Engineering and Technology Journal e-ISSN: 2456-3358
Volume 09 Issue 05 May-2024, Page No.- 3872-3880
DOI: 10.47191/etj/v9i05.01, I.F. – 8.227
© 2024, ETJ

Capsule Attention Based Detection of Contraband in X-Ray Images

YAN Zhiming[1,2], LI Xinwei[1,2*], YANG Yi[1,2]
[1]School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China
[2]Henan Key Laboratory of Intelligent Detection and Control of Coal Mine Equipment, Jiaozuo 454000, China

ABSTRACT: Aiming at the problem of low detection accuracy caused by the different postures, different sizes, complex backgrounds and overlapping occlusions of contraband in the security checking process, a Matrix capsule network based on attention mechanism (MCAM) is designed by introducing attention mechanism into the capsule network. Firstly, a multi-feature-extraction (MFE) module is designed to solve the difficulty of detecting contraband with different sizes and complex backgrounds; then the Conv2d with Attention for ConvCap (CACC) is constructed in the convolutional layer of the capsule, and the weight information is computed in the channel dimensions of the feature map to give the contraband regions higher coefficients to enhance the contraband detection ability when the poses are different and the occlusion is severe; finally, a new capsule detection layer is designed by utilizing the pose matrix of the capsule to give full play to the detection ability of the matrix capsule network. The mAPs of the proposed model on SIXray, SIXray10, and SIXray100 are 85.10%, 64.05%, and 53.67%, respectively, which are 42.79%, 19.73%, and 9.41% higher than those of the original network, higher than those of the current mainstream detection networks, and the number of parameters of the network and the amount of computation are also lower.

KEYWORDS: x-ray optics; contraband detection; matrix capsule networks; attention mechanism; feature enhancement

I. INTRODUCTION

In the field of transportation, security screening of passenger-carried baggage is a necessary part of the process in order to protect the safety of passengers. At present, the commonly used security check is based on dual-energy X-ray screening machine and the use of metal detectors for manual security check. Dual-energy X-ray security machine through the penetrating effect of X-rays and different materials of objects on the X-ray absorption strength of different, so that the X-ray image of the contours and colors of the different, so as to determine whether or not the prohibited goods and their categories. Due to the penetrating characteristics of X-rays, 3-dimensional space in the irrelevant objects will be imaged in 2-dimensional space, so that the X-ray image has the characteristics of mutual obscuration. Meanwhile, the contraband in the actual security check has the problems of many categories, different sizes and different postures, which is a great challenge for manual security check and makes the security check inefficient due to the fatigue and nervousness of the screeners [1].

In order to cope with the above problems, a series of investigations have been carried out for contraband detection, and a series of fast security inspection methods based on image processing have been designed, such as artificial security inspection methods based on image processing and image enhancement [2], feature matching methods based on visual bag of words [3], support vector machine (SVM) detection methods [4], image segmentation methods [5], millimeter wave imaging methods [6], foreground-background separation methods [7] and neural network methods [8]. These image processing-based methods have improved the efficiency of security screening to a certain extent, but there are still problems of slow detection speed, low detection accuracy and the need for personnel involvement. With the development of deep learning, convolutional neural networks have gradually become the most dominant method for contraband detection.S. Akcay et al [9] first introduced deep learning to X-ray contraband detection, and the detection network was migrated from AlexNet network, which has qualitatively improved the detection speed compared to traditional manual and semi-automatic detection methods. The contraband detection network is divided into two types: using the classification network combined with heat map localization to achieve detection and using the detection network combined with calibration frame to achieve detection.

Using a detection network with a calibrated frame, its detection accuracy is high and can accurately locate the position of contraband. Accordingly, Cheng Lang et al [10] constructed an X-ray image contraband detection algorithm based on neural network architecture search, which effectively improved the detection accuracy of contraband in complex backgrounds; Haiqun Wang et al [11] constructed a small-target detection head for the yolov8 network, adaptively adjusted the weights between the different detection layers using the ASFF module, and introduced the EMA attention between the neck and backbone networks through the yolov8 network mechanism and other operations, the detection accuracy of contraband is improved; Zhang Liang et al [12] by introducing Transformer module to yolov5s network, designing sensory field adaptive fusion module in combination with hollow convolution and optimizing the loss function and other operations, the detection accuracy of the model for contraband with complex background and large change of target scale is improved; Yuan Jinhao et al [13] in order to solve the problem of mutual occlusion, target proximity and small target contraband difficult to recognize, based on the yolovx network, through the construction of the attention module, so that the performance of the model has been improved; You Xi et al. [14] based on the Cascade RCNN model, through the introduction of deformable convolution, put forward a spatial adaptive attention module, which effectively reduces the problem of misdetection and omission of the contraband; Dong Yisan et al. [15] through the yolovx network and optimization of the loss function and other operations, so that the model for the complex background and the target scale changes in the detection accuracy of the model has been improved; However, the production of the detection dataset based on calibrated frames is cumbersome and the network detection speed is slow.

Detection using classification networks combined with heat maps to achieve localization is characterized by fast speed and simpler dataset production. Accordingly, Caijing Miao et al [16] proposed the Resnet-CHR (Class Balanced Hierarchical Refinement) method based on the residual network Resnet [17] and improved it, using a hierarchical refinement strategy, which resulted in the improvement of the detection accuracy of contraband when the data is not balanced.

Although the convolutional network improves the detection efficiency of contraband to a certain extent, the convolutional neural network can not actively recognize the posture of contraband, and needs a large amount of data for training to improve the generalization ability of the model. Therefore, Miao Shuo et al [18] used a vector capsule network that can actively recognize the pose of an object and based on this, fused the semantic information of high and low layers by adding a feature enhancement module (dilated convolution multi-scale feature fusion (DMF)), and the added feature screening module (squeeze-and excitation block (SE) to screen the resulting features, which makes the vector capsule network have excellent contraband detection capability compared to the commonly used convolutional network, and combine with heat map localization to realize contraband detection. However, the vector capsule network has a large amount of computation, which makes its detection speed slower.

Therefore, in order to be able to quickly and accurately detect contraband with different postures, different sizes, complex backgrounds and mutual occlusion, a capsule attention network is constructed, based on matrix capsule network [19], by constructing the feature extraction and enhancement structure, the capsule attention mechanism, and the new capsule detection head, so that the network's performance for the detection of contraband can be improved, and it is easy to use easily made categorical dataset, and after the network learns autonomously using the heat map approach to achieve localization, which is fast and convenient compared to the calibrated box approach.

II. CAPSULE ATTENTION MODEL

I. Overall structure of the model

The capsule attention network architecture is shown in Fig. 1 and consists of three main parts, namely, the multi-feature extraction module MFE, the capsule attention module CACC, and the capsule detection layer module convclass.

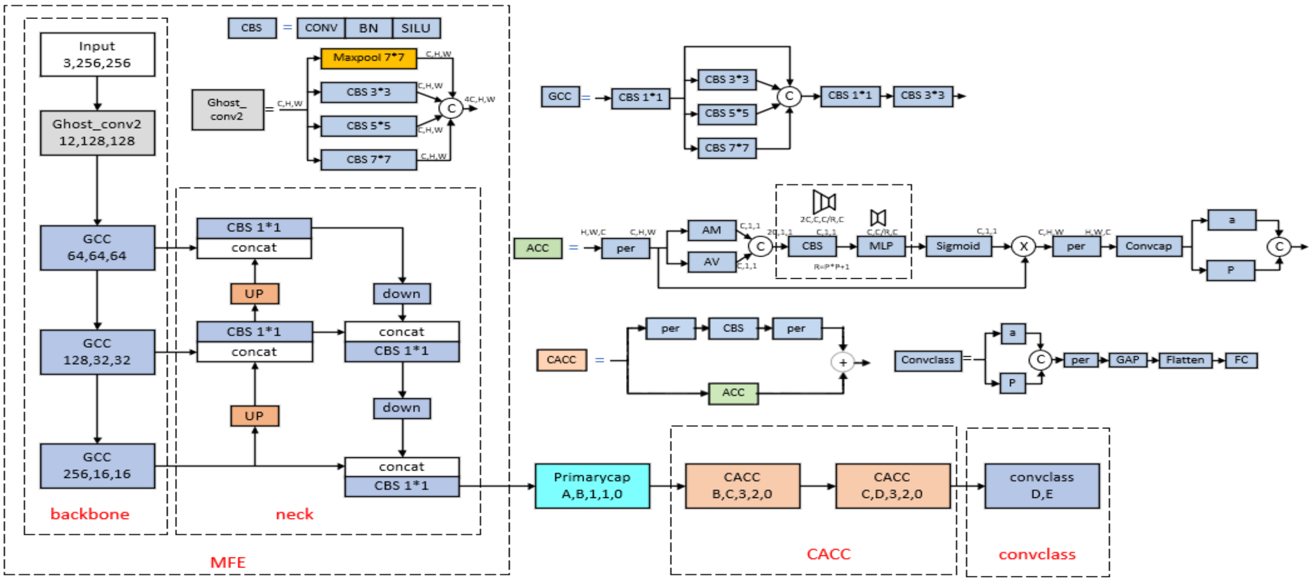


Figure 1 Capsule Attention Modeling Framework

(1) Multi-feature extraction module part, in order to solve the problem of different sizes of contraband and complex background, the multi-branch feature extraction and enhancement structure is designed by drawing on the idea of multi-scale features.

(2) The capsule attention module part, in order to make the network shift the focus of processing to contraband, the attention mechanism is introduced into the convolutional capsule layer, which gives a larger weight to the contraband region and a smaller weight to the other regions, so that the capsule focuses on the contraband region, thus solving the problem of mutual occlusion of contraband.

(3) Capsule detection layer module, in order to use the capsule information more comprehensively, the attitude matrix and activation probability are repackaged, the weight information of the channel dimensions is obtained by using global average pooling, and finally the fully-connected layer is used for detection in order to improve the network detection capability.

II. Designed network modules

Feature Extraction and Enhancement Module

The effect of the underlying feature extraction determines the model's "understanding" and "decision-making" of the image. In order to adapt to the complex background and different sizes of contraband, we need to consider different sizes of features. Borrowing the idea of shadow convolution [20] and improving it, a maximum pooling module is added to process the background information, and its structure is shown in Fig. 2. Four branches are used to jointly extract image information from the input image.

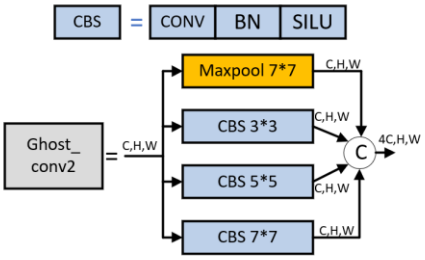


Figure 2 Improved shadow convolution structure

The shadow convolution Ghost_conv is improved by increasing the convolution for dimensionality reduction and improving the nonlinearity of the network, and increasing the convolution to realize feature extraction and reduce the resolution of the feature map to obtain the GCC (Ghost Conv with Conv2d, GCC) feature extraction module, whose structure is shown in Figure 3.

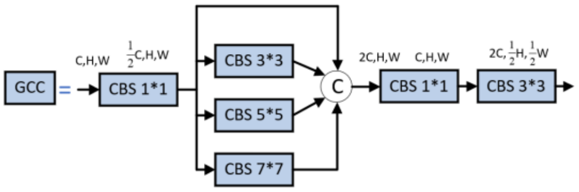


Figure 3 GCC module structure

In the backbone part three GCC modules are used for feature extraction and each module outputs a feature map of size: , , , and the outputs of the three parts are fed into the upsampling network and downsampling network for feature enhancement.

The FPN+PAN structure is used for feature enhancement in the neck part. Since the semantic information of the feature map is gradually enriched in the down-sampling process, but the resolution is gradually reduced, this structure is used to enhance the semantic information and the feature map information.

Capsule Attention Module

Since the matrix capsule network is embedded in the EM algorithm, the error back propagation nests the forward iterations during training, making its computational volume larger. The contradiction between the number of model parameters, the amount of computation and the detection accuracy is difficult to solve.

Drawing on the idea of attention to enhance the capsule's attention to the target contraband, the contraband detection performance is improved while keeping the model unchanged, so as to balance the number of parameters, computation and detection accuracy of the capsule network. Accordingly, a capsule attention (Attention with ConvCap, ACC) module is designed, and its structure is shown in Fig. 4.

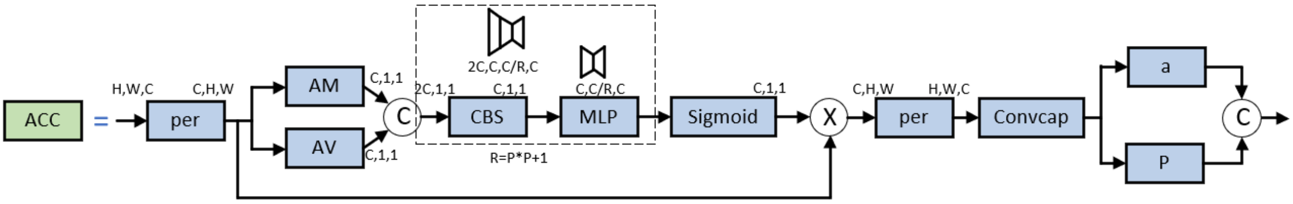


Figure 4 ACC module structure

Let the input feature map be f_{ai}, and the output feature map be f_{ao}, then the operation process of the capsule attention module can be expressed as Equation 1 and Equation 2.

$$f_a = \sigma(\tau(c_1(\varepsilon(a_m(f_{ai}), a_v(f_{ai}))))) \tag{1}$$

$$f_{ao} = \theta(f_a \times f_{ai}) \tag{2}$$

Which a_m represents adaptive maximum pooling operation, a_v represents adaptive average pooling operation; ε represents the connection operation in the channel dimension; c_1 represents the convolution operation; τ represents the multilayer perceptron operation, in which the multiplier of the channel compression is $P \times P + 1$, that is, the capsule dimension is compressed to 1, and then restored to the capsule dimension, so that the capsule filtered out the weight information of the prohibited items through learning; σ represents the use of the Sigmoid function nonlinear activation to get the screening weights f_a in the channel dimension; multiply the screening weights with the input feature map to get the feature map after the excitation; θ denotes the use of capsule convolution for the operation.

The designed capsule attention module makes the capsule screen out the contraband information in the picture, which improves the attention to the contraband, and improves its feature extraction ability for the contraband under occlusion.

The above operation makes the matrix capsule network improve the ability of detecting contraband under occlusion with a smaller model, but the detection speed is still slow due to the loop iteration of the EM algorithm at runtime. Considering that the convolutional layer is much faster than the convolutional capsule layer in the case of processing the same features, the capsule attention module is weight-shared with the convolutional layer to obtain the CACC module, which improves the training speed of the matrix capsule, and at the same time, improves its generalization ability, and the structure is shown in Fig. 5.

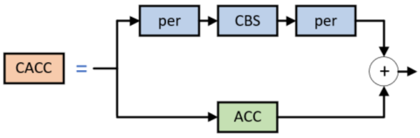


Figure 5 CACC module structure

The CACC module integrates the features of the convolutional layer and the capsule layer, using the 3×3 convolutional layer in the capsule layer to extract the features, and using the capsule attention module in the capsule layer for deep feature extraction, after which the two branches use the matrix summation to share the feature information, and the ablation experiments through 2.4 verified that this structure can improve the speed of detection of the matrix capsule network. Let the input feature map be f_{ai}, and the output feature map be f_{co}, then the operation process of CACC is shown in Equation 3.

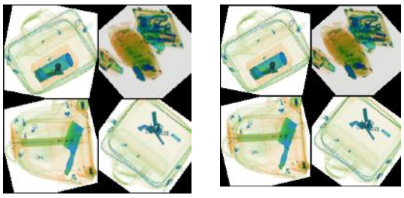
$$f_{co} = c_3(f_{ai}) + f_{ao}(f_{ai}) \qquad (3)$$

where c_3 denotes the 3×3 convolution operation performed and f_{ao} denotes the ACC operation encapsulated above.

Capsule Detection Layer Module

After using the above feature extraction, feature enhancement and capsule attention module, the matrix capsule network obtains the deeper and more abstract contraband feature information, and uses the 1×1 convolutional capsule layer to obtain two parts of information, i.e., the attitude matrix P and the activation probability a, and intercepts the activation probability a to complete the final contraband detection task.

However, through experiments, we know that the detection accuracy obtained in this way is very low. Considering that in the capsule operation project, the attitude matrix and activation probability are being forwarded through the EM algorithm, it is envisioned that the attitude matrix contains the relevant information of contraband, so the image shown in Fig. 6 is obtained by decoding the attitude matrix P.



Groundtruth Reconstruction

Figure 6 Image obtained by decoding the matrix

It can be seen that the gesture matrix reconstructs the input image through decoding, which contains almost all the feature information of contraband, so it is unreasonable to use only the activation probability a as the final detection probability for the detection of complex images.

Therefore, the capsule detection layer convclass (class with conv) is redesigned to integrate the information of the attitude matrix and activation probability, which effectively improves the detection ability of matrix capsule network for complex contraband images, and its structure is shown in Fig. 7.

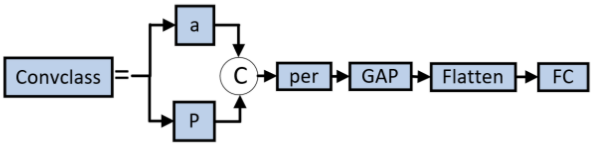


Figure 7 Convolutional classification head structure

The pose matrix P and activation probability a are encapsulated and spliced into a capsule in the channel dimension, then the dimensional transformation is performed, $H\times W\times C$ transformed to $C\times H\times W$, and the Global Average Pooling (GAP) is used to obtain the feature vector (C,1,1) of the channel dimension, which is then flattened using the Flatten function and then sent to the fully connected layer for detection.

III. EXPERIMENTIAL RESULTS AND ANALYSIS

I. Experimental environment and dataset

The experiments are conducted under Ubuntu 20.04 LTS with Intel@ 〖Core〗 ^TM i7-8700K@3.70GHz*12 CPU, 32GB RAM, 4090 GPU, 24GB video memory, Cuda version 12.0, and Pytorch 1.8 network framework, with Epoch set to 50 and batchsize set to 32. The network parameters are tuned using Stochastic gradient descent (SGD). Epoch is set to 50, batch size is set to 32, and Stochastic gradient descent (SGD) is used to adjust the network parameters. The initial learning rate is 0.01, and the learning rate is updated every 20 epochs by multiplying the learning rate by 0.1. The image size is fixed at 256, and random flipping and random cropping are used to improve the generalization of the model during training. To balance the number of parameters and computing speed, the model parameters are set as (A,B,C,D=64,8,16,16).

The dataset used for the experiments is SIXray, which is the X-ray images collected by Miao et al [16] in the field at the security check place, with a total of 1059231 images, some of which are shown in Fig. 8. Among them, there are 8929 images of positive samples containing contraband, including 6 types of contraband, such as knives, guns, wrenches, scissors, pliers and hammers (the number of contraband hammers is small and the dataset is not public, so it will not be used), and the proportion of positive samples to the total number of samples is 0.84%, which can truly reflect the application scenarios of the security check. The dataset is divided into a training set and a test set according to the ratio of 8:2. In order to test the detection performance of the model when the data are unbalanced, the SIXray10 and SIXray100 datasets are recreated, in which there are 100 positive samples and 1,000 and 10,000 negative samples, respectively.

Figure 8 Sample contraband dataset

II. Evaluation indicators

(Mean AP for classifications, *mAP*) is used as an evaluation metric for the model. Multi-tag detection and using the top1 criterion, its calculation formula is shown in Equation 4 and Equation 5:

$$AP = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{4}$$

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{n} \tag{5}$$

Where $\boldsymbol{N_{TP}}$ denotes the number of positive samples actually detected as positive samples, $\boldsymbol{N_{FP}}$ denotes the number of positive samples actually detected as negative samples, n denotes the number of categories, and AP denotes the detection accuracy of a single category.

III. Comparison experiment

In order to compare the detection performance of MCAM models, the current state-of-the-art detection models InceptionV3 [21], MobileNetV3 [22], EfficientNetV2 [23], and CHR-resnet101 [16] networks are selected for comparison. Experiments are conducted on the SIXray, SIXray10, and SIXray100 datasets, respectively, and the detection accuracy of a single category , the average detection accuracy, the number of parameters, the computational effort, the number of images processed per second (Frames Per Second), and the graphics memory footprint (GPU Memory, when batchsize is set to 32) are used as evaluation criteria.

SIXray dataset detection results

The results of the comparison experiments obtained on the SIXray positive sample dataset are shown in Table 1.

Tab. 1 Results of SIXray detection

	AP/%									
	Gun	Knife	Wrench	Pliers	Scissors	mAP/%	Params/M	GFLOPs/G	FPS/n	GM/G
InceptionV3	97.60	90.13	68.17	88.30	79.21	84.68	21.80	11.49	480	8.8
MobileNetV3	96.87	90.22	68.99	86.51	82.04	84.92	478.50	3.90	576	5.3
EfficientNetV2	98.04	88.52	68.91	88.56	77.86	84.38	40.75	6.96	352	29.6
CHR-R	98.94	89.01	62.48	89.21	81.52	84.23	51.96	19.43	512	11.9
MC	75.20	43.39	14.03	58.30	20.61	42.31	1.70	1.60	569	4.7
MC+C	93.43	79.24	45.28	80.74	67.98	73.34	1.68	1.60	581	4.6
MCAM(ours)	**97.81**	**89.45**	**67.38**	**88.99**	**81.88**	**85.10**	**3.22**	**3.10**	**618**	**4.6**

Note: MC is Matrix-Capsules; MC+C denotes the use of a new convolutional detection layer; CHR-R is CHR-Resnet101; all networks are not pre-trained for fair comparison.

As can be seen from Table 1, the *mAP* oforiginal matrix capsule network is 42.31%, and after the introduction of the capsule detection layer into the network, it reaches 73.34%, which is an improvement of 31.03%, indicating that the capsule detection layer is very obvious. After using the multi-feature extraction module, the attention module and the capsule detection layer module, the *mAP* of MCAM reaches 85.10%, which shows that the above design makes the matrix capsule network for contraband detection accuracy has been greatly improved. Compared with other networks, the overall performance of MCAM has obvious advantages, 0.18% higher than MobileNetV3, 0.72% higher than EfficientNetV2, 0.87%

higher than CHR-resnet101, and 0.42% higher than InceptionV3.

In terms of space complexity, MCAM has a parameter count of 3.22M, a computational amount of 3.10M, and 4.6G of video memory for training.Compared with the other four networks, the parameter count and computational amount are lower, with a parameter count that is 48.74M lower than that of CHR-R, 37.53M lower than that of EfficientNetV2, 475.28M lower than that of MobileNetV3 and 18.58M lower than InceptionV3; the computation amount is about 1/6 of CHR-R, 1/13 of EfficientNetV2, 1/4 of InceptionV3, and 0.8G lower than MobileNetV3. As for the memory usage during training, MCAM is 4.6G, much lower than 11.9G of CHR-R, 29.6G of EfficientNetV2, 5.3G of MobileNetV3 and 8.8G of InceptionV3.

In terms of time complexity, the detection speed of MCAM is 618 frames per second, which also has a significant advantage over the other four networks, 106 frames higher than CHR-R, 266 frames higher than EfficientNetV2, 42 frames higher than MobileNetV3, and 138 frames higher than InceptionV3.The spatial and time complexity of MCAM in terms of aspects provides the basis for local deployment and real-time detection of the model.

Positive and negative sample imbalance experiments

Table 2 shows the detection effect of the above model on the SIXray10 and SIXray100 datasets, which demonstrates the performance of the model when the positive and negative samples are unbalanced, simulating the detection effect of contraband in the actual security check.

From the experimental results in Table 2, it can be seen that when there are only 100 positive samples, the detection results of the model show a significant decrease compared to the original, which is due to the fact that the positive sample dataset has fewer samples, and it is difficult for the model to learn the features of the contraband in different poses to achieve generalizability. In the detection results on SIXray10, the mAP of MCAM is 64.05%, which is higher than the detection performance of all the comparison models, and there is nearly 20 percentage points of performance improvement compared to the original matrix capsule network.

In the detection results on SIXray100, where the proportion of positive and negative samples is more obvious, MCAM has a significant advantage over the other four networks, with 53.67%, which is a 9.41% improvement in detection accuracy compared to the original matrix capsule network, and a 10.03, 9.87, 31.59, and 0.73 percentage point improvement in detection accuracy compared to the other four networks, respectively. This highlights the fact that the MCAM model improves the detection ability when the contraband data is unbalanced by actively recognizing the contraband's pose information and thus improves the detection ability when the contraband data is unbalanced.

Tab. 2 SIXray10 and SIXray100 detection results

	Gun/%		Knife/%		Wrench/%		Pliers/%		Scissors/%		*mAP*/%	
	10	100	10	100	10	100	10	100	10	100	10	100
InceptionV3	95.00	75.05	74.11	37.70	62.50	69.17	9.64	20.48	44.13	15.80	57.07	43.64
MobileNetV3	87.50	100	85.38	50.16	41.31	32.68	21.86	10.27	79.80	25.73	63.17	43.77
EfficientNetV2	100	10.12	97.05	10.22	33.79	35.11	12.64	18.37	47.44	36.43	58.18	22.05
CHR-R	100	95	100	90	69.38	51.47	31.99	15.47	14.39	12.77	63.15	52.94
MC+C	95	95	89.61	95	10.38	16.36	9.51	6.29	17.1	8.63	44.32	44.26
MCAM(ours)	**95**	**100**	**90**	**95**	**46.40**	**22.21**	**21.84**	**25.40**	**67.48**	**25.76**	**64.05**	**53.67**

Note: MC+C denotes a matrix capsule network using a new convolutional detection layer; CHR-R is CHR-Resnet101, 10 and 100 denote in SIXray10 and SIXray100 datasets, respectively.

IV. ablation experiment

Firstly, in order to verify that the double branching form of convolutional module and capsule attention module can speed up the model, experiments are carried out in the CACC module part of the MCAM model, which is divided into three cases, firstly, the ACC module, secondly, the convolutional module Conv2d, and thirdly, the combination of the ACC module and the convolutional module. The experimental results are shown in Table 3.

Tab. 3 Experimental results of CACC module

module	*mAP*/%	*FPS*/n
ACC	82.75	540
Conv2d	77.96	630
ACC+Conv2d	85.10(+**2.35**)	618(+**78**)

From the data in Table 3, it can be seen that the use of the combination of the ACC module and the convolution module makes the detection accuracy and detection speed of the network improved to different degrees, which verifies the effectiveness of the combination of the convolution module and the ACC module for improving the detection accuracy and speed.

In order to verify the effectiveness of the designed model to improve the accuracy of contraband detection, an ablation experiment was designed, and the results are shown in Table 4.

Tab. 4 Results of ablation experiments

	convclass	MFE	CACC	***mAP*/%**
Baseline+				42.31
	√			73.34(**+31.03**)
	√	√		80.12(**+6.78**)
	√	√	√	85.10(**+4.98**)

From the results in Table 4, it can be seen that the original network *mAP* is 42.31%; after the introduction of the convclass detection layer, it is enhanced by 31.03% to 73.34%; based on this, the MFE structure is introduced, which is enhanced by 6.78%; and finally, using all the modules, it is achieved by 85.10%, from which it can be concluded that all the three designed modules can effectively enhance the accuracy of the detection of contraband.

V. Actual test results

Fig. 9 shows the actual detection of the network for X-ray contraband, where detection is performed for three types of images, i.e., images of different sizes, mutual occlusion, and complex background. Heat map [24] localization is used to achieve contraband detection to visualize the detection performance of the network.

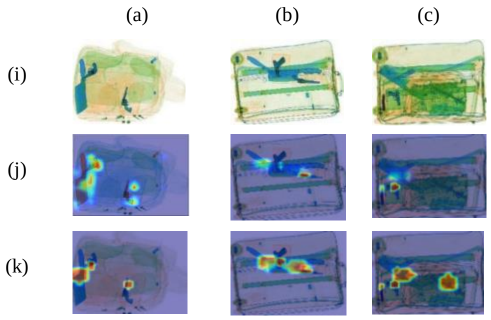


Figure 9 Results of ablation experiments. (a) varying sizes; (b) obscuring each other; (c) complex backgrounds;(i) original image; (j) grad-cam of matrix capsule detection; (k) grad-cam of MCAM detection

From the experimental results in Fig. 9, it can be seen that the original matrix capsule is generally effective in detecting knives and guns with poor attention in the face of contraband images of different sizes; the original network misses a pistol in the face of contraband images with mutual occlusion; and the original network misses a knife in contraband images with complex backgrounds; the MCAM model achieves a better heat map localization in the above contraband images with different sizes, mutual occlusion and complex backgrounds, which reflects that the MCAM model improves the detection capability for contraband in the above cases. The MCAM model achieves better heat map localization in the above contraband images with different sizes, mutual occlusion and complex backgrounds, which reflects that the MCAM model improves the detection capability of contraband in the above cases.

IV. CONCLUSIONS

A new matrix capsule network model MCAM is proposed, which uses multi-feature extraction module and feature enhancement structure to solve the problem of different sizes of contraband and complex background; the attention mechanism is introduced into the matrix capsule to realize the effective extraction of features of contraband and to solve the problem of mutual occlusion of contraband; and finally, the information of the capsule is fused to realize the detection to improve the detection ability of the model for contraband effectively. The model's ability to detect contraband is effectively improved. A large number of experiments show that the MCAM model's contraband detection performance is significantly improved compared with the original network. Although the detection effect of MCAM model on SIXray has a better performance, the detection accuracy still has a large room for improvement, and the lightweight and engineering deployment of the model is also a future research direction.

REFERENCES

1. W Y Q, R J J, YU Q. Impact Analysis of X-ray Machine Conveyor Belt Speed on Airport Security Missing Rate[J].China Transportation Review, 2017,39(05):55-59. (in Chinese).
2. YANG X G, YANG L R.. A Method on X-ray Security Image Enhancement[J]. CT Theory and Applications, 2012, 21(4): 705-712.(in Chinese)
3. Sterchi Y , Hattenschwiler N , Michel S ,et al.[IEEE 2017 International Carnahan Conference on Security Technology (ICCST) - Madrid (2017.10.23-2017.10.26)] 2017 International Carnahan Conference on Sec

urity Technology (ICCST) - Relevance of visual inspection strategy and knowledge about everyday objects for X-ray baggage screening[J]. 2017:1-6.DOI:10.1109/ccst.2017.8167812..

4. YAN W，JING H. Object detection in X-ray images based on object candidate extraction and support vector machine［C］// Ninth International Conference on Natural Computation. May 19，2014，Shenyang，China. IEEE，2014：173-177.
5. Heitz G , Chechik G .Object separation in X-ray image sets[J].IEEE, 2010.DOI:10.1109/CVPR.2010.5539887.
6. Xiao Y , Wang S , Wang Z ,et al.Coordinated-security based on probabilistic shaping and encryption in MMW-RoF system.[J].Optics letters, 2023, 48 11:, 2989-2992. DOI:10.1364/ol.493644.
7. Mery D .Automated detection in complex objects using a tracking algorithm in multiple X-ray views[C]//Computer Vision & Pattern Recognition Workshops.IEEE, 2011.DOI:10.1109/CVPRW.2011.5981715.
8. Hassan T , Akcay S , Bennamoun M ,et al.Trainable Structure Tensors for Autonomous Baggage Threat Detection Under Extreme Occlusion[C]//2020.DOI:10.1007/978-3-030-69544-6_16.
9. Akcay S , Kundegorski M E , Willcocks C G ,et al.Using Deep Convolutional Neural Network Architectures for Object Classification and Detection Within X-Ray Baggage Security Imagery[J].IEEE Transactions on Information Forensics and Security, 2018:2203-2215.DOI:10.1109/TIFS.2018.2812196.
10. CHENG L,JING Ch,CHEN W P. Algorithm for contraband detectedn in X-ray images based on neural network architecture search[J]. Science Technology and Engineering,2024,24(02):665-675.(in Chinese).
11. WANG H Q,WEI P X. X-ray image contraband detection based on improved YOLOv8[J/OL]. Radio Engineering,1-9[2024-02-27].(in Chinese).
12. ZHANG L,XUE ZH CH. X-ray contraband detection based on adaptive multiscale feature fusion[J/OL]. Signal Processing,1-18[2024-02-27].(in Chinese).
13. YUAN J H, ZHANG N F, RUAN J SH, GAO X D. Detection of prohibited items in X-ray images based on modified YOLOX algorithm[J]. *LASER TECHNOLOGY*, 2023, 47(4): 547-552.(in Chinese).
14. YOU X, HOU J, REN D SH, YANG P X, DU M SH. Adaptive Security Check Prohibited Items Detection Method with Fused Spatial Attention[J]. Computer Engineering and Applications, 2023, 59(21): 176-186.(in Chinese).
15. DONG Y SH, LI ZH X,GUO J Y . Improved YOLOv5 Model for X-Ray Prohibited Item Detection[J]. Laser & Optoelectronics Progress, 2023, 60(4): 0415005.(in Chinese).
16. Miao C , Xie L , Wan F ,et al.SIXray: A Large-Scale Security Inspection X-Ray Benchmark for Prohibited Item Discovery in Overlapping Images[J].IEEE, 2019.DOI:10.1109/CVPR.2019.00222.
17. He K , Zhang X , Ren S ,et al.Deep Residual Learning for Image Recognition[J].IEEE, 2016.DOI:10.1109/CVPR.2016.90.
18. M SH, LI X W, YANG Y et al. X-ray image contraband detection based on improved capsule network[J]. Journal of Henan University of Science and Technology(Natural Science Edition),2023,42(03):129-136.DOI:10.16186/j.cnki.1673-9787.2021080065.(in Chinese).
19. Hinton G E , Sabour S , Frosst N .Matrix capsules with EM routing[C]//2018.
20. Wang T , Zhang S .DSC-Ghost-Conv: A compact convolution module for building efficient neural network architectures[J].Multimedia Tools and Applications, 2023.DOI:10.1007/s11042-023-16120-3.
21. SZEGEDY C，VANHOUCKE V，IOFFE S，et al. Rethinking the inception architecture for computer vision［C］// 2016 IEEE Conference on Computer Vision and Pattern Recognition，Jun. 27-30，2016，Las Vegas，NV，USA.IEEE，2016：2818-2826.
22. Howard A , Sandler M , Chu G ,et al. Searching for MobileNetV3[J]. 2019.DOI:10.48550/arXiv.1905.02244.
23. Tan M , Le Q V .EfficientNetV2: Smaller Models and Faster Training[J]. 2021.DOI:10.48550/arXiv.2104.00298.
24. SELVARAJU R R，COGSWELL M，DAS A，et al. Grad-CAM：Visual explanations from deep networks via gradient-based localization［C］// International Conference on Computer Vision，Oct.22-29，2017，Venice，Italy. IEEE，2017：618-626.