第 58 卷 第 22 期/2021 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

复杂背景下多尺度 X 光违禁品检测

张珂，张良*

中国民航大学电子信息与自动化学院，天津 300300

摘要 针对安检 X 光图像中违禁品的自动检测一直存在困难，使用不同尺度的特征比例平衡模块、U 型网络递归模块和残差边注意力模块构建 EM2Det(Enhanced M2Det)模型，进一步提升 M2Det 模型的检测性能。首先考虑主干网络深层中的高语义信息和浅层中的细节特征信息，借鉴特征金字塔思想设计特征融合增强模块，加强模型对主干网络中不同尺度特征的提取能力；然后设计 8 个 U 型网络递归模块，增强其对基本特征不同水平、不同尺度的细节特征提取能力；接着使用 CBAM(Convolutional Block Attention Module)构建残差边注意力模块，使其关注有效特征，抑制无用的背景干扰；最后在 SIXray_OD 数据集上对模型进行验证。实验结果表明，设计的各个模块均有不同程度的提升效果，EM2Det 模型的平均精度比 M2Det 模型提升 6.4 个百分点。

关键词 图像处理；目标检测；安检 X 光图像；EM2Det 模型；特征金字塔；多尺度

中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP202158.2210002

Multi-Scale Detection for X-Ray Prohibited Items in Complex Background

Zhang Ke, Zhang Liang*

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

Abstract Aiming at automatic detection of contraband in security X-ray images is difficult, the EM2Det (Enhanced M2Det) model is constructed using different scale feature proportional balance modules, U-shaped network recursive modules, and residual edge attention modules, which it can further improve the detection performance of the M2Det model. First, considering the high semantic information in the deep layer of the backbone network and the detailed feature information in the shallow layer, the feature fusion enhancement module is designed by referring to the feature pyramid idea to enhance its ability to extract features of different scales in the backbone network. Then, the CBAM (Convolutional Block Attention Module) is used to build a residual edge attention module to focus on effective features and suppress useless background interference. Finally, the model is verified on the SIXray_OD dataset. The experimental results show that each module of the design has different degrees of improvement effects, and the average accuracy of the EM2Det model is 6.4 percentage higher than that of the M2Det model.

Key words image processing; object detection; security X-ray images; EM2Det model; feature pyramid; multiscale

OCIS codes 100.4996; 100.2960; 110.2960

1 引 言

在民航领域，旅客行李的 X 光检查对于旅客生命财产安全的保障具有重要意义。如今，机场安检主要靠安检人员的肉眼来分辨行李中的违禁品，但是安检人员会因工作压力过大[1]或经验不足等因素出现工作失误，即不能及时发现违禁品。基于此，通过计算机视觉算法来自动识别行李中的违禁品，可以提高安检人员的工作效率，以及提升违禁品检查的准确度。

收稿日期：2020-11-30；**修回日期**：2021-01-07；**录用日期**：2021-01-21

基金项目：国家自然科学基金(61179045)

通信作者：*l-zhang@cauc.edu.cn

X 光图像的对比度主要由 X 射线的透射距离、路径上物质的质量衰减系数和物质密度决定[2]，这种特点使得成像对于高密度和强 X 射线吸收能力的物质有显著的辨别能力，同时通过双能量 X 射线的分类着色可以使不同物质更容易区分，其中金属通常以蓝色显示，食品等有机物通常以橙色显示。X 光图像中不同的叠加物体经过分类着色后，物体颜色会出现偏差，叠加越多颜色越深。X 光图像中目标的尺度变化比较明显，同类物品在图片中会呈现不同的尺度，不同位置的镜头监测到的目标也会呈现出尺度差异。不同 X 光安检机器发射的 X 光能量不尽相同，为此其穿透能力也存在差别，所以图像数据在表现形式和特征分布上差别很大[3]。因此，X 光图像中违禁品检测的难点在于物体间的重叠干扰比较严重、目标的多尺度变化以及图像质量等。

随着深度学习技术的发展，卷积神经网络(Convolutional Neural Networks，CNN)在计算机视觉任务上取得了重大突破，如目标检测[4-5]、图像分割和图像分类[6]等。现阶段对于 X 光安检图像的研究，Akçay 等[7]利用 CNN 和迁移学习思想对 X 光行李图像进行检测分类，其中手枪的二分类效果良好，但是识别类别单一，而且对照片质量和数量的要求比较高。Steitz 等[8]提出基于 CNN 的多视角 X 光图像检测算法，相比于单视图的检测效果有所提升，但是图像的尺度变化对结果的影响比较大。Galvez 等[9]使用 YOLO(You Only Look Once)检测器来识别 X 光图像中的危险物品，相比于迁移学习法和训练法的检测精度较低，相比于具有多尺度特性的目标检测精度更低。

本文通过实验选择检测效果比较良好的多尺度目标检测框架 M2Det[10]来进行改进实验，该框架是在 SSD(Single Shot MultiBox Detector)网络[11]的基础上演化而来的，其对多尺度目标有良好的识别精度。但是该框架在特征提取等方面还存在不足，在重叠干扰目标的检测精度方面有提升空间，基于此本文设计 EM2Det(Enhanced M2Det)模型，主要改进如下。1)设计特征融合增强模块，使用该模块代替原有的 FFMv1(Feature Fusion Module v1)生成基本特征，通过融合特征网络中 pool 3、conv 4_3 和 pool 5 层的特征图来构建特征金字塔网络(Feature Pyramid Network，FPN)[12]结构，其可以融合浅层特征层 pool 3 中的小尺度目标特征信息，同时 FPN 结构的输出考虑融合 pool 5 层特征，用来平衡大尺度目标的特征比例。2)融入递归特征金字塔(Recursive Feature Pyramid，RFP)[13]思想将 TUM(Thinned U-shape Module)设计成 U 型网络递归结构，该结构进一步增强不同水平、不同大小特征层的表达能力，本文设计的 U 型网络递归模块只递归一次，提升效果明显。3)在 M2Det 中的 SFAM(Scale-wise Feature Aggregation Module)上使用 CBAM(Convolutional Block Attention Module)[14]设计残差边注意力模块，该模块不仅会关注特征图中通道上的权重，也会注意二维空间上的权重分配，其对于颜色特征明显的 X 光图像有很好的检测效果，同时使用残差边结构可以平衡 CBAM 的影响。

2 相关介绍

2.1 M2Det

M2Det 是近年来非常优秀的一阶目标检测框架。M2Det(图片大小为 320 pixel×320 pixel)由特征提取网络 VGG-16 (Visual Geometry Group-16)、多层特征金字塔(Multi-Level Feature Pyramid Network，MLFPN)和回归分类子网络组成。检测的图片大小为 320 pixel×320 pixel，通道数为 3，特征提取网络 VGG-16 中的特征层分别为 pool 1(160×160×64)、pool 2(80×80×128)、pool 3(40×40×256)、conv 4_3(40×40×512)和 pool 5(20×20×1024)。M2Det 的整体结构如图 1 所示，将特征提取网络 VGG-16 中 conv 4_3 和 pool 5 生成的特征输入到 MLFPN 中，用来对不同水平、不同大小的尺度特征进行提取。将 MLFPN 模块中特征金字塔生成的有效特征分别输入到分类和回归子网络中进行计算，用来对各个特征层上的先验框进行调整，经过得分排序和非极大抑制来筛选得到预测框的位置。

MLFPN 由特征融合模块(Feature Fusion Module，FFM)、TUM 和 SFAM 三部分组成。FFM 有两个子模块，如图 2 所示。第一个子模块 FFMv1 主要是用来融合特征提取网络所提取的特征，其中 512，1×1，1×1，256 表示卷积层的输入通道数为 512，经过大小为 1×1 和步长为 1×1 的卷积核卷积，输出的通道数为 256，2×2 表示上采样操作的步长。pool 5 的通道数通过卷积调整为 512，经过上采样后特征图的大小调整为 40 pixel×40 pixel，conv 4_3 的通道数通过卷积调整为 256，最后对两者调整结果进行特征融合并将融合结果作为基本特征。第二个子模块 FFMv2 是将前一个

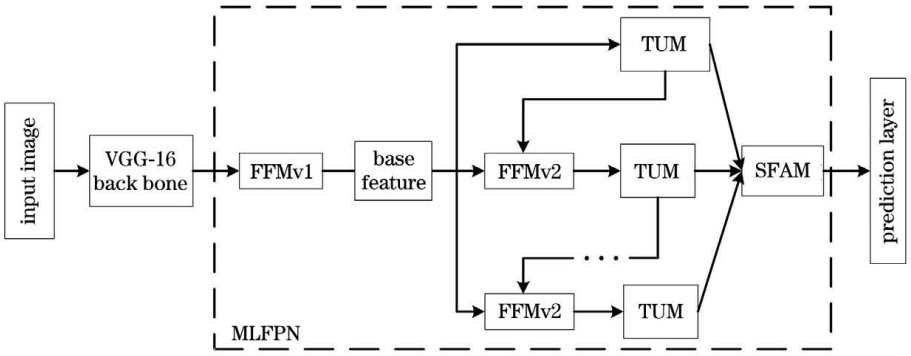

图 1　M2Det 的框架

Fig. 1　Framework of M2Det

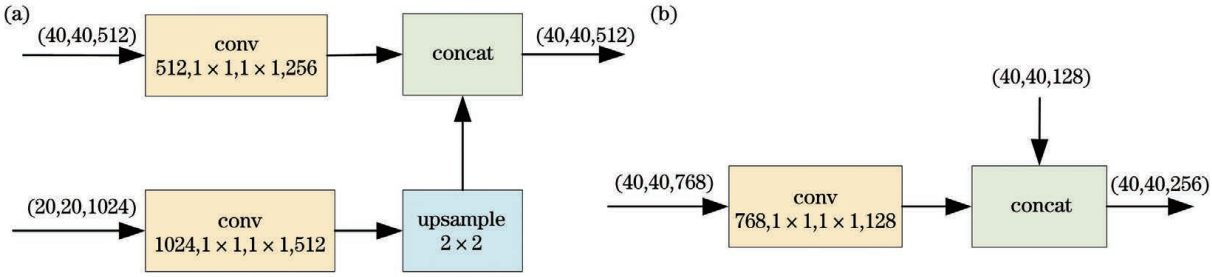

图 2　FFM 中不同模块的结构。(a)FFMv1;(b)FFMv2

Fig. 2　Structure of different modules in FFM. (a)FFMv1; (b)FFMv2

TUM 所输出的最大分辨率特征与基本特征进行特征融合,最后将输出的特征图作为 TUM 的输入,特征图的大小为 40 pixel×40 pixel,通道数为 256。

TUM 结构的主要作用是实现特征的深度提取。M2Det 中有两个超参数,即 TUM 数量 L 和每个 TUM 中 U 型网络结构的数量 i,分别满足 $L=8$ 和 $1\leqslant i\leqslant 6$。TUM 数量为 8 个即输出 8 组不同水平的特征金字塔,每个 TUM 中 U 型网络结构中右半部分特征层的数量为 6 个即每组特征金字塔中包含 6 个不同大小的特征图,其通过 TUM 结构中的 5 次卷积与 5 次上采样来获得。SFAM 主要功能是将不同层的特征按照相同的大小进行特征融合,融合之后的特征还需经过一个注意力模块[15]。

2.2　FPN 结构与 RFP 结构

Lin 等[12]提出的 FPN 结构主要解决的是目标检测中的多尺度问题,通过简单的网络连接可以在基本不增加原有模型计算量的情况下,大幅度提升不同尺度目标的检测性能。FPN 的结构如图 3(a)所示。FPN 首先进行自底向上的正向传播[16],特征图的大小经过卷积核计算后会越变越小,然后将语义更强的高层特征图进行自上而下的上采样,上采样前的特征图需加上自底向上过程的同层次特征图,因此高层特征得到了增强,而且还可以保证每一层特征都有合适的分辨率以及强语义特征。

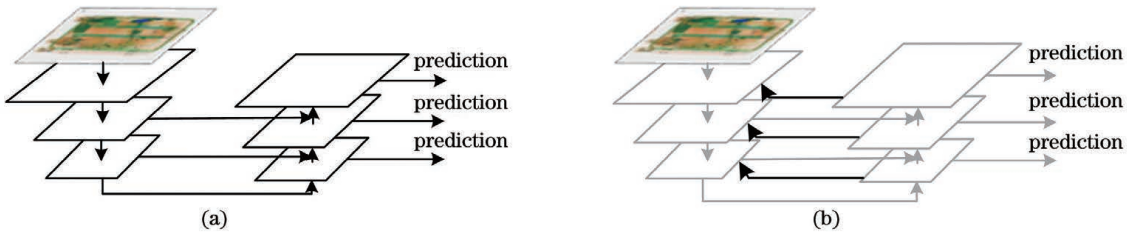

图 3　不同模型的结构。(a)FPN;(b)RFP

Fig. 3　Structure of different models. (a) FPN; (b) RFP

Qiao 等[13]提出的 RFP 结构是在 FPN 的基础上进行改进的,结构如图 3(b)所示。在传统的 FPN 中,将自上而下产生的特征图反馈到自下而上的同层次特征图中进行二次循环,控制循环次数,在最后一次循环后输出。RFP 结构是对 FPN 结构的进一步加强。

本文在 M2Det 模型中设计了基于 FPN 结构的提取模块，用来提升主干特征提取网络对于不同尺度细节特征的提取能力，同时将 M2Det 模型中的 TUM 设计成 RFP 结构，这丰富了不同尺度的细节特征。

2.3 CBAM 注意力

Woo 等[14]提出的 CBAM 是一种综合通道注意力和空间注意力的综合注意力模块，特征层经过 CAM（Channel Attention Module）再经过 SAM（Spatial Attention Module）后，最终完成通道上和空间上的加权调整。

CAM 的结构如图 4(a)所示。首先通过平均池化（AvgPool）和最大池化（MaxPool）操作来聚合特征映射的空间信息，将生成的两个不同通道描述送入一个共享网络中，共享网络由多层感知机（Multi-Layer Perceptron，MLP）和一个隐藏层组成，经过共享网络之后两个通道描述调整各自的权重比例，经过相加和 Sigmoid 激活函数处理后得到通道注意力权重 $M_C \in \mathbb{R}^{(1\times1\times C)}$，其中 C 为通道数。SAM 的结构如图 4（b）所示。首先通过 AvgPool 和 MaxPool 操作来聚合特征映射的通道信息，可以分别生成两个二维映射的空间权重，通过一个标准的卷积层连接和卷积混合可以生成空间注意力权重 $M_S \in \mathbb{R}^{(W\times H\times 1)}$（其中 W 为特征图的宽，H 为特征图的高），设 CBAM 的输入特征图为 $X \in \mathbb{R}^{(W\times H\times C)}$，经过 CAM 之后的输出特征图为 X'，表达式为

$$X' = M_C(X) \otimes X, \tag{1}$$

式中：⊗为点乘符号。X'经过 SAM 后的输出特征图为 X''，表达式为

$$X'' = M_S(X') \otimes X'。 \tag{2}$$

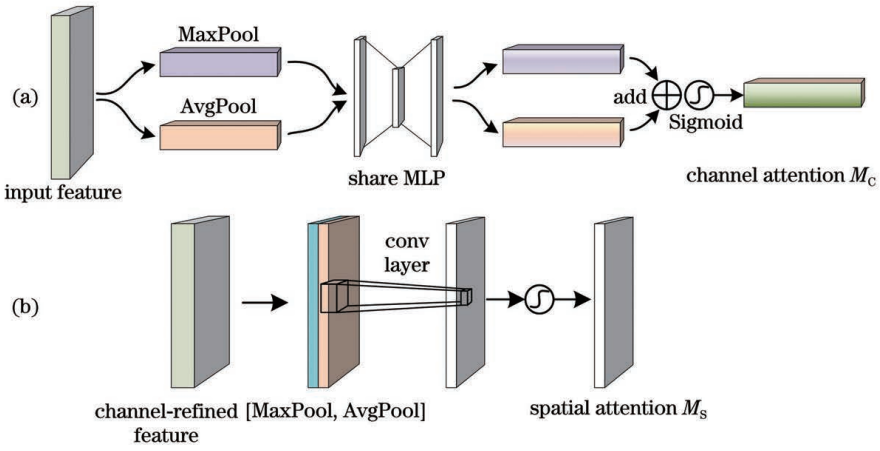

图 4　不同模块的结构。(a)CAM；(b)SAM

Fig. 4　Structure of different modules. (a) CAM; (b) SAM

本文在 M2Det 模型的 MLFPN 中设计了残差边注意力模块，该模块使用残差边来平衡注意机制的影响，从而提升复杂背景下图像中细节特征的分辨能力。

3　EM2Det 模型

本文在 M2Det 模型的基础上设计了目标检测模型 EM2Det。首先对 M2Det 模型中的超参数进行调整，通过权衡模型参数量和平均检测精度来调整 TUM 的数量。根据结构设计和实验结果，发现 M2Det 和 TUM 仍有很大的提升空间，而且 FFMv1 的特征提取能力不足，基于此设计了特征融合增强模块和 U 型网络递归模块以及残差边注意力模块，通过更加优良的 CBAM 来关注通道和空间上的有关特征，可以抑制 X 光图片中背景特征存在的影响。整体的 EM2Det 模型如图 5 所示。

3.1　特征融合增强模块

从结构分析和实验结果发现，FFMv1 对特征提取网络 VGG-16 中 conv 4_3 和 pool 5 所输出的特征进行融合，特征提取并不充分，影响了后续模块的处理效果，因此设计了基于 FPN 结构的特征融合增强模块，结构如图 6 所示。使用特征融合增强模块在特征提取网络中的 pool 3、conv 4_3 和 pool 5 上构建 FPN，用来增加浅层网络中的小尺度目标特征信息，同时将 FPN 结构的两个输出特征与高语义特征层 pool 5 输出的特征进行特征融合，可以进一步保证大尺度目标特征信息的比例。

特征融合增强模块的细节如图 7 所示。特征提取网路中的 pool 3、conv 4_3 和 pool 5 层分别通过 3×3、1×1 和 1×1 的卷积将通道数调整为 256。对调整通道数后的 pool 5 层进行 2×2 的上采样处理，并与conv4_3层调整通道数后的特征图进行特

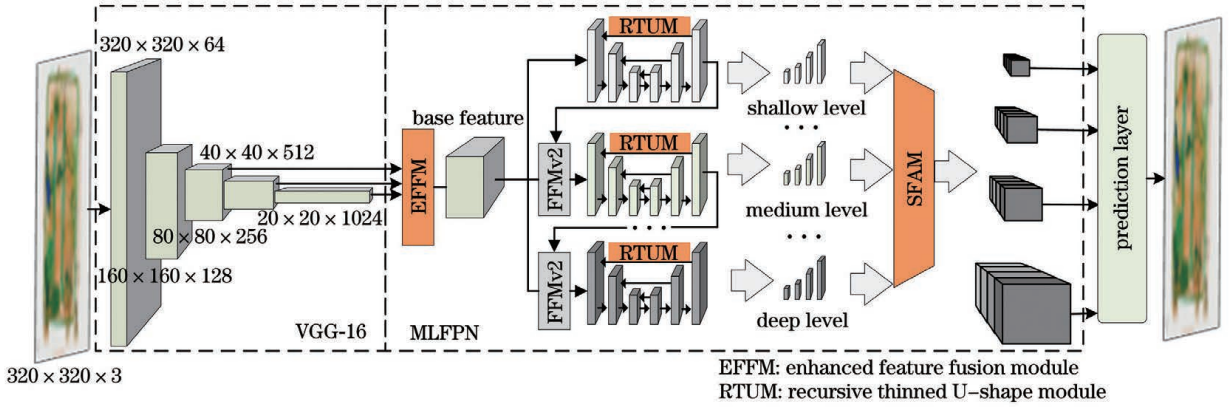

图 5　EM2Det 模型

Fig. 5　EM2Det model

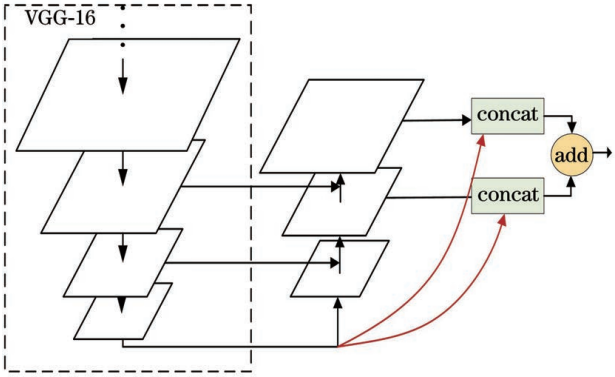

图 6　特征融合增强模块的结构

Fig. 6　Structure of feature fusion enhancement module

征相加，结果记为 f_1；f_1 再与 pool 3 层调整通道数后的特征图进行特征相加，结果记为 f_2；f_1 和 f_2 分别与 pool 5 层调整通道数后的特征图进行特征融合，两者的融合结果再经过特征相加处理后作为一个特征图输出，输出的特征图大小为 40 pixel×40 pixel，通道数为 768。

3.2　U 型网络递归模块

为了进一步提升 TUM 对多尺度目标特征的提取能力，本文设计了 U 型网络递归模块，基本思想是 U 型网络中每次上采样输出需添加到同层卷积操作的输入端。为了防止参数过多，本文设计只递归一次。

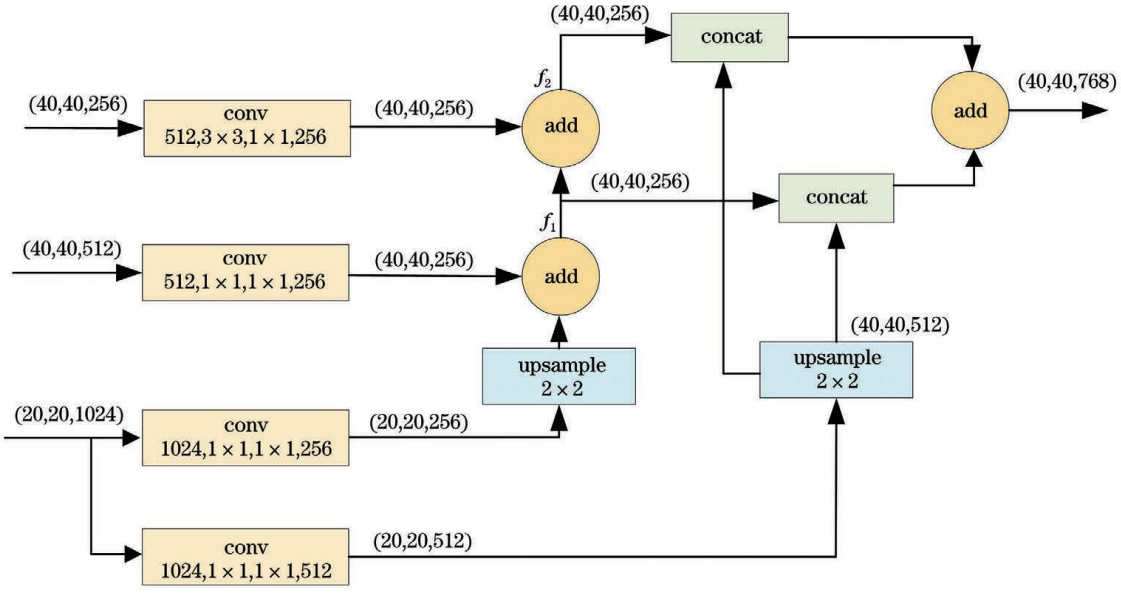

图 7　特征融合增强模块的细节

Fig. 7　Detail of feature fusion enhancement module

U 型网络递归模块的细节如图 8 所示。该模块输入的特征图大小为 40 pixel×40 pixel，依次通过 5 次 3×3 卷积，调整后的大小分别为 20 pixel×20 pixel、10 pixel×10 pixel、5 pixel×5 pixel、3 pixel×3 pixel 和 1 pixel×1 pixel，通道数均为 256，第 5 次卷积后的特征图大小为 1 pixel×

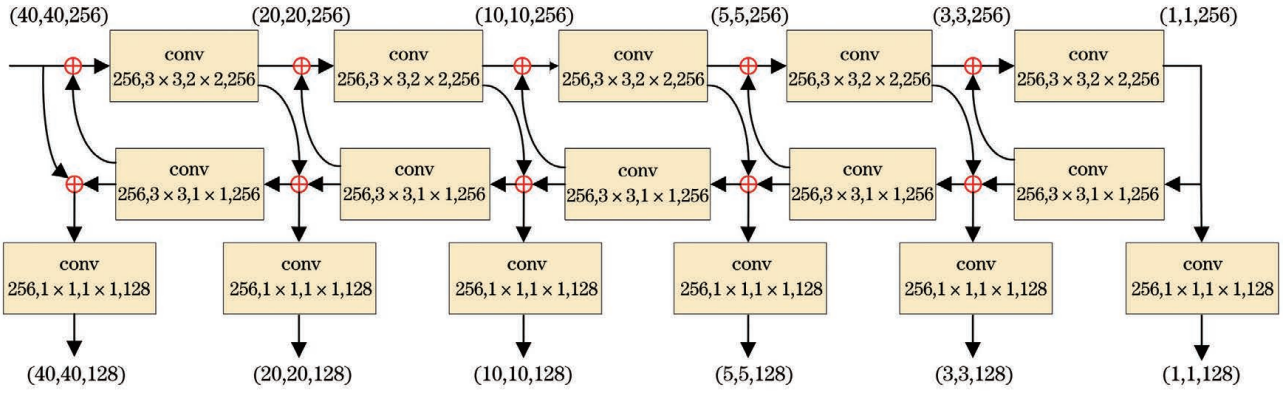

图 8　U 型网络递归模块的细节

Fig. 8　Details of recursive modules for U-shaped network

1 pixel，将该特征图作为输入并依次通过 3×3 卷积进行上采样，大小分别调整为 3 pixel×3 pixel、5 pixel×5 pixel、10 pixel×10 pixel、20 pixel×20 pixel 和 40 pixel×40 pixel，此时并不考虑输出。将上采样调整后的特征图与同层次卷积层的输入作特征相加以完成第一次循环，将相加的结果作为输入进行第二次卷积和上采样操作，第二次循环后将上采样结果与同层的卷积输入作特征相加，之后通过 1×1 的卷积将通道数从 256 调整为 128，并将该层特征输出，从而完成第二次循环。一个 U 型网络递归模块有 6 个大小不同的特征图输出，大小分别为 40 pixel×40 pixel、20 pixel×20 pixel、10 pixel×10 pixel、5 pixel×5 pixel、3 pixel×3 pixel 和 1 pixel×1 pixel，通道数均为 128，而且该模型一共有 8 个 U 型网络递归模块，所以可以输出 8 组不同水平和 6 个不同尺寸的特征图，将其作为 SFAM 的输入。

设特征融合增强模块输出的基本特征为 X_b，T_l 为第 l 个 U 型网络递归模块的递归学习过程，T_l^{out} 为第 l 个 U 型网络递归模块的最后一次学习过程，F 为 FFMv2 的处理过程，x_i^l 为第 l 个 U 型网络递归模块第 i 个尺度的输出，$L=8$，$1\leqslant i\leqslant 6$，则递归一次的 U 型网络递归模块在不同水平、不同尺度的输出特征 O 可以表示为

$$O=\begin{cases}T_l^{\text{out}}T_l(X_b), & l=1\\ T_l^{\text{out}}T_l[F(X_b,x_i^{l-1})], & l=2,\cdots,L\end{cases}。\tag{3}$$

3.3　残差边注意力模块

X 光图片中的目标有很强的颜色和形状等特征，特征图经过卷积后目标的相关特征会体现在特征图的空间和通道中，可以通过卷积块注意力机制来关注图片中的空间特征和通道信息。本文在 SFAM 结构中设计了残差边注意力模块，该模块由 CBAM 所构建的残差边结构组成，CBAM 可以调整空间和通道信息，残差边可以平衡 CBAM 的影响。设计后的残差边注意力模块如图 9 所示。

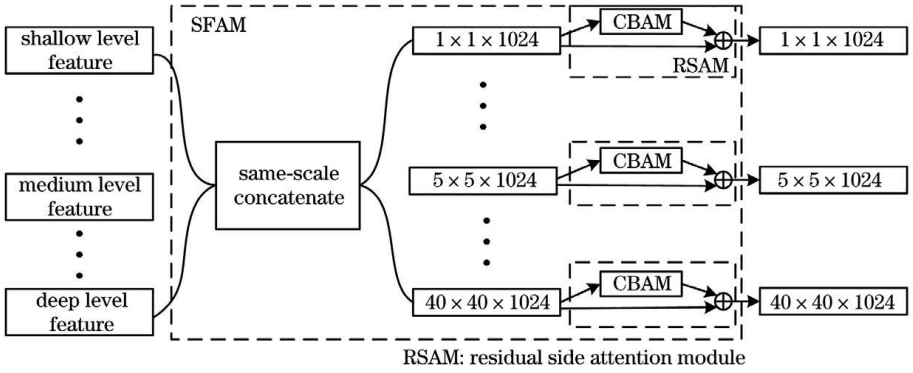

图 9　残差边注意力模块

Fig. 9　Residual side attention module

SFAM 的输入为 8 个 U 型网络递归模块的输出，每个 U 型网络递归模块的输出由 6 个不同尺度大小的特征图组成[(3)式]。SFAM 先对输入的不同层级且相同尺度大小的特征图进行特征融合，表达式为

$$X=\{X_1,\cdots,X_i\},\tag{4}$$

式中：$X_i = \mathrm{Concat}(x_i^1 + x_i^2 + \cdots + x_i^L)$，其中 Concat(•)为特征融合操作，$L=8$，$1 \leqslant i \leqslant 6$。被融合的特征 x_i^l 来自不同水平 l、不同大小 i 的 U 型网络递归模块的输出，X_i 经过残差边注意力模块后的输出可表示为

$$\hat{X}_i = X_i'' \oplus X_i, \tag{5}$$

式中：$X_i'' = M_S[M_C(X_i) \otimes X_i] \otimes [M_C(X_i) \otimes X_i]$。改进后的 SFAM 输出为 $\hat{X} = \{\hat{X}_1, \cdots, \hat{X}_i\}$，通道数均为 128，大小分别为 40 pixel×40 pixel、20 pixel×20 pixel、10 pixel × 10 pixel、5 pixel × 5 pixel、3 pixel×3 pixel 和 1 pixel×1 pixel。

4　实验结果

4.1　数据集与实验环境

实验是在自制的 X 光旅客安检数据集 SIXray_OD 上进行的，该数据集从公开的二分类数据集 SIXray 中挑出 8718 张包含违禁品的图片进行人工标注，构成 VOC 格式的数据集 SIXray_OD。通过 LabelIng 软件来框出图片中违禁品的位置并设置种类，生成包含违禁品位置和类别信息的 XML 文件。手工标注在图片上的框称作真实框，通过模型得到的框称作预测框，通过框与框的交并比来计算模型在训练过程中的损失以及判断匹配的情况。SIXray_OD 数据集共包含 5 类违禁品，如枪、扳手、刀具、钳子和剪刀，其中包含枪的图片数量为 2936 张，包含扳手的图片数量为 2266 张，包含刀具的图片数量为 156 张，包含钳子的图片数量为 3957 张，包含剪刀的图片数量为 1159 张，训练集、验证集和测试集的划分比例为 7∶1∶2。数据集中的部分图片如图 10 所示。

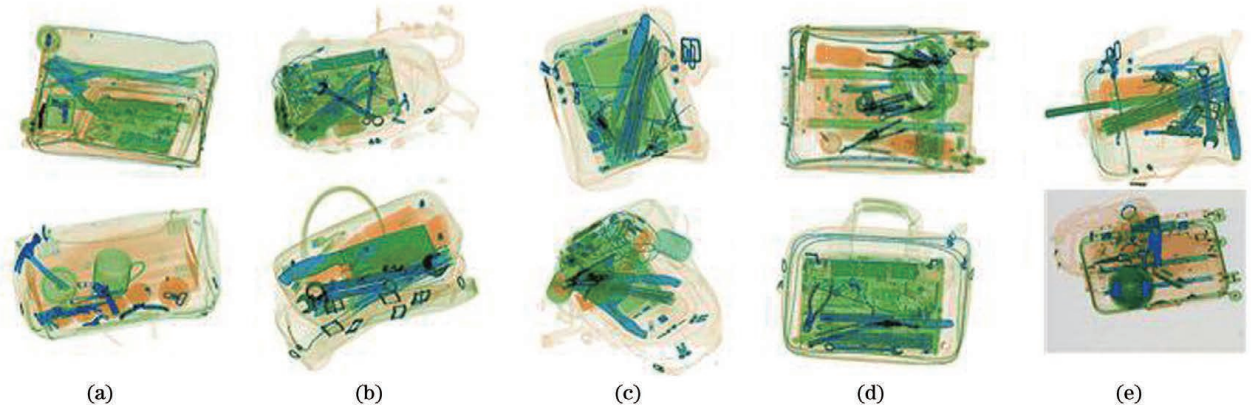

图 10　SIXray_OD 数据集中的部分图片。(a)枪；(b)扳手；(c)刀具；(d)钳子；(e)剪刀

Fig. 10　Some images in SIXray OD dataset. (a) Gun; (b) wrench; (c) knife; (d) pliers; (e) scissors

实验所用的计算机配备两张 NVIDIA GeForce GTX2080Ti 显卡，Ubuntu 16.04 操作系统，CUDA 版本为 10.0，cuDNN 版本为 7.6.0，深度学习框架为 TensorFlow-gpu1.13.1 和 Keras2.1.5。特征提取网络为 VGG-16，训练批处理大小为 16，迭代 100 次。注意力机制的压缩率为 16，阈值(置信度)为 0.5。利用预训练权重加速模型的收敛性和浅层网络提取特征具有的相似性，在前 50 轮训练过程中对模型的前 16 层进行冻结训练，冻结的是部分特征提取网络的参数以加快模型的收敛，在后 50 轮进行解冻训练，即对模型参数进行进一步调整。设置动态学习率，前 50 轮的初始学习率为 3×10^{-4}，如果损失在两轮内都没有改善，那么将学习率降低为原来的 1/2，后 50 轮的初始学习率设置为 1×10^{-4}，若损失在两轮内均没有改善，则将其降低为原来的 1/2。为了防止网络过拟合，设置早停机制，当损失连续 6 轮没有改善时，触发早停机制，使模型训练停止。

4.2　实验结果与分析

对模型的性能进行评价，评价指标为平均检测精度(Mean Average Precision，mAP)、模型的参数量和检测图像的每秒帧数(Frames Per Second，FPS)等。

在改进 M2Det 模型之前，首先对 TUM 的数量进行调整，综合模型的参数量和平均检测精度来确定最适合的 TUM 数量，实验结果如表 1 所示。

表 1　不同 TUM 数量下的指标

Table 1　Index under different number of TUMs

Number of TUMs	Parameters /Mbit	mAP /%
2	182.05	73.34
4	227.88	73.99
6	293.61	76.89
8	**331.62**	**79.03**
10	383.78	79.62

从表 1 可以看到，随着 TUM 数量的增加，模型的平均精度会不断提高，参数量也会随之增加；当 TUM 的数量为 8 时，mAP 达到了 79.03%；当 TUM 的数量增加到 10 时，参数量并未增加，而且平均精度仅仅提升了 0.59 个百分点。综上，本文使用 8 个 TUM（$L=8$）进行接下来的改进实验。

首先对特征融合增强模块进行测试实验，本文在特征提取网络的后三层（pool 3＋conv 4_3＋pool 5）和后 4 层（pool 2＋pool 3＋conv 4_3＋pool 5）上构建 FPN 结构，同时对比增加高语义特征层 pool 5 前后的特征融合结果，实验结果如表 2 所示。

表 2 特征融合增强模块的实验效果

Table 2 Experiment effect of feature fusion enhancement module

Structure	Integration of pool 5	mAP /%
Baseline(conv 4_3＋pool 5)	×	79.03(0)
Three layers (pool 3＋conv 4_3＋pool 5)	×	81.94(2.91↑)
Three layers (pool 3＋conv 4_3＋pool 5)	√	**82.82(3.79↑)**
Four layers (pool 2＋pool 3＋conv 4_3＋pool 5)	×	81.13(2.10↑)
Four layers (pool 2＋pool 3＋conv 4_3＋pool 5)	√	82.05(3.02↑)

从表 2 可以看到，构建多层的 FPN 结构加强了特征提取网络的特征提取能力，对 M2Det 中特征提取网络的后两层 conv 4_3 和 pool 5 进行了特征融合，本文通过构建多层的 FPN 来丰富包含在浅层特征层中的小目标信息，三层（pool 3＋conv 4_3＋pool 5）和 4 层（pool 2＋pool 3＋conv 4_3＋pool 5）结构的 mAP 值分别提升 2.91 个百分点和 2.10 个百分点，在相同的结构下，四层的 mAP 值比三层低约 0.8 个百分点，原因在于 pool 2 的加入导致深层特征层的比例减少，所以大目标的检测精度降低；将 pool 5 的输出特征与 FPN 的浅层输出特征进行特征融合以增加高语义特征信息，增加该层后，三层的 mAP 值提升 3.79 个百分点。

在上述改进的基础上继续实验。首先使用一个 U 型网络递归模块和 7 个 TUM 来验证该模块设计的有效性，然后使用 8 个 U 型网络递归模块进行实验，接着使用 SFAM 中的残差边注意力模块进行实验，最后对引入 CBAM 和 ECA（Efficient Channel Attention）模块[17]的结果进行对比，实验结果如表 3 所示。

表 3 U 型网络递归模块和残差边注意力模块的实验效果

Table 3 Experimental effects of U-shaped network recursive module and residual edge attention module

Structure	Parameters /Mbit	Residual edge attention module	mAP /%
8 TUM	334.50	×	82.82
1 RTUM ＋7 TUM	356.67	×	83.61(0.79↑)
8 RTUM	514.89	×	**84.09(1.27↑)**
8 RTUM＋CBAM	517.89(3.00↑)	×	85.12(2.30↑)
		√	**85.40(2.58↑)**
8 RTUM＋ECA	511.86(3.00↓)	×	84.86(2.04↑)
		√	85.02(2.20↑)

从表 3 可以看到，相比于 TUM，U 型网络递归模块能够进一步增强不同水平、不同大小特征层的表达能力，第一个 TUM 替换为 U 型网络递归模块，mAP 值提升 0.79 个百分点，8 个 U 型网络递归模块的 mAP 值提升 1.27 个百分点，第一个 U 型网络递归模块的引入对模型平均精度的提升比较显著，主要原因是 U 型网络递归模块的一部分输入来自上一个 U 型网络递归模块的输出，因此第一个 U 型网络递归模块的影响最大；引入 CBAM 和 ECA 后的 mAP 都有所提升，其中引入 CBAM 后的 mAP 值提升 1.03 个百分点，引入 CBAM 和残差边结构后的 mAP 值提升 1.31 个百分点，引入

ECA 后的 mAP 值提升 0.77 个百分点，引入 ECA 和残差边结构后的 mAP 值提升 0.93 个百分点；CBAM 的引入会增加模型 3.00 Mbit 的参数量，ECA 能够节省模型 3.00 Mbit 的参数量，但是由于参数量变化较小，最终在残差边注意力模块中使用 CBAM。

本文还测试了常用目标检测模型 SSD、Faster RCNN(Faster Region CNN)[18] 和 YOLOv3[19] 在 SIXray_OD 数据集上的精度，并与本文改进框架所得的结果进行对比，结果如表 4 所示。

表 4　不同检测模型的精度

Table 4　Accuracy of different detection models

Framework	Accuracy /%					mAP /%
	Gun	Pilers	Scissors	Knife	Wrench	
SSD	92	72	66	55	52	70.9
Faster RCNN	93	75	71	65	67	74.2
YOLOv3	92	71	64	53	51	66.2
M2Det	94	83	76	74	68	79.0
M2Det+EFFM	96	86	82	77	73	82.8
M2Det+EFFM+RTUM	96	88	82	78	76	84.0
EM2Det	98	89	84	79	77	85.4

从表 4 可以看到，SSD、Faster RCNN 和 YOLOv3 在 SIXray_OD 数据集上的 mAP 值分别为 70.9%、74.2%和 66.2%；常用 M2Det 模型的检测效果较好，mAP 值能够达到 79.0%，单类别的 mAP 值分别为 94%(gun)、83%(pilers)、76%(scissors)、74%(knife)和 68%(wrench)；本文设计的 EM2Det 模型的 mAP 值达到 85.4%，比 M2Det 模型提升 6.4 个百分点，单类别的 mAP 值分别提升 4,6,8,5,9 个百分点，发现对于扳手和剪刀的检测精度，提升效果最明显。对于具有遮掩干扰和尺度变化的 X 光图像，EM2Det 模型对目标的细节特征提取能力比原 M2Det 模型好，具有更优的多尺度目标检测性能，其中钳子和剪刀常以小目标的形式进行尺度变化，而扳手尺寸比较大，局部遮掩严重，但 EM2Det 模型对这三类均有较大精度的提升。不同情况下的可视化效果如图 11 所示。

从图 11 可以看到，EM2Det 模型在提取细节特征方面有更优秀的表现，在有背景干扰的情况下仍

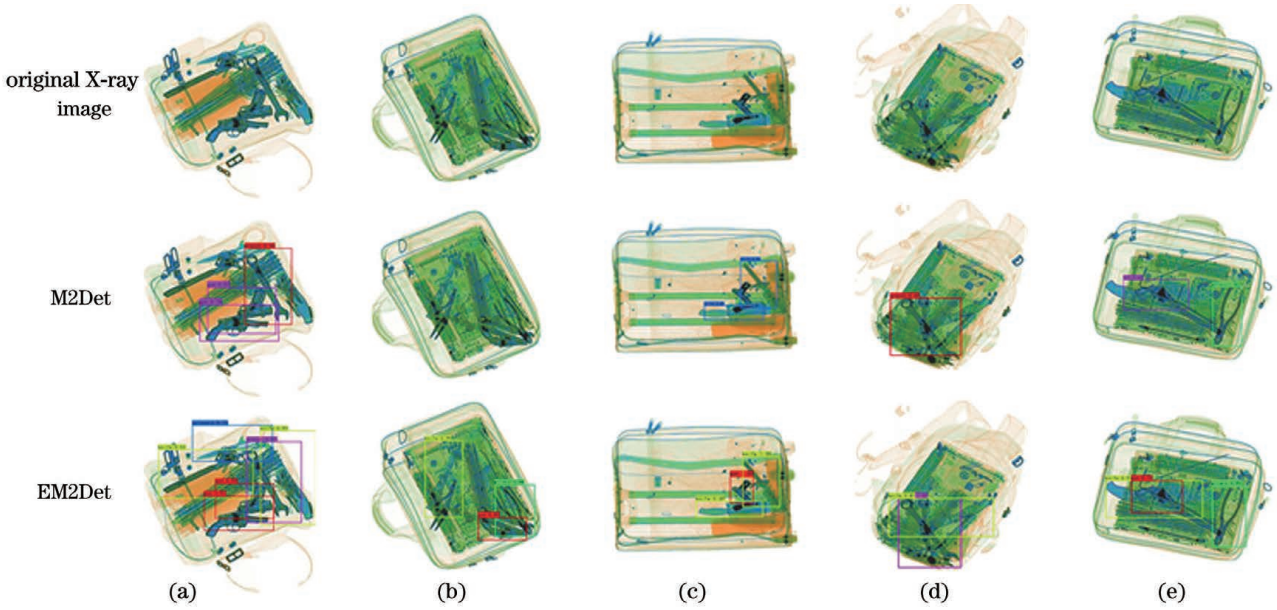

图 11　不同情况下的可视化效果。(a)剪刀；(b)钳子；(c)枪；(d)扳手；(e)刀

Fig. 11　Visualizations in different situations. (a) Scissors; (b) pliers; (c) gun; (d) wrench; (e) knife

能正确识别出违禁品，而且检测框能够准确地框出违禁品。当然检测效果提升的代价是增加模型的参数量，经过实验测试得到 M2Det 模型在 SIXray_OD 数据集上的检测速度约为 11 frame/s，EM2Det 模型的检测速度约为 8 frame/s，但在机场的安检现场中也可以满足实时检查的要求。

5 结 论

本文在 M2Det 模型的基础上设计 EM2Det 模型，增强了在具有遮掩干扰的 X 光图片中多尺度目标的特征提取能力。首先在特征提取网络 VGG-16 中的后三层(pool 3＋conv 4_3＋pool 5)上设计了特征融合增强模块以提取更多的多尺度特征，而且能够同时考虑平衡深层特征中的浅层特征比例。然后设计了 U 型网络递归模块以进一步增强不同水平、不同大小的特征表达能力。最后使用 CBAM 来构造残差边注意力模块以优化 SFAM，使模型更加关注有效特征。本文在 SIXray_OD 数据集上进行改进实验，实验结果表明本文设计的 EM2Det 模型的平均精度提升效果显著。

参 考 文 献

[1] Hou Y Y. Research on the relationship between work stress and safety performance of airport security inspectors[D]. Beijing: Beijing Jiaotong University, 2018: 5-13.
侯彦伊. 机场安检人员工作压力与安全绩效关系研究[D]. 北京: 北京交通大学, 2018: 5-13.

[2] Chen Z Q, Zhang L, Jin X. Recent progress on X-ray security inspection technologies[J]. Chinese Science Bulletin, 2017, 62(13): 1350-1365.
陈志强, 张丽, 金鑫. X 射线安全检查技术研究新进展[J]. 科学通报, 2017, 62(13): 1350-1365.

[3] Zhang C L. Research on X-ray image procession method in the security system[D]. Shenyang: Shenyang University, 2015: 31-42.
张春兰. 安检系统中 X 射线透射图像处理方法研究[D]. 沈阳: 沈阳大学, 2015: 31-42.

[4] 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.

[5] Ji Z, Kong Q K, Wang J. Object detection algorithm guided by dual attention models[J]. Laser & Optoelectronics Progress, 2020, 57(6): 061008.
冀中, 孔乾坤, 王建. 一种双注意力模型引导的目标检测算法[J]. 激光与光电子学进展, 2020, 57(6): 061008.

[6] Ma Y J, Liu P P. Convolutional neural network based on DenseNet evolution for image classification algorithm[J]. Laser & Optoelectronics Progress, 2020, 57(24): 241001.
马永杰, 刘培培. 基于 DenseNet 进化的卷积神经网络图像分类算法[J]. 激光与光电子学进展, 2020, 57(24): 241001.

[7] Akçay S, Kundegorski M E, Devereux M, et al. Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery[C]//2016 IEEE International Conference on Image Processing (ICIP), September 25-28, 2016, Phoenix, AZ, USA. New York: IEEE Press, 2016: 1057-1061.

[8] Steitz J M O, Saeedan F, Roth S. Multi-view X-Ray R-CNN[M]//Brox T, Bruhn A, Fritz M. GCPR 2018: pattern recognition. Lecture notes in computer science. Cham: Springer, 2019, 11269: 153-168.

[9] Galvez R L, Dadios E P, Bandala A A, et al. YOLO-based threat object detection in X-ray images[C]//2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), November 29-December 1, 2019, Laoag, Philippines. New York: IEEE Press, 2019: 19556495.

[10] Zhao Q J, Sheng T, Wang Y T, et al. M2Det: a single-shot object detector based on multi-level feature pyramid network[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 9259-9266.

[11] 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.

[12] 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.

[13] Qiao S, Chen L C, Yuille A. DetectoRS: detecting objects with recursive feature pyramid and switchable atrous convolution[EB/OL]. (2020-06-03)[2020-11-20]. https://arxiv.org/abs/2006.02334.

[14] 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.

[15] 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.

[16] Zhang K, Teng G W, Fan T, et al. FPN multi-scale object detection algorithm based on dense connectivity [J]. Computer Applications and Software, 2020, 37 (1): 165-171, 212.
张宽, 滕国伟, 范涛, 等. 基于密集连接的 FPN 多尺度目标检测算法[J]. 计算机应用与软件, 2020, 37 (1): 165-171, 212.

[17] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.

[18] 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.

[19] Redmon J, Farhadi A. Yolov3: an incremental improvement[EB/OL]. (2018-04-08) [2020-11-20]. https://arxiv.org/abs/1804.02767.