,y,c}=1$代表着该点为物体中心点，反之，若预测点$\hat{Y}_{x,y,c}=0$则认为该点为背景点。

训练阶段，图像中每一个真实物体中心点的位置为$\boldsymbol{p}\in\mathbf{R}^2$。根据下采样率$S$，得到一个低分辨率下的等效坐标$\tilde{\boldsymbol{p}}=\left\lfloor\frac{\boldsymbol{p}}{S}\right\rfloor$。并以该点为高斯中心，使用高斯核$Y=\exp\left[-\frac{(x-\tilde{p}_x)^2+(y-\tilde{p}_y)^2}{2\sigma_p}\right]$计算出

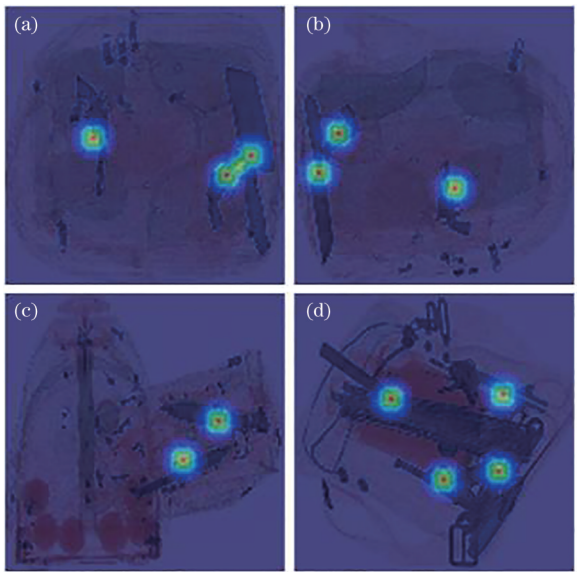

图 1　中心关键点高斯热图。(a)样例 1;(b)样例 2;(c)样例 3;(d)样例 4

Fig. 1　Gauss heat map of central key points. (a) Example 1; (b) example 2; (c) example 3; (d) example 4

图中各点的热力值,其中 σ_p 大小与目标物体尺寸相关。对于非中心点,热力值为根据多个高斯中心计算结果的最大值。而下采样导致的位置偏差则由中心点偏置分支进行预测。

CenterNet 模型结构与推理过程如图 2 所示。网络包含 Hourglass-104 骨干网络、预处理子网络、特征融合子网络及预测子网络 4 大部分。首先,通过由 7×7 大小的卷积层和 3×3 大小的残差卷积块组成的预处理子网络对原图进行 4 倍下采样产生 $\boldsymbol{L}\in\mathbf{R}^{\left(\frac{W}{S}\times\frac{H}{S}\times 256\right)}$。其次,将原图送入编-解码骨干网络 Hourglass-104 中进行特征提取,生成与 $\boldsymbol{L}$ 大小相同的特征图 $\boldsymbol{K}\in\mathbf{R}^{\left(\frac{W}{S}\times\frac{H}{S}\times 256\right)}$。最后,将 $\boldsymbol{K}$ 与 $\boldsymbol{L}$ 传递给特征融合网络,进行特征融合后送入下一个 Hourglass-104 产生特征图,两组特征利用各自检测网络预测中心点热图、尺寸及中心点偏置。

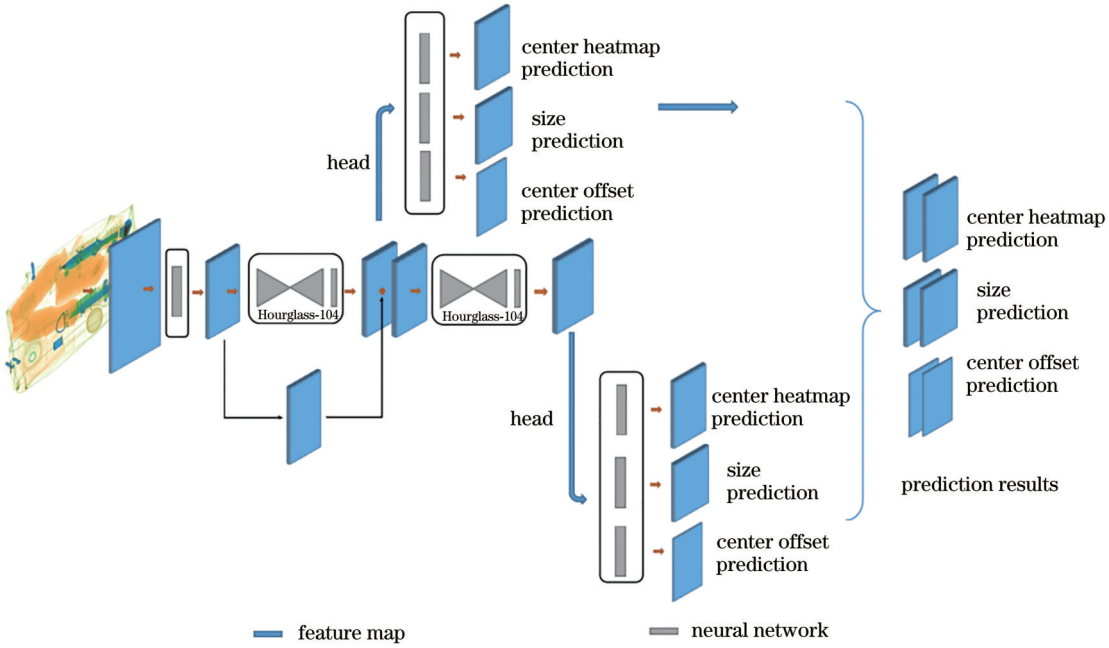

图 2　CenterNet 检测算法

Fig. 2　CenterNet detection algorithm

Hourglass-104 模块主要由提取特征的残差卷积层、2 倍下采样池化层及上采样层组成,如图 3 所示。网络处理流程主要包括上采样过程和下采样过程。根据上采样产生和下采样输入的特征图尺度是否相同,将 Hourglass-104 分为 5 阶 Hourglass 模块。下采样过程,输入 $\boldsymbol{D}\in\mathbf{R}^{W\times H\times C}$ 通过残差卷积块后得到大小不变的特征图 $\boldsymbol{D}_{\mathrm{b}}\in\mathbf{R}^{W\times H\times C}$,在特征提取的同时保留细节特征;同时对输入进行池化操作,通过另外一个卷积残差块,得到结果 $\boldsymbol{D}'\in\mathbf{R}^{\frac{W}{2}\times\frac{H}{2}\times C}$,并输入到下一阶 Hourglass 模块。上采样过程,输入 $\boldsymbol{U}\in\mathbf{R}^{W\times H\times C}$ 经过上采样层处理后得到 $\boldsymbol{U}'\in\mathbf{R}^{2W\times 2H\times C}$,与同一阶的 $\boldsymbol{D}_{\mathrm{a}}\in\mathbf{R}^{W\times H\times C}$ 相加后输入下一阶 Hourglass 模块。Hourglass-104 包含 5 阶 Hourglass 模块,输出结果的大小为 128×128。

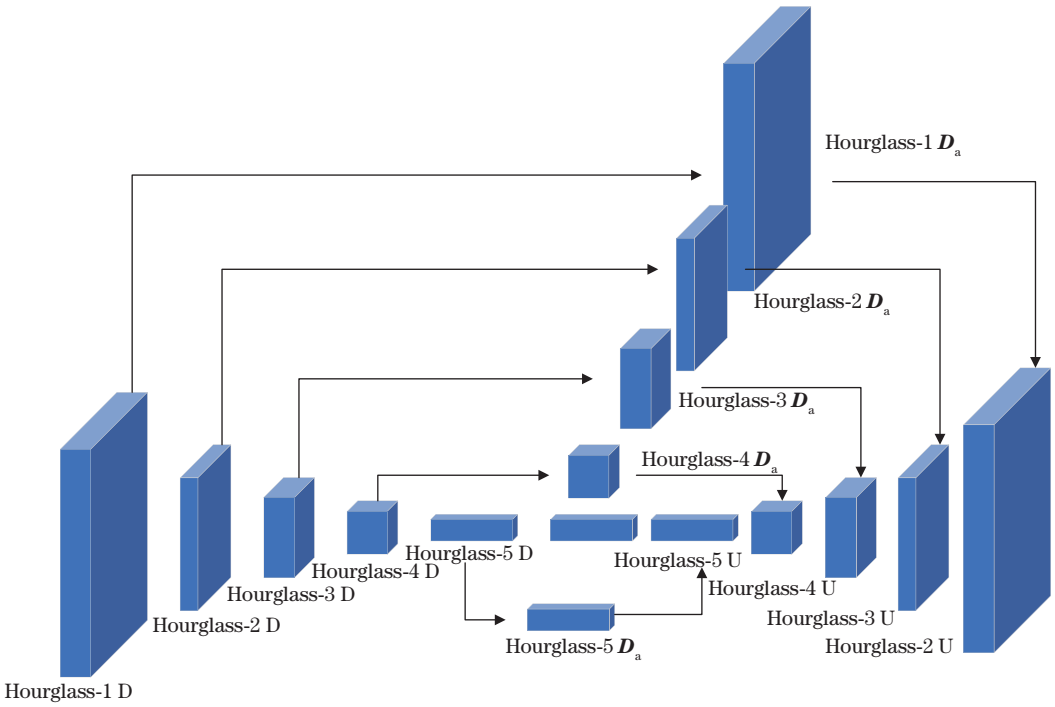

图 3　Hourglass-104 结构

Fig. 3　Structure of Hourglass-104

3　主要工作

3.1　金字塔卷积

3.1.1　金字塔卷积

Luo 等[15]指出，在实际模型中感受野内所有像素对输出特征图的贡献并不相同，有效感受野仅仅是理论感受野的一部分。Li 等[16]指出，深度网络模型对于不同尺寸物体的检测性能与网络的感受野呈正相关，即较大的感受野对于大尺度物体的检测性能会更好，对于小物体而言，较小的感受野更有利。2020 年，Duta 等[5]首次提出金字塔卷积的概念，并在 ResNet 中加入金字塔卷积，增强了感受野并丰富了特征。而 Hourglass 相较于 ResNet，具有特殊的编-解码结构。针对 Hourglass-104，本实验组提出金字塔沙漏网络，探索更适宜 Hourglass-104 骨干网络的金字塔卷积使用方案。

金字塔结构如图 4 所示，与标准卷积相比，卷积核自上而下，大小、深度依次增加，每一层均对整个输入进行处理。卷积核尺度的不同导致每一层的感受野各不相同，能够提取到不同层次的特征，既有 1×1、3×3 大小的卷积核来学习细节信息，又有 5×5、7×7 大小的卷积核学习较大视野中的语义信息。另外，网络能够感知不同层次的特征在空间上的联系，这也使得卷积层具有保留细节特征的能力，有效缓解下采样导致的局部信息丢失问题。

网络预测过程中，特征图尺度按照 2 倍下采样率依次下降，再随着 2 倍上采样逆序递增。显然，对于浅层特征图而言，细节信息丰富，与原图像较接近，金字塔卷积中每一层可以使用较大的卷积核来增强感受野。而对于深层特征，大小为 7×7、9×9 的卷积核均与输入特征图大小相近，特征提取效果不理想。为此，引入了 4 种金字塔卷积层，依次为浅层、浅中层、中层、深层，组成相应的金字塔残差卷积块（PyConv_residual_block），如图 5 所示。

将 Hourglass-1 模块中残差连接中的卷积块用浅层金字塔残差卷积块替换，Hourglass-2 模块中的残差块用浅中层浅层金字塔残差卷积块替换，Hourglass-3、Hourglass-4 中的残差块用中层金字塔残差卷积块替换，Hourglass-5 中的残差块用深层金字塔残差卷积块替换。所得新的金字塔沙漏网络如图 6 所示。

3.2　带状池化

3.2.1　带状池化（strip pooling）

能够有效理解 X 光图像中的复杂场景，对图像中像素点间关系进行建模的方法通常有两种。1）在网络模型中引入 self-attention[17]或者 non-local[18]机制，但每个空间位置之间的关系矩阵计算所带来的大计算量削弱了算法的实时性；2）空洞卷

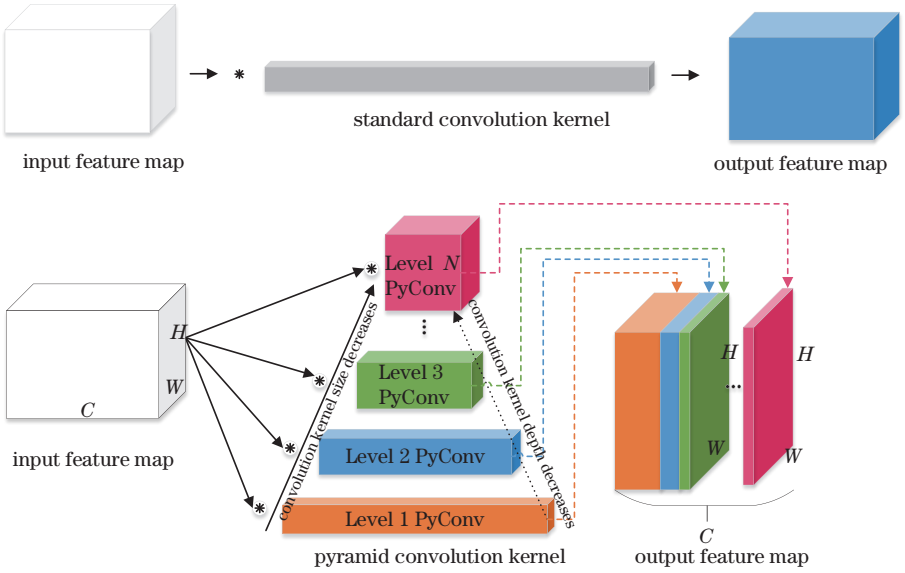

图 4　标准卷积和金字塔卷积

Fig. 4　Standard convolution and Pyramid convolution

积[19]的感受野让模型可以在训练中学习上下文信息。这些方法均是在正方形窗口内对输入特征图进行特征提取的。当存在尺度较大的带状物体(扳手、钳子)时,方形滑动窗口会引入无关区域的干扰信息,限制了网络在捕获各向异性上下文信息的灵活性。Hou 等[6]于 2020 年首次提出带状池化的概念,将其加入 ResNet 中,为场景图像分割视觉任务设计了 SPNet。针对目标检测,本实验组没有使用带状池化进行特征提取,而是将带状池化加入检测网络中,设计了带状池化检测子网络。

3.2.2　带状池化模块(strip pooling module)

带状池化核能够有效地建立窗口间的依赖关系,并对窗口内的特征图进行编码。同时,与其他维度的特征图融合能够对特征图的细节信息进行建模。带状池化模块如图 7 所示。

首先,使用水平池化核和垂直池化核分别对输入 $\boldsymbol{x}\in\mathbf{R}^{W\times H\times C}$ 进行平均池化操作,输出分别为 $\boldsymbol{y}^{h}_{c,i}\in\mathbf{R}^{C\times H}$、$\boldsymbol{y}^{v}_{c,j}\in\mathbf{R}^{C\times W}$;然后,使用卷积核为 3 的一维卷积对一维卷积窗口内特征进行编码,扩展后生成 $\boldsymbol{y}^{H}_{c,i,j}\in\mathbf{R}^{W\times H\times C}$、$\boldsymbol{y}^{W}_{c,i,j}\in\mathbf{R}^{W\times H\times C}$;将其进行特征融合,有

$$\boldsymbol{y}_{c,i,j}=\boldsymbol{y}^{H}_{c,i,j}+\boldsymbol{y}^{W}_{c,i,j}。\tag{1}$$

通过 1×1 卷积与 sigmoid 激活层后,与输入原特征图相乘,所得输出特征图 $\boldsymbol{y}$ 上每个位置的特征值与输入的多个位置相关,从而达到对上下文建模的目的。带状池化的感受野长而窄,避免了在长距离像素点间建立过多的不必要的连接,这样也有效地抑制无关信息的引入。

3.2.3　带状池化预测子网络(strip pooling head)

模型引入带状池化,它在提取长距离依赖关系的同时,防止无关区域的信息干扰,能够兼顾局部细节信息。在关键点、尺度、中心点偏置 3 分支预测网络中加入带状池化模块,形成带状池化检测网络(strip pooling head)。

图 8 为所设计的两种方案,区别在于带状池化模块是否由 3 个分支共享。在性能方面,共享方案在极大程度上保持了原算法的运算速率,精度提升不高;而不使用参数共享的方案会使精度大幅度提升,但运算速率会有下降。

3.3　Intersection over union (IoU)损失

3.3.1　原网络训练损失

原网络在训练过程中,对于中心点热图的预测,学习过程采用逐像素点计算的 focal 损失进行监督。

$$L_k=-\frac{1}{N}\sum_{xyc}\begin{cases}\left(1-\hat{Y}_{xyc}\right)^{\alpha}\log\hat{Y}_{xyc}, & Y_{xyc}=1\\ \left(1-Y_{xyc}\right)^{\beta}\hat{Y}_{xyc}^{\ \alpha}\log\left(1-\hat{Y}_{xyc}\right), & Y_{xyc}\neq1\end{cases},\tag{2}$$

式中:α 和 β 为 focal 损失的超参数。而对于尺度预测及偏置预测,学习过程则采用 L1 损失[20]进行监督。

$$L_{\text{off}}=\frac{1}{N}\left|\hat{O}_{\tilde{p}}-\left(\frac{p}{R}-\tilde{p}\right)\right|,\tag{3}$$

$$L_{\text{size}}=\frac{1}{N}\sum_{k=1}^{N}\left|\hat{S}_{pk}-S_k\right|,\tag{4}$$

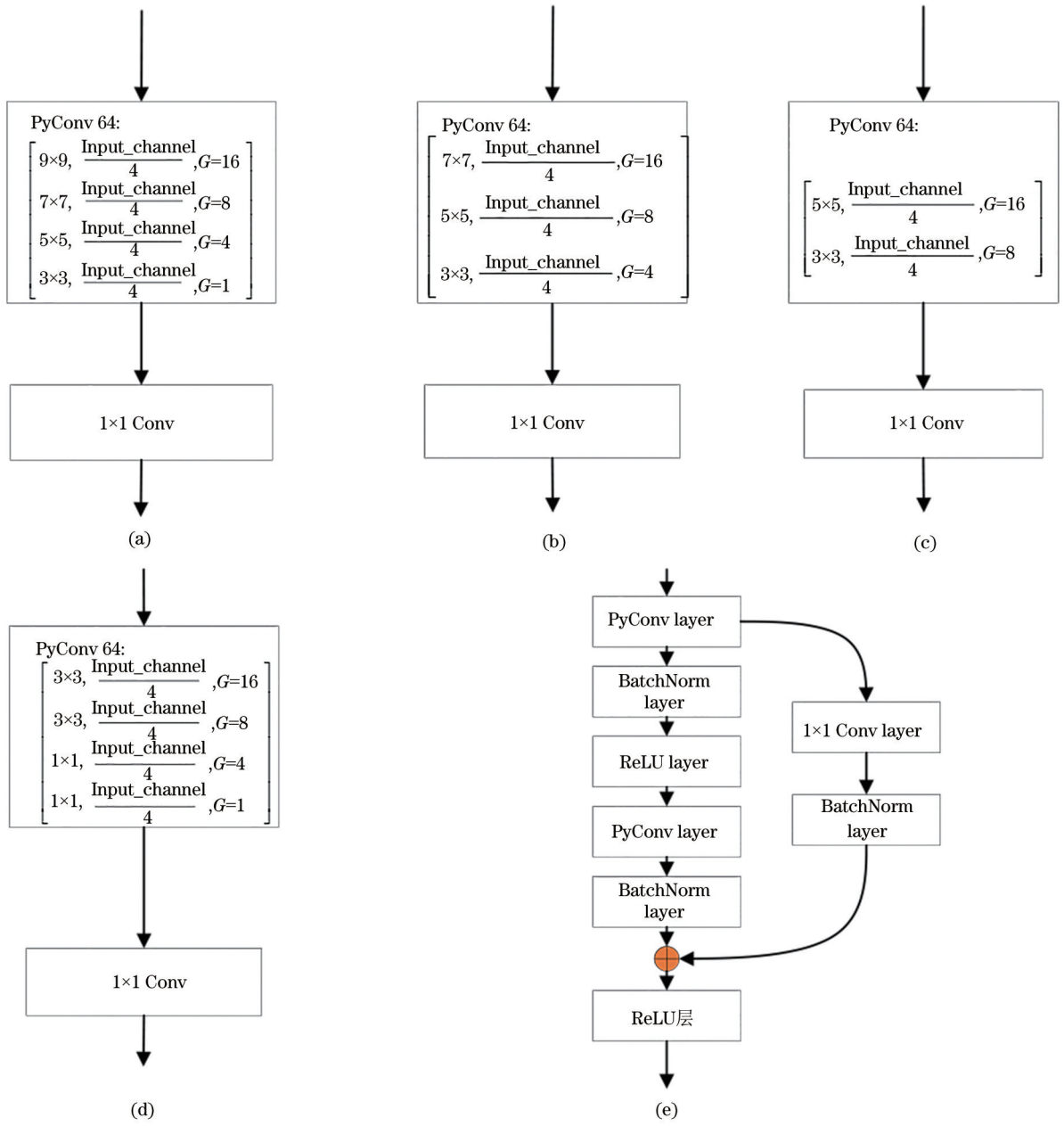

图 5　金字塔卷积核结构与金字塔卷积残差块结构。(a)浅层金字塔卷积;(b)浅中层金字塔卷积;(c)中层金字塔卷积;(d)深层金字塔卷积;(e)金字塔卷积残差,

Fig. 5　Pyramid convolution kernel structure and pyramid convolution residual block structure. (a) Shallow layer pyramid convolution;(b) shallow middle layer pyramid convolution; (c) middle layer pyramid convolution; (d) deep layer pyramid convolution; (e) pyramid convolution residual block

式中:$\hat{O}_p$为偏置分支模型预测的值;$\frac{p}{R}$为位置坐标除以下采样率的位置;$\tilde{p}$为特征图目标中心点的位置;$\hat{S}_{pk}$为尺度分支预测的高、宽;S_k为真实框的高、宽,对于任一类别的物体,它的宽高 $S_k=\left(x_{\max}^{(k)}-x_{\min}^{(k)},y_{\max}^{(k)}-y_{\min}^{(k)}\right)$。而总体监督学习的损失,则是各项损失的加权和。

$$L_{\det}=L_{\mathrm{k}}+\lambda_{\mathrm{size}}L_{\mathrm{size}}+\lambda_{\mathrm{off}}L_{\mathrm{off}},\tag{5}$$

式中:λ_{size}、λ_{off}是对应损失的权重,通常取 $\lambda_{\mathrm{size}}=0.1$、$\lambda_{\mathrm{off}}=1$。

3.3.2　IOU 损失

在训练过程中对边界框大小的回归分为高、宽两分支。两分支在学习过程中相互独立,分别与对应长度真值和宽度真值计算各自的损失并求和。但在现实中,长度和宽度特征并不独立,往往与物体类别息息相关,包含着丰富的特征。另外,判断

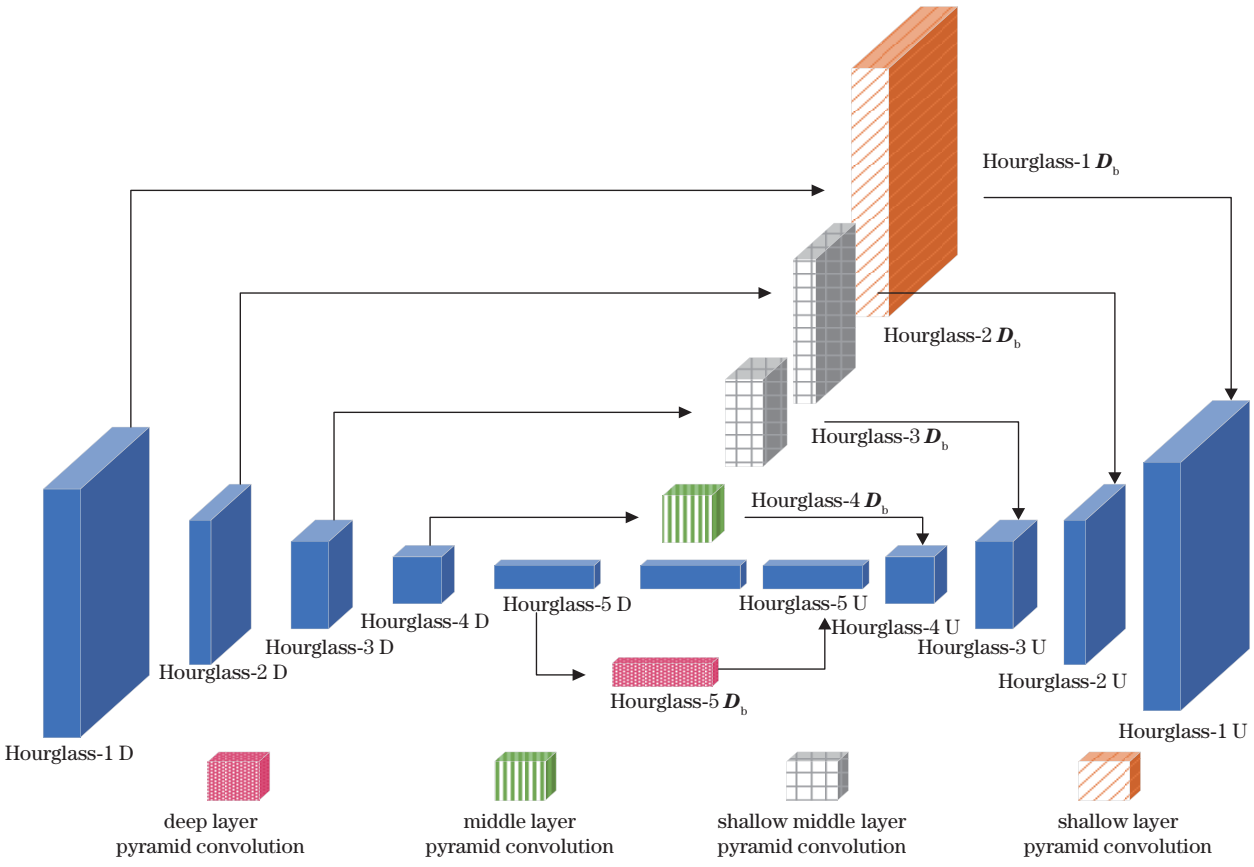

图 6　金字塔沙漏网络结构

Fig. 6　Pyramid Hourglass-104 network structure

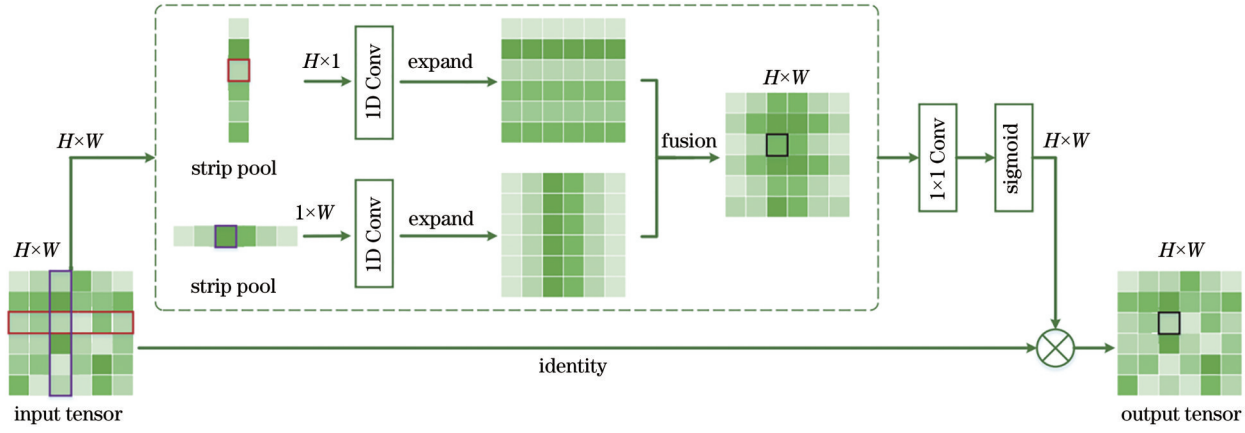

图 7　带状池化模块

Fig. 7　Strip pooling module

物体是否检测及精准与否，往往通过检测框与真实框的交并比与阈值的比较来进行，大于则认为检测成功，反之则失败。对于这种评价标准，高度、宽度独立学习显然并不适合。

为此，将监督尺度分支学习的损失函数 L1 变为更适合尺度学习的 IoU 损失。

$$\mathrm{IoU}(A,B)=\frac{A\cap B}{A\cup B}=\frac{A\cap B}{|A|+|B|-A\cap B},\tag{6}$$

式中：A 为真实框所包围区域的面积 S，$S=W\times H$；B 为预测框所包围区域的面积。

则 IoU 损失的表达式为

$$L_{\mathrm{loss}}=1-\mathrm{IoU}(A,B)。\tag{7}$$

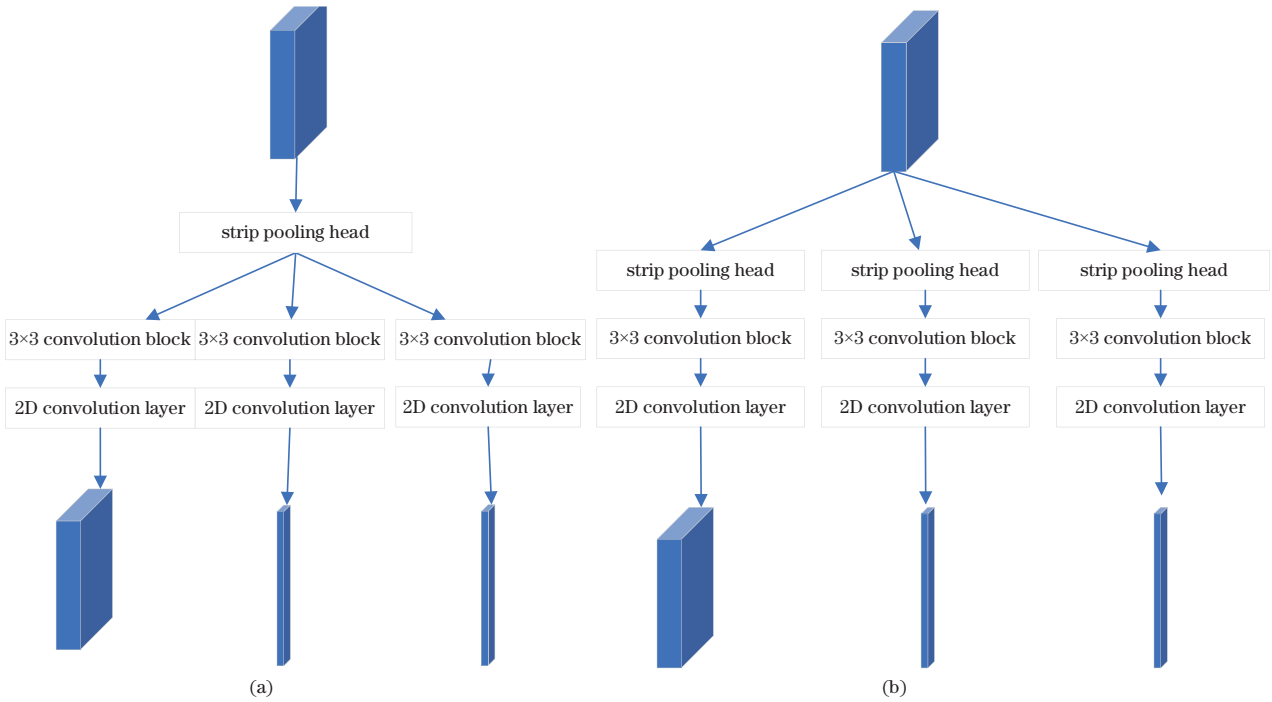

图 8　带状池化检测网络。(a)共享方案;(b)非共享方案

Fig. 8　Strip pooling head. (a) Sharing scheme; (b) unshared scheme

4　分析与讨论

4.1　实验设计与数据集介绍

首先,将 CenterNet 模型与经典目标检测模型进行对比,验证 CenterNet 在检测任务的精度优势;其次,针对所提 3 处设计分别进行消融实验,探索改动是否对算法有益;然后,对所提金字塔卷积核使用方案进行验证;最后,对带状池化模块的引入位置进行对比实验研究。

实验均是在惠普深度学习工作站、Ubuntu 16.04 系统下进行的。CPU 为 Intel®Xeon(R) Silver 4110,处理频率为 2.10 GHz,GPU 为 Nvidia GeForce GTX 1080 Ti×2,并搭配 CUDA 9.0 与 CUDNN 7.0.5 进行网络模型的运算加速。使用 Python 3.6 实现算法,pytorch 0.4.1 深度学习框架搭建网络模型。

训练轮数设为 140,输入图片尺寸默认为 512×512,网络的训练学习率设为 1.25×10^{-4},梯度下降算法采用批量梯度下降,批大小设为 7,每隔 4 轮对模型进行测试,比较损失,寻找最优模型。训练过程中采用随机剪裁、随机旋转及随机缩放的数据增强策略,由于 X 光图像的特殊性,颜色与物体类别相关联,因而并没有采用原算法中的颜色增强策略。

实验中所采用的数据集为 SIXray_OD 数据集[21],该数据集中的图片来自二分类公共数据集 SIXray。SIXray_OD 数据集中总共包含 5 类违禁品:枪(Gun)、刀(Knife)、扳手(Wrench)、剪刀(Scissors)、钳子(Pliers)。为了提高模型的泛化能力,通过反转、旋转、平移、亮度调节及对比度调节等数据增强方法,将数据集扩充。最终将 SIXray_OD 按照 4∶1 的比例分为训练集和测试集。包含各类违禁物图像数量统计信息如表 1 所示,图片样例如图 9 所示。

表 1　SIXray_OD 数据集统计信息

Table 1　Statistics of SIXray_OD dataset

Category	Knife	Scissors	Wrench	Gun	Pliers	Total
Number of images	12488	9272	31608	23488	18128	69744

4.2　Hourglass-104 网络模型预训练

为加快模型收敛,实验中没有使用随机初始化的网络参数,而是将原网络在 MS COCO 2017 数据集上进行复现,用 0.5 阈值下的平均精度(mAP)来

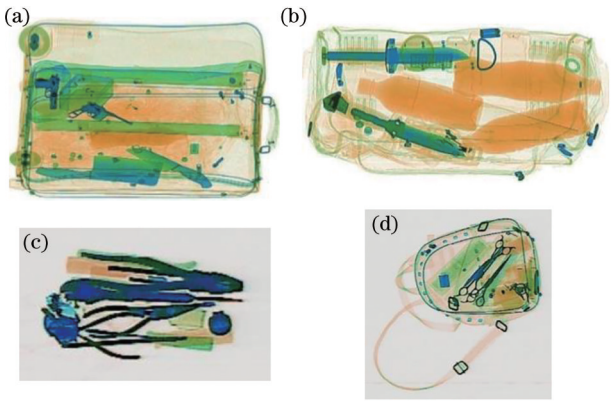

图 9　SIXray_OD 数据集。(a)样例 1;(b)样例 2;(c)样例 3;(d)样例 4

Fig. 9　SIXray_OD dataset. (a) Example 1; (b) example 2; (c) example 3; (d) example 4

表示算法的检测精度。复现模型的 mAP50 为 58.84%,mAP50 表示在 0.5 阈值下的平均精度。同时,将复现结果作为预训练模型,对本实验组所有实验进行初始化,加快其收敛速度。

4.3　消融实验

4.3.1　不同模型对比试验

首先将 CenterNet 网络与目前主流的目标检测网络在相同设备和工作环境下进行实验对比,表 2 为 SSD、YOLO、Faster RCNN、CenterNet 4 种目标检测网络在 SIXray_OD 数据集上的检测精度。从表 2 中可以看出,CenterNet 的检测精度达到 86.6%,比 YOLOv3、SSD、Faster R-CNN 高。

表 2　不同检测网络对比实验结果

Table 2　Comparative experimental results of different detection networks

Network	Backbone	mAP50 /%
SSD	VGG16	71.89
YOLOv3	DarkNet-53	64.34
Faster R-CNN	VGG16	78.41
CenterNet	Hourglass-104	86.6

4.3.2　各项调整对比实验

针对所提改动:金字塔卷积(Py_Hourglass_104)、带状池化检测网络(strip pooling head)及 IoU 损失替换的消融实验结果如表 3 所示.

实验结果表明,引入金字塔卷积与带状池化使模型的 mAP50 显著提升。Py_Hourglass_104 的单独引入可以使网络在 mAP50 上提升 0.7 个百分点,带状池化的加入会带来 0.9 个百分点的提升,二者相结合,所提升的精度达到 1.4 个百分点。因此,金字塔卷积和带状池化在一定程度上提升了模型对关键点的检测精度,模型可以更好地拟合出违禁品中心关键点在 X 光图像中的分布。此外,损失函数的单独调整也使得精度有所提升。

虽然大尺度卷积核能够增强网络感受野,但也会带来计算量增大、收敛难度大的问题。为探索最优的卷积核使用方案,对 3 种使用方案进行消融实验,方案细节如表 4 所示。

表 3　各改进点消融实验结果

Table 3　Ablation experimental results of each improvement point

Experiment	Py_Hourglass_104	Strip pooling	IoU loss	AP /%					
				mAP50	Gun	Knife	Pliers	Wrench	Scissors
CenterNet				86.6	95.69	90.44	88.21	83.62	75.04
Experiment 1	✓			87.3	96.00	91.23	88.65	84.53	76.09
Experiment 2		✓		87.5	96.21	91.16	89.12	84.07	76.94
Experiment 3			✓	87.4	95.86	90.64	89.54	85.72	75.25
Experiment 4	✓	✓		88.0	96.38	91.12	89.86	85.92	76.73
Experiment 5	✓		✓	87.7	96.12	91.53	89.05	85.18	76.62
Experiment 6		✓	✓	87.9	96.15	91.85	89.24	85.10	77.06
Experiment 7	✓	✓	✓	88.3	96.40	91.88	89.90	85.90	77.43

表 4　金字塔卷积方案

Table 4　Pyramid convolution scheme

Scheme	Pyramid convolution residual block used
Scheme 1	[shallow、shallow middle、shallow middle,middle、deep]
Scheme 2	[shallow、shallow、shallow middle、shallow middle、deep]
Scheme 3	[shallow、shallow、shallow middle、middle、deep]

从 1 阶至 5 阶，依次将 Hourglass-104 中每一阶 $\boldsymbol{D}_a$ 位置上的标准卷积块替换为方案中相应金字塔卷积残差块，$\boldsymbol{D}_a$ 为第 i 阶 Hourglass 模块中处理 Hourglass-$i\boldsymbol{D}$、产生 Hourglass-$i\boldsymbol{D}_a$ 的残差卷积块，其中 $i\in[1,2,3,4,5]$。实验结果如表 5 所示。

表 5　不同金字塔卷积方案对比结果

Table 5　Comparison results of different pyramid convolution schemes

Experiment	Scheme 1	Scheme 2	Scheme 3	mAP50 / %
CenterNet				86.6
Experiment 1	✓			87.3
Experiment 2		✓		87.0
Experiment 3			✓	86.9

从表中可以看出，3 种方案均对算法有积极的影响，方案 1 效果最佳，mAP50 提升 0.7 个百分点。

金字塔卷积残差块替换的位置对模型的精度有着很大的影响。为此同样需要进行消融实验探究，实验采用上述金字塔卷积使用方案中的方案 1。在 Hourglass-104 骨干网络的不同位置，将原有卷积块替换为金字塔卷积残差块。实验结果如表 6 所示，其中 $\boldsymbol{D}$ 位置为第 i 阶 Hourglass 模块中处理 Hourglass-$i\boldsymbol{D}$、产生 Hourglass-$(i+1)\boldsymbol{D}$ 的卷积残差块，其中 $i\in[1,2,3,4,5]$。在这两个位置中进行对比实验，从表 6 中可以看出：在 $\boldsymbol{D}_a$ 位置中加入金字塔卷积、$\boldsymbol{D}_a$ 位置与 $\boldsymbol{D}$ 位置同时加入金字塔卷积均会使模型的精度提升；而在 $\boldsymbol{D}$ 位置的修改会使算法 mAP50 降低 1.9 个百分点；实验 3 中的涨幅比实验 1 中少 0.3 个百分点。因此，替换 $\boldsymbol{D}_a$ 位置上的残差卷积块能够较好地发挥金字塔卷积的优势，对骨干网络特征表示能力提升最大。

表 6　金字塔卷积使用位置对比实验结果

Table 6　Experimental results of pyramid convolution using position comparison

Experiment	$\boldsymbol{D}_a$	$\boldsymbol{D}$	mAP50 /%
CenterNet			86.6
Experiment 1	✓		87.3
Experiment 2		✓	84.7
Experiment 3	✓	✓	87.0

同样，带状池化的位置设定也十分重要，为此，尝试在网络其他位置中加入带状池化模块。在网络中的 3 个位置引入带状池化模块，包括替换骨干网络位置 $\boldsymbol{D}_a$ 中的卷积层、预处理子网络及在检测网络。检测网络中采用图 8 中的两种方案，方案 1 共享带状池化卷积模块权重，而方案 2 各自带状池化模块相互独立。实验结果如表 7 所示，从表中可以看出，方案 1 与方案 2 均对骨干网络提取的特征图进行上下文语义信息的建模，为后续网络提供检测依据，对算法起到积极的作用，mAP50 分别提升 0.8 个百分点和 1.2 个百分点。在预处理子网络、骨干网络中加入带状池化模块后，算法表现不佳。

表 7　带状池化使用位置对比实验结果

Tab. 7　Experimental results of strip pooling using position comparison

Experiment	Backbone	Preprocessing subnet	Strip pooling head 1	Strip pooling head 2	mAP50 /%
No					86.6
Experiment 1	✓				85.6
Experiment 2		✓			86.4
Experiment 3			✓		87.4
Experiment 4				✓	87.8

经过实验测试，CenterNet 网络在 SIXray_OD 数据集上的检测速度约为 11 frame/s，PS-CenterNet 网络的单张图片检测速度约为 6 frame/s，在机场安检应用场景中可以满足实时检查的要求。

综上所述，金字塔卷积使用方案对比实验表明，所提算法使用方案效果最佳，与原算法相比，mAP 提升了 0.7 个百分点。而金字塔卷积位置对比实验表明，将 Hourglass-104 中所有下采样中全部卷积替换为金字塔卷积并不能为算法带来最大增益。原因在于原网络中特殊的编-解码结构，针对目标检测任务而言，在检测网络中引入带状池化对模型性能有益，而使用带状池化进行特征提取的效果不佳。此外，表 3 的多个消融实验说明，所提 3 项改动均是有效的。

图 10 为所提算法与原算法检测结果对比图。从图 10 中可以看出，原算法会漏检挡在刀、扳手后的剪刀，所提算法可以改善这种漏检现象。尽管图像内遮挡严重，所提算法也可以较好地完成检测任务。

图 10　检测结果对比图。(a) CenterNet;(b)所提算法

Fig. 10　Comparison of detection results. (a) CenterNet; (b) proposed algorithm

5　结　　论

在 CenterNet 检测框架基础上，针对 X 光安检图像特性，提出了一种基于金字塔卷积和带状池化的 X 光目标检测算法。所提算法通过引入金字塔卷积，丰富骨干网络 Hourglass-104 的感受野，提升网络的特征表示能力；通过引入带状池化模块，建立像素间的上下文关系，为后续检测网络提供检测依据。同时，在训练过程中采用 IoU 损失来监督生成框的尺度大小。通过一系列的改进，所提算法的精度取得了明显的提升，更加适宜 X 光图像目标检测。另外，所提金字塔沙漏网络是对骨干网络的感受野增强和特征丰富后的结果，对使用 Hourglass-104 网络进行特征提取的其他算法模型同样有益，可以扩展到其他视觉任务。

参　考　文　献

[1] Zhou B, Li R X, Shang Z H, et al. Object detection algorithm based on improved Faster R-CNN[J]. Laser & Optoelectronics Progress, 2020, 57(10): 101009.
周兵, 李润鑫, 尚振宏, 等. 基于改进的 Faster R-CNN 目标检测算法[J]. 激光与光电子学进展, 2020, 57(10): 101009.

[2] Guo R H, Zhang L, Yang Y, et al. X-ray image controlled knife detection and recognition based on improved SSD[J]. Laser & Optoelectronics Progress, 2021, 58(4): 0404001.
郭瑞鸿, 张莉, 杨莹, 等. 基于改进 SSD 的 X 光图像管制刀具检测与识别[J]. 激光与光电子学进展, 2021, 58(4): 0404001.

[3] Hosoya Y, Suganuma M, Okatani T. Analysis and a solution of momentarily missed detection for anchor-based object detectors[C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV), March 1-5, 2020, Snowmass, CO, USA. New York: IEEE Press, 2020: 1399-1407.

[4] Hosang J, Benenson R, Schiele B. Learning non-maximum suppression[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6469-6477.

[5] Duta I C, Liu L, Zhu F, et al. Pyramidal convolution: rethinking convolutional neural networks for visual recognition[EB/OL]. (2020-06-20)[2021-02-01]. https://arxiv.org/abs/2006.11538.

[6] Hou Q B, Zhang L, Cheng M M, et al. Strip pooling: rethinking spatial pooling for scene parsing [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 4002-4011.

[7] Rezatofighi H, Tsoi N, Gwak J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[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: 658-666.

[8] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[9] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-28) [2021-02-01]. https://arxiv.org/abs/1804.02767.

[10] Womg A, Shafiee M J, Li F, et al. Tiny SSD: a tiny single-shot detection deep convolutional neural network for real-time embedded object detection[C]// 2018 15th Conference on Computer and Robot Vision (CRV), May 8-10, 2018, Toronto, ON, Canada. New York: IEEE Press, 2018: 95-101.

[11] 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.

[12] Law H, Deng J. CornerNet: detecting objects as paired keypoints[J]. International Journal of Computer Vision, 2020, 128(3): 642-656.

[13] Zhou X Y, Zhuo J C, Krähenbühl P. Bottom-up object detection by grouping extreme and center points[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: 850-859.

[14] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points[EB/OL]. (2019-04-16)[2021-02-01]. https://arxiv.org/abs/1904.07850.

[15] Luo W J, Li Y J, Urtasun R, et al. Understanding the effective receptive field in deep convolutional neural networks[EB/OL]. (2017-01-15) [2021-02-01]. https: //arxiv.org/abs/1701.04128.

[16] Li Y H, Chen Y T, Wang N Y, et al. Scale-aware trident networks for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6053-6062.

[17] Perreault H, Bilodeau G A, Saunier N, et al. SpotNet: self-attention multi-task network for object detection[C]//2020 17th Conference on Computer and Robot Vision (CRV), May 13-15, 2020, Ottawa, ON, Canada. New York: IEEE Press, 2020: 230-237.

[18] Wang X L, Girshick R, Gupta A, et al. Non-local neural 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: 7794-7803.

[19] Wang Z Y, Ji S W. Smoothed dilated convolutions for improved dense prediction[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19-23, 2018, London, United Kingdom. New York: ACM, 2018: 2486-2495.

[20] Zhao H, Gallo O, Frosio I, et al. Loss functions for image restoration with neural networks[J]. IEEE Transactions on Computational Imaging, 2017, 3(1): 47-57.

[21] Guo S X, Zhang L. Yolo-C: one-stage network for prohibited items detection within X-ray images[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810003.
郭守向, 张良. Yolo-C: 基于单阶段网络的X光图像违禁品检测[J]. 激光与光电子学进展, 2021, 58(8): 0810003.