Improved Terahertz Image-Based Detection of Concealed Aviation Threats via DeepLabv3+ with DSConv and YOLOv5s

Chao Yang
Civil Aviation School, Zhejiang Guangsha Vocational and Technical University of Construction, Dongyang, 321000, China
E-mail: ynljyc@163.com

Keywords: Deep LABv3, YOLOv5s, dangerous goods, detection, civil aviation

Received: October 14, 2025

In recent years, with the growth of civil aviation passenger volume, the importance of dangerous goods detection in ensuring airport and aviation safety has become increasingly prominent. To address the issue of low accuracy in existing dangerous goods detection methods, we propose an improved method for civil aviation dangerous goods detection. The method integrates an enhanced DeepLabv3+ model and a YOLOv5s model with an introduced Coordinate Attention (CA) mechanism. The DeepLabv3+ model is optimized by incorporating Depthwise Separable Convolution (DSConv) and Squeeze-and-Excitation (SE) attention mechanisms to enhance feature extraction capabilities. Meanwhile, the YOLOv5s model improves detection accuracy by incorporating the CA mechanism. We employ the HiXray, PIDray, and a self-collected terahertz dataset for training and validation. Using a methodology that involves semantic segmentation followed by object detection, experimental results demonstrate that the proposed method achieves an average precision (mAP@0.5) of 97.66% and a frame rate (FPS) of 27.23 f/s, outperforming comparison methods such as Faster R-CNN, EA-YOLOv8, and SSD. Additionally, an analysis of the application effectiveness of the proposed method shows detection accuracy rates of 97.8%, 96.6%, and 97.4% for knives, pistols, and lighters, respectively, with a CPU usage of 42.71% and a detection time of 17.89 ms, all of which are superior to the comparison methods. The above research results indicate that the developed hazardous material detection method is effective and practical. This method can provide a theoretical basis for research in the field of hazardous substance detection.

Povzetek: Izboljšana metoda z združenima modeloma DeepLabv3+ in YOLOv5s zelo učinkovito poveča natančnost in hitrost zaznavanja nevarnih predmetov v letalstvu.

1 Introduction

With the rapid development of economic globalization and socio-economic factors, civil aviation, as an important hub for domestic and international passenger transportation, has seen a continuous increase in its passenger flow. People conduct frequent business and travel activities through civil aviation airports, and airport security checks have become increasingly important. In this context, dangerous goods (DGs) detection has become an important part of ensuring airport and aviation safety [1]. However, the current methods for detecting DGs have the problem of low accuracy. Many experts have conducted relevant research, such as Wei et al. who developed a synthesis method for X-ray safety inspection images to address the issue of low detection accuracy of dangerous objects. Comparative analysis results showed that this method improved the recognition performance of DGs [2]. Daud et al. suggested a system grounded on software defined radio technology to address the high cost and low accuracy of traditional liquid hazardous material detection methods (HMDMs). The system was found to have accurately identified over 95% of both suspicious and non-suspicious liquids [3]. To deal with the problem of low real-time detection accuracy of DGs hidden under human clothing, the Jayachitra team proposed a weighted improved YOLOv5 framework, and experimental results showed that this framework was superior to existing frameworks [4]. Fahad et al. proposed a robust denoising framework that combines discrete wavelet transform (DWT) and stationary wavelet transform (SWT) to address the problem of decreased target recognition rate in airport DGs detection due to noise in dual energy X-ray imaging. Outcomes denoted that the framework strengthened the reliability of DGs detection [5].

Although significant progress has been made in luggage security checks, the accuracy of detecting hidden DGs in human body security checks is still relatively low compared to luggage security checks due to factors such as clothing obstruction, low resolution of terahertz images, and high noise. The protection of people's lives and property is of the utmost importance. Therefore, it is crucial to establish a precise and efficient technique to detect concealed DGs in clothing. DeepLabv3+ is a deep learning model designed for semantic segmentation (SS). It offers the advantages of multi-scale feature fusion and

high computational efficiency. This model is commonly used in fields such as smart security systems and autonomous vehicles [6]. YOLOv5s is a real-time object detection model that has outstanding real-time performance and lightweight model advantages, and is widely used in security monitoring and autonomous driving scenarios [7]. Many experts have conducted relevant research. For instance, Chen et al. suggested using an enhanced DeepLabv3+ lightweight neural network to tackle the problem of inaccurate SS in remote sensing images of complex scenes. Their findings revealed that the neural network exhibited strong robustness [8]. Zhao et al. raised an enhanced DeepLabV3+ model to address the issue of low resolution in remote sensing images caused by building shadows. Outcomes demonstrated that the model enhanced the detection accuracy of buildings [9]. In response to the issue of low tomato recognition accuracy caused by multiple environmental factors affecting the picking robot, the Gao G team proposed an improved YOLOv5s. The findings indicated that the improved YOLOv5s model outperformed the original one, with a 2.72% increase in recognition accuracy and a 1.29% increase in mAP [10]. Xiao et al. developed an enhanced YOLOv5s algorithm to deal with the issue of the inability of the intelligent detection model for porosity in circular flower centers to simultaneously consider speed and accuracy. Comparative research findings denoted that the proposed improved YOLOv5s algorithm outperformed the comparative algorithm [11]. Summarizing the above research content, Table 1 can be obtained.

Table 1: Comparison of research methods

Literature	Technology	Advantage	Disadvantage
Wei et al. [2]	soft-non-maximum suppression (combined with soft-nms) and Mask RCNN	The mAP has increased by 3.4%	Increased computation time
Daud et al. [3]	A platform based on software-defined radio (SDR) technology for radio frequency (RF) sensing and the most advanced machine learning (ML) algorithms	More than 95% of the suspicious and non-suspicious liquids were successfully classified	The generalization ability is relatively low
Jayachitra team. [4]	Improved YOLOv5	good results with high mAP@. 5 and mAP@.5:95.	Poor adaptability
Fahad et al. [5]	Robust denoising techniques of DWT and SWT	The proposed system achieved an average PSNR of 35.23 and an MSE of 19.52 for 256×256 DEXI images,	The computation time is too long
Chen et al. [8]	Improved DeepLabv3+	The Mean Pixel Accuracy (MPA) and Mean Intersection over Union (MIoU) are generally best than DeepLabv3+ , U-Net, and PSP-Net, which are respectively improved by 1.22%, − 0.22%, and 2.22% and 2.17%, 1.35%, and 3.42%. ==	Requires high hardware support
Zhao et al. [9]	Convolutional Block Attention Module (CBAM) and DeepLabV3+	Achieving an overall accuracy, precision, recall, and F1 score of 94%, 87%, 90%, and 89%.	Weak adaptability of the plan
Gao team. [10]	YOLOv5s and CBAM	The detection accuracy and recall rate were 92.08% and 82.42% respectively, and the mAP was 92.75%	Increased computation time
Xiao et al. [11]	Slimneck and YOLOv5s	mAP@ 0.5 of the verification set reaches 99.17%	The adaptive ability is relatively weak

From Table 1, existing research has made progress in scenarios such as luggage and liquids. However, for the high missed detection rate caused by "clothing obstruction, low resolution, and high noise" in terahertz human body security checks, there is still a lack of a complete solution that simultaneously meets the requirements of "lightweight, high recall, and low false alarm". Therefore, this paper raises and verifies the following core question: In terahertz images, can the human mask generated by the improved DeepLabv3+ be used as a prior, in conjunction with the Ca-enhanced lightweight YOLOv5s, to significantly improve the detection accuracy of hidden DGs? Therefore, we employ the Deep Labv3+ model and improve it by using Depthwise Separable Convolution (DSConv) and Squeeze-and-Excitation (SE) Attention Mechanisms (AMs) to solve its problems of large parameter count and misjudgment of human contour regions. A terahertz human security image processing algorithm is constructed based on improved Deep Labv3+. At the

same time, it is combined with the improved YOLOv5s to establish a civil aviation DGs detection method based on Deep Labv3+and YOLOv5s. The novelty of this study lies in integrating DSConv and SE AMs into the Deep Labv3+ model, and improving the YOLOv5s algorithm to enhance the accuracy of DGs detection in civil aviation airports. It is expected to provide a certain theoretical basis for the research field of hazardous substance detection.

2 Methods

2.1 Construction of terahertz human security inspection image processing model based on improved deep Labv3+

In the last few years, with the increasing flow of civil aviation passengers, the detection of DGs has become increasingly important [12]. To achieve accurate detection of DGs in civil aviation, it is necessary to segment the human background [13]. Terahertz imaging is a non-destructive testing technology that has advantages such as high penetration, high-precision imaging, and low radiation. It has been widely used in security inspection fields such as airports and train stations. However, the current methods for detecting hidden DGs in civil aviation have reduced detection accuracy due to issues such as low resolution and high noise in terahertz imaging. Therefore, we utilize the Deep Labv3+ model to process security check images of civil airports to enhance detection accuracy. The Deep Labv3+ model is shown in Figure 1 [14].

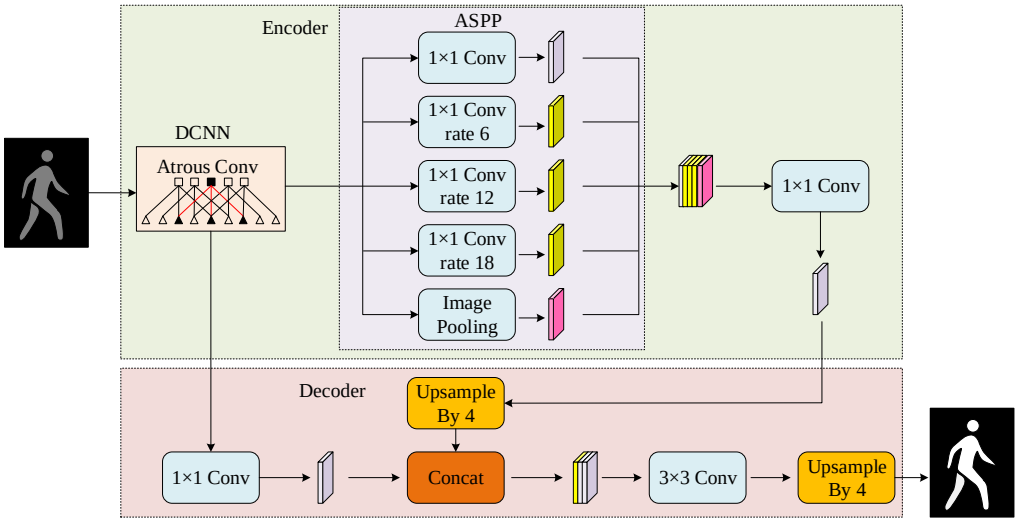


Figure 1: Deep Labv3+ model

From Figure 1, the DeepLabv3+ model contains an encoder and a decoder. The encoder first extracts feature from the input image with deep convolutional neural network (DCNN). Secondly, the obtained features are input in parallel into a convolutional layer (Conv) with a size of 1 × 1, atrous convolutions (Atrous Conv) with dilation rates of 6, 12, and 18, and four branches of global average pooling. Each branch is compressed through a 1 × 1 Conv channel and concatenated to form an atrous spatial pyramid pooling (ASPP) module, thereby achieving multi-scale context fusion. The decoder upsamples the fused features by a factor of 4. Simultaneously, the features obtained from the DCNN undergo dimension reduction via 1×1 convolutions and are concatenated. Then, a 3×3 convolution is employed to fuse semantic information and fine details. After fusion, the resolution is restored through 4x upsampling, and the final output is pixel level SS results. Among them, dilated convolution is a deep learning technique that expands the receptive field of convolutional kernels by introducing dilation rate parameters. It has advantages such as multi-scale information aggregation and maintaining feature map (FM) resolution, and is broadly utilized in areas such as image segmentation and object detection. If there is an image that undergoes dilated convolution with respect to the input feature j, the spatial position p present in the output FM k can be expressed by equation (1).

$$k[p] = \sum_{l} j[p + r \cdot l]w[l] \tag{1}$$

In equation (1), r is the expansion rate, l means the size of the visual field, and w means the convolution kernel (CK). Due to the variation of the dilation rate of the CK, the actual size of the dilated CK can be represented by equation (2).

$$Z = z + (z-1)(r-1) \tag{2}$$

In equation (2), z denotes the size of the original CK. Z stands for the actual size of the CK after dilation. Thus, the area of input data that the dilated kernel of a dilated convolution can process is expressed by equation (3).

$$F_{i+1} = F_i + (Z-1) \cdot S \tag{3}$$

In equation (3), F_i and F_{i+1} represent the respective input data regions that the CKs of the previous layer and the current layer can process after expansion. S denotes

the stride distance by which the CK moves during the convolution operation. Therefore, the Deep Labv3+ model avoids information loss and enhances the accuracy of image segmentation through dilated convolution. However, the DeepLabv3+ model has the problem of excessively large parameters. DSConv is a convolution algorithm that reduces computational complexity and parameter count by splitting spatial and channel dimension correlations. It has the advantages of significantly reducing parameter and computational complexity, improving model efficiency and performance, and can effectively compensate for the shortcomings of DeepLabv3+ models [15]. The calculation steps of standard convolution and DSConv are shown in Figure 2.

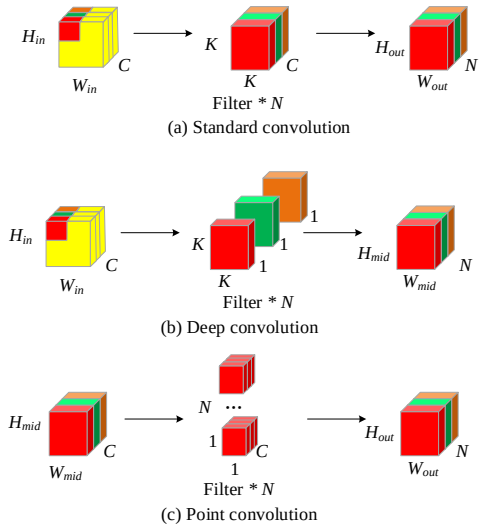


Figure 2: The calculation steps of standard convolution and DSConv

In Figure 2, H_{in}, W_{in}, and C indicate the height, width, and the amount of channels of the input FM. N means the amount of channels in the output FM, K means the height and width of the ck, H_{out}, $W_{H_{out}}$, and H_{out} denote the length, width, and amount of channels in the output FM, while H_{mid} and W_{mid} mean the length and width of the intermediate FM. From Figure 2 (a), the standard convolution is utilized to the input FM of $H_{in} \times W_{in} \times C$, and after calculating the convolution with N $K \times K \times N$ s, the output of $H_{out} \times W_{out} \times N$ is obtained. The calculation cost can be represented by equation (4).

$$C_s = H_{in} \times W_{in} \times C \times N \times K \times K \quad (4)$$

In equation (4), C_s is the standard convolution. The calculation steps of DSConv are divided into two steps. Firstly, the deep convolution in Figure 2 (b) is used to separately perform $K \times K \times 1$ spatial convolution on each input channel, while keeping the amount of channels C constant, to obtain the intermediate feature $H_{mid} \times W_{mid} \times N$. Subsequently, the convolution with N $1 \times 1 \times C$ s in Figure 2 (c) is used to linearly combine the intermediate features, resulting in the final N channel outputs. The DSConv calculation cost C_{DSConv} can be represented by equation (5).

$$C_{DSConv} = H_{in} \times W_{in} \times C \times K \times K + C \times N \times K \times K \quad (5)$$

Comparing equations (5) and (4), the computational complexity decreases by about 8 to 9 times, while the accuracy remains almost unchanged [16]. Thus, the standard convolution in the DeepLabv3+ model can be replaced with DSConv to achieve lightweighting. The SE AM has been introduced into the encoder of the DeepLabv3+ model to strengthen its ability to extract human detail features [17]. Thus, a terahertz human security inspection image processing model based on improved Deep Labv3+ is constructed. The model is shown in Figure 3.

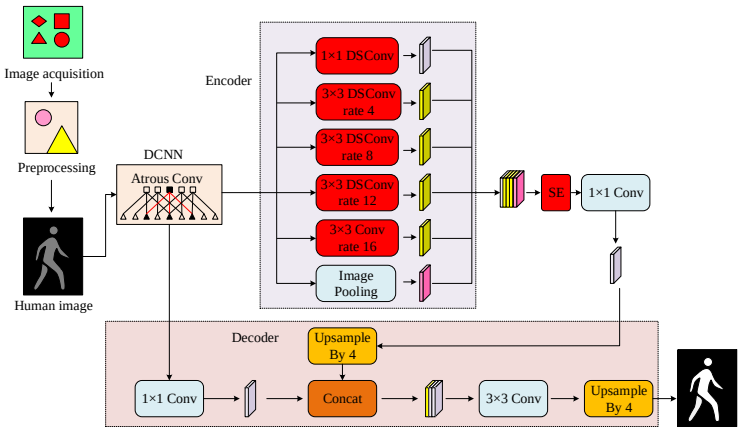
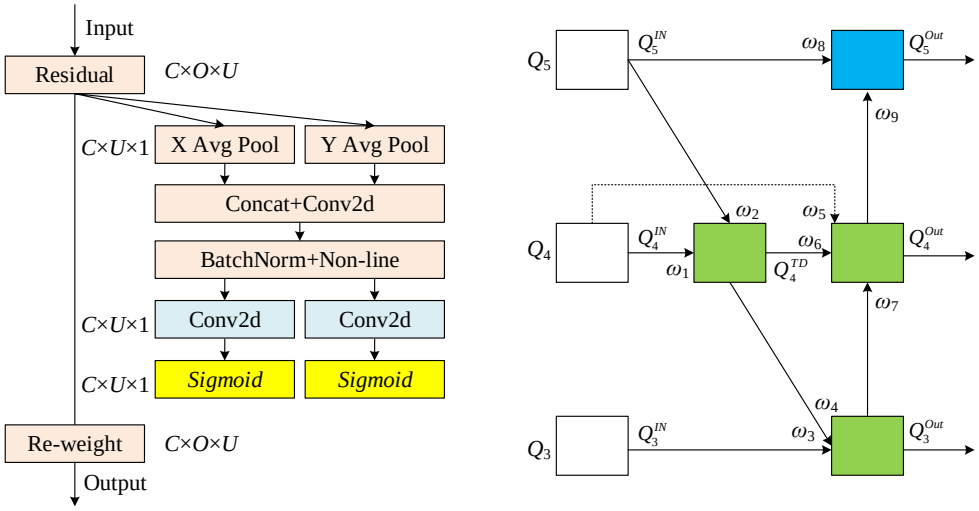


Figure 3: Terahertz human body security inspection image processing model based on improved Deep Labv3+

From Figure 3, the operation steps of the model can be seen. Firstly, civil aviation security personnel use terahertz security systems to sample individuals carrying different hazardous materials under different environmental and weather conditions. After sampling, the video is converted into an image and saves in JPG format with a resolution of 304 * 140 [18]. Secondly, the image is preprocessed. During preprocessing, median filtering is used to denoise the image. Logarithmic transformation method is used to enhance the image. Next, the processed images are input into the improved Deep Labv3+ model. Subsequently, the preprocessed image enters the encoder of the Deep Labv3+ model. The encoder uses DCNN to preliminarily extract image features. After extraction, multi-scale features are extracted using 3 × 3 DSConv layers with expansion rates of 4, 8, 12, and 16, respectively. Simultaneously, a pooling layer is utilized to capture global contextual information. Then, the SE AM is used to reinforce important features. Finally, the decoder of the original Deep Labv3+ model is used to convert the output features of the encoder into images and output the results.

2.2 Construction of hazardous material detection method based on improved Deep Labv3+ and improved YOLOv5s

After building the improved Deep Labv3+ model, to further ENHANCE the performance of DGs detection in civil aviation human security inspection images, the YOLOv5s network model is introduced to complement it. The YOLOv5s is a lightweight object detection model offering fast detection speeds and high accuracy. It is employed in various fields, including video surveillance and autonomous driving [19]. However, when facing civil aviation security check scenarios, the DGs carried by the human body are relatively small, and YOLOv5s network has the ISSUE of low detection accuracy. Therefore, we utilize coordinate attention (CA) and Lightweight bidirectional feature pyramid network (BiFPN) to strengthen it. The CA is an efficient feature extraction method that has the advantages of enhancing feature extraction capabilities and reducing computational complexity, and is widely used to improve network performance. Compared to other AMs, the CA not only focuses on channel information, but also sensitively captures orientation and position sensitive information. BiFPN is an improved feature pyramid network structure that has advantages such as bidirectional information transmission and structural optimization, and is widely used to enhance the detection accuracy of object detection models for objects of different sizes. The CA and lightweight BiFPN structure are denoted in Figure 4.



(a) The operation process of the CA attention mechanism (b) Lightweight BiFPN structure

Figure 4: CA and lightweight BiFPN structure

In Figure 4, O and U denote the width and height of the FM. From Figure 4 (a), the operation steps of the CA are to first input the FM into a residual connection to alleviate the gradient vanishing problem. Secondly, feature representations are obtained in both horizontal and vertical directions by averaging and pooling the FMs globally. The output of global average pooling can be represented by equation (6).

$$\begin{cases} z_c^h(h) = \dfrac{1}{W} \sum\limits_{0 \le i \le W} x_c(h,i) \\ z_c^w(w) = \dfrac{1}{H} \sum\limits_{0 \le i \le H} x_c(j,w) \end{cases} \tag{6}$$

In equation (6), i and j denote coordinate indices, $x_c(j,u)$ and $x_c(o,i)$ denote the feature matrices, and $\varphi_c^o(o)$ and $\varphi_c^u(u)$ denote the eigenvalues. Next, the

output of the Conv is normalized. Then, two convolutions and the Sigmiod activation function are used to generate attention weights in two directions. Finally, the attention weights generated are applied to the original FM input, producing the final FM output. The output of CA can be represented by equation (7).

$$\psi_c(i,j) = x_c(i,j) \times f_c^o(i) \times f_c^u(j) \quad (7)$$

In equation (7), $x_c(i,j)$ indicates the original input feature matrix, $f_c^o(i)$ denotes the horizontal-direction output, $f_c^u(j)$ signifies the vertical-direction output, and $\psi_c(i,j)$ is the final feature matrix generated by the CA. In Figure 4 (b), Q_3, Q_4, and Q_5 are FMs from different levels input to BiFPN, ω_1 to ω_9 are weight coefficients, Q_3^{IN}, Q_4^{IN}, and Q_5^{IN} are the original FMs input, Q_4^{TD} is the upsampled FM, and Q_3^{Out}, Q_4^{Out}, and Q_5^{Out} are the FMs processed by BiFPN. As shown in Figure 4 (b), the lightweight BiFPN only has three feature input layers. Compared to the original BiFPN, it lacks two feature input layers, which can improve computational efficiency without affecting feature fusion performance. Q_4^{TD} can be represented by equation (8).

$$Q_4^{TD} = Conv\left(\frac{\omega_1 \cdot Q_4^{IN} + \omega_2 \cdot Resize(Q_5^{IN})}{\omega_1 + \omega_2 + \varepsilon}\right) \quad (8)$$

In equation (8), ε is a very small constant, and $Resize$ is the operation of adjusting the size of the FM. So, Q_3^{Out}, Q_4^{Out}, and Q_5^{Out} can be represented by equation (9).

$$\begin{cases} Q_3^{Out} = Conv\left(\dfrac{\omega_3 \cdot Q_3^{IN} + \omega_4 \cdot Resize(Q_4^{TD})}{\omega_3 + \omega_4 + \varepsilon}\right) \\ Q_4^{Out} = Conv\left(\dfrac{\omega_5 \cdot Q_4^{IN} + \omega_6 \cdot Q_4^{TD} + \omega_7 \cdot Resize(Q_3^{Out})}{\omega_5 + \omega_6 + \omega_7 + \varepsilon}\right) \\ Q_5^{Out} = Conv\left(\dfrac{\omega_7 \cdot Q_5^{IN} + \omega_9 \cdot Resize(Q_4^{Out})}{\omega_8 + \omega_9 + \varepsilon}\right) \end{cases} \quad (9)$$

Although the detection accuracy of YOLOv5 has been enhanced with CA and lightweight BiFPN, the commonly used Generalized Intersection over Union (GIoU) loss function (LF) still has limitations. It cannot determine the relative position when the predicted box completely encloses the real box. In view of this, we apply the α-Complete Intersection over Union (Alpha-OU) LF and the Complete Intersection over Union (CIoU) entropy LF to improve it. Among them, the loss value L_{CIoU} of the CIoU LF can be represented by equation (10).

$$L_{CIoU} = 1 - IoU + \frac{\rho(a,b)}{d^2} + \beta v \quad (10)$$

In equation (10), a denotes the center coordinates of the model-annotated bounding box, β is a hyperparameter, b means the center coordinates of the model-predicted bounding box, L_{CIoU} is the loss value, $\rho(\cdot)$ is the Euclidean distance, d is the minimum diagonal length of the mini bounding box, v is the additional penalty for aspect ratio difference, and IoU is the intersection-over-union ratio. By introducing the α parameter, Alpha-OU further improved CIoU, enabling it to better balance confidence and positional accuracy. By combining Alpha-IoU and CIoU, a new LF $\alpha - CIoU$ can be obtained. The loss value $Loss$ can be expressed by equation (11).

$$Loss = 1 - IoU^{\alpha} + \frac{\rho^{2\alpha}(a,b)}{d^{2\alpha}} + (\beta v)^{\alpha} \quad (11)$$

Thus, the improved YOLOv5s network model has been constructed, and its structure is shown in Figure 5.

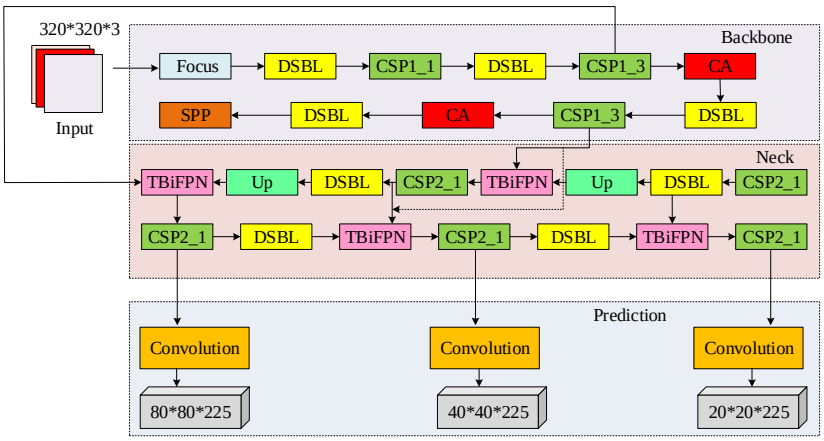


Figure 5: Improved YOLOv5s network model structure

From Figure 5, the enhanced YOLOv5s first utilizes the backbone network to extract input image features. The backbone network includes the Focus module, Darknet Squeeze-and-Excitation Block with Depthwise Convolution (DSBL) module, Cross Stage Partial (CPS), and CA mechanism. The role of DSBL, CPS, and CA mechanisms is to improve the network's feature extraction capability. The Spatial Pyramid Pooling (SPP) module is used for multi-scale feature extraction. Secondly, the neck network is employed to further perform multi-scale fusion of features. The neck network includes multiple lightweight BiFPN (TBiFPN), upsampling (Up), DSBL, and CSP modules. Finally, the network predicts through different Convs and outputs FMs with sizes of 80x80x225, 40x40x225, and 20x20x225, respectively.

Finally, we combine improved Deep Labv3+ and improved YOLOv5s to construct a method for detecting DGs at civil airports. The detection method first generates terahertz human body images through the terahertz security system during civil aviation security checks. Secondly, the median filtering method is employed to denoise the obtained image. Moreover, the logarithmic transformation method is employed to enhance the image. Then, using an improved Deep Labv3+ model for human segmentation, the human body is separated from the background for easier detection of clothing. Finally, the improved YOLOv5 is used to detect DGs in the segmented images and output the detection results. To overcome the problem of insufficient number of self-collected terahertz images, we conduct offline re-augmentation of the samples. First, rotate randomly at $\pm 15^{\circ}$ and simultaneously apply a horizontal flip with a 50% probability. Then, we utilize superimpose Gaussian noise (σ =0.01) to simulate the random disturbance of the equipment. Subsequently, within the Saturation and Valu Spaces, the luminance variation is $\pm$30%, combined with Gamma correction in the [0.8,1.2] interval. Finally, contrast-constrained adaptive histogram equalization is adopted for local contrast enhancement. 8 new samples are generated for each original image. Experiments are conducted on an Ubuntu 18.04 workstation equipped with an Intel Xeon E5-2680 v3 CPU at 3.30 GHz and two RTX 3090 GPUs, accelerated by PyTorch 1.13 and CUDA 11.7. Input images are resized to 640 $\times$ 640 pixels using bilinear interpolation while preserving the aspect ratio and padded with a gray value of 114 on the edges. For the improved YOLOv5s, stochastic gradient descent is applied with a batch size of 16, an initial learning rate of 0.007 that is cosine-annealed to 0.0007, a 3-epoch warm-up, momentum of 0.937 and weight decay of 0.0001. The loss function combined CIOU, classification, and objectness terms with equal weights and the backbone is initialized with ImageNet pretrained weights. Training ran for 300 epochs with early stopping patience of 30 epochs. For the improved DeepLabv3+, SGD is also used with a batch size of 16, an initial learning rate of 7×10^{-4} decayed by a poly factor of 0.9, a 5-epoch warm-up, momentum of 0.9 and weight decay of 1×10^{-4}. The loss consists of Dice and Focal terms with Focal alpha of 0.75 and Focal gamma of 2, and the backbone is initialized from ImageNet pretrained weights for 300 epochs. Both networks employ gradient clipping with max_norm = 10 and mixed-precision training. The best model is selected according to the highest validation mAP@0.5. The pseudo-code of the detection method is shown in Table 2.

Table 2: Pseudo-code of the detection method

DGs detection method based on improved Deep Labv3+ and improved YOLOv5s
Import the necessary libraries import torch import torchvision # Define the DeepLabv3+ model with DSConv def create_deeplabv3_plus_with_dsconv():   model = torchvision.models.segmentation.deeplabv3plus_resnet50(pretrained=False)   # Here, the replacement logic of DSConv is added   # For example: Replace some Conv with DSConv   return model # Define the YOLOv5s model def create_yolov5s():   model = torch.hub.load('ultralytics/yolov5', 'yolov5s', pretrained=True)   return model # Data loading and preprocessing def load_and_preprocess_terahertz_images(image_paths):   images = []

```
    for path in image_paths:
        image = torchvision.io.read_image(path)
        # Carry out necessary preprocessing, such as normalization, etc
        image = image / 255.0
        images.append(image)
    images = torch.stack(images)
    return images
# Detect hidden aviation threats
def detect_concealed_aviation_threats(images, deeplabv3_plus, yolov5s):
    # Generate human masks using DeepLabv3+
    deeplab_output = deeplabv3_plus(images)['out']
    human_mask = (deeplab_output > 0.5).float()
    # Use YOLOv5s for object detection
    yolov5_output = yolov5s(images * human_mask)
    # Process the test results
    threats = []
    for result in yolov5_output.pred:
        # Filter out threat targets
        threat_boxes = result[result[:, 4] > 0.5]  # The confidence threshold is
0.5
        threats.append(threat_boxes)
    return threats, human_mask
# Main function
def main():
    # Create a model
    deeplabv3_plus = create_deeplabv3_plus_with_dsconv()
    yolov5s = create_yolov5s()
    # Load terahertz images
    image_paths = ['path/to/image1.jpg', 'path/to/image2.jpg']
    images = load_and_preprocess_terahertz_images(image_paths)
    # Detect hidden aviation threats
    threats, human_mask = detect_concealed_aviation_threats(images,
deeplabv3_plus, yolov5s)
    return threats, human_mask
# Run the main function
threats, human_mask = main()
```

3 Results and analysis

3.1 Performance analysis of dangerous goods detection methods

To test the efficacy of the developed detection method, a performance comparison experiment was conducted against other methods. The comparative methods included the HMDM based on Faster R-CNN, the weight decay rate and batch size were 1×10^{-4} and 16, respectively. The momentum was 0.9, and the optimizer was stochastic gradient descent. The human body image dataset was subjected to terahertz security checks by 15 participants carrying DGs. Data preprocessing first used median filtering to remove noise. Secondly, logarithmic transformation was used for enhancement processing. Finally, the Labelme labeling software was used to label the human body contour and generate a "human body" label. After completing the annotation, the original image and corresponding label images were stored in the database. There were a total of 4892 images, with 70% being the training set, 20% being the testing set, and 10% being the validation set. The hazardous materials dataset was sourced from HiXray and PIDray, with a total of 13462 images. 80% of the dataset was the training set and 20% was the testing set. There are 45,403 original images in the HiXray dataset and 12,600 original images in the PIDray dataset. Both databases contain dangerous items such as firearms, knives, liquid bottles, lighters, and scissors. HiXray provides COCO JSON annotation, and PIDray provides YOLO txt annotation. To align with the concealed carrying scenarios, only the above five types of targets are retained, uniformly converted to the COCO JSON format, and the images are scaled to 640× 640 pixels and converted to single-channel grayscale. After category filtering, a total of 13,462 images were obtained, including 10,362 HiXray and 3,100 PIDray.

HMDM based on EA-YOLOv8, and the HMDM based on SSD. Network parameter settings of improved YOLOv5: batch size and weight decay rate were 16 and 0.0001, respectively. The optimizer utilized stochastic gradient descent method. The training epochs were 300 times. The initial learning rate was 0.007. The training iterations and initial learning rate of the improved Deep Labv3+ model were 300 and 7×10^{-4}, respectively. The

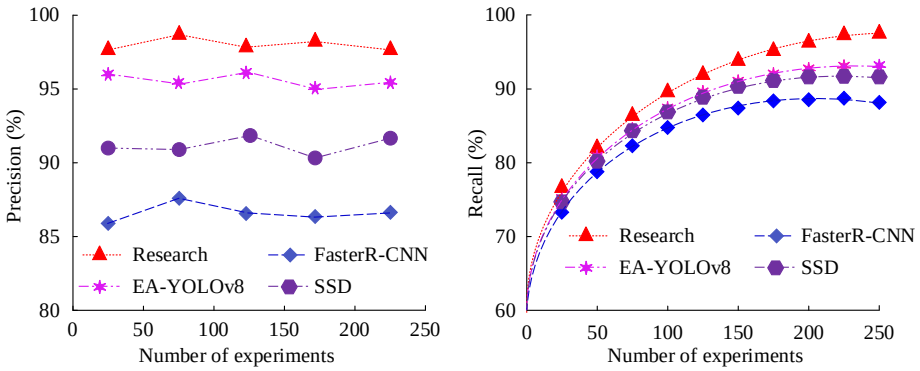
The preprocessing flow first performs CLAHE local contrast enhancement, and then applies rotation, flipping, noise addition and luminance-gamma combined augmentation, expanding each image by 8 times. The experimental comparison indicators were precision, recall, and F1 score, etc. The experimental environment is denoted in Table 3.

Table 3: Experimental environment

Parameter names	Parameter
CPU	Intel®Xeon® Processor E5-2680 v3
Constant frequency	3.30 GHz
Random access memory	16GB
Hard disk drive	500GB
Operating system	Ubuntu 18.04
Matrix laboratory version	2021a
Programming language	Python3.7
Development framework	PyTorch1.7
Data analysis software	Spss24.0

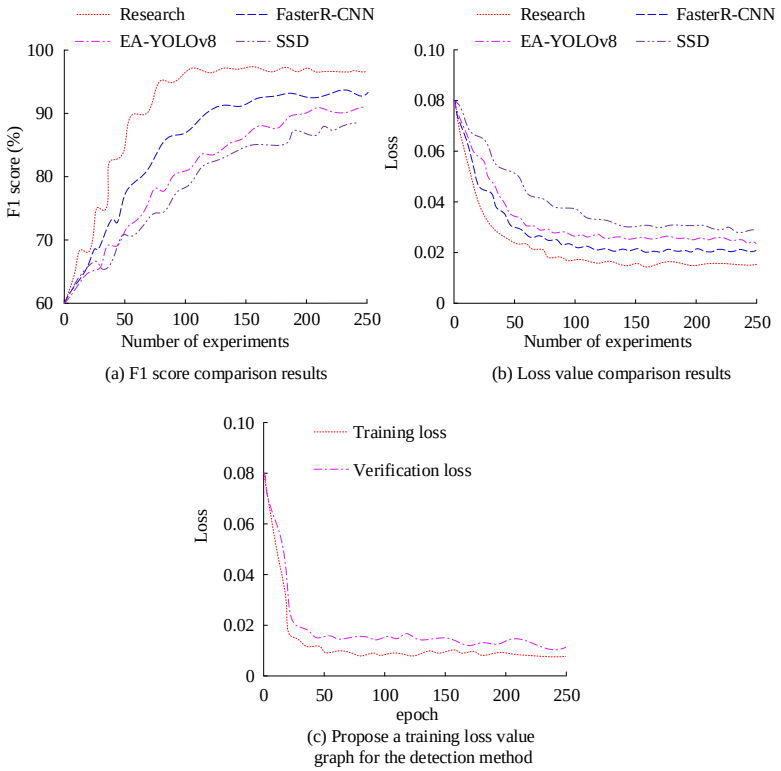
The tests of FPS and CPU usage were conducted on a system equipped with an Intel Xeon E5-2680 v3 processor. The test environment was the Ubuntu 18.04 operating system, equipped with 16GB random access memory (RAM), and was executed under the condition of ensuring that no other high-load processes were running. To ensure the repeatability of the test results, we conducted 100 inference iterations on the model, processing a 640×640 pixel image in each iteration. All tests were conducted under the same system Settings to ensure the consistency of the results. Figure 6 illustrates the experimental results of the study, which are based on a comparative analysis of the precision and recall of various detection methods in the aforementioned environment.

From Figure 6 (a), the proposed detection method had a precision of 97.83%, which was much higher than Faster R-CNN's 95.34%, EA-YOLOv8's 91.49%, and SSD's 87.12%. According to Figure 6 (b), the recall rates of the proposed detection method, Faster R-CNN, EA-YOLOv8, and SSD were 97.17%, 90.03%, 89.76%, 88.75%, and 87.88%, respectively. The research method had the highest recall rate. The above findings suggest that, compared to other detection methods, the research method offers superior precision and recall. The comparison results of F1 score loss values and the training loss graph of the proposed detection method for various detection methods are shown in Figure 7.



(a) Precision comparison results (b) Recall rate comparison results

Figure 6: Comparison results of recall rate and precision rate



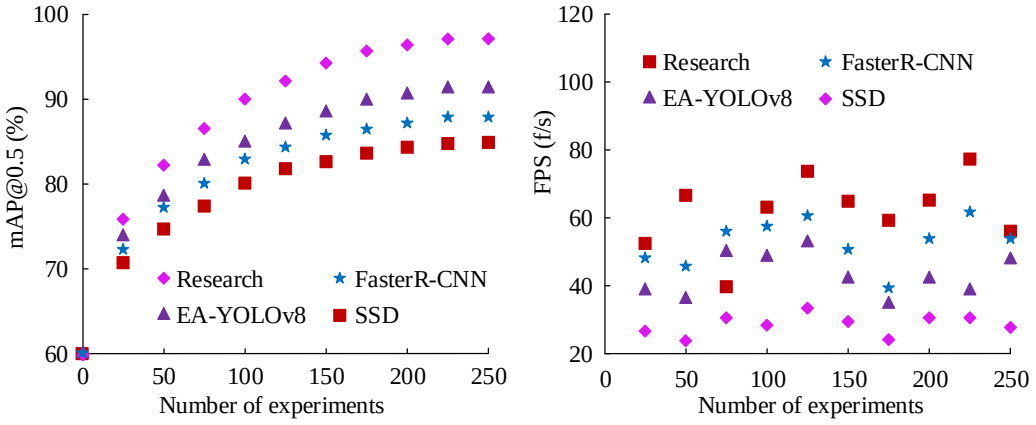
(a) F1 score comparison results

(b) Loss value comparison results

(c) Propose a training loss value graph for the detection method

Figure 7: Comparison results of F1 score, loss values, and the training loss graph of the proposed detection method

According to Figure 7 (a), the F1 score of the proposed detection method was 98.89%, the F1 score of Faster R-CNN was 92.57%, the F1 score of EA-YOLOv8 was 90.12%, and the F1 score of SSD was 88.11%. The research method had the highest F1 score. From Figure 7 (b), the loss values of the proposed detection method, Faster R-CNN, EA-YOLOv8, and SSD were 0.18, 0.22, 0.32, and 0.38, respectively. The research proposed detection method had the lowest loss value. The above findings denote that the research method is more effective than the comparative methods in terms of both the F1 score and the loss value. Furthermore, as shown in Figure 7 (c), the validation loss continued to decline and stabilize without any rebound, indicating that the model has not experienced significant overfitting. Figure 8 showcases the comparison results of the mean precision (mAP@0.5) and frames per second (FPS) for each method when the IoU threshold was set to 0.5.



(a) mAP@0.5 comparison results

(b) FPS comparison results

Figure 8: mAP@0.5 and FPS comparison results

In Figure 8(a), the proposed detection method achieved the highest mAP@0.5 of 97.66%, surpassing Faster R-CNN (88.79%), EA-YOLOv8 (90.08%), and SSD (84.26%). In Figure 8(b), the research method achieved the highest average FPS at 59.68 fps. In comparison, FasterR-CNN achieved 47.32 fps, EA-YOLOv8 achieved 42.14 fps, and SSD achieved 27.23 fps. These findings reveal that the research method outperforms the comparison methods in terms of both mAP@0.5 score and FPS. To verify the reliability of the experiment,

statistical analysis was conducted on various experiments involving hazardous materials. The results of the statistical test are shown in Table 4.

Table 4: Statistical test results

DGs	Indicator	Research	Faster R-CNN	EA-YOLOv8	SSD	*p* value
Cutting tools	Precision	97.83%	95.34%	91.49%	87.12%	$p<0.001$
Pistols	Recall	97.17%	90.03%	89.76%	88.75%	$p<0.001$
Lighters	F1 score	98.89%	92.57%	90.12%	88.11%	$p<0.001$
Cutting tools	Loss values	0.18	0.22	0.32	0.38	$p<0.001$
Pistols	mAP@0.5	97.66%	88.79%	90.08%	84.26%	$p<0.001$
Lighters	FPS	59.68 fps	47.32 fps	42.14 fps	27.23 fps	$p<0.001$

From Table 4, the *p*-values of the comparative experiments with Precision, Recall, F1 score, Loss values, mAP@0.5, and FPS were all less than 0.001, indicating that the experimental results are reliable. To verify the performance contribution of the α-CIoU loss function, the SE module and the CA module to the detection method, ablation experiments were conducted on them. The experimental results are shown in Table 5.

Table 5: Results of the ablation experiment

Project	Precision	Loss value
α-CIoU	97.83%	0.18
GIoU	90.14%	0.21
CIoU	89.86%	0.26
Remove the SE module	89.96%	0.24
Remove the SE module and the CA module	84.14%	0.27

As shown in Table 5, the Precision rate of using the α-CIoU loss function was significantly higher than that of GIoU and CIoU, and the loss value was lower than that of using GIoU and CIoU. After removing the SE module and the CA module, the Precision rate of model checking decreased significantly, and the loss value gradually increased. This result indicates that the introduction of the α-CIoU loss function, SE module and CA module significantly improves the performance of the model.

3.2 Detection performance analysis

After testing the effect of the raised detection method, we conducted an analysis of its application effectiveness compared to other methods. We conducted separate inspections on civil aviation security personnel carrying knives, pistols, and lighters. The test results are shown in Figure 9.

From Figure 9 (a), the detection accuracy of the research method for cutting tools, pistols, and lighters was 97.8%, 96.6%, and 97.4%, respectively. From Figures 9 (b), 9 (c), and 9 (c), Faster R-CNN, EA-YOLOv8, and SSD had the highest detection accuracies of 86.2%, 88.4%, and 89.7% for the three DGs, respectively, which were lower than the lowest detection accuracy of 96.6% proposed by the research method. The above findings indicate that the proposed detection method can accurately detect DGs carried by the human body and has practical value. The results of CPU occupancy, memory size, detection time, Average Precision (AP), Mean Square Error (MSE), and Root Mean Square Error (RMSE) for each method are shown in Table 6.

From Table 6, the AP of the developed detection method was 97.62%, significantly lower than the 89.21% of Faster R-CNN, 90.62% of EA-YOLOv8, and 82.87% of SSD. In addition, the detection methods proposed in the study showed MSE value, RMSE value, memory usage size, CPU usage rate, and detection time of 1.257, 1.121, 13.2 MB, 42.71%, and 17.89 ms, respectively. The above results indicate that compared with the comparative methods, the detection method proposed in the study exhibits better performance in AP value, MSE value, RMSE value, memory usage size, CPU usage rate, and detection time. To verify the reliability of the detection effect, we conducted a statistical analysis of the experiment, and the analysis results are shown in Table 7.

From Table 7, the *p*-values of the detection effect experiments were all less than 0.001. This result indicates that the conducted experiments have a significant difference at the statistical 0.1% level, and the experimental results are reliable. To verify the performance of the proposed model under different conditions, experiments were conducted on it in various environments. The experimental results are shown in Table 8.

From Table 8, the detection methods proposed in the study were applied in Indoor airport security check, Outdoor airport security check, and High-humidity airport security. The missed detection rates of check in the three different environments were 1.51%, 1.78%, and 1.98% respectively. The false detection rates were 2.29%, 2.06% and 2.21% respectively. All were lower than the comparison methods. The above results indicate that the detection method proposed in the research has superior detection performance in different environments and possesses good universality and practical value.

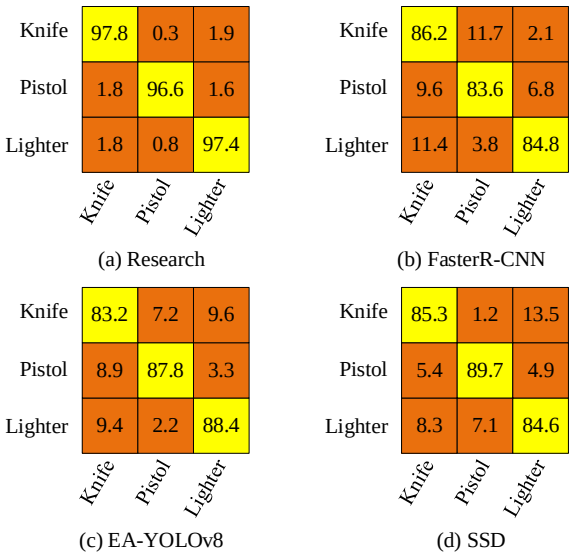


Figure 9: Detection effect of hazardous materials

Table 6: Comparison results of each method

Indicator	Research	Faster R-CNN	EA-YOLOv8	SSD
AP	97.62%	89.21%	90.62%	82.87%
MSE	1.257	2.796	2.893	4.671
RMSE	1.121	1.672	1.701	2.161
Memory usage size	13.2 MB	147.3 MB	168.9 MB	203.2 MB
CPU	42.71%	57.38%	61.22%	56.37%
Detection time	17.89 ms	48.92 ms	59.94%	46.27%

Table 7: Statistical analysis results of the detection experiment

Category	Indicator	Research	Faster R-CNN	EA-YOLOv8	SSD	*p* value
All types of hazards	AP	97.62%	89.21%	90.62%	82.87%	$p<0.001$
All types of hazards	MSE	1.257	2.796	2.893	4.671	$p<0.001$
All types of hazards	RMSE	1.121	1.672	1.701	2.161	$p<0.001$
All types of hazards	Memory usage size	13.2 MB	147.3 MB	168.9 MB	203.2 MB	$p<0.001$
All types of hazards	CPU	42.71%	57.38%	61.22%	56.37%	$p<0.001$
All types of hazards	Detection time	17.89 ms	48.92 ms	59.94%	46.27%	$p<0.001$
Cutting tools	Accuracy	97.8%	86.2%	83.2%	85.3%	$p<0.001$
Pistols	Accuracy	96.6%	83.6%	87.8%	89.7%	$p<0.001$
Lighters	Accuracy	97.4%	84.8%	88.4%	84.6%	$p<0.001$

Table 8: Test results under different conditions

Environmental conditions	Model name	Missed detection rate	False detection rate	*p* value
Indoor airport security check	Research	1.51%	2.29%	$p<0.001$
	Faster R-CNN	2.02%	3.04%	$p<0.001$
	EA-YOLOv8	1.83%	2.58%	$p<0.001$

Environmental conditions	Model name	Missed detection rate	False detection rate	*p* value
	SSD	2.21%	2.87%	$p<0.001$
Outdoor airport security check	Research	1.78%	2.06%	$p<0.001$
	Faster R-CNN	2.55%	3.57%	$p<0.001$
	EA-YOLOv8	2.04%	2.32%	$p<0.001$
	SSD	2.37%	2.73%	$p<0.001$
High-humidity airport security check	Research	1.98%	2.21%	$p<0.001$
	Faster R-CNN	3.06%	3.82%	$p<0.001$
	EA-YOLOv8	2.24%	2.63%	$p<0.001$
	SSD	2.53%	2.95%	$p<0.001$

4 Discussion

This study conducted comparative experimental analysis on the performance of DGs detection methods for civil airports based on improved Deep Labv3+ and improved YOLOv5s, and conducted application effect analysis experiments on this detection method. The findings denoted that the precision of this detection method, Faster R-CNN, EA-YOLOv8, and SSD were 97.83%, 95.34%, 91.49%, and 87.12%, respectively. The proposed detection method had the highest precision. This result indicates that the introduction of DSConv and SE improves the detection accuracy of the detection method. This result coincides with the relevant research findings of the Kuo T C team [20]. In the recall comparison experiment, the recall rates of the proposed detection method, Faster R-CNN, EA-YOLOv8, and SSD were 97.17%, 90.03%, 89.76%, 88.75%, and 87.88%, respectively. The proposed detection method had the highest recall rate. This result indicates that the introduction of CA improves the accuracy of the detection method. The results obtained by Peng H et al. in related studies are similar [21]. In the loss value comparison experiment, the loss values of the proposed detection method, Faster R-CNN, EA-YOLOv8, and SSD were 0.18, 0.22, 0.32, and 0.38, respectively. The proposed research method had the lowest loss value. This indicates that combining a lightweight BiFPN with an improved DeepLabv3+ model enhances the detection performance. This result is consistent with the relevant research findings of Kim W's team [22]. In addition, the proposed detection method had a recall rate of 97.17% and an F1 score of 92.57%, mAP@0.5 value was 97.66%, and the average FPS was 27.23 f/s, both of which were better than the comparison methods. This further demonstrates the superior performance of the raised detection method. In the comparative analysis experiment of application effects, the detection accuracy of the developed detection method for cutting tools, pistols, and lighters were 97.8%, 96.6%, and 97.4%, respectively, all of which were superior to the comparative methods. Moreover, the AP value, MSE value, RMSE value, memory usage size, CPU usage rate, and detection time were 97.62%, 1.257%, 1.121%, 13.2 MB, 42.71%, and 17.89 ms, respectively, all of which were better than the comparative detection method. This result indicates that the proposed detection method has good practical value. This finding is similar to the research finding of Gallo G et al. in 2022 [23]. The above research results indicate that the proposed method for detecting DGs in civil aviation is effective and has practical application value. The limitation of this study is that it only conducted the detection of hidden DGs in civil aviation human bodies, without considering the luggage security check process. The future research direction is to combine terahertz images of the human body with corresponding X-ray images of luggage, and use graph neural networks to achieve cross modal feature fusion detection, thereby improving the success rate of collaborative detection of DGs in the human body and luggage. For the practical application of airport security inspection, the proposed method can make use of the existing terahertz security inspection equipment and other detection devices at the airport in terms of hardware integration, and connect high-performance computing devices (GPU servers or edge computing devices) externally to accelerate the model operation, meeting the low latency requirements of real-time security inspection. In terms of software integration, API interfaces are developed to achieve seamless communication between the detection results and the existing security inspection system. At the same time, the results are converted into compatible formats (such as XML, JSON) through a data conversion module to ensure that the data can be accepted and processed by the existing system. To meet the real-time requirements, operations such as compressing the model and optimizing the algorithm are carried out to reduce the detection time. Lightweight model architectures (such as Tiny-YOLO) are utilized to enhance the detection speed. In terms of hardware, dedicated AI chips (FPGA, ASIC) are adopted to accelerate computing. Code optimization and parallel processing further improve efficiency. Considering compatibility, middleware is developed to ensure seamless integration of the new system with the existing communication network. Standard network protocols (such as TCP/IP) are used to guarantee the security and reliability of data transmission. Integration with the passenger information management system is explored to quickly obtain detection results and respond.

Facing hardware limitations and system compatibility constraints, model quantization and pruning are adopted to reduce the computational load. A lightweight architecture is selected to adapt to the existing hardware. Virtualization or containerization technology (Docker) is utilized to ensure the compatibility of the new system and achieve seamless integration. The integration with the passenger information management system is explored to respond quickly to the detection results.

5 Summary

To solve the problem of low accuracy in current HMDMs, an improved Deep Labv3+ model incorporating DSConv and SE AMs was introduced. At the same time, it was combined with an improved YOLOv5s network model to establish a method for detecting DGs in civil aviation. A performance comparison analysis was performed to assess the research method against other methods. The findings revealed that the proposed method outperformed the others in terms of precision, recall and F1 score. Subsequently, the proposed detection method was analyzed for its application effectiveness, and it was found that the method exhibited high accuracy in detecting three types of DGs. In addition, the proposed detection method has demonstrated superior performance in terms of MSE, RMSE, AP value, CPU usage, and detection time. The above findings reveal that the developed detection method has effectiveness and practical value. This research has achieved certain results in the detection of DGs in terahertz images, but there are also limitations. The generalization ability of the model has not been fully verified in other imaging modes or diverse environments, which limits the evaluation of its wide applicability. The limited size of the dataset may lead to overfitting of the model, affecting its predictive performance on new data. In addition, the research mainly focuses on detection in non-occluded scenarios, and the detection capability under occluded conditions such as human clothing has not been fully evaluated. At present, the model only covers three specific types of DGs. For a wider range of prohibited items, its detection efficiency needs further investigation. Future work will involve validating models in more complex environmental Settings and exploring the integration of multiple imaging techniques to enhance the robustness and adaptability of the models. Meanwhile, expanding and diversifying datasets is also an important direction for enhancing the generalization ability of models and reducing the risk of overfitting.

Fundings

The research is supported by: Zhejiang Guangsha Construction Vocational and Technical University, 2025 New Major Connotation Construction Project – Civil Aviation Safety Technology Management Major, (NO. ZK09-04).

References

[1] Islam A, Othman F, Sakib N, Babu, H. M. H. Prevention of Shoulder-Surfing Attack Using Shifting Condition with the Digraph Substitution Rules. Artificial Intelligence and Applications. 2023, 1(1): 58-68. DOI: 10.47852/bonviewAIA2202289.

[2] Wei Q, Ma S, Tang S, Li, B., Shen, J., Xu, Y., Fan, J. A deep learning-based recognition for dangerous objects imaged in X-ray security inspection device. Journal of X-ray science and technology, 2023, 31(1): 13-26. DOI: 10.3233/xst-221210.

[3] Daud A, Khan M B, Khattak A B, Tanoli, S. A. K., Mustafa, A., Rehman, M., López, O. L. Next-generation security: Detecting suspicious liquids through software defined radio frequency sensing and machine learning. IEEE Sensors Journal, 2024, 24(5): 7140-7152. DOI: 10.1109/JSEN.2024.3351226.

[4] Jayachitra J, K S D, Manisekaran S V, Satti, S. K. An optimal deep learning model for recognition of hidden hazardous weapons in terahertz and millimeter wave images. Earth Science Informatics, 2023, 16(3): 2709-2726. DOI: 10.1007/s12145-023-01056-x.

[5] Fahad M, Zhang T, Khan S U, Albanyan, A., Siddiqui, F., Iqbal, Y., Geng, Y. Optimizing dual energy X-ray image enhancement using a novel hybrid fusion method. Journal of X-Ray Science and Technology, 2024, 32(6): 1553-1570. DOI: 10.3233/XST-240227.

[6] Cheng L, Xiong R, Wu J, Yan, X., Yang, C., Zhang, Y., He, Y. Fast segmentation algorithm of USV accessible area based on attention fast DeepLabV3. IEEE Sensors Journal, 2024, 24(15): 24168-24177. DOI: 10.1109/JSEN.2024.3410403.

[7] Yu K, Tang G, Chen W, Hu, S., Li, Y., Gong, H. MobileNet-YOLO v5s: An improved lightweight method for real-time detection of sugarcane stem nodes in complex natural environments. IEEE Access, 2023, 11(1): 104070-104083. DOI: 10.1109/ACCESS.2023.3317951.

[8] Chen H, Qin Y, Liu X, Wang, H., Zhao, J. An improved DeepLabv3+ lightweight network for remote-sensing image semantic segmentation. Complex & Intelligent Systems, 2024, 10(2): 2839-2849. DOI: 10.1007/s40747-023-01304-z.

[9] Zhao X, Li S, Sun Z, Ma, H., Kong, X., Xu, Y. Detection of building shadows in high-resolution remote sensing images by using improved DeepLabV3+. Remote Sensing Letters, 2025, 16(3): 290-301. DOI: 10.1080/2150704x.2025.2452310.

[10] Gao G, Shuai C, Wang S, Ding, T. Using improved YOLO V5s to recognize tomatoes in a continuous working environment. Signal, Image and Video Processing, 2024, 18(5): 4019-4028. DOI: 10.1007/s11760-024-03010-w.

[11] Xiao Z, Zhu Z, Wei G, X., Liang, S. D., Yang, C. C., Zheng, X., He, F. An improved model based on YOLO v5s for intelligent detection of center

porosity in round bloom. ISIJ International, 2024, 64(1): 76-83. DOI: 10.2355/isijinternational.isijint-2023-341.

[12] Boulkroune A, Hamel S, Zouari F, Boukabou, A., Ibeas, A. Output - Feedback Controller Based Projective Lag - Synchronization of Uncertain Chaotic Systems in the Presence of Input Nonlinearities. Mathematical Problems in Engineering, 2017, 2017(1): 8045803. DOI: 10.1155/2017/8045803.

[13] Boulkroune A, Zouari F, Boubellouta A. Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems. Journal of Vibration and Control, 2025: 10775463251320258. DOI: 10.1177/10775463251320258.

[14] Gao X, Chen C, Gan Y, Liu, Y. Comparison of Extraction accuracy of Sugarcane from different resolution satellite images using Deep lab V3+ Mode. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2024, 48(1): 179-184. DOI: 10.5194/isprs-archives-xlviii-1-2024-179-2024.

[15] Jang J G, Quan C, Lee H D, Kang, U. Falcon: lightweight and accurate convolution based on depthwise separable convolution. Knowledge and Information Systems, 2023, 65(5): 2225-2249. DOI: 10.1007/s10115-022-01818-x.

[16] Zouari F, Saad K B, Benrejeb M. Adaptive backstepping control for a class of uncertain single input single output nonlinear systems. 0th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13). IEEE, 2013: 1-6. DOI: 10.1109/ssd.2013.6564134.

[17] Zouari F, Saad K B, Benrejeb M. Robust neural adaptive control for a class of uncertain nonlinear complex dynamical multivariable systems. International Review on Modelling and Simulations, 2012, 5(5): 2075-2103.

[18] Rigatos G, Abbaszadeh M, Sari B, Siano, P., Cuccurullo, G., Zouari, F. Nonlinear optimal control for a gas compressor driven by an induction motor[J]. Results in Control and Optimization, 2023, 11: 100226. DOI: 10.1016/j.rico.2023.100226.

[19] Merazka L, Zouari F, Boulkroune A. High-gain observer-based adaptive fuzzy control for a class of multivariable nonlinear systems. 6th International Conference on Systems and Control (ICSC). IEEE, 2017: 96-102. DOI: 10.1109/icosc.2017.7958728.

[20] Kuo T C, Cheng T W, Lin C K, Chang, M. C., Cheng, K. Y., Cheng, Y. C. Using DeepLab v3+-based semantic segmentation to evaluate platelet activation. Medical & Biological Engineering & Computing, 2022, 60(6): 1775-1785. DOI: 10.1007/s11517-022-02575-3.

[21] Peng H, Xiang S, Chen M, Li, H., Su, Q. DCN-Deeplabv3+: A novel road segmentation algorithm based on improved Deeplabv3+. IEEE Access, 2024, 12(1): 87397-87406. DOI: 10.1109/access.2024.3416468.

[22] Kim W, Jun S, Kang S, Lee, C. O-Net: Dangerous Goods Detection in Aviation Security Based on U-Net. IEEE Access, 2020, 8: 206289-206302. DOI: 10.1109/access.2020.3037719.

[23] Gallo G, Di Rienzo F, Garzelli F, Ducange, P., Vallati, C. A smart system for personal protective equipment detection in industrial environments based on deep learning at the edge. IEEE Access, 2022, 10: 110862-110878. DOI: 10.1109/ACCESS.2022.3215148.