

# 多尺度特征融合的安检图像危险品检测

王昱晓, 张良\*

中国民航大学天津市智能信号与图像处理重点实验室, 天津 300300

**摘要** 现有的目标检测算法检测 X 光安检图像中较小尺寸的危险品精度较低, 为此提出一种多尺度特征融合检测网络, 即 MFFNet (Multi-scale Feature Fusion Network), 其以 SSD 检测模型为基础并采用更深的特征提取网络, 即 ResNet-101。通过跳跃连接的方式将网络的高层语义丰富特征与低层边缘细节特征进行融合, 为小尺度危险品的检测添加上下文信息, 可以有效提升对小尺度目标的识别与定位精度。将融合得到的新特征层与 SSD 扩展卷积层一起送入检测。实验结果表明, MFFNet 能够使 X 光安检图像中的危险品特别是较小尺寸的危险品, 检测精度得到较大的提升, 同时能够保持相对较快的检测速度, 满足现代化安检的要求。

**关键词** 图像处理; 特征融合; X 光安检图像; 危险品检测

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.0810012

## Dangerous Goods Detection Based on Multi-Scale Feature Fusion in Security Images

Wang Yuxiao, Zhang Liang\*

Tianjin Key Laboratory of Intelligent Signal and Image Processing, Civil Aviation University of China, Tianjin 300300, China

**Abstract** Existing target detection algorithms have low accuracy in detecting smaller-sized dangerous goods in X-ray security inspection images. Therefore, a multi-scale feature fusion detection network called MFFNet (Multi-scale Feature Fusion Network) is proposed, which is based on the SSD detection model and uses a deeper feature extraction network, namely ResNet-101. The high-level semantic rich features of the network are merged with the low-level edge detailed features through the jump connection method, and contextual information is added for the detection of small-scale dangerous goods, which can effectively improve the identification and positioning accuracy of small scale targets. The new feature layer obtained by fusion and the SSD extended convolution layer are sent into detection together. Experimental results show that MFFNet can greatly improve the detection accuracy of dangerous goods in X-ray security inspection images, especially those of smaller sizes, while maintaining a relatively fast detection speed to meet the requirements of modern security inspection.

**Key words** image processing; feature fusion; X-ray security image; dangerous goods detection

**OCIS codes** 100.2000; 100.3008; 100.4996

## 1 引言

公共场所的安全问题一直备受关注, 不管是过去还是现在, 不管是国内还是国外, 在出行场所均需要对旅客行李进行安全检查, 而检查所用的设备大

都是双能 X 射线安检设备<sup>[1]</sup>。当物品通过该设备时, 利用了 X 射线的穿透性对行李内的物品进行检测并以图片的形式自动呈现出来, 但是对行李内物品的判别仍然需要人工操作。由于 X 光安检图像常常复杂多变, 大量的交通需求也给安检人员带来

收稿日期: 2020-08-10; 修回日期: 2020-09-04; 录用日期: 2020-09-15

基金项目: 国家自然科学基金(61179045)、民航安全能力建设项目(20600523)

\* E-mail: l-zhang@cauc.edu.com

了巨大的工作压力,稍有不慎就容易造成错检和漏检。因此,X光安检图像中危险品的自动检测研究具有重要意义。

传统的 X 射线图像处理方法是利用人工设计的特征对物体进行分类。早期,Bastan 等<sup>[2]</sup>采用了视觉词袋(BoVW)方法对 X 射线行李图像进行分类,但其分类性能仍与普通可见光图像差距较大。2015 年,Bastan<sup>[3]</sup>对此前关于 X 射线危险品图像的检测工作进行了回顾和总结,指出神经网络将会成为解决 X 射线图像检测的方法之一。

近年来,随着人工智能技术在日常生活中的广泛应用,以及卷积神经网络(CNN)的进一步发展,X 射线图像中危险品的识别从图像分类发展到了目标检测<sup>[4]</sup>。目前,目标检测框架大致可以分为两类:一类是以 R-CNN(Region-CNN)<sup>[5]</sup>为代表的两阶段检测模型,另一类是端到端的一阶段检测模型,如 YOLO(You Only Look Once)<sup>[6]</sup>和 SSD(Single Shot Multibox Detector)<sup>[7]</sup>等。SSD 模型借鉴了 Faster R-CNN<sup>[8]</sup>与 YOLO 的思想,在一阶网络的基础上使用了固定框进行区域生成,而且利用了多层的特征信息,这在速度与检测精度上都了一定的提升<sup>[9]</sup>。但 SSD 模型用于检测的浅层特征层只有一层,特征细节信息较少,多个特征层之间的信息并未得到综合的利用,这对于小尺度目标的检测效果不佳。

提高小尺度违禁物品的检测精度一直是 X 光图像检测领域的主要技术挑战之一。在提升 SSD 模型对小尺度目标的检测能力方面,已经有部分学者进行了相应的研究。Fu 等<sup>[10]</sup>提出了 DSSD(Deconvolutional SSD)模型,其最大的改进就是加

入了反卷积模块,将网络模型的上下层信息融合到一起,有效提升了小尺度目标的检测精度,但该模型结构过于复杂,使得检测速度大大降低。FSSD(Feature Fusion SSD)模型<sup>[11]</sup>同样是基于 SSD 模型进行改进,通过对网络中的两个低层特征进行融合,可以改善模型的多尺度特征的表达能力,为网络增加了特征细节信息,虽然在一定程度上提升了小目标的检测效果,但仍然有待提高。

针对复杂背景下 X 光安检图像中小尺度目标检测效果不佳的问题,本文以 SSD 模型为基础,提出一种新的用于 X 光安检图像检测的特征融合框架,即 MFFNet(Multi-scale Feature Fusion Network)。MFFNet 选用 ResNet-101 作为骨干网络以更好地提取特征图信息,对不同深度的高语义特征层的特征进行放大,可以得到更高分辨率的特征,接着将高语义特征层与低层细节特征丰富的特征层通过跳跃连接的方式进行融合。为了充分利用上下文信息,设计特征融合模块,新的融合特征具有丰富的语义信息和较高的分辨率,能够在有效提高小尺度目标检测精度的同时,也能实现不同尺度和不同特征层多元信息的分类预测与位置回归。在 X 光安检图像数据集 SIXray\_OD 中对模型进行测试,获得优于现有模型的检测结果。

## 2 基本原理

### 2.1 SSD 模型分析

SSD 模型是一种端到端的单阶段检测模型,其从多个角度对物体检测进行创新,是一个速度与精度兼具的检测模型。

SSD 模型如图 1 所示。输入图像首先需要经过

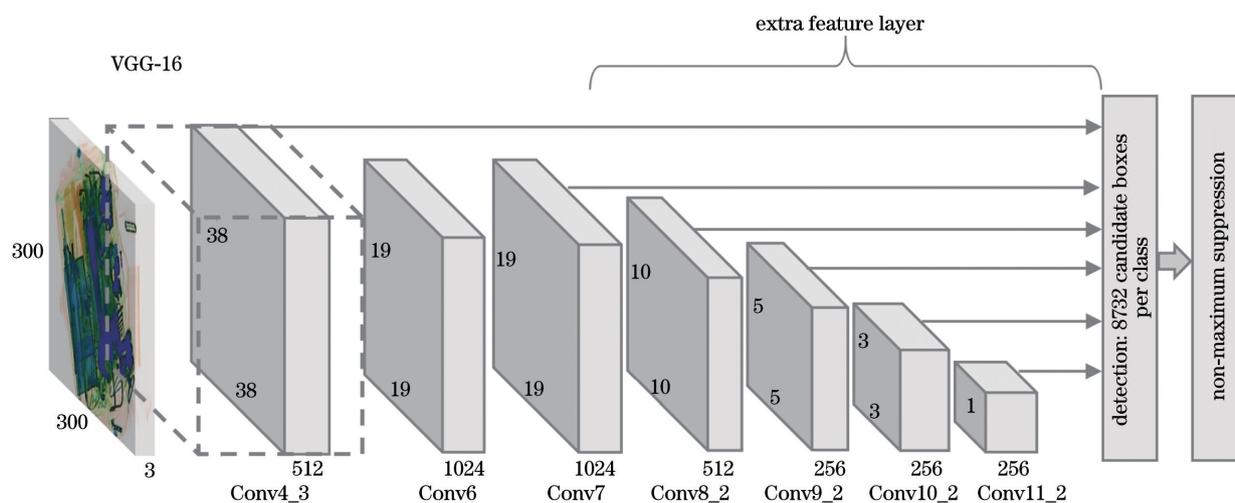


图 1 SSD 模型的结构

Fig. 1 Structure of SSD model

SSD 模型中的基础网络,即 VGG-16 (Visual Geometry Group-16)<sup>[12]</sup>。为了提取更高层的语义特征,增加 4 个卷积层来检测不同尺度的物体。与两阶段检测网络 Faster R-CNN 中的锚框类似,SSD 模型利用固定大小与宽高的先验框进行区域生成。但与 Faster R-CNN 不同的是,SSD 模型不只是一个特征图上设定预选框,而是在 6 个不同尺度的特征图上都设定预选框。SSD 模型在浅层特征图上设定较小的先验框,用来负责检测小物体,在深层特征图上设定较大的先验框,用来负责检测大物体。利用  $3 \times 3$  的卷积在 6 个特征图上进行特征提取<sup>[13]</sup>,并将提取的特征分为分类与回归两个分支,将计算得到每个分支的预测值作为所有预选框的预测值。随后对预选框与真实框进行匹配,利用交并比(IoU)来筛选出正样本与负样本,接着计算出分类损失与回归损失。采用非极大值抑制<sup>[14]</sup>算法去掉大多数重复的框,从而获取最终的检测结果。

SSD 模型虽然在同期的检测模型中表现较为优秀,但对于小目标的检测效果一般。小尺度目标通常会由感受野较小的网络中的浅层部分检测出来,而 SSD 模型中用于小尺度目标检测的特征层只有 Conv4\_3 层<sup>[15]</sup>,虽然浅层特征图的分辨率大但语义信息不足,所以小目标的特征提取不充分。此外,模型虽然能够在不同的特征图上进行预测,但特征图之间语义信息的相互关联性未得到充分利用,从而导致网络对小尺度目标特征的特征能力较弱,检测效果不佳。

## 2.2 多尺度特征融合

在 CNN 提取特征的过程中,随着网络层次的加深,不同层级上的特征图呈现出不同的特点<sup>[16]</sup>。

通常低层特征的分辨率较高,更多地保留了图像的边缘信息,但语义信息相对欠缺;高层特征层经过多次卷积之后,语义信息较为丰富,但像素在卷积过程中的缺失使其对细节的感知能力变差。若充分利用网络各层和各个尺度的特征信息,以便对各类目标物体进行预测,通常需要对不同尺度的特征进行融合。2019 年, Dlouhy 等<sup>[17]</sup>提出的 FPN (Feature Pyramid Network)就利用了特征融合的方法,通过横向连接来构建自下而上和自上而下的金字塔特征,较好地解决了多尺度检测的问题。FCN (Fully Convolutional Network)<sup>[18]</sup>则是通过对称结构和跳跃连接的方式来关联低层特征和高层特征。

与传统的网络相比,上述网络在精度上有了显著提高,然而这些网络大都是利用了最顶层的反卷积层,这对于小目标的精细细节表达较少。另外,基于融合特征的体系结构虽能够实现每个预测层的连接,但更多的附加层无疑会引起更多的计算开销,这对检测速度要求较高的实际应用变得不现实。

## 3 MFFNet 模型

基于第 2 节的研究,提出 MFFNet 模型并用于 X 光安检图像中危险品的检测,该模型融合了高层和低层特征,为小尺度目标添加了上下文信息,通过将目标的高层语义特征有效地加入到不同分辨率的特征层中,不同分辨率的特征层可以具有更高层的语义信息,从而有效提高网络对小尺度目标的识别和定位精度。

### 3.1 MFFNet 模型的构建

MFFNet 模型结构如图 2 所示。为了更好地提取特征,将 SSD 模型中用来提取特征的 VGG-16 替

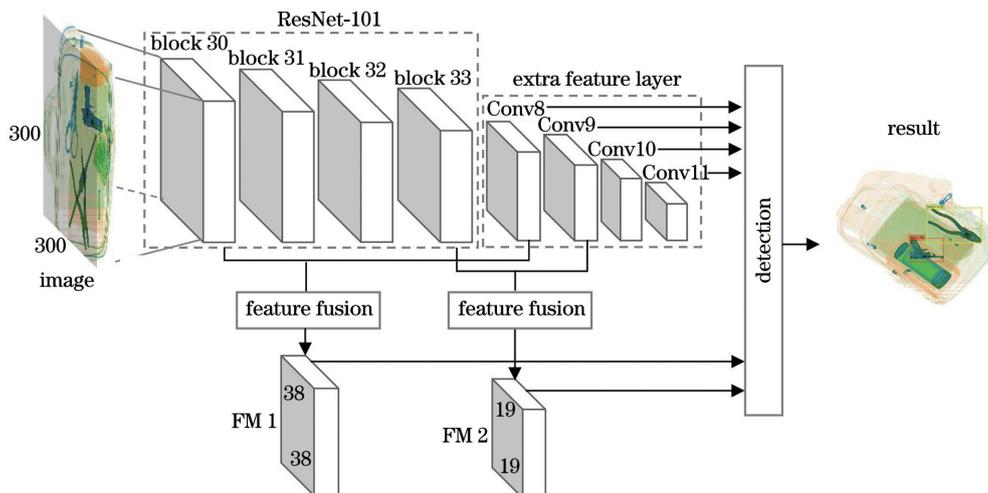


图 2 MFFNet 模型的结构

Fig. 2 Structure of MFFNet model

换为 ResNet-101。此外,相较于原始的 SSD 模型, MFFNet 模型新增了两个特征融合模块:一个是 ResNet-101 的 block 30 层与 Conv8\_2 层融合而成的融合模块 1,将其定义为 FM 1;另一个是 ResNet-101 的 block 33 层与 Conv9\_2 层融合而成的融合模块 2,将其定义为 FM 2。使用 FM 1 代替原来的 Conv4\_3 层,使用 FM 2 代替原来的 Conv7 层,再与后面的 Conv8\_2、Conv9\_2、Conv10\_2 以及 Conv11\_2 层一起送入预测模块中并进行检测。

MFFNet 模型中共有 6 个不同尺度的预测层,其中包括两个融合模块(FM 1 和 FM 2)和 4 个原始的 SSD 预测层(Conv8\_2、Conv9\_2、Conv10\_2 和 Conv11\_2)。此外,SSD 模型中用来检测的最低网络特征层为 Conv4\_3 层,而 MFFNet 模型却用到了两个融合特征层。与此同时,跳跃连接的方式增加了网络中特征图的种类,这在一定程度上增加了模型的运算复杂度,使得检测速度有所降低,但仍然能够较快地为小目标物体的分类及定位提供更准确的语义信息和几何信息,而且充分利用了高低层网络中上下文的语义信息,使检测结果得到了有效提升。

### 3.2 融合模块

MFFNet 模型有两个融合模块,融合操作过程如图 3 所示,其中 $\oplus$ 为元素相加运算。为了使高层特征能够与低层特征融合,而且能够保留低层特征的细节信息,首先需要对高层特征进行上采样处理,使其二维尺寸与低层特征相同。常用的上采样方法主要有双线性插值方法和反卷积方法两种。反卷积方法能够为网络提供学习的参数,但在提升网络性能的同时也会大大增加计算的复杂度。双线性插值方法能够在保持通道数不变的前提下增加特征的二维尺寸,在不大量增加计算量的同时也能保证较好的性能,因此实验选择双线性插值方法。

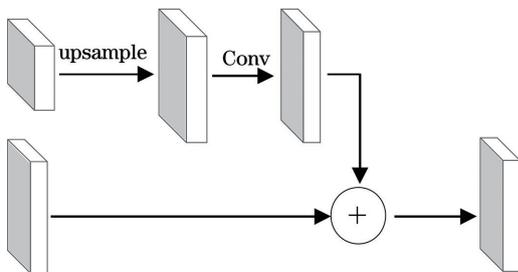


图 3 融合模块的处理过程

Fig. 3 Process of fusion module

高层特征经过上采样处理后,还需要对其进行降维处理。具体操作过程:引入一个卷积核大小为  $1 \times 1$  的卷积,其能够在保持特征尺度不变的前提下

实现降维,还可以利用后续非线性激活函数来增加网络的非线性特性<sup>[19]</sup>,从而进一步提升网络的表达能力,最终实现跨通道的特征信息融合。

经过上采样与降维操作处理后,高、低层特征图的通道数与尺寸已经完全相同,接下来直接通过对对应元素相加运算后就能够得到新的融合特征图。新的融合特征图不仅保留了低层的细节信息,还额外增加了高层的语义信息,这更有利于较小尺寸目标的检测。

融合模块在充分利用上下文语义信息的同时,也会在一定程度上增加模型的复杂度。相比于 SSD 模型,融合模块多出两个跳跃连接方式,其采用对应元素进行相加运算,该运算过程涉及的特征图有多个尺度,这在一定程度上影响了模型的总体实时性,但是在牺牲较少实时性的同时较大地提升了检测精度。

### 3.3 候选框设置

CNN 中不同的尺度特征图有不同的感受野,其对应检测图像中不同尺度的区域。为了处理不同图像中尺寸不同和形状各异的目标,对于网络产生的不同特征图需要设置不同尺度与比例的候选框,候选框的设置包含尺度和长宽比两个方面。相应的理论推导公式为

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{m - 1}(k - 1), k \in [1, m], \quad (1)$$

式中: $m$  表示特征图的个数; $S_k$  表示候选框的大小相对于特征图的比例; $S_{\min}$  和  $S_{\max}$  分别表示比例的最小值与最大值。中间层的候选框可以使用(1)式计算得出。根据 SIXray\_OD 数据集中各类违禁品的尺度信息,在此将尺度系数  $S$  的范围设为  $\{0.1, 0.9\}$ ,即  $S_{\min} = 0.1, S_{\max} = 0.9$ 。同时,在同一特征图中使用纵横比例不同的候选框,可以增强候选框对物体形状的鲁棒性。为此,实验设计相应尺度与纵横比的候选框,且每个特征图中设置不同数量的候选框。候选框的纵横比  $r$  通常设为  $\{1, 2, 3, 1/2, 1/3\}$ ,根据面积和长宽比可以得到候选框的宽和高,表达式为

$$w_k = s_k \cdot \sqrt{r}, \quad (2)$$

$$h_k = s_k / \sqrt{r}. \quad (3)$$

默认情况下,每个特征图中除了上面 5 个比例的候选框,还会设置一个尺度  $s'_k = \sqrt{s_k \cdot s_{k+1}}$  且  $r=1$  的先验框,这样每个特征图中都设置两个长宽

比为 1, 但尺寸不同的正方形先验框。实验发现 FM 1、Conv10\_2 和 Conv11\_2 层中是否加入纵横比为 3 和 1/3 的候选框对结果的影响较小, 因此为了简化计算, 节约时间成本, 上述特征层中只设置 4 个不同尺寸的候选框, 其余设置 6 个。

## 4 实验结果与分析

### 4.1 实验设计

实验平台配置: CPU 为 Intel i7-6700, 内存为 32 G; GPU 为 GTX1080Ti; 操作系统为 64 位 Ubuntu 16.04 LTS; 实验框架为 PyTorch 开源框架。

#### 4.1.1 实验数据集及设置

实验使用 SIXray\_OD 数据集, 其源于开源数据

集 (Security Inspection X-ray Benchmark, SIXray)<sup>[20]</sup>。SIXray 数据集共包含 1059231 张图片, 其中有 8929 张图片包含 6 类违禁品, 其余为不包含违禁品作为负样本。由于实验是对违禁品进行检测, 故只选取其中包含违禁品的部分图片。其中锤子一类因为包含样本数量过少而被弃用, 后使用图像标注工具重新对这些图片进行手动标注并作为实验所用的 SIXray\_OD 数据集, 如图 4 所示。从图 4 可以看到, SIXray\_OD 数据集中包含 5 类违禁品, 分别为枪、刀、扳手、钳子和剪刀, 共有 8718 张 X 光图片。训练过程中, 数据集的训练测试分配比例以及统计信息如表 1 所示。

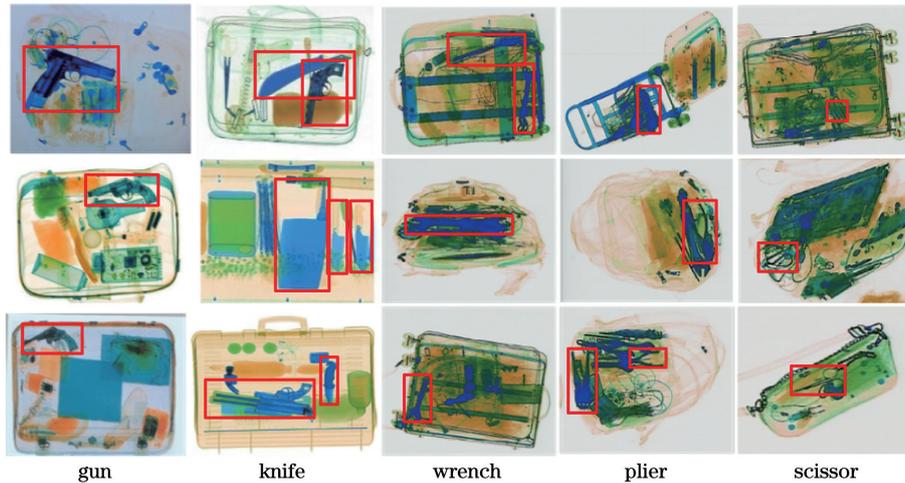


图 4 SIXray\_OD 数据集中的部分图片

Fig. 4 Some pictures in SIXray\_OD dataset

表 1 数据集中不同类型的图片数量

Table 1 Number of images of different types in datasets

Type	Number of images					Total
	Gun	Knife	Wrench	Plier	Scissor	
Training	2055	1092	1586	2765	811	6102
Test	881	469	680	1186	348	2616
Total	2936	1561	2266	3951	1159	8718

#### 4.1.2 实验参数设置

为了节省训练时间以及加快收敛速度, 将预训练好的 SSD 模型中低层结构的权值参数共享。此外, 实验采用随机梯度下降 (SGD) 法来优化参数, 优化器的动量设为 0.9, 批处理大小设为 16, 初始学习率设为  $1 \times 10^{-3}$ , 权重衰减参数设为  $5 \times 10^{-4}$ 。

#### 4.1.3 性能评估指标

对于 X 光安检图像中违禁品的检测, 首先需

要对各个物品类别单独计算其平均检测精度 (AP), 然后对各个类别的 AP 值求和并求解平均值, 即可得到综合检测指标的平均精度均值 (mAP)。mAP 衡量的是在所有类别上的平均好坏程度, 可以用来评估目标检测模型的性能, 有效避免某些极端化类别对其他类别的影响。mAP 的表达式为

$$x_{\text{mAP}} = \frac{1}{N} \sum_{C=1}^N x_{\text{AP}, C}, \quad (4)$$

式中:  $N$  表示总的类别数目。mAP 值越大, 则表示模型的检测精度越高, 相反则越低。

检测速度是另外一个非常重要的性能评价指标, 其对于实时性要求高的检测应用场景极其重要。常用的评估检测速度的指标是每秒帧率 (FPS), 即每秒内可以处理的图片数量。在单位时间内处理的图片数量越多, 则表示速度越快。所有对比实验均在同一硬件的条件下进行。

## 4.2 实验结果分析

### 4.2.1 对比实验

为了验证不同卷积层的融合对模型效果的影响,分别进行三种不同组合的实验,具体的组合方式如表 2 所示。

表 2 不同融合特征层的组合方式

Combination	Based layer		Extra layer		
	Block 30	Block 33	Conv8	Conv9	Conv10
1	✓		✓		
2	✓		✓	✓	
3	✓			✓	✓

从表 2 可以看到,每个组合方式都有两个不同的融合模块;对于组合 1 来说,block 30 与 Conv8 融合而成的新特征层即为 FM 1,block 33 与 Conv9 融合而成的新特征层即为 FM 2;对于组合 2 来说,block 30 与 Conv8 融合而成的新特征层即为 FM 1,block 33 与 Conv10 融合而成的新特征层即为 FM 2;对于组合 3 来说,block 30 与 Conv9 融合而成的新特征层即为 FM 1,block 33 与 Conv10 融合而成的新特征层即为 FM 2。在此基础上,对每种组合方式进行元素相加、元素相乘以及级联处理,结果如表 3 所示。

表 3 不同融合方式的检测精度结果

Combination	methods		
	Sum	Product	Concat
1	78.27	77.76	78.10
2	77.84	77.53	77.92
3	78.05	77.02	77.54

从表 3 可以看到,当融合模块为组合 1 或者组合 3 时,元素相加的融合方式都得到相对最优的检测结果;组合 2 的级联方式得到的结果最优,但综合来看还是选择元素相加的融合方式,原因在于元素相加可以最大程度地利用多尺度的特征信息,特征互为补充,联合作用,从而获得对预测目标更完备的特征表达,使训练更有效<sup>[21]</sup>;元素相乘则会导致输入元素的梯度不稳定,进而造成网络不容易收敛,实

验证明其效果最差。

对数据集中的各类违禁品分别在 SSD 原始模型和改进后的 MFFNet 模型上进行对比,结果如表 4 所示,其中 MFFNet 模型的数据即为添加组合 1 中的两个融合模块,并采用元素相加方式进行融合后得到的结果。

表 4 各类违禁品的检测精度结果

Type	unit: %	
	SSD	MFFNet
Gun	89.91	90.42
Knife	73.31	75.29
Wrench	69.46	71.17
Plier	75.57	82.22
Scissor	63.61	72.24
mAP	74.37	78.27

从表 4 可以看到,MFFNet 模型的 mAP 值比 SSD 模型提升 3.9 个百分点;对于较大尺寸的目标来说,MFFNet 模型检测枪的 AP 值达到 90% 以上;对于较小尺寸的目标来说,如小钳子和小剪刀的检测结果更是有近 10 个百分点的提升,说明改进模型对危险品的检测效果,特别是对于小尺度物体的检测效果提升显著。此外,对 MFFNet 模型和 SSD 模型的收敛性进行对比,并在数据集上对其进行损失分析。两个模型的训练损失随迭代次数的变化曲线如图 5 所示,最终的迭代次数为  $1.2 \times 10^5$ 。从图 5 可以看到,MFFNet 模型在起始阶段的损失值小于 SSD 模型,随着迭代次数的增加两个模型虽然都能够迅速收敛,并随着迭代的进行而稳步降低直到趋于稳定,但最终改进模型的收敛损失值还是略小于 SSD 模型。

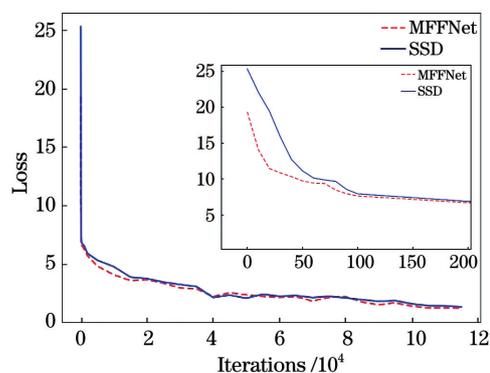


图 5 网络的训练损失函数曲线

Fig. 5 Training loss function curves of network

使用 SIXray\_OD 数据集分别在 SSD 模型和 MFFNet 模型上进行可视化对比,如图 6 所示。从图 6 可以看到,SSD 模型存在小剪子和小钳子漏检的情况,而 MFFNet 模型却能够将其检测出来,这也进一步证明 MFFNet 模型能够在一定程度上解决 X 光安检图像中小尺度违禁品的检测问题。

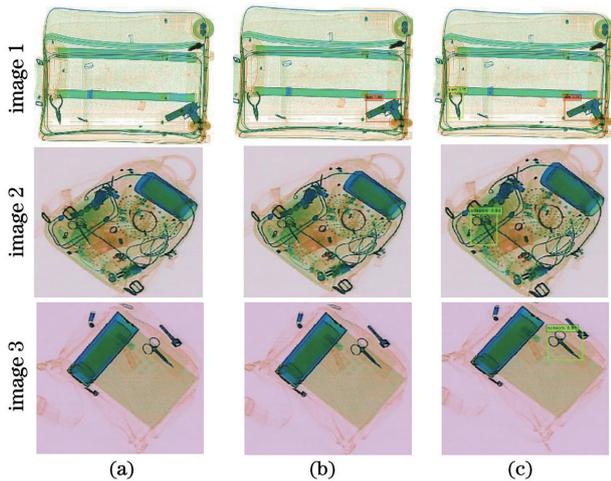


图 6 不同模型的可视化检测结果。(a)原始图像;  
(b)SSD 模型;(c)MFFNet 模型

Fig. 6 Visual detection results of different models.

(a) Original images; (b) SSD model; (c) MFFNet model

#### 4.2.2 消融实验

本节以 SSD 网络为基础模型进行消融实验,添加不同模块后对 MFFNet 模型检测效果的影响进行对比。FM 1 由 block 30 与 Conv8 融合而成,FM 2 由 block 33 与 Conv9 融合而成。该部分实验参数设置与对比实验一致,实验结果如表 5 所示。

表 5 消融实验的结果

Table 5 Results of ablation experiment

Model	Backbone	mAP/%
SSD	VGG-16	74.37
SSD	ResNet-101	76.18
SSD+FM 1	ResNet-101	77.38
SSD+FM 2	ResNet-101	77.75
SSD+FM 1+FM 2	ResNet-101	78.27

从表 5 可以看到,当骨干网络由 VGG-16 替换为 ResNet-101 时,检测结果有 1.81 个百分点的提升;加入 FM 1 和 FM 2 两个融合模块后,与原始 SSD 模型相比精度分别有 3.01 个百分点和 3.38 个百分点的提升。两个融合模块的同时加入,使检测精度的提升更明显,较原 SSD 模型提升 3.9 个百分点,实验结果进一步证明设计的融合模块对模型检

测性能的提升效果显著。

为了更好地评估 MFFNet 模型,将其与其他模型进行性能对比,结果如表 6 所示。

表 6 不同模型的检测结果

Table 6 Detection results of different models

Model	mAP / %	FPS
SSD	74.37	56
FSSD	75.75	41
Faster R-CNN	77.81	2
YOLO-v3	70.49	70
MFFNet	78.27	19

从表 6 可以看到,MFFNet 模型在检测精度方面达到目前最优的检测结果,速度方面虽然相比于一阶段 SSD 和 YOLO-v3 模型有所下降,但仍然领先于精度相对高的二阶段 Faster R-CNN 模型,而且能够满足智能安检过程中对时间的要求。

## 5 结 论

提出一种新的违禁品检测模型,即 MFFNet。在 SSD 模型的基础上将更深的 ResNet-101 作为骨干网络以更好地提取特征,新增两个轻量级的特征融合模块并生成新的特征金字塔,用于 X 光安检图像中危险品的检测。对特征融合模块的组合方式和不同特征层间的融合方式进行对比,并通过详细的消融和对比实验得到最佳方案。实验结果表明,MFFNet 模型对 X 光安检图像中的违禁品检测具有有效性。相较于现有的一阶段和两阶段模型, MFFNet 模型的 mAP 值都有不同程度的提升,且每秒内可以处理的图片数量为 19 张,这符合安检工作的实际要求,为智慧机场的建设提供新思路。

## 参 考 文 献

- [1] Han N. A deep learning-based dangerous goods detection and tracking algorithm from X-ray images [D]. Lanzhou: Lanzhou University, 2018.  
韩宁. 基于深度学习的 X 射线图像危险品检测与跟踪算法研究[D]. 兰州: 兰州大学, 2018.
- [2] Bastan M, Yousefi M R, Breuel T M. Visual words on bag-gage X-ray images[M]//Real P, Diaz-Pernil D, Molina-Abril H, et al. Computer analysis of images and patterns. Lecture notes in computer science. Heidelberg: Springer, 2011, 6854: 360-368.
- [3] Baştan M. Multi-view object detection in dual-energy X-ray images[J]. Machine Vision and Applications,

- 2015, 26(7/8): 1045-1060.
- [4] Su Ha S, Liu T T, Liu G H, et al. Algorithm for student behavior detection based on neural network [J]. *Laser & Optoelectronics Progress*, 2020, 57(22): 221016.  
苏寒松, 刘腾腾, 刘高华, 等. 基于神经网络的学生行为检测算法研究[J]. *激光与光电子学进展*, 2020, 57(22): 221016.
- [5] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [7] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.
- [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] Ji X L, Wu J, Yi J B, et al. Automatic detection algorithm for controlled items based on deep learning [J]. *Laser & Optoelectronics Progress*, 2019, 56(18): 180402.  
吉祥凌, 吴军, 易见兵, 等. 基于深度学习的管制物品自动检测算法研究[J]. *激光与光电子学进展*, 2019, 56(18): 180402.
- [10] Fu C Y, Liu W, Ranga A, et al. DSSD: deconvolutional single shot detector[EB/OL]. (2017-01-23) [2020-08-09]. <https://arxiv.org/abs/1701.06659>.
- [11] Li Z X, Zhou F Q. FSSD: feature fusion single shot multibox detector[EB/OL]. (2018-05-17) [2020-08-09]. <https://arxiv.org/abs/1712.00960>.
- [12] Zhao Y N, Wu L M, Chen Q. Small object detection algorithm based on multi-scale fusion SSD [J]. *Computer Engineering*, 2020, 46(1): 247-254.  
赵亚男, 吴黎明, 陈琦. 基于多尺度融合 SSD 的小目标检测算法[J]. *计算机工程*, 2020, 46(1): 247-254.
- [13] Chen P F. Real-time object detection based on multi-scale feature fusion[D]. Xi'an: Xidian University, 2018.  
陈鹏飞. 基于多尺度特征融合的实时目标检测[D]. 西安: 西安电子科技大学, 2018.
- [14] Neubeck A, van Gool L. Efficient non-maximum suppression [C] // 18th International Conference on Pattern Recognition (ICP R'06), August 20-24, 2006, Hong Kong, China. New York: IEEE Press, 2006: 850-855.
- [15] Chen H J, Wang Q Q, Yang G W, et al. SSD object detection algorithm with multi-scale convolution feature fusion[J]. *Journal of Frontiers of Computer Science and Technology*, 2019, 13(6): 1049-1061.  
陈幻杰, 王琦琦, 杨国威, 等. 多尺度卷积特征融合的 SSD 目标检测算法[J]. *计算机科学与探索*, 2019, 13(6): 1049-1061.
- [16] Huang H J, Duan X H, Huang X C. Research and improvement of fruits detection based on deep learning[J]. *Computer Engineering and Applications*, 2020, 56(3): 127-133.  
黄豪杰, 段先华, 黄欣辰. 基于深度学习水果检测的研究与改进[J]. *计算机工程与应用*, 2020, 56(3): 127-133.
- [17] Dlouhy A C, Bailey D K, Steimle B L, et al. Fluorescence resonance energy transfer links membrane ferroportin, hephaestin but not ferroportin, amyloid precursor protein complex with iron efflux [J]. *Journal of Biological Chemistry*, 2019, 294(11): 4202-4214.
- [18] 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.
- [19] Zhu M M, Xu Y L, Ma S P, et al. Airplane detection based on feature fusion and soft decision in remote sensing images[J]. *Acta Optica Sinica*, 2019, 39(2): 0210001  
朱明明, 许悦雷, 马时平, 等. 基于特征融合与软判决的遥感图像飞机检测[J]. *光学学报*, 2019, 39(2): 0210001.
- [20] 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.
- [21] Jiang W T, Zhang C, Zhang S C, et al. Multiscale feature map fusion algorithm for target detection[J]. *Journal of Image and Graphics*, 2019, 24(11): 1918-1931.  
姜文涛, 张驰, 张晟翀, 等. 多尺度特征图融合的目标检测[J]. *中国图象图形学报*, 2019, 24(11): 1918-1931.