

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2023)02-0430-11

论文引用格式: Han P, Yang H and Fang C. 2023. Region enhancement and multi-feature fusion for contraband recognition in X-ray images. Journal of Image and Graphics, 28(02): 0430-0440 (韩萍, 杨慧, 方澄. 2023. 区域增强和多特征融合的 X 光图像违禁品识别. 中国图象图形学报, 28(02): 0430-0440) [DOI:10.11834/jig.211134]

## 区域增强和多特征融合的 X 光图像违禁品识别

韩萍, 杨慧, 方澄\*

中国民航大学智能信号与图像处理天津市重点实验室, 天津 300300

**摘要:** **目的** 对旅客行李进行安全检查是维护公共安全的措施之一, 安检智能化是未来的发展方向。基于 X 光图像的安检因不同的安检机成像方式不同, 同一类违禁品在不同设备上的 X 光图像在颜色分布上有很大的差异, 导致安检图像智能识别算法在训练与测试数据分布不同时, 识别性能明显降低, 同时 X 光行李图像中物品的混乱复杂增加了违禁品识别的难度。针对上述问题, 本文提出一种区域增强和多特征融合模型。**方法** 首先, 通过注意力机制的思想提取一种区域增强特征, 消除颜色分布不同的影响, 保留图像整体结构并增强违禁品区域信息。然后, 采用多特征融合策略丰富特征信息, 使模型适用于图像中物品混乱复杂情况。最后, 提出一种三元损失函数优化特征融合。**结果** 在公开数据集 SIXray 数据集上进行整体识别性能和泛化性能的分析, 即测试本文方法在相同和不同颜色分布样本上的性能。在整体识别性能方面, 本文方法在平均精度均值 (mean average precision, mAP) 上相较于基础模型 ResNet18 和 ResNet34 分别提升了 4.09% 和 2.26%, 并优于一些其他识别方法。对于单类违禁品, 本文方法在枪支和钳子类违禁品上的平均识别精度为 94.25% 和 90.89%, 相较于对比方法有明显优势。在泛化性能方面, 本文方法在 SIXray\_last101 子数据集上可正确识别 26 张含违禁品样本, 是基础模型能够正确识别数量的 4.3 倍, 表明本文方法在颜色分布不同样本上的有效性。**结论** 本文方法根据 X 光安检图像颜色差异的特点设计出一种区域增强特征, 并与彩色和边缘特征融合, 以获取多元化信息, 在枪支、刀具、钳子等违禁品的识别任务中表现出较好效果, 有效缓解了图像颜色分布差异导致的性能下降问题。

**关键词:** X 光安检图像; 违禁品识别; 区域增强; 注意力机制; 多特征融合

## Region enhancement and multi-feature fusion for contraband recognition in X-ray images

Han Ping, Yang Hui, Fang Cheng\*

Tianjin Key Laboratory for Advanced Signal Processing, Civil Aviation University of China, Tianjin 300300, China

**Abstract:** **Objective** X-ray security screening technology is widely used in public transportation infrastructures. The real-time security images are generated via X-ray-related scanning for checking. Due to manual inspection mechanism has its hidden risks, it is required to develop prohibited items-related intelligent recognition based on X-ray security check images.

**Method** The convolutional neural network based (CNN-based) technique has been developing dramatically in the field of computer vision tasks. The CNN-based intelligent recognition model is restricted by a huge amount of label-manual X-ray images for training. Current recognition model is just suitable for homogeneous data-distributed between the training and tes-

收稿日期: 2021-12-06; 修回日期: 2022-03-09; 预印本日期: 2022-03-16

\* 通信作者: 方澄 cfang@cauc.edu.cn

基金项目: 中国民航安全能力基金项目 (14002500000019J012, 14002500000019J014)

Supported by: Safety Capacity Construction Funds for Civil Aviation Administration of China (14002500000019J012, 14002500000019J014)

ting sets. When the color distribution of X-ray images in the testing set is inconsistent in the training set, it is difficult to identify the target for the model. However, the problem of multiple datasets distribution is more prominent in practice for such application scenarios. Dual-energy X-ray imaging technique allows the scanner to distinguish different colors in terms of the item's effective atomic number. The heterogeneity problem of X-ray images is challenged in color distribution. The performance of X-ray image intelligent recognition algorithm will be lower intensively when the distribution of training and test data is inconsistent. So, we develop a heterogeneity-alleviated multi-feature fusion model further. First, to alleviate the influence of different color distribution of prohibited items, the attention mechanism is adopted to extract a newly pixel-level feature, called region-enhanced feature, which are trained in terms of overall feature distribution. The generalization ability is improved for multicolor-distributed X-ray images. Then, multi-feature fusion strategy is used to enrich the feature information like color, shape and outline. The features of color, edge and region-enhanced are melted into a centralized manner. The balanced weight parameters are added to the three kinds of features. Multi-feature fusion can be used to realize more effective feature information and optimal robustness in the case of chaotic objects in an image. Finally, a ternary loss functions are illustrated in relevant to fusion, edge and regional enhancement. To get feature fusion better, weight of three losses are set to balance the weighted feature-parameters. **Result** The experimental analysis is carried out on the public dataset for the performance evaluation of entirety and generalization (i. e., performance on test samples with the same and different color distributions), called SIXray. Our mean average precision (mAP) can be improved by 4.09% and 2.26% of each in comparison with the ResNet18 and ResNet34. For a single class prohibited items, the average accuracy of our method can reach 94.25% and 90.89% in the identification of guns and pliers-prohibited. We can identify 26 samples in SIXray\_ last101 dataset in generalization, which is 4.3 times beyond benchmark. The demonstration shows the effectiveness is improved in terms of multicolor-distributed samples. Additionally, ablation experiments are conducted to verify the effects of multiple features and hyper-parameter settings. The experimental results show that the overall recognition performance can be improved based on the richer multiple features (each of edge features and regional enhancement features improve the overall recognition performance by 1.32 and 1.05 percentage points). **Conclusion** A region-enhanced multi-feature fusion method is developed to deal with rich color and different distribution and chaotic and complex objects through X-ray security images-relevant feature analysis. The enhanced region features are obtained in terms of feature distribution overall. Multi-feature fusion strategy is implemented for the optimization of color, shape and contour details. And, a ternary loss function is used to improve the fusion effect and its heterogeneity. Our analyses demonstrate that the performance of the model can be improved for prohibited items checking. The effectiveness and robustness of the proposed method are verified as well. The multi-branch structure of the model is required to be developed further due to its limitations of computational cost and recognition efficiency.

**Key words:** X-ray security image; contraband recognition; region enhancement; attention mechanism; multi-feature fusion

## 0 引言

X光安检机广泛应用于机场、火车站、地铁站等交通运输和各种大型公共活动场所,是维护公共安全和运输安全的重要设施。旅客行李通过X光机扫描并实时产生安检图像,安检员通过检查安检图像判断行李中是否存在违禁品,从而保障公共安全(Mery等,2017)。然而,人工检查机制存在漏检和错检的风险(侯彦伊,2018)。对X光安检图像中的违禁品进行智能识别具有重要的现实意义。

X光安检图像违禁品的识别旨在判断X光图

像中是否存在违禁品及违禁品的类别。随着深度学习的发展,违禁品智能识别多依赖于基于深度学习的智能识别算法,Akça等人(2016)将卷积神经网络(convolutional neural networks,CNN)应用于安检图像研究,采用迁移学习和视觉词袋模型(bag of visual word,BoVW)(Baştan等,2011)检测含有枪支的安检图像取得了较好结果。该研究表明,卷积神经网络方法明显优于传统方法。随后出现了更多的采用深度学习方法进行检测的研究(Jaccard等,2017; Zhao等,2018; Morris等,2018)。Caldwell等人(2017)研究了卷积神经网络在不同X光数据集训练模型的泛化能力。首先用X光货物数据集训练

网络,然后用 X 光包裹数据集测试,但效果并不理想,当采用两个数据集的样本训练时,模型性能得到很大改善。该研究表明,CNN 很难识别与训练数据来源不同的目标数据。近期,一项较有影响力的研究公开了大规模安检图像数据集 SIXray (Miao 等, 2019),并针对数据集类别极不平衡和背景复杂的特点提出了类不平衡分层细化(class-balanced hierarchical refinement, CHR)的方法,该方法提高了模型的精度与速度,有效缓解了数据类别不平衡导致的识别性能下降问题。

智能识别模型需要在大量有标注的 X 光图像训练集上学习得到,然后根据模型在测试集上的表现验证识别性能,当前的识别模型仅适用于测试集与训练集保持同一分布的数据(何彦辉 等, 2021)。本文在实验中发现,当测试集中出现与训练集图像颜色分布不一致的含有违禁品的 X 光图像时,即便是人工很容易识别的违禁品,模型都难以正确识别。实际上,数据集不同分布的情况在实际应用场景下表现得更为突出,安检设备采用双能量 X 射线成像技术,使扫描可以根据物体的有效原子序数呈现不同颜色。例如,金属显示为蓝色,有机物显示为橙色等,随着技术的发展,安检设备不断更迭,各站点使用的安检设备都不完全相同,由于不同安检设备应用不同的图像处理方法,采集到的图像颜色分布存在较大差异。如图 1 所示,①和⑤均含有手枪类违禁品,在 X 光扫描后呈现蓝色,但二者的蓝色存在一定差异。同时,旅客行李中物品种类繁多,所成 X 光图像颜色丰富,由图像颜色分布差异引起的识别性能下降问题对安检系统的智能化造成挑战。因此智能识别模型应既能适用于各种设备采集的 X 光安检图像,又能利用安检图像的多彩颜色特征,解决混乱复杂导致图像中的违禁品易与相同材质的其他物品混淆的问题。

针对以上问题,本文提出一种区域增强和多特征融合模型(region enhanced multi-feature fusion model, REMF),融合了颜色和边缘等更多元化的信息,实现了对物品杂乱的 X 光行李图像中的违禁品进行识别。本文主要贡献如下:1)定义了一种区域增强特征。特征提取器应用注意力机制的思想通过学习训练集样本的整体分布,将注意力集中在图像中的违禁品部分。区域增强特征可消除颜色分布不同的影响,保留图像几何结构并对违禁品区域进行

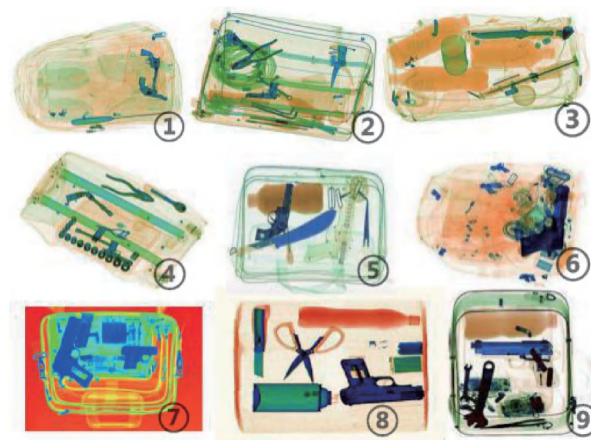


图 1 不同颜色分布安检图像样例

Fig. 1 Samples of original images of security inspection with different color distribution

增强,突出违禁品的形状信息,提高模型对不同颜色分布的 X 光图像的泛化能力。2)采用多特征融合策略。安检人员检查 X 光行李图像通常是在颜色信息的基础上通过观察物品呈现的形状轮廓进行判断,因此图像的颜色、形状和轮廓信息都具有重要意义。本文将彩色特征、边缘特征和区域增强特征进行集中式融合,并且在 3 种特征上增加可调权重参数,使不同特征发挥各自优势,得到更加丰富的特征信息,提高模型在图像中物品混乱复杂情况下的鲁棒性。3)根据本文方法多特征的特点引入一种三元损失函数,引导特征自适应融合。三元损失函数由融合损失、边缘损失和区域增强损失组合而成,通过设置 3 种损失的权重,引导模型通过调整特征权重参数更好地进行特征融合。

## 1 相关基础理论

### 1.1 注意力机制

注意力机制(attention mechanism)是深度学习领域中广泛应用的一种思想(Xiao 等, 2015; Vaswani 等, 2017; Wang 等, 2017),借鉴了人类在感知事物过程中的思维方式。人眼在扫过视野范围时可以迅速获取需要关注的目标区域,极大提高视觉信息的处理效率及准确性。注意力机制通常以加权的方式实现,其计算式为

$$A = F(X) \quad (1)$$

$$X_A = A \otimes X \quad (2)$$

式中,  $X \in \mathbf{R}^{W \times H \times C}$  表示输入特征图,  $F(\cdot)$  为由神经



网络构成的注意力模块,  $\mathbf{A}$  为注意力权重,  $\otimes$  表示将权重赋予输入特征图的操作, 包括矩阵数乘或点乘,  $\mathbf{X}_A$  表示输出特征图。

注意力机制按照特征的不同维度可以分为通道注意力和空间注意力(刘可文等, 2020), CBAM(convolutional block attention module)方法(Woo等, 2018)将通道注意力与空间注意力串联起来, 相比于仅关注通道域, CBAM 可以达到更好的效果, 其框架如图2所示。

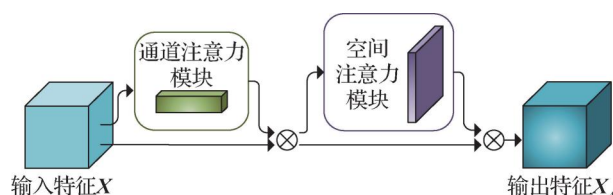


图2 CBAM 结构图

Fig.2 Structure diagram of CBAM

通道注意力模块和空间注意力模块的处理过程为

$$F_C(\mathbf{X}) = \sigma(MLP(f_{AvgPool}(\mathbf{X})) + MLP(f_{MaxPool}(\mathbf{X}))) \quad (3)$$

$$F_S(\mathbf{X}) = \sigma(Conv([f_{AvgPool}(\mathbf{X}); f_{MaxPool}(\mathbf{X})])) \quad (4)$$

式中,  $F_C(\cdot)$  表示通道注意力模块,  $F_S(\cdot)$  表示空间注意力模块。 $f_{AvgPool}$  和  $f_{MaxPool}$  分别表示平均池化和最大池化操作, 用以聚合特征映射的空间信息;  $MLP$  和  $Conv$  分别为全连接网络和卷积操作, 对特征进行非线性映射;  $\sigma$  为 sigmoid 函数, 将权重值映射到 0 ~ 1 之间。通道注意力权重  $F_C(\mathbf{X}) \in \mathbf{R}^{1 \times 1 \times C}$  表示输入特征在每个通道上的重要程度, 空间注意力权重  $F_S(\mathbf{X}) \in \mathbf{R}^{W \times H \times 1}$  表示输入特征在每个区域的重要程度。

两个注意力模块在训练中学习寻找输入特征中需要关注的部分, 输入特征  $\mathbf{X}$  在经过注意力模块加权后得到输出特征  $\mathbf{X}_A$ ,  $\mathbf{X}_A$  相较于  $\mathbf{X}$  在通道和空间维度的关键部分增强, 噪声与其他无意义信息减弱, 注意力机制的引入提高了模型的学习效率。

## 1.2 多特征融合

生活中的图像较为复杂, 仅依赖某种特征进行判断往往无法达到很好的效果, 需要将蕴含各种信息的特征融合, 发挥其优势, 提升模型性能。特征融合分为分布式融合和集中式融合。分布式融合先根

据不同的特征集指定个体决策, 然后将其组合到一个决策中。集中式融合先将不同的特征进行同化处理后集成到一个决策中, 再根据任务需求进行特征选择。对于不同表现方式的特征, 集中式融合方法既不会在融合过程中损失或破坏原有的特征信息, 又能将不同特征的优势结合起来。Liu 等人(2019)提出的回归模型采用了集中式融合方法, 其结构如图3所示。首先, 对输入图像进行不同特征的提取, 得到  $N$  种不同表现方式的特征, 记为特征  $i$ , 然后将其送入卷积神经网络提取相应的深度特征, 记为深度特征  $i$ , 将深度特征  $i$  进行拼接后送入全连接网络进行特征分析与选择, 最后根据任务需求得到输出结果。

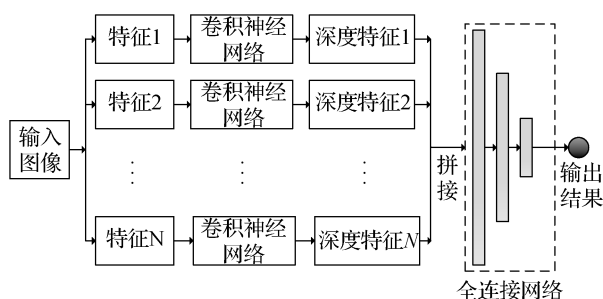


图3 集中式融合结构图

Fig.3 Structure diagram of centralized fusion

## 2 本文方法

不同安检机生成的图像颜色分布存在偏差, 如图4(b)所示。当出现与训练集颜色分布不同样本时, 智能检测效果会受到影响, 意味着当前的智能方法在识别违禁品时过多依赖了颜色信息。因此在设计违禁品智能识别算法时, 需加强能够区别违禁品与其他物品除颜色以外的特征。安检人员检查 X 光图像时, 通常根据颜色信息判断物体材质, 并在此基础上通过观察物品的形状轮廓来判断物品类别。因此, 颜色、形状和轮廓都是违禁品识别的重要信息。基于上述分析, 本文提出了区域增强和多特征融合模型 REMF, 应用注意力机制设计一种像素级的区域增强特征, 消除图像颜色分布差异的影响, 增强违禁品区域信息, 在保留图像几何结构的同时凸显违禁品的形状, 将彩色特征、边缘特征和区域增强特征进行集中式融合, 通过三元损失函数对特征融合进行优化。

本文方法分为两个阶段, 一阶段模型的训练可

以得到区域增强特征提取器 (region-enhanced feature extractor, RFE), 二阶段模型融合了多种特征并完成违禁品识别任务。本文方法的实现分为训练过

程和测试过程。训练时需训练一阶段模型和二阶段模型; 测试时仅需将图像输入二阶段模型中进行违禁品识别。REMF 方法结构如图 5 所示。

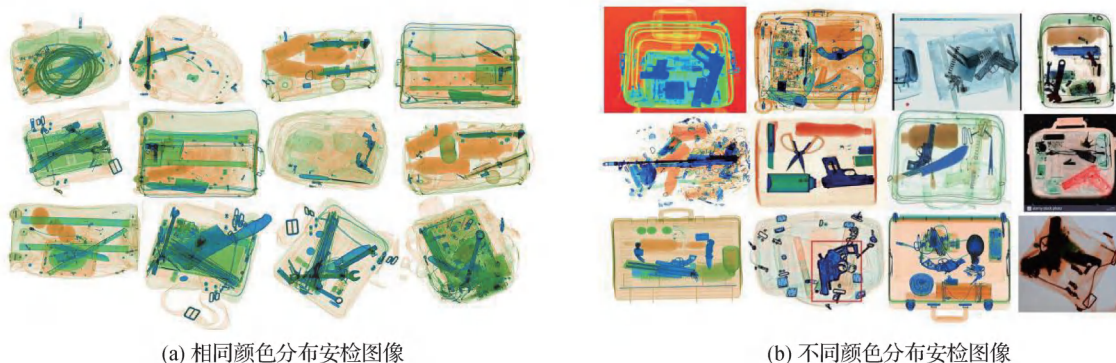


图 4 相同颜色分布和不同颜色分布安检图像样例

Fig. 4 Examples of security images with the same color distribution and different color distribution

((a) security images with the same color distribution; (b) security images with different color distribution)

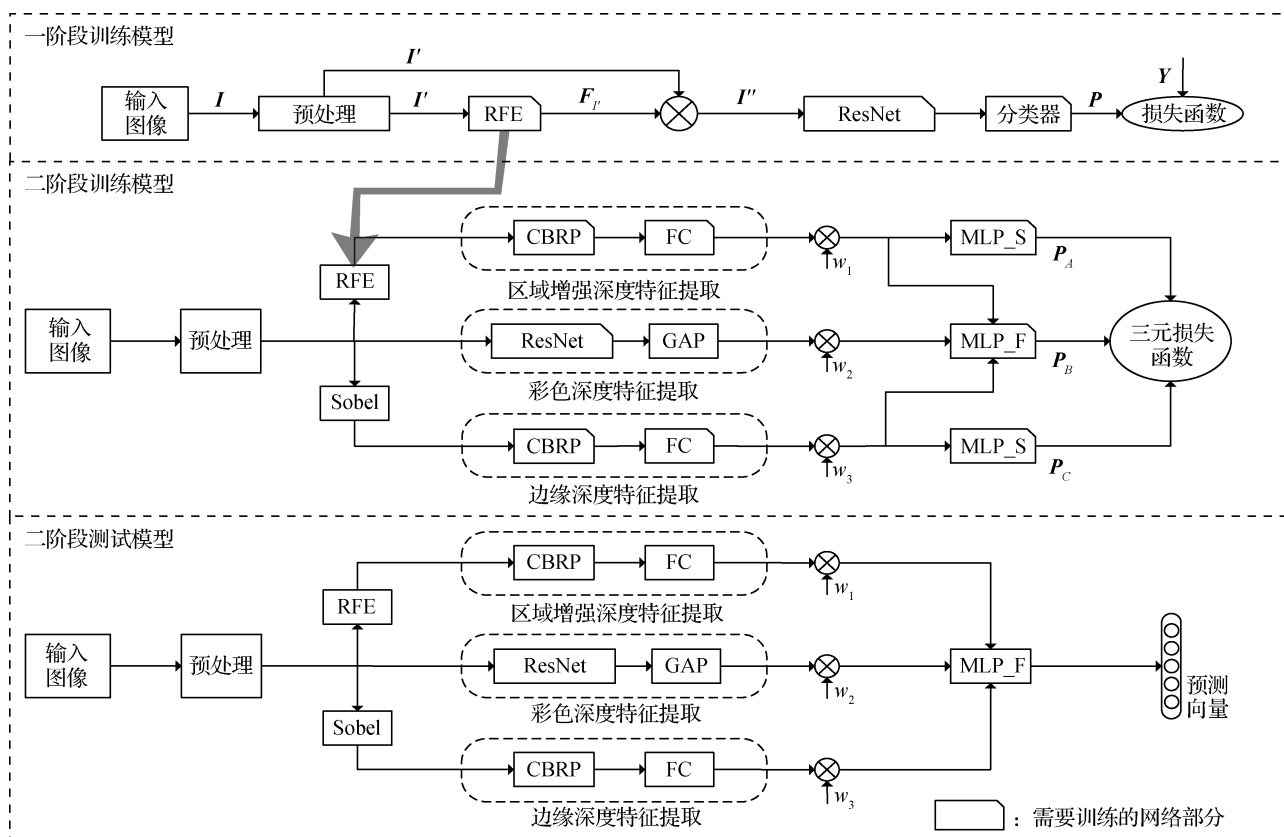


图 5 REMF 方法框图

Fig. 5 Structure diagram of REMF

## 2.1 一阶段模型

### 2.1.1 RFE 模块

RFE 模块采用 CBAM 的注意力结构, 如图 6 所示。通过归一化和标准化预处理后的输入图像

$I' \in \mathbf{R}^{W \times H \times 3}$  ( $W$  和  $H$  分别为  $I'$  的宽和高), 通过两层卷积操作, 得到尺寸为  $W/2 \times H/2$ 、通道数为 32 的特征图  $I'_1$ , 将其输入通道注意力模块, 得到尺寸为  $1 \times 1$ 、通道数仍为 32 的通道注意力掩码  $A_c$ , 表示各

个通道特征的权重系数;将 $A_c$ 与 $I'_1$ 按照对应通道相乘得到新的特征 $I'_2$ 。将 $I'_2$ 输入空间注意力模块,得到尺寸为 $W/2 \times H/2$ 、通道数为1的空间注意力

掩码 $A_s$ ,采用双线性插值的方法对其进行上采样,使之与输入图像尺寸相同,得到区域增强特征 $F_r$ 。

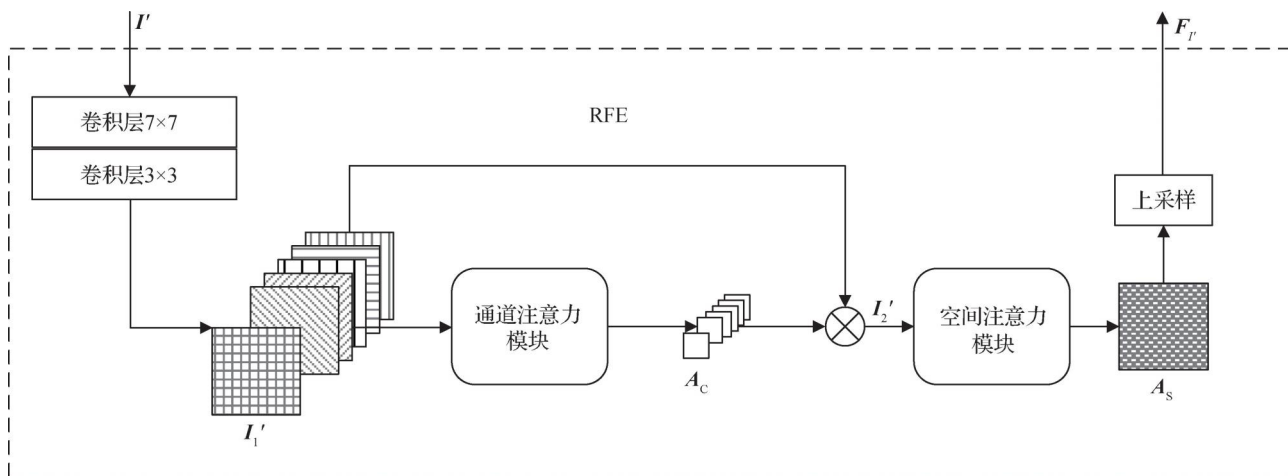


图6 RFE 模块结构图

Fig.6 Structure of region-enhanced feature extractor

区域增强特征的本质为输入图像的注意力分布掩码,每个数值表示图像中对应像素的重要程度,图7为图1的可视化区域增强特征图。与图1对比可以看出,区域增强特征图消除了颜色分布不同的影响,可以很好地保留图像的几何结构,违禁物品区域相较于背景和其他物品得到明显增强。

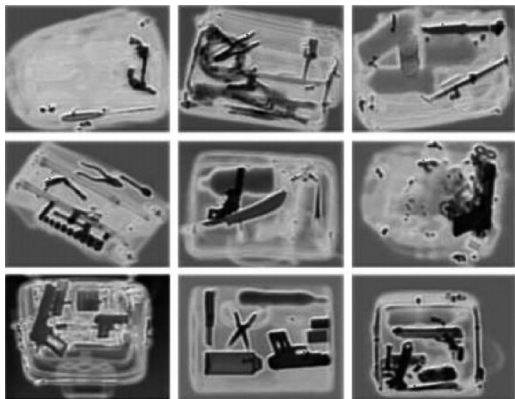


图7 不同颜色分布的安检图像区域增强特征样例

Fig.7 Samples of region-enhanced feature of security inspection with different color distribution

### 2.1.2 一阶段模型结构

在一阶段模型中,输入图像 $I$ 通过预处理后得到 $I'$ ,将 $I'$ 送入RFE模块进行特征提取,得到像素级区域增强特征 $F_r \in \mathbf{R}^{W \times H \times 1}$ ,将 $F_r$ 与 $I'$ 的每个通道按照对应位置相乘得到 $I''$ 。将 $I''$ 送入ResNet模型(He等,2016)进行深度特征提取,后接全连接分

类器得到预测向量 $P$ ,即每类违禁品存在的可能性。标签向量 $Y$ 表示该图像中每类违禁品是否存在,采用交叉熵损失函数(Rubinstein,1999)计算预测向量 $P$ 与标签向量 $Y$ 的损失,通过反向传播更新模型参数进行训练,当损失不再下降时终止训练。一阶段模型的训练调整了RFE模块的参数,使之能够找到图像中与识别结果重要相关的区域。由于第2阶段模型结构中RFE模块不再对输入图像进行加权,为防止第2阶段模型的训练对RFE模块的性能产生影响,训练结束后将RFE模块中的参数进行冻结,在第2阶段直接使用该模块提取区域增强特征。

## 2.2 二阶段模型

### 2.2.1 二阶段模型结构

二阶段模型是多特征融合的模型,将图像的彩色特征、边缘特征以及区域增强特征用神经网络模块进行深度特征提取并融合用于识别图像中的违禁品。如图5所示,输入图像经过图像预处理后进入3个并行的网络分支,从上到下依次是区域增强特征提取网络、彩色特征提取网络和边缘特征提取网络。

在彩色特征分支中,将预处理后的RGB三通道彩色图像送入卷积神经网络模型进行深度特征提取,选择基础的残差网络ResNet作为图像彩色特征的提取器,图像进入网络模型后,经过4个残差层提取到 $7 \times 7$ 大小的特征图,通过全局平均池化



(global average pooling, GAP) 操作, 得到 1 024 元 1 维特征向量。

对信息量更少的单通道区域增强特征和边缘特征, 使用层数较浅的 CBRP (convolution batchnormal relu pooling) 卷积模块, CBRP 卷积模块的结构如图 8 所示, 由 layer1—layer4 堆叠而成, 每层包括卷积、归一化、激活和池化操作。

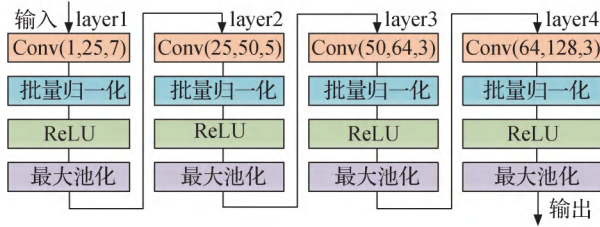


图 8 CBRP 卷积模块结构图

Fig. 8 Structure diagram of CBRP convolution module

在区域增强特征分支中, 先将预处理后的图像送入 RFE 模块提取区域增强特征, 将其输入 CBRP 模块进行深度特征提取, 最后将深度特征通过全连接 (fully connected, FC) 层映射为 1 024 元 1 维特征向量。在边缘特征分支中, 先用 sobel 算子对预处理后的图像进行边缘提取, 后接 CBRP 卷积模块进行深度特征提取, FC 层后得到从边缘特征中提取到的 1 维特征向量。

由于 3 种特征在不同违禁品识别场景下对最终决策的重要程度不同, 在 3 种特征向量上分别增加了可调权重参数  $w_1$ 、 $w_2$  和  $w_3$ 。权重参数可在训练过程中自适应调整, 使 3 种特征发挥不同程度的作用, 更好地进行融合。图 5 中的 MLP\_F 模块采用集中式融合方法完成特征融合任务, 将 3 种 1 维特征向量拼接后输入两层全连接网络, 最终输出每类违禁品的得分。全连接网络融合方法的本质是对 3 种特征提取出的 1 维特征向量进行特征选择, 通过激活函数选择对识别违禁品任务有意义的特征表现, 最后综合特征表现判断图像中是否含有违禁物品。

### 2.2.2 三元损失函数

实验中发现, 若仅采用 MLP\_F 模块输出预测分数计算损失进行训练, 模型在相同分布图像上的性能虽有提升, 但对于不同颜色分布的图像, 模型的识别情况没有明显提高。寻找问题时发现此训练方式会使区域增强特征与边缘特征分支的权重  $w_1$  和  $w_3$

过小, 意味着在特征融合时, 彩色特征在网络的学习过程中被更多地依赖, 而其他特征只发挥了微小的作用。

为了加强标签信息对于被弱化特征的监督作用, 本文分别在边缘特征和区域增强特征后连接了由两层全连接层构成的 MLP\_S 模块, 输出仅由单一特征得到的预测向量, 因此在训练过程中模型的输出为  $P_A, P_B, P_C$ , 分别代表边缘特征、融合特征和区域增强特征的预测向量, 每个输出向量对应一个交叉熵损失函数, 即  $L_A, L_B$  和  $L_C$ 。本文构建了一种三元损失函数, 计算为

$$Loss = \alpha \times L_A + \beta \times L_B + \gamma \times L_C \quad (5)$$

式中,  $\alpha, \beta$  和  $\gamma$  为 3 个输出向量对应损失函数的系数, 且  $\alpha + \beta + \gamma = 1$ 。实验时, 根据网络表现设置 3 个系数的值, 引导网络在训练过程中更好地进行特征融合。

交叉熵损失函数计算为

$$L(P, Y) = -[Y \times \log(P^T) + (E - Y) \times \log(E - P)^T] \quad (6)$$

$$Y = [y_1, y_2, \dots, y_n]_{1 \times n} \quad (7)$$

$$P = [p_1, p_2, \dots, p_n]_{1 \times n} \quad (8)$$

$$E = [1, 1, \dots, 1]_{1 \times n} \quad (9)$$

式中,  $P$  表示模型输出的预测向量,  $Y$  表示样本的标签向量,  $n$  为违禁品类数量, 若样本图像中含有第  $i$  个类别的违禁品, 则  $y_i = 1$ , 反之  $y_i = 0$ 。

通过损失函数引导特征权重参数  $w_1, w_2$  和  $w_3$  动态调整的解决方案既不会固定权重参数, 失去特征自适应融合的意义, 也不会使参数在自适应调整过程中失去控制, 而是在给定学习方向的情况下寻找参数的最优值。

## 3 实验及分析

为了对上述方案进行验证, 先通过对比实验与其他方法进行比较, 测试该方案的整体性能及泛化能力, 即对不同颜色分布安检图像的识别性能, 然后通过消融实验检测各部分对分类性能的影响。

### 3.1 数据集构建及参数设置

实验采用 SIXray 公开数据集。该数据集是目前公开的规模最大的 X 光安检图像数据集, 已广泛用于违禁品检测研究 (Hassan 等, 2021; Zhang 等, 2021; Akcay 和 Breckon, 2022), 共 1 059 231 幅 X 光

安检图像,采自不同地铁站的安检设备。实验时,将含有违禁品的图像定义为正样本,反之为负样本。正样本共8 929幅,包含枪支、刀具、扳手、钳子和剪刀等5种违禁品,且单幅图像中可包含1种或多种违禁品,负样本共1 050 302幅。SIXray数据集包含3个不平衡程度不同的子数据集SIXray10、SIXray100和SIXray1000。从SIXray10子数据集中进行采样得到子数据集SIXray\_sub。SIXray10子数

据集由训练集和测试集构成,SIXray\_sub中的训练样本与测试样本分别来自SIXray10中的训练集与测试集,且正负样本比约为1:3。SIXray\_sub子数据集的信息如表1所示。在观察数据集的过程中,发现正样本中编号从P08828到P08929的101幅图像明显与其他图像颜色分布不同,将其设为子数据集SIXray\_last101,用于验证模型的泛化性能测试。SIXray\_sub中的测试集包含SIXray\_last101子数据集。

表1 SIXray\_sub子数据集设置  
Table 1 SIXray\_sub sub dataset setting

| 数据集 | 正样本   |       |       |       |     |       | 负样本    | 总样本    |
|-----|-------|-------|-------|-------|-----|-------|--------|--------|
|     | 枪     | 刀     | 扳手    | 钳子    | 剪刀  | 样本数   |        |        |
| 训练集 | 2 705 | 1 748 | 2 012 | 3 434 | 807 | 7 496 | 22 504 | 30 000 |
| 测试集 | 425   | 205   | 192   | 529   | 188 | 1 433 | 4 567  | 6 000  |
| 总样本 | 3 130 | 1 953 | 2 204 | 3 963 | 995 | 8 929 | 27 071 | 36 000 |

实验采用的软硬件环境为深度学习框架Pytorch, ubuntu16.04系统, GPU( graphics processing unit)显卡型号为Tesla M60。参数设置如表2所示。

表2 参数设置  
Table 2 Parameter setting

| 设置                      | 参数值                                       |
|-------------------------|-------------------------------------------|
| 归一化尺寸 $W \times H$ (训练) | $256 \times 256$                          |
| 归一化尺寸 $W \times H$ (测试) | $224 \times 224$                          |
| 归一化均值                   | 0.763, 0.785, 0.707                       |
| 归一化方差                   | 0.267, 0.196, 0.258                       |
| 批量大小                    | 64                                        |
| 优化动量                    | 0.9                                       |
| 权重衰减                    | 0.000 1                                   |
| 初始学习率                   | 0.01                                      |
| 三元损失系数                  | $\alpha = 0.3, \beta = 0.2, \gamma = 0.5$ |

预处理操作中归一化采用的均值和方差参数由该数据集计算得到,表2中均值和方差的3个参数值分别对应RGB的3个通道。训练过程中采用随机梯度下降(stochastic gradient descent,SGD)优化方法,学习率随训练情况衰减,当损失不下降时停止训练。针对第2.2.2节提到的过多依赖彩色特征、需减弱彩色分支的权重,以及需加强区域增强特征以提高模型泛化能力,将三元损失系数设置为 $\alpha =$

$0.3, \beta = 0.2, \gamma = 0.5$ 。

### 3.2 对比实验

对比实验分为整体性能对比和泛化能力对比两部分。整体性能对比实验主要验证本文方法在颜色分布相同的数据集上对各类违禁品识别性能的表现,该实验在SIXray10数据集上进行,将ResNet18和ResNet34网络作为实验的基础模型,对比方法为采用多尺度卷积与组卷积的金字塔卷积(pyramid convolution, PyConv)(Duta等,2020)融合高中低3层特征并采用高层监督低层的分层细化方法(hierarchical refinement,HR)(Miao等,2019)、将通道注意力与空间注意力串联的卷积注意力方法(CBAM)(Woo等,2018)以及本文提出的区域增强和多特征融合模型(REMF)。由于SIXray10数据集为类不平衡数据集,为了缓解数据的不平衡程度,该实验在所有方法上均增加类平衡处理(class-balanced,C)(Miao等,2019)。采用平均精度(mean average precision,mAP)作为评估分类性能的指标(尉婉青等,2021),将每秒帧率(frame per second, FPS)即每秒可处理图像数量作为评估识别效率的指标。各方法在ResNet18和ResNet34模型上的表现如表3和表4所示。可以看到,在整体的识别效果上,本文方法相较于基础模型分别提升了4.09%和2.26%,相较于其他3种方法也有所提升。在单类的识别效果上,本文方法在5类违禁品的识别效



果上均高于基础模型,尤其在枪支类的识别效果上有明显优势,相对于其他方法最少提升了 2.24% 和

3.09%,并且在钳子类的识别效果上优于其他方法。在识别效率方面,本文方法可达到实时处理的效果。

表 3 不同方法在 ResNet18 模型上的整体性能对比  
Table 3 Overall performance comparison of different methods on ResNet18 model

| 模型                         | 枪支/%         | 刀具/%         | 扳手/%  | 钳子/%         | 剪刀/%         | mAP/%        | FPS/(帧/s) |
|----------------------------|--------------|--------------|-------|--------------|--------------|--------------|-----------|
| C + Baseline               | 88.32        | 86.47        | 72.49 | 84.54        | 61.29        | 78.63        | <b>44</b> |
| C + PyConv( Duta 等,2020)   | 90.14        | 89.08        | 68.86 | 85.22        | 60.20        | 78.70        | 28        |
| C + HR( CHR)( Miao 等,2019) | 88.10        | <b>89.74</b> | 72.99 | 85.63        | 64.49        | 80.19        | 36        |
| C + CBAM( Miao 等,2019)     | 88.92        | 87.96        | 77.83 | 88.32        | <b>68.01</b> | 82.21        | 36        |
| C + REMF( 本文)              | <b>92.38</b> | 87.85        | 76.25 | <b>90.54</b> | 66.56        | <b>82.72</b> | 34        |

注:加粗字体表示各列最优结果。

表 4 不同方法在 ResNet34 模型上的整体性能对比  
Table 4 Overall performance comparison of different methods on ResNet34 model

| 模型                         | 枪支/%         | 刀具/%         | 扳手/%         | 钳子/%         | 剪刀/%         | mAP/%        | FPS/(帧/s) |
|----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------|
| C + Baseline               | 91.16        | 86.89        | 75.42        | 87.62        | 66.06        | 81.43        | <b>34</b> |
| C + PyConv( Duta 等,2020)   | 89.77        | 88.39        | 74.09        | 86.78        | 62.16        | 80.24        | 17        |
| C + HR( CHR)( Miao 等,2019) | 91.05        | <b>89.75</b> | 77.73        | 88.11        | <b>67.41</b> | 82.81        | 26        |
| C + CBAM( Miao 等,2019)     | 90.53        | 89.39        | <b>79.64</b> | 90.87        | 64.34        | 82.96        | 27        |
| C + REMF( 本文)              | <b>94.25</b> | 88.76        | 78.41        | <b>90.89</b> | 66.15        | <b>83.69</b> | 25        |

注:加粗字体表示各列最优结果。

泛化能力对比实验主要验证本文方法在不同颜色分布图像上的泛化能力。将已经训练好的表 3 和表 4 中模型在 SIXray\_last101 数据集上进行对比,由于 SIXray\_last101 数据中都含有枪支类违禁品,因此实验可以通过比较正确识别的枪支样本数量( correct number/total number, CN/TN) 进行泛化能力对比,实验结果如图 9 所示。

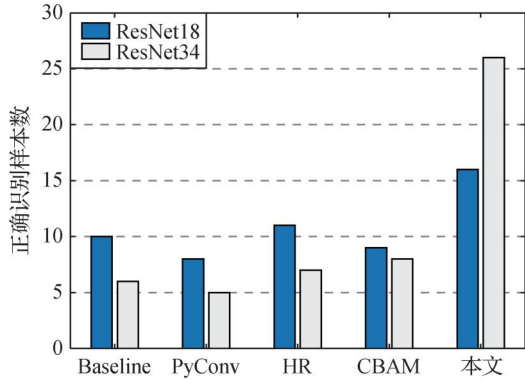


图 9 泛化能力对比实验结果

Fig. 9 Comparative experimental results of generalization ability

从图 9 可以看出,由于在训练模型过程中,训练样本的颜色符合同一分布,当模型面对与训练样本颜色分布不一致的 SIXray\_last101 测试集时,很难表现出好的效果。ResNet18 模型能正确识别 10 幅含有枪支的安检图像,而 ResNet34 的性能低于 ResNet18,仅能正确识别 6 幅,在 PyConv、HR 和 CBAM 方法上也看同样的规律。本文提出的 REMF 方法在泛化能力方面相较于其他方法有所提升,并且在 ResNet34 上效果更加明显,以 ResNet34 为基础的 REMF 可以正确识别 26 幅含枪支违禁品的图像。图 10 为 REMF 方法相较于其他方法额外识别出来的部分图像,尽管对于全部的测试样本,本文方法只识别出了一小部分,但相对于基础模型提升了 4.3 倍,表明本文方法提升了在不同颜色分布图像上的泛化能力,与理论分析保持一致。

3.3 消融实验

为了更加直观地检验不同特征以及超参数设置对模型性能的影响,实验选择 ResNet34 网络为基础模型、SIXray\_sub 为实验数据进行消融实验。实验结果如表 5 所示。



图 10 仅 REMF 方法正确识别的图像

Fig. 10 Only images correctly identified by REMF method

表 5 消融实验结果

Table 5 Ablation results

| 特征              | 超参数设置    |         |          | 性能           |               |
|-----------------|----------|---------|----------|--------------|---------------|
|                 | $\alpha$ | $\beta$ | $\gamma$ | mAP/%        | CN/TN         |
| 彩色              | -        | 1       | -        | 85.21        | 11/101        |
| 边缘 + 彩色         | 1        | 1       | -        | 86.53        | 22/101        |
| 边缘 + 彩色 + 区域增强  | 1        | 1       | 1        | <b>87.58</b> | 17/101        |
| 边缘 + 彩色 + 区域增强* | 0.3      | 0.2     | 0.5      | 86.02        | <b>30/101</b> |

注:加粗字体表示各列最优结果,“-”表示无需设置该参数。

第 1 组实验为基础模型,仅采用彩色特征分支进行预测;第 2、3 组实验增加边缘特征和区域增强特征,并采用自适应方式进行融合,即超参数默认为 1。第 4 组实验引入三元损失函数进行引导式融合,即对超参数进行设置,减弱彩色特征分支权重,增强区域增强特征分支权重。

从表 5 可以看出,边缘特征和区域增强特征使整体识别性能分别提升了 1.32% 和 1.05%,由于训练样本单一的颜色分布,多特征的自适应融合会提升模型在同样颜色分布图像上的性能,但面对不同颜色分布的图像,训练得到的自适应参数并不完全适用,因此模型在泛化性能的表现并不稳定。三元损失函数的引入提升了模型的泛化性能,基础模型仅能识别出 11 幅颜色分布不同的含枪支违禁品的图像,自适应融合可以正确识别 17 幅含枪样本,引导式融合可正确识别出 30 幅含枪样本。尽管整体识别性能相对于自适应融合方法有所欠缺,但相较于只提取彩色特征的基础模型仍然有所提高。

## 4 结 论

本文根据 X 光安检图像颜色丰富且分布不同以及物品混乱复杂的特点,提出一种区域增强和多特征融合模型。引用注意力机制的思想并依据数据的整体特征分布训练得到区域增强特征,可以消除颜色差异的影响并增强违禁品区域信息。根据安检员进行安全检查关注的色彩、形状和轮廓细节等信息,采用多特征融合策略,并提出一种三元损失函数优化融合效果。在公开的来自不同安检设备的实测数据集上进行实验,结果表明本文方法提升了模型识别违禁品的性能,并缓解了图像颜色分布不同引起的识别性能下降问题,验证了本文方法的有效性及其鲁棒性。本文方法相较于实验中的其他方法,在整体识别性能和泛化性能方面均有所提升。但本文方法的多分支结构增加了模型计算量,因此在识别效率方面没有明显优势。未来将进一步优化网络结构,在保证识别效果的情况下减小模型的计算量,提高模型的识别效率。

## 参考文献 (References)

- Akçay S and Breckon T. 2022. Towards automatic threat detection: a survey of advances of deep learning within X-Ray security imaging. *Pattern Recognition*, 122: #108245 [DOI: 10.1016/J. PATCOG. 2021.108245]
- Akçay S, Kundegorski M E, Devereux M and Breckon T P. 2016. Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery//*Proceedings of 2016 IEEE International Conference on Image Processing (ICIP)*. Phoenix, USA: IEEE: 1057-1061 [DOI: 10.1109/ICIP. 2016. 7532519]
- Baştan M, Yousefi M R and Breuel T M. 2011. Visual words on baggage X-ray images//*Proceedings of the 14th International Conference on Computer Analysis of Images and Patterns*. Seville, Spain: Springer: 360-368 [DOI: 10.1007/978-3-642-23672-3\_44]
- Caldwell M, Ransley M, Rogers T W and Griffin L D. 2017. Transferring x-ray based automated threat detection between scanners with different energies and resolution//*Proceedings Volume 10441, Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies*. Warsaw, Poland: SPIE: #104410F [DOI: 10.1117/12.2277641]
- Duta I C, Liu L, Zhu F and Shao L. 2020. Pyramidal convolution: rethinking convolutional neural networks for visual recognition [EB/

- OL]. [2021-08-28]. <https://arxiv.org/pdf/2006.11538v1.pdf>
- Hassan T, Akçay S, Bennamoun M, Khan S and Werghi N. 2021. Unsupervised anomaly instance segmentation for baggage threat recognition. *Journal of Ambient Intelligence and Humanized Computing*, 1-12 [DOI: 10.1007/S12652-021-03383-7]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, USA; IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- He Y H, Wu G X and Wu Z Q. 2021. Domain alignment based object detection of X-ray images. *Computer Science*, 48(1): 175-181 (何彦辉, 吴桂兴, 吴志强. 2021. 基于域适应的X光图像的目标检测. *计算机科学*, 48(1): 175-181) [DOI: 10.11896/jsjx.200200023]
- Hou Y Y. 2018. Research on the Relationship between Work Stress and Safety Performance of Airport Security Inspectors. Beijing: Beijing Jiaotong University (侯彦伊. 2018. 机场安检人员工作压力与安全绩效关系研究. 北京: 北京交通大学)
- Jaccard N, Rogers T W, Morton E J and Griffin L D. 2017. Detection of concealed cars in complex cargo X-ray imagery using deep learning. *Journal of X-Ray Science and Technology*, 25(3): 323-339 [DOI: 10.3233/XST-16199]
- Liu K W, Ma Y, Xiong H X, Yan Z J, Zhou Z J, Liu C Y, Fang P P, Li X J and Chen Y L. 2020. Medical-image super-resolution reconstruction method based on residual channel attention network. *Laser and Optoelectronics Progress*, 57(2): #021014 (刘可文, 马圆, 熊红霞, 严泽军, 周志军, 刘朝阳, 房攀攀, 李小军, 陈亚雷. 2020. 基于残差通道注意力网络的医学图像超分辨率重建方法. *激光与光电子学进展*, 57(2): #021014) [DOI: 10.3788/LOP57.021014]
- Liu Y, Zhang C, Cheng J, Chen X and Wang Z J. 2019. A multi-scale data fusion framework for bone age assessment with convolutional neural networks. *Computers in Biology and Medicine*, 108: 161-173 [DOI: 10.1016/j.combiomed.2019.03.015]
- Mery D, Svec E, Arias M, Rizzo V, Saavedra J M and Banerjee S. 2017. Modern computer vision techniques for X-ray testing in baggage inspection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(4): 682-692 [DOI: 10.1109/TSMC.2016.2628381]
- Miao C J, Xie L X, Wan F, Su C, Liu H Y, Jiao J B and Ye Q X. 2019. SIXray: a large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, USA; IEEE: 2114-2123 [DOI: 10.1109/CVPR.2019.00222]
- Morris T, Chien T and Goodman E. 2018. Convolutional neural networks for automatic threat detection in security X-ray images//*Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. Orlando, USA; IEEE: 285-292 [DOI: 10.1109/ICMLA.2018.00049]
- Rubinstein R. 1999. The cross-entropy method for combinatorial and continuous optimization. *Methodology and Computing in Applied Probability*, 1(2): 127-190 [DOI: 10.1023/A:1010091220143]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I. 2017. Attention is all you need//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, USA; Curran Associates Inc.: 6000-6010
- Wang F, Jiang M Q, Qian C, Yang S, Li C, Zhang H G, Wang X G and Tang X O. 2017. Residual attention network for image classification//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, USA; IEEE: 6450-6458 [DOI: 10.1109/CVPR.2017.683]
- Wei W Q, Yu J, Bo M Y and Xiao C B. 2021. Video object detection using fusion of SSD and spatiotemporal features. *Journal of Image and Graphics*, 26(3): 542-555 (尉婉青, 禹晶, 柏曼曼, 肖创柏. 2021. SSD与时空特征融合的视频目标检测. *中国图象图形学报*, 26(3): 542-555) [DOI: 10.11834/jig.200020]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany; Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2\_1]
- Xiao T J, Xu Y C, Yang K Y, Zhang J X, Peng Y X and Zhang Z. 2015. The application of two-level attention models in deep convolutional neural network for fine-grained image classification//*Proceedings of 2015 Conference on Computer Vision and Pattern Recognition*. Boston, USA; IEEE: 842-850 [DOI: 10.1109/CVPR.2015.7298685]
- Zhang W F, Ni J J, Liu L B and Hu Q M. 2021. Symmetric triangle network for object detection within X-ray baggage security imagery//*Proceedings of 2021 International Joint Conference on Neural Networks (IJCNN)*. Shenzhen, China; IEEE: 1-7 [DOI: 10.1109/IJCNN52387.2021.9533991]
- Zhao Z H, Zhang H G and Yang J F. 2018. A GAN-based image generation method for X-ray security prohibited items//*Proceedings of the 1st Chinese Conference on Pattern Recognition and Computer Vision*. Guangzhou, China; Springer: 420-430 [DOI: 10.1007/978-3-030-03398-9\_36]

## 作者简介

韩萍,女,教授,主要研究方向为图像处理与模式识别。

E-mail: hanpingcauc@163.com

方澄,通信作者,男,讲师,主要研究方向为数据挖掘、大数据和计算机视觉。E-mail: cfang@cauc.edu.cn

杨慧,女,硕士研究生,主要研究方向为图像处理和深度学习。E-mail: 893045044@qq.com