



OPEN ACCESS

EDITED BY
Huafeng Li,
Kunming University of Science and
Technology, China

REVIEWED BY
Lulu Wang,
Kunming University of Science and
Technology, China
Neng Dong,
Nanjing University of Science and
Technology, China
Xiaosong Li,
Foshan University, China

*CORRESPONDENCE
Jiaxu Leng,
✉ lengjx@cqupt.edu.cn

SPECIALTY SECTION
This article was submitted to Radiation
Detectors and Imaging,
a section of the journal
Frontiers in Physics

RECEIVED 06 December 2022
ACCEPTED 19 December 2022
PUBLISHED 05 January 2023

CITATION
Chen J, Leng J, Gao X, Mo M and Guan S
(2023), Atomic number prior guided
network for prohibited items detection
from heavily cluttered X-ray imagery.
Front. Phys. 10:1117261.
doi: 10.3389/fphy.2022.1117261

COPYRIGHT
© 2023 Chen, Leng, Gao, Mo and Guan.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Atomic number prior guided network for prohibited items detection from heavily cluttered X-ray imagery

Jinwen Chen, Jiaxu Leng*, Xinbo Gao, Mengjingcheng Mo and Shibo Guan

School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China

Prohibited item detection in X-ray images is an effective measure to maintain public safety. Recent prohibited item detection methods based on deep learning has achieved impressive performance. Some methods improve prohibited item detection performance by introducing prior knowledge of prohibited items, such as the edge and size of an object. However, items within baggage are often placed randomly, resulting in cluttered X-ray images, which can seriously affect the correctness and effectiveness of prior knowledge. In particular, we find that different material items in X-ray images have clear distinctions according to their atomic number Z information, which is vital to suppress the interference of irrelevant background information by mining material cues. Inspired by this observation, in this paper, we combined the atomic number Z feature and proposed a novel atomic number Z Prior Guided Network (ZPGNet) to detect prohibited objects from heavily cluttered X-ray images. Specifically, we propose a Material Activation (MA) module that cross-scale flows the atomic number Z information through the network to mine material clues and reduce irrelevant information interference in detecting prohibited items. However, collecting atomic number images requires much labor, increasing costs. Therefore, we propose a method to automatically generate atomic number Z images by exploring the color information of X-ray images, which significantly reduces the manual acquisition cost. Extensive experiments demonstrate that our method can accurately and robustly detect prohibited items from heavily cluttered X-ray images. Furthermore, we extensively evaluate our method on HiXray and OPIXray, and the best result is 2.1% mAP_{50} higher than the state-of-the-art models on HiXray.

KEYWORDS

object detection, X-ray image, prohibited items detection, prior knowledge, public safety

1 Introduction

As society develops, the flow of people on public transport is increasing. X-ray security machine is widely used in the security inspection of railway stations and airports, which is a critical facility for maintaining public safety and transportation safety. However, traditional security checks mostly rely on manual identification methods. After prolonged work hours, security inspectors easily cause fatigue, significantly increasing the risk of missed and false detection and laying many hidden dangers for public safety. Therefore, it is increasingly necessary to identify prohibited items through intelligent algorithms.

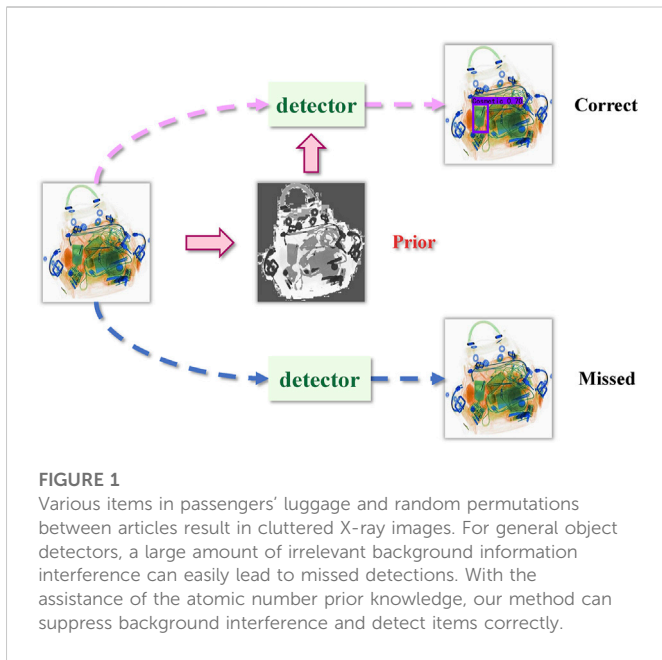


FIGURE 1
 Various items in passengers' luggage and random permutations between articles result in cluttered X-ray images. For general object detectors, a large amount of irrelevant background information interference can easily lead to missed detections. With the assistance of the atomic number prior knowledge, our method can suppress background interference and detect items correctly.

Different from traditional detection tasks, in this scenario, there are various items in the passenger's luggage and random permutations between items, resulting in heavily cluttered X-ray images [1–4]. Therefore, object detection algorithms for general natural images do not perform well on cluttered X-ray images as in Figure 1. Fortunately, the tremendous success of deep learning [5–11] has made the intelligent detection of prohibited items possible by

transforming it into an object detection task in computer vision [12–14]. Hence, many researchers have applied deep learning methods to prohibited object detection. Flitton et al. [15] explored 3D feature descriptors with application to threat detection in Computed Tomography (CT) airport baggage imagery. Bhowmik et al. [16] investigated the difference in detection performance achieved using real and synthetic X-ray training imagery for CNN architecture. Gaus et al; [17] evaluated several leading variants spanning the Faster R-CNN, Mask R-CNN, and RetinaNet architectures to explore the transferability of such models between varying X-ray scanners. Hassan et al; [18] presented a cascaded structure tensor framework that automatically extracts and recognizes suspicious items in multi-vendor X-ray scans. Zhao et al; [19] established the associations between feature channels and different labels and adjust the features according to the assigned labels (or pseudo labels) to tackle the overlapping object problem. These methods all improve detection performance to a certain extent but do not use the unique imaging characteristics of X-ray images to improve the algorithm.

Recently, some works have tried adding prior information about X-ray images to guide network learning, as shown in Figure 2 [20]. Obtained edge images by using the traditional edge detection algorithm Sobel. Chang et al. [4] found that different classes of prohibited objects have a clear distinction in physical size and used Otsu's threshold segmentation algorithm [21] to segment the original image into foreground and background, treating the foreground region as the approximate size of the detected object. Although these two methods improve the detection accuracy to a certain extent by introducing such prior information, the obtained prior information is easily disturbed by other irrelevant information due

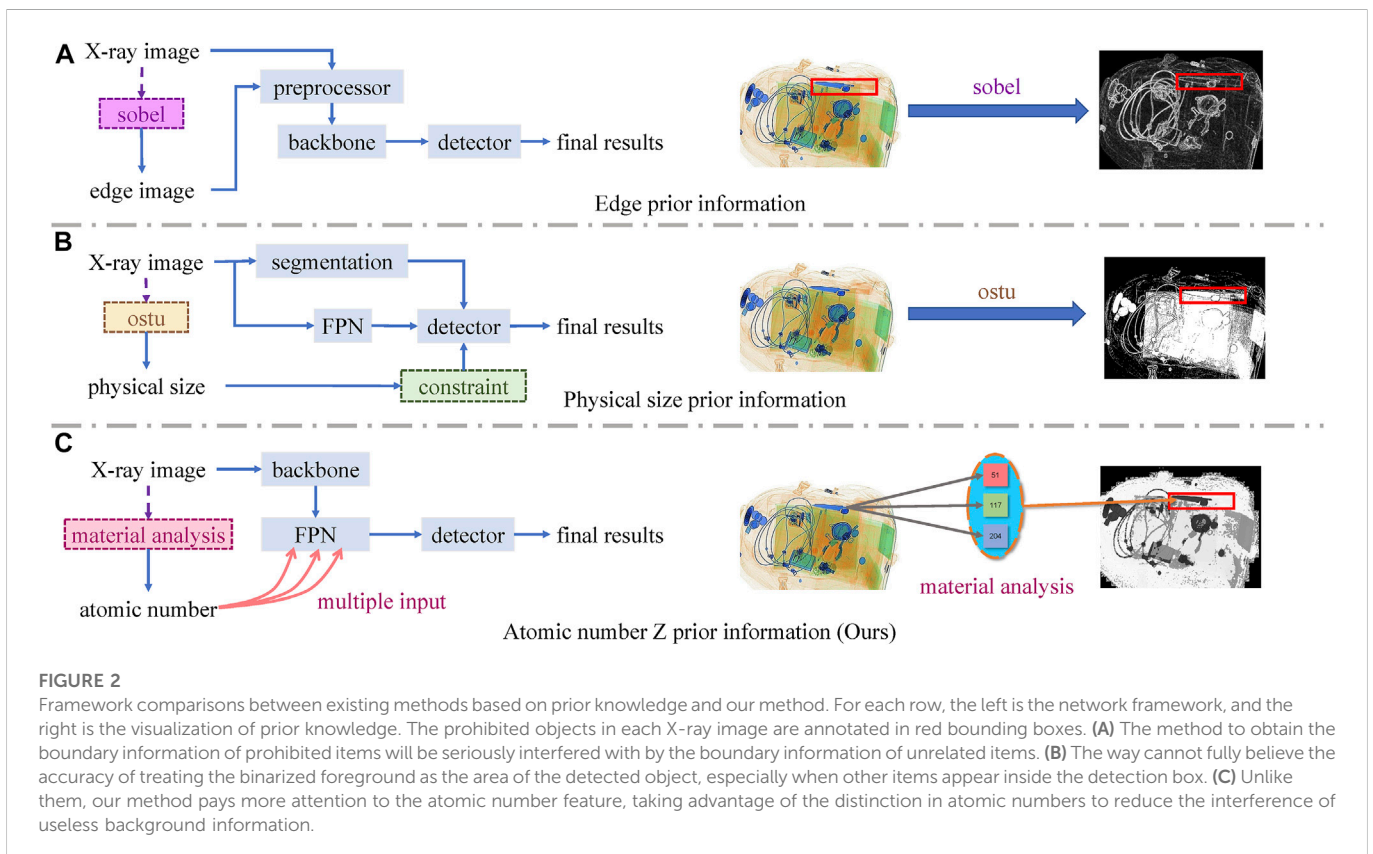


FIGURE 2
 Framework comparisons between existing methods based on prior knowledge and our method. For each row, the left is the network framework, and the right is the visualization of prior knowledge. The prohibited objects in each X-ray image are annotated in red bounding boxes. (A) The method to obtain the boundary information of prohibited items will be seriously interfered with by the boundary information of unrelated items. (B) The way cannot fully believe the accuracy of treating the binarized foreground as the area of the detected object, especially when other items appear inside the detection box. (C) Unlike them, our method pays more attention to the atomic number feature, taking advantage of the distinction in atomic numbers to reduce the interference of useless background information.

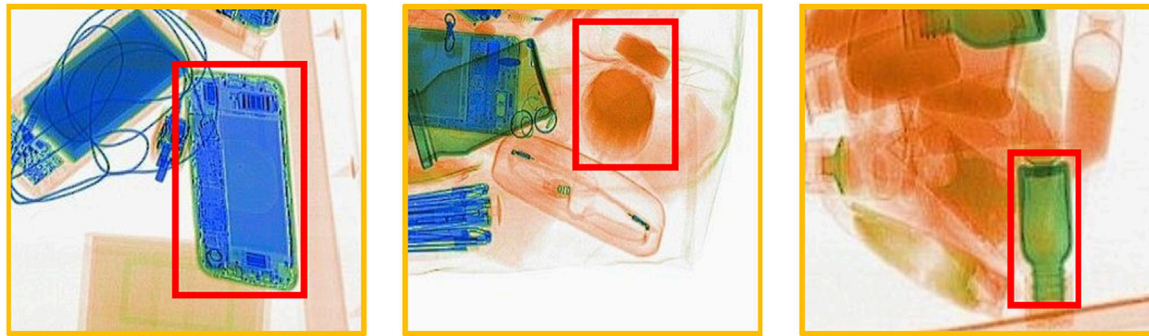


FIGURE 3
From left to right are inorganic matter, organic matter, and mixture.

to the messy distribution of prohibited items, which hinders further performance improvement. Specifically, in the presence of cluttered items, the former method to obtain the boundary information of prohibited terms is severely interfered with by the boundary information of irrelevant items. Furthermore, the latter cannot fully believe the accuracy of treating the binarized foreground as the area of the detected items, especially when other items appear inside the detection region.

In this paper, we propose a novel atomic number Z Prior Guided Network (ZPGNet) for heavily cluttered X-ray images, which can remove irrelevant background information by effectively incorporating the atomic number feature. Unlike optical images, X-ray images are generated by illuminating objects with X-ray. X-ray security inspection machine is based on the object difference in absorbing X-ray to detect the effective atomic number and then show distinct colors [22]. Specifically, the color information in X-ray images represents material information, where blue represents inorganic material, orange represents organic material, and green represents mixture [23], as shown in Figure 3. Atomic number images of X-ray image variants can directly reflect the material type of an item, which is the dominant information in X-ray images. This characteristic motivates us to explore this critical information to improve detection accuracy by removing irrelevant background information. Bhowmik et al; [24] examined the impact of atomic number images *via* the use of CNN architectures for the object detection task posed within X-ray baggage security screening and obviously illustrated a vital insight into the benefits of using atomic number images for object detection and segmentation tasks. However, they only simply connect atomic number images with RGB images and do not fully use atomic number images. In order to make full use of the atomic number features of items, we designed a Material Activation (MA) module. It cross-scale flows atomic number information through the network to mine deep material clues, which is beneficial to reduce irrelevant information interference in detecting prohibited items.

Atomic number images need to be collected manually, which increases the costs. In particularly, X-ray imaging systems render different materials in different colors. Blue represents inorganic material, orange represents organic material, and green represents mixture, as shown in Figure 3. Therefore, we can obtain the material classification of each pixel by analyzing the color. Thus, we propose an atomic number Z Prior Generation (ZPG) module, which

automatically generates the atomic number feature according to the imaging color of X-ray images, as those shown in Figure 4.

Overall, the contributions of our work can be summarized as follows:

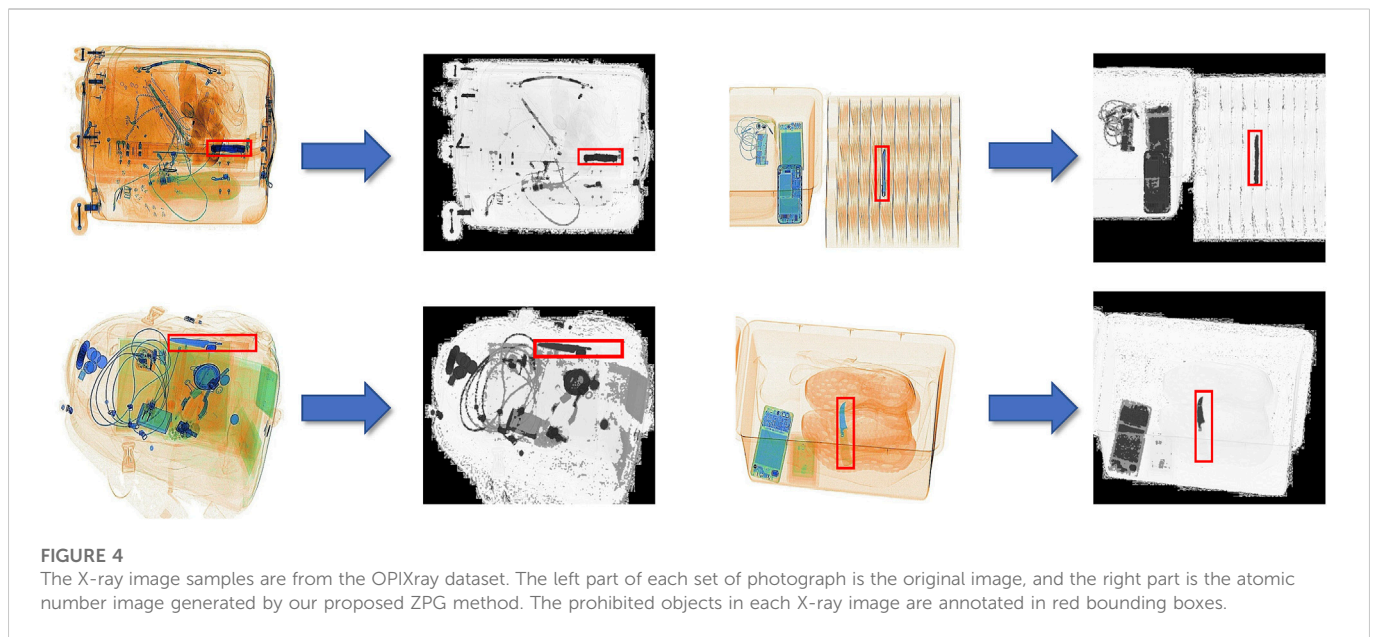
- We propose a novel atomic number Z Prior Guided Network (ZPGNet) to improve the detection accuracy of cluttered items by effectively incorporating the atomic number feature. In addition, the proposed method is generic and can be easily embedded into existing detection frameworks as a module.
- We propose an atomic number Z Prior Generation (ZPG) module, which automatically generates the atomic number feature according to the imaging color of X-ray images. Compared with the manual collection, the costs are significantly reduced.
- We design a Material Activation (MA) module to cross-scale fuse image features with the atomic number feature and then flow the fused features from high-level to low-level to enhance the ability of the model to mine deep material clues.
- We evaluate ZPGNet on the HiXray and OPIXray datasets and demonstrate that the performance of our ZPGNet is superior to state-of-the-art methods in identifying prohibited objects from cluttered X-ray baggage images.

2 Related work

In this section, we first introduce the existing public datasets for detecting prohibited items in X-ray images and then describe some generic object detection methods and some strategies to solve the clutter problem in X-ray images.

2.1 Security inspection image dataset

X-ray security inspection machines show different colors for different material items by the object distinction in absorption X-ray [22]. Therefore, it has many applications in many tasks, such as security inspection [4, 25–27] and medical imaging analysis [8, 28–33]. However, there are very few X-ray image datasets due to the particularity of security inspection scenes. To our knowledge, four recently published datasets are GDXray [22], SIXray [26], OPIXray



[20], and HiXray [34]. The GDxray dataset has 19,407 images containing three prohibited items, namely, guns, darts, and razors. However, the GDxray dataset only contains grayscale images, which are far from realistic scenarios. The SIXray includes 1,059,231 X-ray images, which only have 8,929 labeled images. The pictures in the SIXray dataset are obtained by real security machines from several subway stations, which is more in line with the data distribution of real scenes. The OPIXray dataset is the first high-quality security target detection dataset, which contains five categories of prohibited items, namely, folding knives, straight knives, scissors, utility knives, and multitool knives, with a total of 8885 X-ray images. The HiXray dataset contains 44,364 X-ray images from daily security checks at international airports, which contain eight categories of prohibited items such as lithium batteries, liquids, and lighters that are common in daily life. Each image in the HiXray dataset is annotated by airport staff, which ensures the accuracy of the data.

2.2 Generic object detection

Object detection is an essential part of computer vision tasks, which supports many downstream tasks [35–38]. Methods based on convolutional neural networks can be summarized into two categories: single-stage [39–43] and multi-stage [44–46]. In recent years, compared with multi-stage detection methods, single-stage detection methods have been widely adopted due to their simple design and powerful performance. YOLOv3 [42] considers both real-time and accuracy by using the region proposal method. RetinaNet [41] improves the detection accuracy while maintaining the inference speed by solving the problem of class balance. It is far higher in real-time performance and accuracy than general multi-stage detection methods. FCOS [43] is anchor box free, as well as proposal free, to solve object detection in a per-pixel prediction fashion. In addition, YOLOv5 [47] makes several improvements based on YOLOv3, which significantly improves the detection speed and accuracy. However, so far, most object detection methods are for natural images. In the security check scene, various items in the passenger's luggage and

random permutations between the objects resulted in heavily cluttered X-ray images, so the detection effect is often unperformed.

2.3 Solutions to heavily cluttered problems

Previous works have mainly focused on solving the problem of highly cluttered X-ray images. Shao et al. [48] proposed a foreground and background separation X-ray prohibited item detection framework that separates prohibited items from other items to exclude irrelevant background information. Tao et al. [34] proposed a lateral inhibition module to eliminate the influence of noisy neighboring regions on the interest object regions and activate the boundary of items by intensifying it.

3 Proposed method

Atomic number images of X-ray image variants can directly reflect item material, which is the dominant information in X-ray images. Inspired by this, we propose a novel atomic number Z Prior Guided Network (ZPGNet) for cluttered X-ray images, as shown in Figure 5. The ZPGNet consists of three main components: 1) an atomic number Z Prior Generation (ZPG) module automatically generates atomic number images, which reduces the cost of manually collecting atomic number images, 2) a Material Activation (MA) module fuses the atomic number feature to remove irrelevant background information, 3) a Bidirectional Enhancement (BE) module enriches feature expression through bidirectional information flow.

Specifically, we first design the ZPG module, combining the characteristics that different materials will show different colors, to map a three-channel (RGB) color image to a single-channel atomic number image. Then, we repeatedly pass the atomic number feature generated by the ZPG module into the network to pay more attention to item material information. To effectively fuse the extracted image features and the atomic number feature, MA cross-scale flows the atomic number feature under the extracted multi-scale features and uses a channel

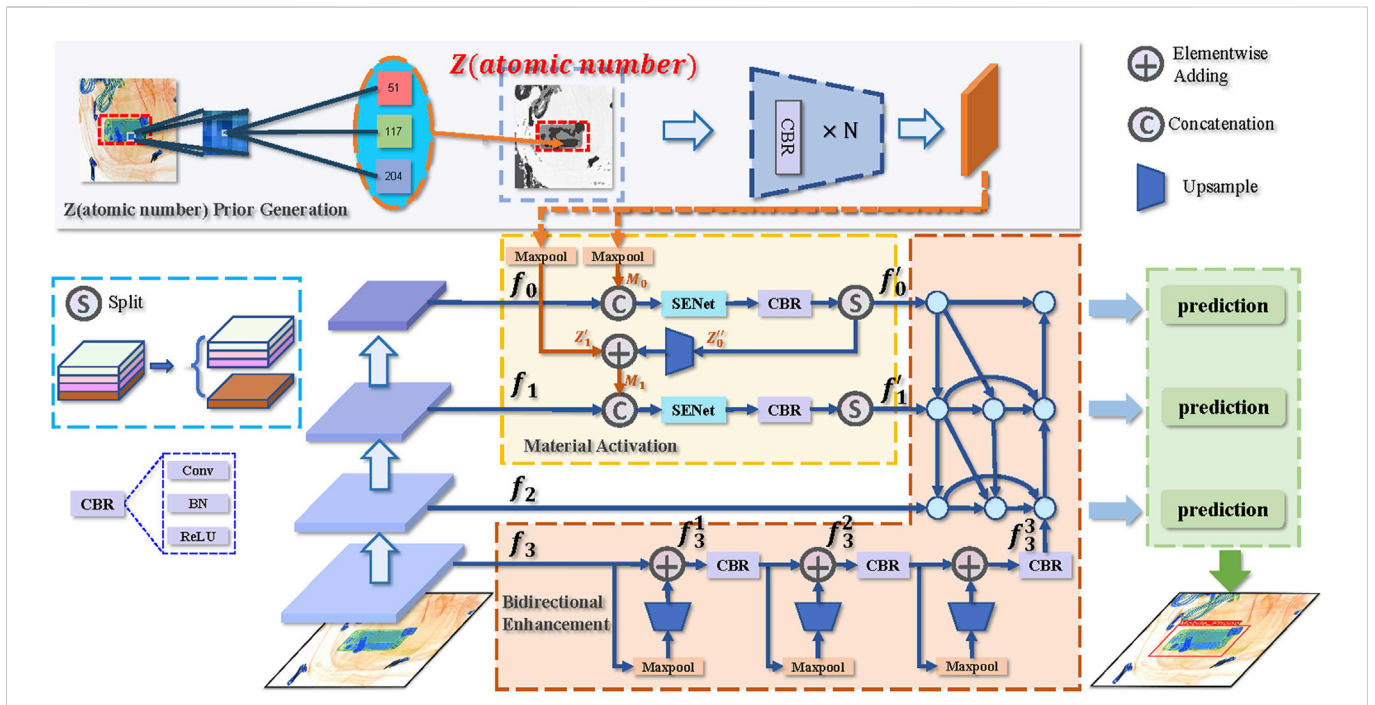


FIGURE 5 Overall framework of the proposed atomic number Z Prior Guided Network (ZPGNet). The network consists of three key modules, i.e., an atomic number Z Prior Generation (ZPG) module generating the atomic number feature, a Material Activation (MA) module cross-scale fusing the image features with the atomic number feature, and a Bidirectional Enhancement (BE) module mining contextual semantics for enhancing feature representation. CBR is composed of a convolution layer, a batch normalization layer, and a relu activation function. SENet stands for Squeeze-and-Excitation Networks [49].

attention module to self-adapt the importance of different features. Finally, we add a layer of low sampling rate features to obtain more detailed information and mine contextual semantics for enriching feature expression.

3.1 Z Prior Generation

Unlike optical images, X-ray images are generated by illuminating objects with X-rays, whose penetration is related to the material's density, size, and composition [22]. X-ray security machines detect the atomic number of objects based on the difference in absorbing X-rays, which then display a distinct color. Bhowmik et al. [24] proved that the introduction of atomic number images is an effective method to improve detection performance *via* large experiments. Inspired by this, the designed ZPG module compresses three-channel X-ray images into a single-channel to generate atomic number images that can highlight material differences. Compared with manually collecting atomic number images, it significantly reduced costs.

For each pixel in the RGB image, the maximum of the three channels will render its corresponding color. We use its subscripts to classify different materials.

$$g_{ij} = \text{argmax}(x_{ijk}) \tag{1}$$

where x_{ijk} denotes the value of the k -channel at position (i, j) the input image. $\text{argmax}(\bullet)$ denotes the index corresponding to finding the maximum value of an element.

Materials of the same class tend to present different depths of color due to different thicknesses. We introduce two variables, base-value B , and

width-value W . The former is used to distinguish different materials, and the latter reflects the difference between the same materials.

$$B_{ij} = g_{ij} + \alpha \tag{2}$$

$$W_{ij} = \left(\sum x_{ij} - x_{ijg_{ij}} \right) * (1 - \beta) * (1 - \alpha) / (255 + 255) + x_{ijg_{ij}} * \beta * (1 - \alpha) / 255 \tag{3}$$

Where α and β are hyperparameters that respectively control basis-value B and width-value W .

Finally, the basis-value B and width-value W are added and normalized, and then passed through a series of convolutional layers to obtain the atomic number feature Z .

$$Z_{ij} = \begin{cases} 0, & \text{if } x_{ij} = (255, 255, 255) \\ (B_{ij} + W_{ij}) / 3, & \text{if others} \end{cases} \tag{4}$$

$$Z = \phi_n(Z) \tag{5}$$

where $\phi_n(\bullet)$ denotes the n -layer ‘‘Conv-BN-ReLU’’ operation. Since no items are in the white area, we specially treat for the pixel $(255, 255, 255)$.

3.2 Material activation

In particular, different material items in X-ray images have clear distinctions according to their atomic number information, which is vital to suppress the interference of background information by mining deep material cues.

In cluttered X-ray images, the boundary and color information of prohibited items are easily interfered with by background

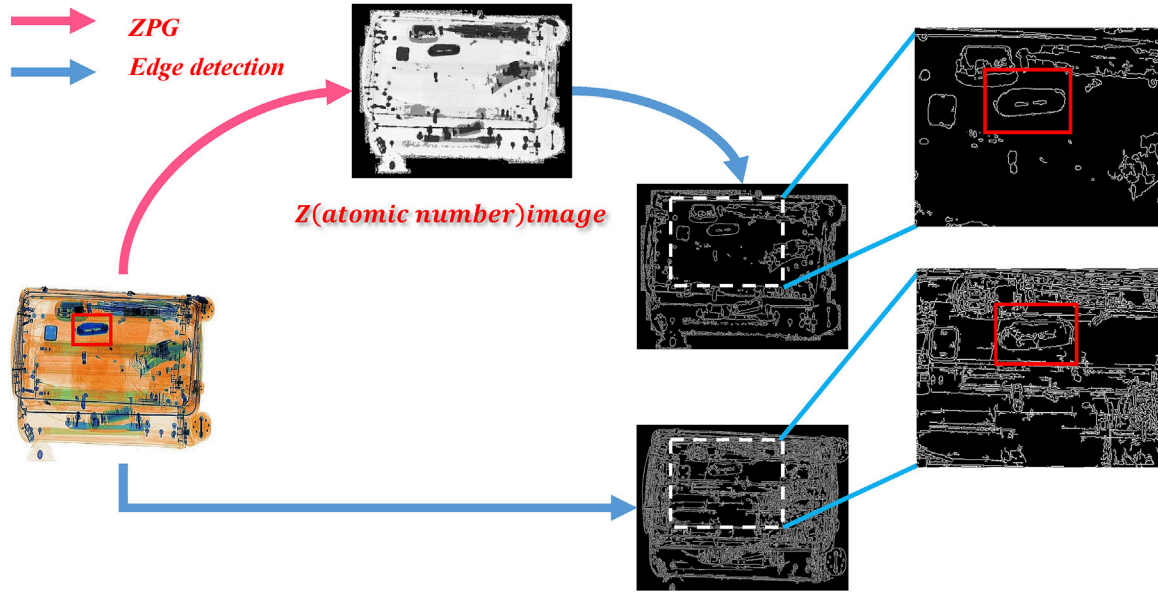


FIGURE 6
The bottom part shows the edge detection results obtained directly by the Canny algorithm [50], and the top part is obtained by first passing through the ZPG module and then through the Canny detection. It is intuitive to see that the edges of the items processed by the ZPG module are more evident than the original. The prohibited objects in each X-ray image are annotated in red bounding boxes.

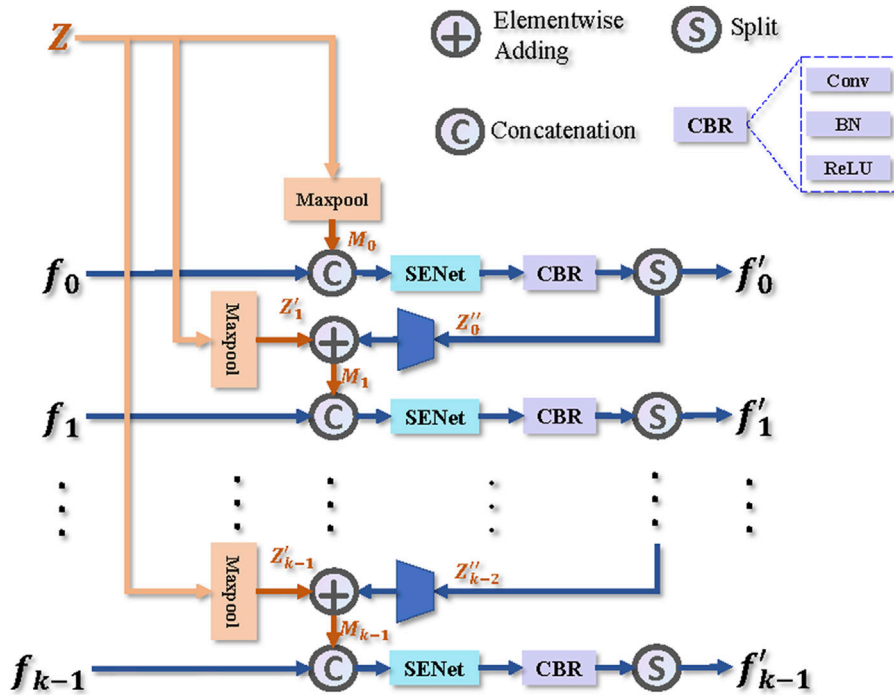


FIGURE 7
Illustration of the proposed Material Activation (MA) module, where k indicates that the input of the MA module has k different-scale feature maps.

information. MA introduces the atomic number feature to mine material cues, which is beneficial to reduce useless background information interference in detecting prohibited items, as shown in Figure 6.

Specifically, the backbone network has n feature map outputs $F = \{f_0, \dots, f_{n-1}\}$. As shown in Figure 7, the MA structure makes the former k layers of F as the input. For Z and F feature maps, which are output by ZPG and Backbone, we pool the atomic number feature Z to increase the receptive field

and then add Z flowing down from the previous layer to get a more robust feature M . Furthermore, we concatenate them with F for information fusion and apply channel attention operation (Squeeze-and-Excitation Networks [49]) $SE(\bullet)$ on the fused features to adapt the importance between the material feature and other image features (edge, texture, size, etc.).

$$Z'_i = \mathcal{D}_i(Z) \quad (6)$$

$$F_{ei} = \phi_1(SE(f_i \| M_i)) \quad (7)$$

where $\|$ represents the operation of concatenating, $\mathcal{D}_i(\bullet)$ denotes the Pooling operation.

Separate F_{ei} into f'_i and Z''_i along the channel dimension, whose dimensions are the same as f_i and Z'_i , respectively, where the f'_i is used as the input of the next BE module, and the Z''_i is passed to the next layer of the MA module as an enhanced atomic number feature to obtain the more robust feature.

$$\begin{cases} f'_i = F_{ei}^0 \\ Z''_i = F_{ei}^1 \end{cases} \quad (8)$$

$$M_i = Z'_i + \mathcal{U}(Z''_{i-1}) \quad (9)$$

where F_{ei}^0 and F_{ei}^1 denote the two features obtained by separating F_{ei} along the channel, $\mathcal{U}(\bullet)$ denotes the Upsample operation. Especially, $M_0 = Z'_0$.

3.3 Bidirectional Enhancement

When the down-sampling rate is high, it is easy to obtain larger receptive fields and more large-scale item information, which is beneficial for detecting large-scale prohibited objects. However, for some minor prohibited items, too large a downsampling rate tends to lose too much detail feature information of small-scale objects.

In the HiXray [34] high-quality prohibited items dataset, the average resolution of images is 1,200*900, with the largest resolution being 2000*1,024. The resolution of some small lighters is only 21*57, which is about 1/1,000 the size of the original image. After excessive downsampling, the feature information of lighters is seriously missing, resulting in poor detection in SSD [51], LIM [34], DOAM [20], and other detection models.

BE module adds a low sampling rate feature to obtain more detailed information about the tiny-size prohibited items. However, the low sampling rate feature often contains additional noise information. We remove noisy information by performing multiple pooling operations.

$$f_3^{i+1} = \phi_1(\mathcal{U}(\mathcal{D}_i(f_3^i)) + f_3^i) \quad (10)$$

where f_3^3 is the finally denoised low-sampling rate feature, and specific $f_3^0 = f_3$.

Finally, the material activation feature $\{f'_0, \dots, f'_{k-1}\}$ obtained by the MA module, Backbone output feature $\{f_k, \dots, f_2\}$, and f_3^3 are streamed bidirectionally, which mines contextual semantics to enrich feature expression.

4 Experiments

4.1 Datasets and evaluation Metrics

We conduct extensive experiments to evaluate our proposed model on two prohibited item detection datasets, HiXray [34] and

OPIXray [20]. HiXray dataset consists of 45,364 X-ray images from routine security checks at international airports, which contains 8 categories of 102,928 everyday prohibited items commonly seen in daily life, such as lithium batteries, liquids, lighters, etc. Each image in the HiXray dataset was annotated by an airport employee, which ensures the accuracy of the data. OPIXray dataset is the first high-quality object detection dataset for security, which focused on the widely-occurred prohibited item “cutter”, annotated manually by professional inspectors from the international airport. The dataset contains five categories of prohibited objects with a total of 8885 X-ray images (7,109 for training and 1,776 for testing).

Average Precision (AP) denotes the area under the precision-recall curve of the detection results for a single category of objects. To fairly evaluate the performance of all models, we compute the mean average precision (mAP) with an IOU threshold of .5. In addition, we calculate AP for all categories for each model to see the improvement for each category.

4.2 Implementation details

All our experiments were done in Pytorch and trained on one NVIDIA RTX 3090 GPU with the initial learning rate set to 1e-2. The parameters were optimized through stochastic gradient descent (SGD). The momentum and weight decay are set to .937 and .0005, respectively. Besides, two new hyperparameters were introduced with respect to the module ZPG, i.e., α and β , which respectively control base-value B and width-value W, and values are set to .4 and .5.

4.3 Quantitative results

We test the model performance on HiXray [34] and OPIXray [20] datasets. Specifically, we embedded ZPGNet into YOLOv3 [42] and YOLOv5s [47] and compared it with the state-of-the-art methods DOAM [20] and LIM [34]. Table 1 presents the experimental results of DOAM, LIM, and the proposed ZPGNet on HiXray and OPIXray datasets. In order to illustrate the effectiveness of our method and better compare it with the existing state-of-the-art (SOTA) models, we use YOLOv3 and YOLOv5s as this baseline.

4.3.1 Results on HiXray dataset

The experimental results of different algorithms on the HiXray [34] dataset are shown in Table 1. For a fair comparison, we adopt the same baseline YOLOv5s [47] as DOAM [20] and LIM [34], which performs the best results on both DOAM and LIM. The proposed method ZPGNet with YOLOv5s baseline improves to 83.9% in mean average prediction, outperforming DOAM and LIM by 1.7% mAP_{50} and .7% mAP_{30} , respectively. In order to further verify the effectiveness of our model, we also adopted the YOLOv3 [42] baseline, which is still 1.2% mAP_{50} higher than the SOTA method (YOLOv5s + LIM).

The (YOLOv3+ZPGNet) experiment results show that our method is lower than some methods in some categories Water, Laptop, Mobile Phone, and Tablet, but has an 8.0% AP and 4.8% AP improvement in the cosmetics and lighter categories, respectively, compared to the SOTA method LIM. Cosmetics belong to the mixtures category, commonly disturbed by organic substances (such as plastics), resulting in decreased detection confidence or even missed detection. The significant

TABLE 1 Quantitative evaluation results on the HiXray dataset and OPIXray dataset. Where PO1, PO2, WA, LA, MP, TA, CO, and NL denote “Portable Charger 1 (lithium-ion prismatic cell)”, “Portable Charger 2 (lithium-ion cylindrical cell)”, “Water,” “Laptop,” “Mobile Phone,” “Tablet,” “Cosmetic” and “Non-metallic Lighter” in the HiXray dataset. FO, ST, SC, UT, and MU donate “Folding Knife,” “Straight Knife,” “Scissor,” “Utility Knife,” and “Multi-tool Knife” in the OPIXray dataset, respectively.

Method	HiXray									OPIXray					
	mAP_{50}	PO1	PO2	WA	LA	MP	TA	CO	NL	mAP_{50}	FO	ST	SC	UT	MU
SSD [51]	71.4	87.3	81.0	83.0	97.6	93.5	92.2	36.1	.01	70.9	76.9	35.0	93.4	65.9	83.3
SSD + DOAM [20]	72.1	88.6	82.9	83.6	97.5	94.1	92.1	38.2	.01	74.0	81.4	41.5	95.1	68.2	83.8
SSD + LIM [34]	73.1	89.1	84.3	84.0	97.7	92.4	92.4	42.3	0.1	74.6	81.4	42.4	95.9	71.2	82.1
Xdet [4]	—	—	—	—	—	—	—	—	—	86.7	90.4	76	91.5	84.3	91.3
FCOS [43]	75.7	88.6	86.4	86.8	89.9	88.9	88.9	63.0	13.3	82.0	86.4	68.5	90.2	78.4	86.6
FCOS + DOAM [20]	76.2	88.6	87.5	87.8	89.9	89.7	88.8	63.5	12.7	82.4	86.5	68.6	90.2	78.8	87.7
FCOS + LIM [34]	77.3	88.9	88.2	88.3	90.0	89.8	89.2	69.8	14.4	83.1	86.6	71.9	90.3	79.9	86.8
ATSS [19]	—	—	—	—	—	—	—	—	—	86.6	92.3	72.0	96.6	80.38	91.7
ATSS + DOAM [19]	—	—	—	—	—	—	—	—	—	85.6	90.7	66.8	96.2	81.8	92.5
ATSS + Lacs [19]	—	—	—	—	—	—	—	—	—	88.3	90.0	75.0	97.6	85.7	93.0
YOLOv5s [47]	81.7	95.5	94.5	92.8	97.9	98.0	94.9	63.7	16.3	87.8	93.4	67.9	98.1	85.4	94.1
YOLOv5s + DOAM [20]	82.2	95.9	94.7	93.7	98.1	98.1	95.8	65.0	16.1	88.0	93.3	69.3	97.9	84.4	95.0
YOLOv5s + LIM [34]	83.2	96.1	95.1	93.9	98.2	98.3	96.4	65.8	21.3	90.6	94.8	77.6	98.2	88.9	93.8
YOLOv5s + ZPGNet (Ours)	83.9	95.7	95.2	92.5	96.5	97.7	94.4	66.4	33.0	90.7	95.0	79.3	98.0	86.8	94.2
YOLOv3 [42]	83.0	96.7	94.9	91.9	97.9	97.7	94.0	71.9	18.6	78.2	92.5	36.0	97.3	70.8	94.4
YOLOv3+ZPGNet (Ours)	84.4	96.6	95.2	92.7	97.7	98.0	95.2	73.8	26.1	85.4	88.5	65.1	96.7	83.5	93.3

Bold values represent the best performance in the same evaluation index.

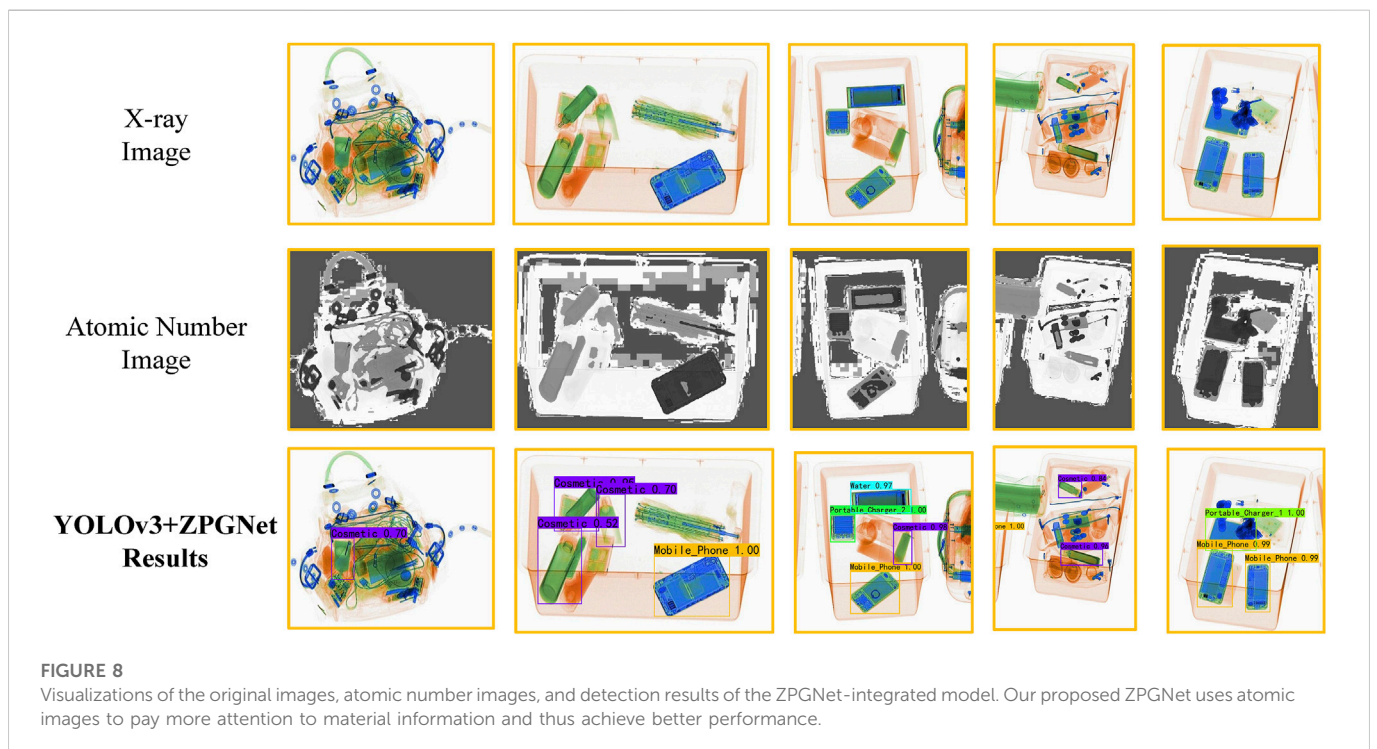


FIGURE 8 Visualizations of the original images, atomic number images, and detection results of the ZPGNet-integrated model. Our proposed ZPGNet uses atomic images to pay more attention to material information and thus achieve better performance.

improvement in cosmetics indicates that our method, introducing the atomic number feature map, can better reduce the interference of useless information in Figure 8. This advantage is facilitated by

our method of paying extra attention to the material information using atomic number features. Lighters in luggage are tiny in size and prone to profound feature loss after downsampling. Our

TABLE 2 Comparisons between the ZPGNet-integrated network and three object detection methods.

Method	mAP_{50}	FO	ST	SC	UT	MU
RetinaNet [41]	87.4	89.4	69.2	98.2	86.3	94.0
RetinaNet + ZPGNet	88.1	91.3	72.1	98.7	85.8	92.6
YOLOv5s [47]	87.8	93.4	67.9	98.1	85.4	94.1
YOLOv5s + ZPGNet	90.7	95.0	79.3	98.0	86.8	94.2
YOLOv3 [42]	78.2	92.5	36.0	97.3	70.8	94.4
YOLOv3+ZPGNet	85.4	88.5	65.1	96.7	83.5	93.3

We embedded our method into three different baseline models respectively and divided the models embedded with and without our method into a group, where the bold figures represent the best performance in a group.

method achieves 11.7% AP improvement over LIM [34] with the same baseline YOLOv5s in the lighter category, which is due to the fact that we use a low sampling rate feature map in the BE module to increase the information of small prohibited items.

4.3.2 Results on OPIXray dataset

Table 1 represents the performance of our method on the OPIXray [20] dataset. With the same baseline YOLOv5s [47], ZPGNet outperforms DOAM [20] and LIM [34] by 2.7% mAP_{50} and .1% mAP_{50} , respectively. In particular, ZPGNet has the highest score on mAP_{50} among all the models. It can be clearly seen that the proposed method ZPGNet achieves significant performance improvement based on YOLOv3 [42], especially on AP of the

severely occluded prohibited items named “straight knife” improved by 29.1%. This benefits from the fact that our method effectively removes the interference of irrelevant background information.

4.4 Generality verification

To further evaluate the effectiveness of the proposed model ZPGNet and verify that ZPGNet can be applied to various detection networks, we choose the classical detection models YOLOv3 [42], RetinaNet [41], and YOLOv5s [47] to use our method. Experiments were performed on the OPIXray dataset [20]. As shown in Table 2, our approach ZPGNet improves YOLOv3 by 7.2% mAP_{50} , RetinaNet by .7% mAP_{50} , and YOLOv5s by 2.9% mAP_{50} , respectively. Many objects are commonly disturbed by useless items, quickly resulting in low confidence or even miss detection on the general detection model. As shown in Figure 9, the comparison plot of the experimental results in the first and second rows shows that even with high confidence, there is a particular improvement after introducing the atomic number features. Embedding ZPGNet makes the network pay more attention to object material information to reduce the interference of ineffective information and alleviate the problems of low confidence and missed detection. This indicates that our model can be embedded into most detection networks as a plug-and-play component to minimize the interference of useless background information and achieve better performance.

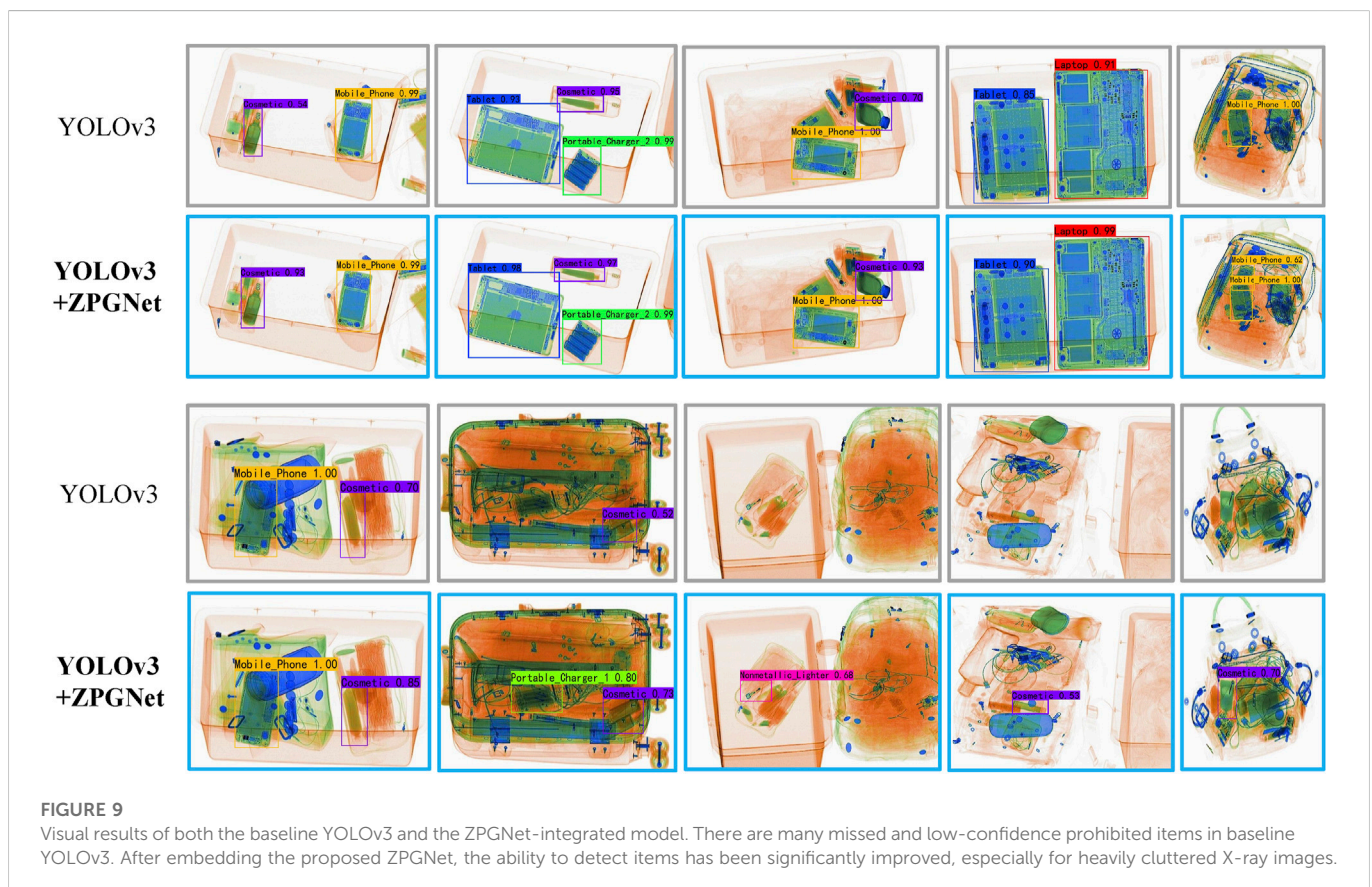


TABLE 3 Ablation results of the proposed ZPG, MA, and BE on the HiXray dataset.

Method	mAP_{50}	PO1	PO2	WA	LA	MP	TA	CO	NL
YOLOV5s [47]	81.7	95.5	94.5	92.8	97.9	98.0	94.9	63.7	16.3
+ZPG + MA	83.1	95.3	95.5	92.4	94.9	97.7	93.6	69.0	26.0
+ZPG + MA + BE	83.9	95.7	95.2	92.5	96.5	97.7	94.4	66.4	33.0

Bold values represent the best performance in the same evaluation index.

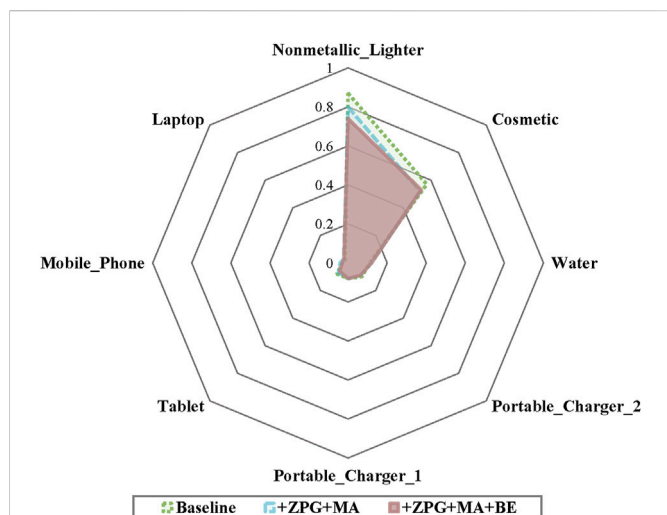


FIGURE 10 Performance comparison of different categories. The number on the gray line indicates the log-average miss rate. Useless background information interference can easily lead to prohibited item missed detections. With the proposed ZPG, MA, and BE, the log-average miss rate of prohibited items (i.e., cosmetic and lighter) is significantly reduced.

4.5 Ablation study

In this subsection, we conduct a series of ablation experiments to analyze the influence of involved hyperparameters and the contribution of critical components of the proposed ZPGNet. In the ablation study, all experiments were performed on the HiXray dataset [34].

4.5.1 Effectiveness of ZPG, MA, and BE

ZPG, MA, and BE are essential modules in ZPGNet, and we embed them one by one into YOLOv5s [47] to evaluate their performance. The insertion of ZPG requires the support of MA, so unity replaces ZPG and MA together into the model. All experiments here uniformly set the number of MA layers to 2. As shown in Table 3, the network embedded with ZPG and MA modules improves its performance by 1.4% mAP_{50} compared to the base model, especially in the cosmetics category, where it improves by 5.3% mAP_{50} . Cosmetics are commonly disturbed by organic substances (such as plastics), resulting in low confidence and missed detection. The significant improvement in cosmetics indicates that our method, introducing the atomic number features, can better reduce the interference of useless information, as shown in Figure 10. After applying the Bidirectional Enhancement (BE) module, the performance is

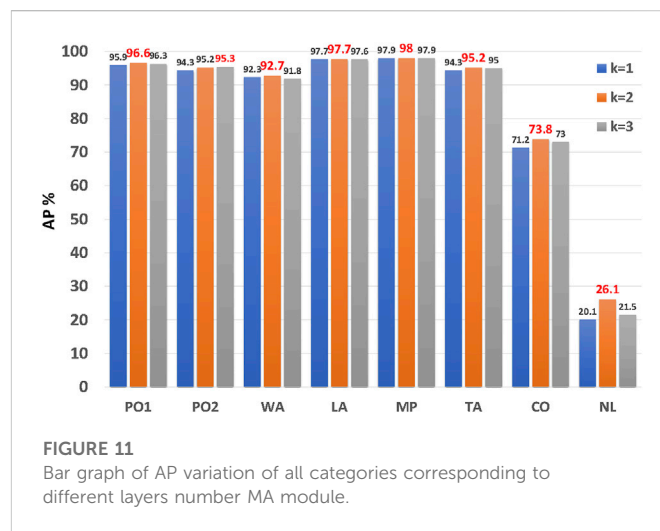


FIGURE 11 Bar graph of AP variation of all categories corresponding to different layers number MA module.

2.2% mAP_{50} higher than the basic module and .8% mAP_{50} higher than that embedded with MA and ZPG, which proves the effectiveness of the BE module.

4.5.2 Number of layers in MAs

We also show the effects of different layer numbers in the proposed MA, as shown in Figure 11. The model performs best when the layer numbers equal 2. The excessive number of layers can lead to performance degradation of the MA module. We believe that the possible reason is that the over-introduction of the atomic number feature leads to the suppression of other essential cues, which leads to a degradation in performance. When MA layers are equal to 2, it can well balance the importance between the atomic number feature and other features. So, in other experiments, we set the layer numbers in each MA to 2.

5 Conclusion

Prohibited item detection in X-ray images is an effective measure to maintain public safety. The interference of a large amount of useless background information caused by object disordered placement is an urgent problem to be addressed in prohibited item detection. Inspired by the imaging characteristics of X-ray images, this paper proposes an atomic number Z Prior Generation (ZPG) method, which can automatically generate atomic number images and reduce the cost of manual acquisition. Furthermore, we designed an atomic number Z Prior Guided Network (ZPGNet) to solve useless background information interference in prohibited item detection. The

proposed ZPGNet method cross-scale flows the atomic number Z information through the network to mine deep material clues to reduce irrelevant background information interference. We comprehensively evaluate ZPGNet on HiXray and OPIXray datasets, and this result shows that ZPGNet can be embedded into most detection networks as a plug-and-play module and achieve higher performance. There is still a severe occlusion problem in X-ray images, but this paper does not solve the occlusion problem. In the future, we intend to use features such as contour and scale to solve the occlusion problem between items.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/OPIXray-author/OPIXray>.

Author contributions

Conceptualization, JC, JL, MM, XG, and SG; methodology, JC; software, MM, and SG; validation, JC; investigation, JL and SG; writing—original draft preparation, JC and JL; writing—review and editing, XG, MM, and SG; visualization, JC; funding acquisition, JL and XG. All authors have read and agreed to the published version of the manuscript.

References

- Gaus YFA, Bhowmik N, Akçay S, Guillén-García PM, Barker JW, Breckon TP. Evaluation of a dual convolutional neural network architecture for object-wise anomaly detection in cluttered x-ray security imagery. In: 2019 international joint conference on neural networks (IJCNN); July 14-19, 2019; Budapest, Hungary (2019).
- Hassan T, Bettayeb M, Akçay S, Khan S, Bennamoun M, Werghi N. Detecting prohibited items in x-ray images: A contour proposal learning approach. In: 2020 IEEE International Conference on Image Processing (ICIP); October 25-28, 2020 (2020).
- Isaac-Medina BK, Willcocks CG, Breckon TP. Multi-view object detection using epipolar constraints within cluttered x-ray security imagery. In: 2020 25th International Conference on Pattern Recognition (ICPR); 10 - 15 January 2021; ITALY (2021). p. 9889.
- Chang A, Zhang Y, Zhang S, Zhong L, Zhang L. Detecting prohibited objects with physical size constraint from cluttered x-ray baggage images. *Knowledge-Based Syst* (2022) 237:107916. doi:10.1016/j.knsys.2021.107916
- Xia S, Wang F, Xie F, Huang L, Wang Q, Ling X. An efficient and robust target detection algorithm for identifying minor defects of printed circuit board based on phfe and fl-rfcn. *Front Phys* (2021) 9:661091. doi:10.3389/fphy.2021.661091
- Franzel T, Schmidt U, Roth S. Object detection in multi-view x-ray images. In: *Joint DAGM (German association for pattern recognition) and OAGM symposium*. Charn: Springer (2012). p. 144–54.
- Gao R, Sun Z, Huyen J, Li W, Xiao L, Yao B, et al. Small foreign metal objects detection in x-ray images of clothing products using faster r-cnn and feature pyramid network. *IEEE Trans Instrumentation Meas* (2021) 70:1–11. doi:10.1109/tim.2021.3077666
- Luz E, Silva P, Silva R, Silva L, Guimarães J, Míozzo G, et al. Towards an effective and efficient deep learning model for Covid-19 patterns detection in x-ray images. *Res Biomed Eng* (2022) 38:149–62. doi:10.1007/s42600-021-00151-6
- Akçay S, Breckon T. Towards automatic threat detection: A survey of advances of deep learning within x-ray security imaging. *Pattern Recognition* (2022) 122:108245. doi:10.1016/j.patcog.2021.108245
- Tian C, Fei L, Zheng W, Xu Y, Zuo W, Lin C-W. Deep learning on image denoising: An overview. *Neural Networks* (2020) 131:251–75. doi:10.1016/j.neunet.2020.07.025
- Goudet O, Grelier C, Hao J-K. A deep learning guided memetic framework for graph coloring problems. *Knowledge-Based Syst* (2022) 258:109986. doi:10.1016/j.knsys.2022.109986
- Wei X, Liu S, Xiang Y, Duan Z, Zhao C, Lu Y. Incremental learning based multi-domain adaptation for object detection. *Knowledge-Based Syst* (2020) 210:106420. doi:10.1016/j.knsys.2020.106420
- Pérez-Hernández F, Tabik S, Lamas A, Olmos R, Fujita H, Herrera F. Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance. *Knowledge-Based Syst* (2020) 194:105590. doi:10.1016/j.knsys.2020.105590
- Zuo S, Xiao Y, Chang X, Wang X. Vision transformers for dense prediction: A survey. *Knowledge-Based Syst* (2022) 253:109552. doi:10.1016/j.knsys.2022.109552
- Flitton G, Breckon TP, Megherbi N. A comparison of 3d interest point descriptors with application to airport baggage object detection in complex ct imagery. *Pattern Recognition* (2013) 46:2420–36. doi:10.1016/j.patcog.2013.02.008
- Bhowmik N, Wang Q, Gaus YFA, Szarek M, Breckon TP (2019). The good, the bad and the ugly: Evaluating convolutional neural networks for prohibited item detection using real and synthetically composited x-ray imagery. arXiv preprint arXiv:1909.11508
- Gaus YFA, Bhowmik N, Akçay S, Breckon T. Evaluating the transferability and adversarial discrimination of convolutional neural networks for threat object detection and classification within x-ray security imagery. In: 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA); December 16-19, 2019; Boca Raton, Florida, USA (2019). p. 420–5.
- Hassan T, Khan SH, Akçay S, Bennamoun M, Werghi N (2019). Cascaded structure tensor framework for robust identification of heavily occluded baggage items from multi-ventor x-ray scans. arXiv preprint arXiv:1912.04251
- Zhao C, Zhu L, Dou S, Deng W, Wang L. Detecting overlapped objects in x-ray security imagery by a label-aware mechanism. *IEEE Trans Inf Forensics Security* (2022) 17:998–1009. doi:10.1109/tifs.2022.3154287
- Wei Y, Tao R, Wu Z, Ma Y, Zhang L, Liu X. Occluded prohibited items detection: An x-ray security inspection benchmark and de-occlusion attention module. In: Proceedings of the 28th ACM International Conference on Multimedia; October 12 - 16, 2020; Seattle WA USA (2020). p. 138–46.
- Otsu N. A threshold selection method from gray-level histograms. *IEEE Transactions Systems, Man, Cybernetics* (1979) 9:62–6. doi:10.1109/tsmc.1979.4310076
- Mery D. *Computer vision for x-ray testing*, 10. Switzerland: Springer International Publishing (2015). p. 973–8.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grants No. 62102057 and No. 62036007, in part by the Natural Science Foundation of Chongqing under Grand No. CSTB2022NSCQ-MSX1024, in part by the Chongqing Postdoctoral Innovative Talent Plan under Grant No. CQBX202217, in part by the Postdoctoral Science Foundation of China under Grant No. 2022M720548, in part by the Special Project on Technological Innovation and Application Development under Grant No. cstc2020jscx-dxwtB0032, and in part by Chongqing Excellent Scientist Project under Grant No. cstc2021ycjh-bgzxm0339.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

23. Liang KJ, Heilmann G, Gregory C, Diallo SO, Carlson D, Spell GP, et al. Automatic threat recognition of prohibited items at aviation checkpoint with x-ray imaging: A deep learning approach. *Anomaly Detect Imaging X-Rays (Adix) (Spie)* (2018) 10632:1063203. doi:10.1117/12.2309484
24. Bhowmik N, Gaus YFA, Breckon TP. On the impact of using x-ray energy response imagery for object detection via convolutional neural networks. In: 2021 IEEE International Conference on Image Processing (ICIP); 19–22 September, 2021; Alaska, USA (2021). p. 1224.
25. Viriyasaronon T, Chae S-H, Choi J-H. Mfa-net: Object detection for complex x-ray cargo and baggage security imagery. *Plos one* (2022) 17:e0272961. doi:10.1371/journal.pone.0272961
26. Miao C, Xie L, Wan F, Su C, Liu H, Jiao J, et al. Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 20 2021 to June 25 2021; Nashville, TN, USA. (2019). p. 2119–28.
27. Mery D, Riffo V, Zscherpel U, Mondragón G, Lillo I, Zuccar I, et al. Gdxray: The database of x-ray images for nondestructive testing. *J Nondestructive Eval* (2015) 34:42–12. doi:10.1007/s10921-015-0315-7
28. Talamonti C, Kanxheri K, Pallotta S, Servoli L. Diamond detectors for radiotherapy x-ray small beam dosimetry. *Front Phys* (2021) 9:632299. doi:10.3389/fphy.2021.632299
29. Fourcade A, Khonsari R. Deep learning in medical image analysis: A third eye for doctors. *J Stomatol, Oral Maxill Surg* (2019) 120:279–88. doi:10.1016/j.jormas.2019.06.002
30. Zhou Z, Sodha V, Rahman Siddiquee MM, Feng R, Tajbaksh N, Gotway MB, et al. Models Genesis: Generic autodidactic models for 3d medical image analysis. In: International conference on medical image computing and computer-assisted intervention; September 18th to 22nd 2022; Singapore (2019). p. 384–93.
31. Yang Y, Yan T, Jiang X, Xie R, Li C, Zhou T. Mh-net: Model-data-driven hybrid-fusion network for medical image segmentation. *Knowledge-Based Syst* (2022) 248:108795. doi:10.1016/j.knsys.2022.108795
32. Tang P, Yang P, Nie D, Wu X, Zhou J, Wang Y. Unified medical image segmentation by learning from uncertainty in an end-to-end manner. *Knowledge-Based Syst* (2022) 241:108215. doi:10.1016/j.knsys.2022.108215
33. Liu Y, Wang H, Chen Z, Huangliang K, Zhang H. TransUNet+: Redesigning the skip connection to enhance features in medical image segmentation. *Knowledge-Based Syst* (2022) 256:109859. doi:10.1016/j.knsys.2022.109859
34. Tao R, Wei Y, Jiang X, Li H, Qin H, Wang J, et al. Towards real-world x-ray security inspection: A high-quality benchmark and lateral inhibition module for prohibited items detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 27 October - 2 November 2019; Seoul, South Korea (2021). p. 10923.
35. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, June 27–30, 2016 (2016). 779–88.
36. Zhao Z-Q, Zheng P, Xu S-t, Wu X. Object detection with deep learning: A review. *IEEE Trans Neural networks Learn Syst* (2019) 30:3212–32. doi:10.1109/tnnls.2018.2876865
37. Zou Z, Shi Z, Guo Y, Ye J (2019). Object detection in 20 years: A survey. *arXiv preprint arXiv:1905.05055*
38. Velayudhan D, Hassan T, Damiani E, Werghi N. Recent advances in baggage threat detection: A comprehensive and systematic survey. *ACM Comput Surv* (2022). doi:10.1145/3549932
39. Zheng W, Tang W, Jiang L, Fu C-W. Se-ssd: Self-ensembling single-stage object detector from point cloud. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 20 2021 to June 25 2021; Nashville, TN, USA (2021). p. 14494–503.
40. Zhang Y, Xie F, Huang L, Shi J, Yang J, Li Z. A lightweight one-stage defect detection network for small object based on dual attention mechanism and pafpn. *Front Phys* (2021) 9:708097. doi:10.3389/fphy.2021.708097
41. Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision; October 13 - 16, 2017; NW Washington, DC, United States (2017). p. 2980–8.
42. Redmon J, Farhadi A (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*
43. Tian Z, Shen C, Chen H, He T. Fcos: Fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF international conference on computer vision; Oct. 27 2019 to Nov. 2 2019; Seoul, Korea (2019). p. 9627.
44. Ouyang W, Luo P, Zeng X, Qiu S, Tian Y, Li H, et al. (2014). Deepid-net: Multi-stage and deformable deep convolutional neural networks for object detection. *arXiv preprint arXiv:1409.3505*
45. Du L, Zhang R, Wang X. Overview of two-stage object detection algorithms. *J Phys Conf Ser* (2020) 1544:012033. doi:10.1088/1742-6596/1544/1/012033
46. Xie X, Cheng G, Wang J, Yao X, Han J. Oriented r-cnn for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 11–17 Oct. 2021 (2021). p. 3520–9.
47. Jocher G, Stoken A, Borovec J, Nano Code012 C, Changyu L, Wang M, et al. (2020). ultralytics/yolov5: v3
48. Shao F, Liu J, Wu P, Yang Z, Wu Z. Exploiting foreground and background separation for prohibited item detection in overlapping x-ray images. *Pattern Recognition* (2022) 122:108261. doi:10.1016/j.patcog.2021.108261
49. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 18–20 June 1996; San Francisco, CA, USA (2018). p. 7132.
50. Ding L, Goshtasby A. On the canny edge detector. *Pattern recognition* (2001) 34: 721–5. doi:10.1016/s0031-3203(00)00023-6
51. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. Ssd: Single shot multibox detector. In: European conference on computer vision; 8–16 October; Amsterdam, The Netherlands (2016). p. 21–37.