文章编号：1001－9014(2019)01－0032－07

DOI:10.11972/j.issn.1001－9014.2019.01.006

A low-complexity method for concealed object detection in active millimeter-wave images

WANG Chong-Jian[1*], SUN Xiao-Wei[2], YANG Ke-Hu[1]

(1. State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China;
2. Key Laboratory of Terahertz Technology, Shanghai Institute of Microsystem and Information Technology, CAS, Shanghai 200050, China)

Abstract: Active millimeter wave imaging (AMWI) is an efficient way to detect dangerous objects concealed under clothes. However, because the images acquired by AMWI are often obscure and some of concealed objects are small in size, the automatic detection and localization of the objects remain as a challenging problem. Yao[1] first employed convolutional neural networks(CNNs) and used a dense sliding window method to detect concealed objects. In this paper, the author presents two improvements over Yao's work: 1) Using contextual information to suppress interference and improve detection probability; 2) Using a two-step search method instead of exhaustive search to reduce the computational complexity. To reduce the computational complexity, the author first uses a CNN in vertical direction to filter the interference and obtain the vertical position of the concealed object, then uses another CNN to determine the horizontal position of the concealed object. To make use of big window containing contextual information, the author uses IoG (intersection-over-ground-truth) instead of IoU (Intersection-over-Union) to define positive and negative samples in training and testing process. Experimental results show that the proposed method will make the length of computational time reduced to about 30% of that of the exhaustive search while achieving better detection performance.

Key words: active millimeter-wave image, concealed object detection, CNN, contextual information
PACS: 84.40.Xb

一种用于主动式毫米波图像的低复杂度隐匿物品检测方法

王崇剑[1*], 孙晓玮[2], 杨克虎[1]

(1. 西安电子科技大学 综合业务网理论与关键技术国家重点实验室,陕西 西安 710071;
2. 中国科学院上海微系统与信息技术研究所 中科院太赫兹固态技术重点实验室,上海 200050)

摘要:主动式毫米波成像(AMWI)技术是检测隐藏在衣服下的危险物体的有效方法.但AMWI获取的图像通常很模糊,而且一些隐匿物体的尺寸较小,因此隐匿物品的自动检测和定位仍然是一个具有挑战性的问题.姚家雄等[1]首先使用卷积神经网络(CNNs)结合穷举滑动窗口方法来检测隐藏物体.做了两点改进:(1)使用上下文(背景)信息抑制干扰,(2)使用两步搜索方法代替穷举搜索来降低计算复杂度.首先在垂直方向上使用一个CNN来过滤干扰,得到隐藏物体的垂直位置,然后用另一个CNN来确定水平位置.为了充分利用上下文信息,使用IoG(交集和真值的比)代替IoU(交并比)来定义训练和测试过程中的正负样本.实验结果表明,该方法将计算时间减小到约30%,同时实现更好的检测性能.

关 键 词:主动式毫米波图像;隐匿物品检测;卷积神经网络;上下文信息

中图分类号:TP751　　**文献标识码**:A

Received date: 2018-04-18, **revised date**: 2018-07-05　　**收稿日期**:2018-04-18, **修回日期**:2018-07-05
Foundation items: Supported by National Natural Science Foundation of China (61731021)
Biography: WANG Chong-Jian (1981-), male, Xi'an, China, PhD. candidate. Research fields include machine learning and object detection. E-mail: xd_wangcj@126.com
* **Corresponding author**: E-mail: xd_wangcj@126.com

Introduction

It is well known that millimeter electromagnetic wave is able to penetrate clothing and does not cause harm to human body with low power. Recently, millimeter-wave imaging radar has been widely used in human security check areas[2-3]. However, due to the low signal to noise ratio (SNR) and low contrast of the AMWI images, as shown in Fig. 1, the detection and localization of concealed objects in those images still remain as a challenging problem.

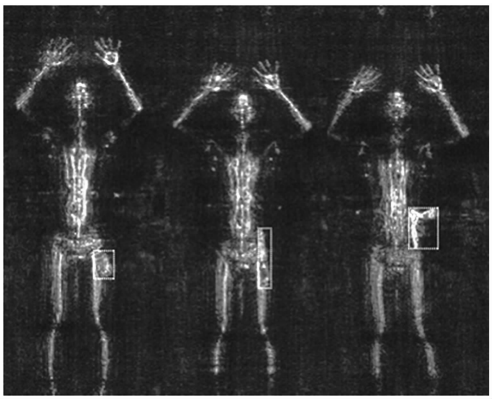

Fig. 1　AMWI images of human body with concealed objects located in the mentioned box
图 1　AMWI 人体图像,方框内为隐匿物品

Object detection is a fundamental and challenging problem in the field of computer vision, where great progresses have been made in the field of object detection by combining machine learning with searching techniques, e. g. , the works in Ref. [4] and Ref. [5]. In Ref. [6-8], new detection methods are proposed to search for objects and the CNNs are used to classify the objects in the region of interest.

There are some works on the concealed object detection for AMWI images[1], where the exhaustive search with dense sliding windows is used. Similar to the work in Ref. [5], the authors use the CNN to multiple locations and accumulate the evidence at each location in an image. Here, however, two problems would arise: 1) interference will be mistakenly considered as concealed objects when sliding a window over an image; 2) the densely window sliding method makes computational complexity significantly increased.

Because visual context plays an important role in visual perception of object, exploiting contextual information[9-11] in images to improve object detection performance become an increasing interest. In Ref. [9-11], Ref. [10] employs contextual information outside the region of interest using spatial recurrent neural networks and shows improvements on small objects detection and Ref. [11] proposes to detect objects in a coarse to fine manner and give candidate region that may contain objects in the coarse step, which can be regarded as another way of using contextual information. Inspired by those ideas, we consider using contextual information in concealed object detection.

In this paper, we propose a low-complexity method for concealed object detection by two steps. First, we use a big window with the width same as that of the AMWI image and the height comparable with the object size, then we slide the window pixel by pixel from top to bottom, the image within each sliding window is sent to a CNN to detect the concealed object. The merit of the big window is that the context of the object is preserved. The detection results are merged to give the region of interest (RoI). Second, we slide another window from left to right over the RoI, where the concealed object is detected and localized according to the image in the window by another CNN. This implies that the exhaustive search and detection over an entire image[1] are no longer required.

The paper is organized as follows. In Section 1, we give the procedures of our method and the associated CNNs. In Section 2, the post process after the CNN is given in detail. In Section 3, we show the experimental results. Section 4 concludes the paper.

1　Detection algorithm and the associated CNNs

1.1　The detection algorithm

In fact, a concealed object could be located at any place in an AMWI image. To detect the object, it is natural to search over the entire image without *a priori* information. However, as mentioned in Ref. [12], this makes the length of the computational time significantly increased. In order to reduce the computational complexity, it would be better to design a method to output a set of proposal regions which are likely to contain the concealed objects for a given AMWI image.

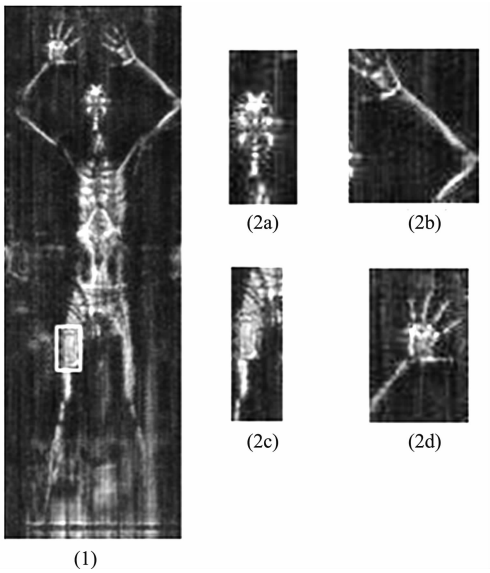

Fig. 2　An example of AMWI image and part of its selective search results. (1) is the AMWI image (ground-truth is in the box). The selective search results are: (2a) head, (2b) left arm, (2c) right leg, (2d) right hand
图 2　AMWI 图像和对应的选择性搜索结果. (1)为 AMWI 图像(方框内为隐匿物品),选择性搜索结果为,(2a)头部,(2b)左胳膊,(2c) 右腿,(2d) 右手

Unfortunately, most of commonly used region proposal methods do not work well for detection of small objects[13] in optical images. When such methods are used to AMWI images, similar phenomenon would happen. For example, for an AMWI image, part of the search results by selective search[12] are shown in Fig. 2, the result closest to the ground-truth is shown in 2(c), where the largest IoU is only 0.09.

In order to reduce the complexity, we here reconsider the sliding-window method by elaborately choosing the window size. By observing the objects in AMWI images, we find that the size of a concealed object is within a certain range. This will help us to choose a suitable window size for the sliding window. For convenience, we define the height and width range of a concealed object as H_R and W_R, respectively. In order to reduce miss detection probability, we choose a window ensuring that it can contain the largest concealed object both in vertical and horizontal directions. As shown in Fig. 3, our method consists of two steps:

Step 1: Set the width of the window to the width of an AMWI image and the height of the window to H_R, sliding the window pixel by pixel from the top to the bottom of the AMWI image. The image within each sliding window is sent to a CCN, called V_CNN, which outputs a series of detection results. The detection results are merged to find the region of interest for further detection process in the next step.

Step 2: Define another window with height H_R and width W_R, sliding this window from left to right over the region of interest found in Step 1. The image within each sliding window is sent to another CCN, called H_CNN. The detection results are merged and analyzed to obtain the final localization result.

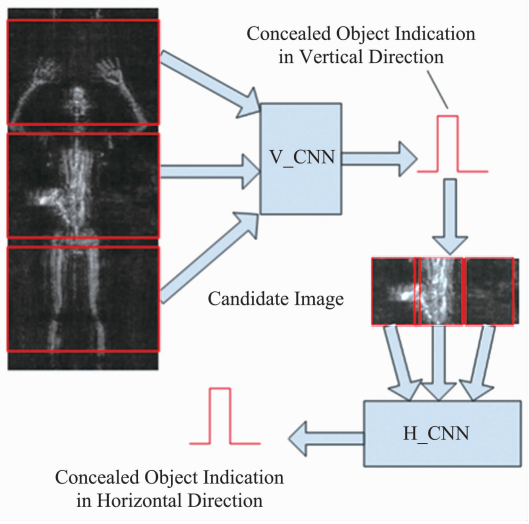

Fig. 3 Two-step one-dimensional search and detection process
图3 一维两步法目标搜索和检测流程

1.2 Performance analysis of the algorithm

When performing two-dimensional exhaustive search in AMWI images, as shown in Fig. 4(a), the following difficulties would be encountered for the small-sized window, where one is that interferences in the window looks similar to the concealed object, which causes false alarm, and another is that the image in the small-sized window loses the contextual information, which eventually increase the probability of missed detection[10]. Here, we show an example in Fig. 4(b).

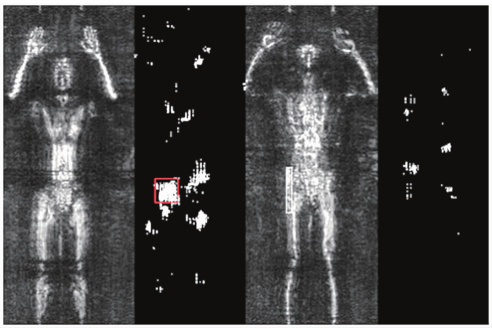

Fig. 4 AMWI images and their corresponding detection results when using a dense sliding window method. Using the method in Ref [1], we accumulate the probability of the connected region and compare it with the threshold. In Fig (a), the accumulated probability of the connected region in the red box is greater than the threshold, which causes false alarm; In Fig (b), the accumulated probability is less than the threshold in the ground-truth region, which causes the detection missed
图4 AMWI图像以及使用密集滑动窗口方法的检测结果. 我们使用文献[1]的方法累积连通区域的概率并和阈值相比较. 在图(a)中,红色框内连通区域的累积概率超过了阈值,从而产生了虚警;在图(b)中,隐匿物品区域的累积概率小于阈值而导致了漏检

Both of the two difficulties could be overcome by employing a big window which contains the object and its background. As shown in Fig. 5, the big window image contains the object (interference) and its background. By observing the contrast between small window images (shown in the red box in Fig. 5), there is clearer contrast between the object image and interference image, so the interference can be classified more easily.

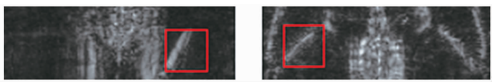

Fig. 5 When a concealed object is placed with its background, there are clear contrasts between the object (knife) and the Interference (arm)
图5 当把隐匿物品放置于背景中时,隐匿物品(刀)和干扰(胳膊)的对比就很清晰了

1.3 The Architecture of CNNs

The architecture of V_CNN is summarized in Fig. 6, the input of V_CNN are the images with size 60 × 152. From the input layer, the first convolutional layer contains 40 5 × 5 convolutional kernels, the output of the first convolutional layer is inputted to the first max pooling layer, the pooling kernel size is 2 × 2 and striding size is 2 × 2. The second convolutional layer contains 30 5 × 5 convolutional kernels and the second max pooling layer is also with 2 × 2 pooling kernel size and 2 × 2 striding size. The third convolutional layer contains 30 5 × 5 convolutional kernels and outputs to the full connect lay-

er, the full connect layer connects to the final softmax layer. It has 2 outputs, corresponding to the positive and negative category.

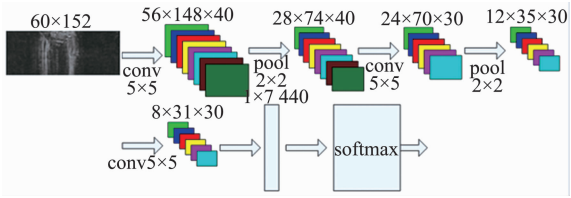

Fig. 6　The architecture of V_CNN
图 6　V_CNN 的结构

The architecture of H_CNN is quite similar to that of V_CNN. H_CNN also contains 3 convolutional layers, 2 max pooling layers, one full connecting layer, and one softmax layer. The convolutional kernel size is 5×5 and max pooling kernel size is 2×2.

1.4　Training Details

Labeled positive and negative samples are used to train CNN, and IoU is usually used to distinguish positive and negative samples. For example, an image sample is labeled positive with IoU > 0.5, else negative. Here we propose a new metric called IoG to define positive and negative samples. It will be proved that it is easier to detect object under big window by using IoG, especially for small-size object. IoG and IoU are defined as:

$$IoG = \frac{Area(G) \cap Area(B)}{Area(G)}, \quad (1)$$

$$IoU = \frac{Area(G) \cap Area(B)}{Area(G) + Area(B) - Area(G) \cap Area(B)}. \quad (2)$$

In (1) and (2), B is the sliding window inputted to CNN and G is the ground-truth. We define the sample with IoG > 0.8 as a positive sample, otherwise, a negative one.

We compare the detection performance of IoU and IoG under different sizes of windows. The IoU and IoG decision thresholds are given as follows:

$$IoU = \frac{Area(G) \cap Area(B)}{Area(G) + Area(B) - Area(G) \cap Area(B)} > 0.5, \quad (3)$$

$$IoG = \frac{Area(G) \cap Area(B)}{Area(G)} > 0.8. \quad (4)$$

Formula (3) and (4) can be rewritten as:

$$Area(G) \cap Area(B) > \frac{1}{3} Area(G) + \frac{1}{3} Area(B), \quad (5)$$

$$Area(G) \cap Area(B) > \frac{4}{5} Area(G). \quad (6)$$

If using IoG is more likely to detect the object, then formula (7) will hold.

$$\frac{4}{5} Area(G) \leqslant \frac{1}{3} Area(G) + \frac{1}{3} Area(B) \equiv Area(G) \leqslant \frac{5}{7} Area(B). \quad (7)$$

From formula (7), we can get that if the size of the object is less than $\frac{5}{7}$ of the window size, it is easier to detect object using IoG than using IoU. Our large-size window can ensure formula (7) holds, so that the use of IoG is more advantageous for object detection. The details of the comparison are shown in Section 3.1.

Training of H_CNN is more like that of V_CNN by using IoG to classify the training samples. The difference is that the training samples of V_CNN are obtained throughout the entire image whereas the training samples of H_CNN are only around the object. The fact is that V_CNN has filtered out the samples far away from the concealed object and only samples near or containing the concealed object needs to be classified by H_CNN. By contrast, due to the use of exhaustive search in Ref. [1], there is seriously imbalance between positive and negative samples, and a large number of negative samples have to be discarded to balance the proportion of positive and negative samples. Since we have filtered lots of negative samples in vertical direction, we do not need to discard any negative samples when training H_CNN.

2　Detection of concealed object

2.1　Detection

When using a window sliding over the image, CNN will output a series of classification results. There are three kinds of results: the pulse-free output, the output with only single pulse, and the output with a number of pulses. It is evident that the pulse-free output is associated with none of the concealed objects detected while the output with only single pulse implies that a concealed object is present.

The ideal result can be represented by

$$S_{p,l}(i) = \begin{cases} 1, p \leqslant i \leqslant p + l \\ 0 \quad else \end{cases}, \quad (8)$$

where p is the starting position of the pulse, l is the pulse length. Assuming W is the window width, N indicates the number of sliding times of the window. Then p and l are constrained by (9).

$$1 \leqslant p \leqslant N, 0 \leqslant l \leqslant W. \quad (9)$$

Assuming that the real result is Y, the square error between Y and S is denoted by E, E can be considered as the function of (p, l) and is defined by (10).

$$E(p,l) = \sum_{i=1}^{N} (Y(i) - S(i))^2. \quad (10)$$

It is seen that by minimizing $E(p, l)$, we can find a solution (p_{LS}, l_{LS}) and obtain the final detection result. However, it is hard to obtain an analytic solution of the problem by minimizing (10). Due to the fact that (p, l) belongs to a finite set, exhaustive search can be used. An example of the solution is shown in Fig. 7. We slide a window from top to bottom over an AMWI image, the images in all the sliding windows are inputted to V_CNN and a series of classification results are obtained (denoted by the black line). The horizontal axis is the vertical position of each window, and the vertical axis is detection score associated with the position. The detection score is Y in equation (10). By minimizing (10), we have the solution (p_{LS}, l_{LS}) and plot the final result (denoted by the red line) in Fig. 7. It can be seen that even if the detection results have some errors, the final result is still close to the ideal result.

After (p_{LS}, l_{LS}) is obtained, the width of the pulse can be determined by $K = l_{LS}$. An intuitive inference is

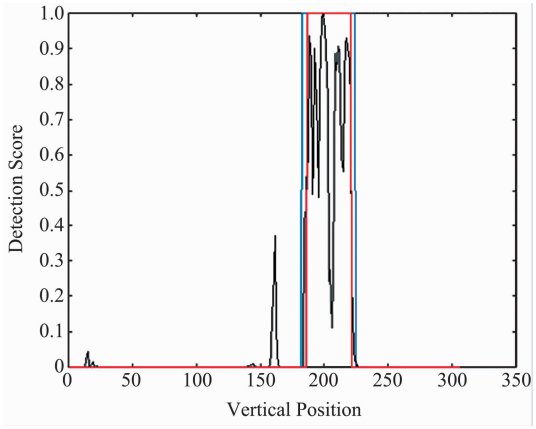

Fig. 7 Detection result by minimizing (10) (The black line is the detection score from V_CNN and the red line denotes the result using the parameter obtained by minimizing (10), the blue line represents an ideal output generated by ground-truth)
图7 最小化公式(10)得到的检测结果(黑色曲线为使用V_CNN的检测曲线;红色曲线为通过最小化公式(10)得到的参数绘制的曲线,蓝色曲线为理想输出曲线)

that the smaller K is, the more likely it is to be a false alarm. Here we formulate this issue under the framework of distributed detection[14]. The issue can be considered as a two hypothesis detection problem with individual detector decisions being the observations. In the problem, the window images are employed as distributed inputs and CNN is local decision maker. $\mu_1, \mu_2, \cdots, \mu_K$ are the local decisions sent to the fusion center. The output of the final decision is μ_0, representing whether a concealed object is detected. Assuming the event that an object is present is H_1, the opposing event is H_0. Their probabilities are P_1 and P_0.

According to Bayes' theorem, the posterior probability of detection can be written as:

$$P(H_1 \mid \mu_1\mu_2\cdots\mu_K) = \frac{P_1 \times P(\mu_1\mu_2\cdots\mu_K \mid H_1)}{P_1 \times P(\mu_1\mu_2\cdots\mu_K \mid H_1) + P_0 \times P(\mu_1\mu_2\cdots\mu_K \mid H_0)} . \quad (11)$$

Because each local detector makes a local decision based on its observation, there is no communication among them, $\mu_1, \mu_2, \cdots, \mu_K$ are conditionally independent. Eq. (11) can be written as:

$$P(H_1 \mid \mu_1\mu_2\cdots\mu_K) = \frac{P_1 \times \Pi_{i=1}^{K} P(\mu_i \mid H_1)}{P_1 \times \Pi_{i=1}^{K} + P_0 \times \Pi_{i=1}^{K} P(\mu_i \mid H_0)} . \quad (12)$$

In Eq. (12), $P(\mu_i \mid H_j)$ is obtained from CNN, P_i is the prior probability of H_i. By checking the value of equation (12), it is easy to identify whether the pulse is a false alarm or representing a concealed object.

2.2 Size estimation of concealed objects

By using window sliding over the image, a typical detection result is shown in Fig. 8. In fact, the bounding box is too large for the object in most cases. Here, we use an analytic result to estimate its size in each dimension.

As shown in Fig. 9, assuming that the width of the sliding window is W_b, the object width is W_o, and the detected pulse width is W_p, then we can have Eq. (13).

$$W_p = W_o + 2 \times (W_b - 0.8 \times W_o) - W_b = W_b - 0.6 \times W_o . \quad (13)$$

From Eq. (13), the size of the concealed object can be estimated.

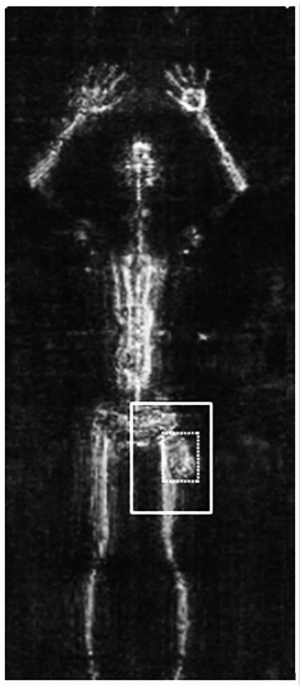

Fig. 8 Detection result by using V_CNN and H_CNN (ground-truth is the dashed box)
图8 V_CNN和H_CNN得到的检测结果(隐匿物品在图中的虚线框内)

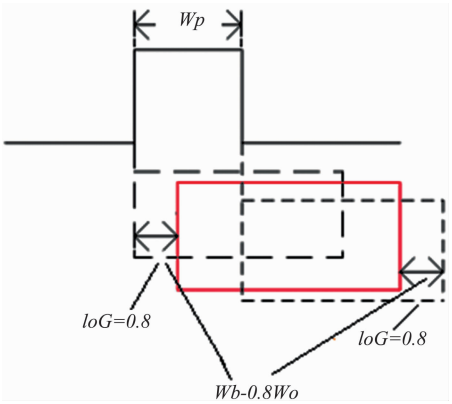

Fig. 9 Spatial relation between the sliding windows and the object (object is indicated in red box and sliding windows in dashed box)
图9 滑动窗口和目标之间的空间关系(目标以红色框表示,滑动窗以虚线框表示)

3 Experimental results

3.1 Comparison of detection performance using IoU and IoG

Before comparing the final experimental results, we compare the detection performance of IoU and IoG under different window sizes and object sizes by experiments. For simplicity, we use an exhaustive search method such as that in Ref. [1] and assume the detection result is ideal. We choose three kinds of objects with large (pistol), medium (cell phone) and small (knife) size and two sizes of windows (28×28 and 36×36) for comparison.

Test results using 28×28 window size

As shown in Fig. 10, we slide the window over the entire image and show the detection results by using IoU and IoG. It can be seen that when the window size is small, formula (7) does not hold for big object like pistol, the detection result using IoG is worse than that using IoU (fewer detection times). For medium object like cell phone (in Image2), the detection result using IoG is better for that formula (7) holds. The image with knife (Image3) is an extreme case; formula (3) and (4) do not hold for the entire image, which will lead to miss detection.

Test results using 36×36 window size

As shown in Fig. 11, when we increase the window size, formula (7) holds for all the three kinds of concealed objects and the performance is better by using IoG. In Image2, formula (3) is no longer valid for cell phone with the increase of the window, so it leads to miss detection by using IoU. It should be noted that in Image3, even if we increase the window size, knife still cannot be detected using IoU. The reason is that the aspect ratio of the window and the aspect ratio of the object are seriously mismatched. However, IoG is not sensitive to the aspect ratio mismatch, so the slender object can be detected.

3.2 Comparison of detection performance

We use the same dataset as in Ref. [1], the dataset is collected from the SimImage system of Chinese Academy of Sciences. There are 440 images in the dataset, and among those there are 400 ones with concealed objects including knives, pistols and cell phones. 308 images are used for training and 132 images for test. There are 114 images with all the three kinds of concealed objects and 18 images without concealed objects in the test dataset.

Some of the detection results are shown in Fig. 12. Using the same criterion as in Ref. [1], miss detection is defined that a window output with IoU less than 0.3. The statistical data of this experiment is shown in Table 1.

Table 1　Detection Results
表 1　检测结果

Images with concealed object	Number	False Alarm	Miss Detection
Y	114	0	2
N	18	2	0
Total	132	2	2

We use *precision*, *accuracy* and *recall* to compare our method with [1]. Their definitions are as follows:

$$Precision = \frac{True\ Positives}{True\ Positives + Flase\ Positives}$$

$$Accuracy = \frac{True\ Positives + True\ Negtives}{Positives + Negtives}$$

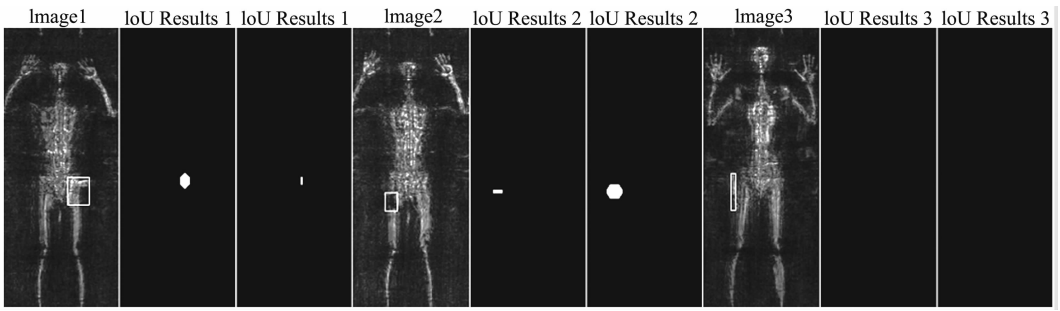

Fig. 10　Images with three kinds of concealed objects (in the box) and their corresponding detection results using IoU and IoG. The concealed objects are pistol (Image1), cell phone (Image2) and knife (Image3). The window size is 28×28

图 10　三种隐匿物品的图像(以实线框表示)以及使用 IoU 和 IoG 作为检测标准时的检测结果. 隐匿物品为手枪(图 1),手机(图 2)和刀(图 3). 窗口尺寸为 28×28

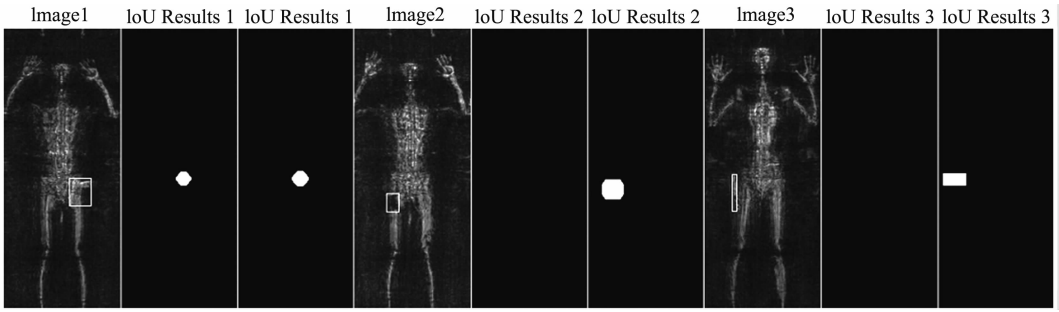

Fig. 11　The same images as in Fig. 10 and their corresponding detection results using IoU and IoG. The window size is 36×36

图 11　使用和图 10 相同的图像,将窗口尺寸修改为 36×36,分别使用 IoU 和 IoG 作为检测标准时的检测结果

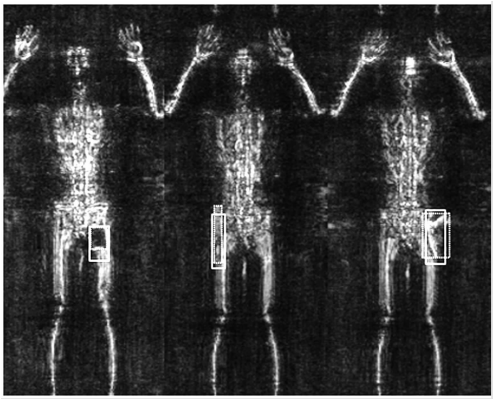

Fig. 12 Localization results for different objects, the ground-truth are inside the dashed boxes (from left to right, the concealed objects are cell phone, knife and pistol)
图12 对不同目标的检测结果,隐匿物品在图中虚线框内(从左到右,隐匿物品分别为:手机,刀和手枪)

$$Recall = \frac{True\ Positives}{Positives}$$

The comparisons with Ref. [1] are shown in Table 2, where the *precision*, *accuracy* and *recall* are all improved using our method. The better results lie in the advantages of our two CNNs method: 1) using V_CNN to filter out most of the interference, which can effectively reduce false alarm probability; 2) using big window to preserve the context for object and using IoG to classify objects, which can reduce miss detection probability.

Table 2 Detection Results Comparison
表2 检测结果比较

Items	*Precision*	*Accuracy*	*Recall*
Dense Sliding Window Method[1]	95.5%	93.2%	93%
Our Method	98.2%	96.9%	98.2%

3.3 Comparison of computational efficiency

Compared with the dense sliding window method[1], another advantage of this method lies in its lower computational complexity. Unlike Ref. [1], our method only performs exhaustive search in vertical direction, so the length of the computational time can be significantly reduced.

We use GTX 650 from Nvida to test the testing time (seconds/image). Our method process images 3.5 × faster than Ref. [1]. The comparison is shown in Table 3.

Table 3 Run Time Comparison
表3 运行时间比较

Items	Max time /img	Min time /img	Mean time /img
Dense Sliding Window Method[1]	2.11 s	1.82 s	1.93 s
Our Method	0.58 s	0.53 s	0.55 s
Time Speedup	3.6 ×	3.43 ×	3.55 ×

4 Conclusions

A low complexity method is presented to detect concealed object in active millimeter-wave image. This method employs two CNNs for detection in vertical and horizontal direction and achieves better detection performance than the method using dense sliding window. This method also has the advantage in computational complexity, which reduces the length of the computational time to lower than 30% of that of the exhaustive search. In addition, without size estimation, our method can be used as a region proposal method to generate RoI for object detection.

References

[1] Yao JX, Yang MH, Zhu YK, *et al*. Use Convolutional Neural Network to Localize Forbidden Object in Millimeter-wave Image [J]. *Journal of Infrared and Millimeter Waves*, 2017, **36**(3): 354 - 360.

[2] Chen H. M, Lee S, Rao R. M, *et al*. Imaging for concealed weapon detection: a tutorial overview of development in imaging sensors and processing [J]. *IEEE Signal Processing Magazine*, 2005, **22**(2): 52 - 61.

[3] Appleby R, Anderton R. N. Millimeter-wave and submillimeterwave imaging for security and surveillance [J]. *Proceedings of the IEEE*, 2007, **95**(8):1683 - 1690.

[4] Vaillant R, Monroc C, and LéCun Y. Original approach for the localisation of objects in images [J]. *IEE Proc on Vision, Image, and Signal Processing*, 1994, **141**(4):245 - 250.

[5] Sermanet P, Eigen D, Zhang X, *et al*. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks [J]. *Eprint Arxiv*, 2013.

[6] Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 580 - 587.

[7] Girshick R, Donahue J, Darrell T, *et al*. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(1): 142 - 158.

[8] Girshick R, Fast R-CNN [J]. *Computer Science*, 2015.

[9] Hu P, Ramanan D. Finding Tiny Faces [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 951 - 959.

[10] Bell S, Zitnick C. L, Bala K, *et al*. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks. CVPR, 2016: 2874 - 2883.

[11] Li XB, Wang SJ. Object Detection Using Convolutional Neural Networks in a Coarse-to-Fine Manner [J]. *IEEE Geoscienc and Remote Sensing Letters*. 2017, **14**(11):2037 - 2041.

[12] Uijlings J. R. R, Sande K. E, Gevers T, *et al*. Selective Search for Object Recognition [J]. *International Journal of Computer Vision*, 2013, **104**(2):154 - 171.

[13] Hosang J, Benenson B, Dollár P, *et al*. What makes for effective detection proposals? [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(4):814 - 830.

[14] Varshney P. K, Distributed Detection and Data Fusion [M]. Springer New York, 1997.