

Research Article

On Using XMC R-CNN Model for Contraband Detection within X-Ray Baggage Security Images

Yong Zhang ^{1,2}, Weiwu Kong,² Dong Li,² and Xudong Liu¹

¹School of Computer Science and Engineering, Beihang University, Beijing, China

²First Research Institute of the Ministry of Public Security, Beijing, China

Correspondence should be addressed to Yong Zhang; zhangyong@fiscan.cn

Received 18 July 2020; Accepted 8 September 2020; Published 16 September 2020

Academic Editor: Chenxi Huang

Copyright © 2020 Yong Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present an X-ray material classifier region-based convolutional neural network (XMC R-CNN) model for detecting the typical guns and the typical knives in X-ray baggage images. The XMC R-CNN model is used to solve the problem of contraband detection in overlapped X-ray baggage images by the X-ray material classifier algorithm and the organic stripping and inorganic stripping algorithm, and better detection rate and the miss rate are achieved. The detection rates of guns and knives are 96.5% and 95.8%, and the miss rates of guns and knives are 2.2% and 4.2%. The contraband detection technology based on the XMC R-CNN model is applied to X-ray baggage images of security inspection. According to user needs, the safe X-ray baggage images can be automatically filtered in some specific fields, which reduces the number of X-ray baggage images that security inspectors need to screen. The efficiency of security inspection is improved, and the labor intensity of security inspection is reduced. In addition, the security inspector can screen X-ray baggage images according to the boxes of automatic detection, which can improve the effect of security inspection.

1. Introduction

In recent years, with the increasing seriousness of terrorist activities, the safety of air transport has been paid more and more attention by all countries in the world. At present, there are three pain points of the security inspection, which need to be solved urgently: first, how to improve the effect of security inspection; second, how to improve the efficiency of security inspection; and third, how to reduce labor intensity of security inspection. The contraband detection technology is not used in the traditional technical solution, because the baggage contents are complex and highly varying. How to accurately identify the contraband in the X-ray baggage image is the most important and most difficult challenge for human operators. In addition, during peak periods, the human operators have limited time to screen images.

As the most popular machine learning method, deep learning has achieved excellent results in object classification and detection. For the task of X-ray image classification, the previous work proposed the traditional machine learning

method. In this paper, the XMC R-CNN method based on deep learning will be used to detect the contraband within X-ray baggage security images.

In the ImageNet Large-Scale Visual Recognition Challenge 2012 (ILSVRC12), Hinton's team won the championship with the AlexNet model constructed by convolutional neural networks (CNNs), which ignited the enthusiasm of academic and industrial for deep learning. In order to use the deep learning method to detect the contraband in X-ray baggage security images, the framework of deep learning, the backbone model of deep learning, and the detection model of deep learning should be understood, as shown in Figure 1.

With the upsurge of deep learning research, various open-source frameworks of deep learning emerge in endlessly. The main framework is introduced in [1–4], such as TensorFlow, Caffe, MXNet, Keras, CNTK, Torch, and Theano. Different deep learning frameworks usually have different model design, interface, deployment, performance, and architecture design, so their advantages and disadvantages are different.

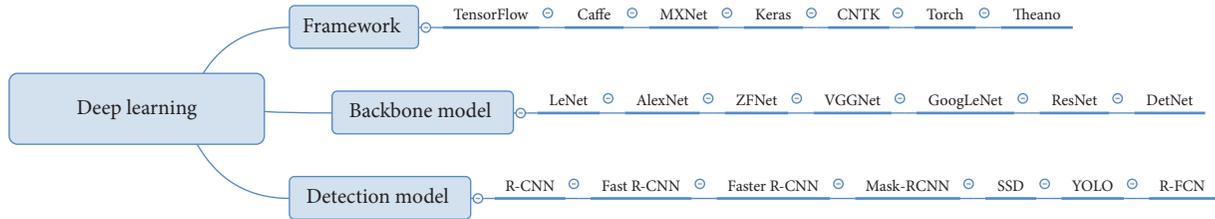


FIGURE 1: Deep learning map.

The classic backbone models mainly include LeNet [5], AlexNet [6], ZFNet [7], VGGNet [8], GoogleNet [9], ResNet [10], and DetNet [11]. Although the algorithms of object detection are different, the convolutional neural networks are usually used to process the input image, generate the feature map, and then use various algorithms to complete the region generation and loss calculation. The convolutional neural networks are the backbone of the whole detection algorithm. The basic components of the backbone include convolution layer, activation function layer, pooling layer, dropout layer, batch normalization layer, and full connection layer.

The main backbone models are introduced above, which are usually used for object classification models. We study that the contraband detection belongs to classification and detection, and only classification is not enough. There is much contraband in X-ray baggage security images, so it is not only necessary to classify different contraband but also to determine the locations and sizes of the contraband. The classic classification and detection models include R-CNN [12], Fast R-CNN [13], Faster R-CNN [14], Mask-CNN [15], SSD [16], YOLO [17], and R-FCN [18]. These models have developed from two stages to one stage, from bottom-up only to top down, from the single-scale network to the feature pyramid network, and many algorithms have achieved excellent results on the ImageNet dataset.

Computer-aided screening (CAS) [19] has also been widely used in the automatic detection of the contraband in X-ray baggage images; however, this largely remains an unsolved problem. The contraband detection based on multiview X-ray images is carried out in each X-ray image, and then, the constraint between multiview images is used to improve the detection accuracy in [20–27]. The contraband detection in computed tomography (CT) images is to extend the detection method of the single X-ray 2D image to 3D image in [28–35]. As mentioned above, both the contraband detections in multiview images and CT images are traditional technologies. The contraband detection by deep learning is proposed in [19, 36–38], and the accuracy of contraband detection is improved through deep learning methods. At present, there are still some problems in using these techniques to detect the contraband in overlapped X-ray baggage images. For example, a knife or a gun is covered by other objects in the baggage, and the outline of the knife or the gun is not clear in the original X-ray image, and its color is shown as orange (it is not metal color), so it cannot be correctly detected by the prior detection algorithm. Also, an explosive is hidden under several steel plates

in the baggage, and it cannot be seen at all in the original image. It is difficult for the security inspector to determine that there is a dangerous object in this area. The XMC R-CNN model is used to solve the problem of contraband detection in overlapped X-ray baggage images by the X-ray material classifier algorithm and the organic stripping and inorganic stripping algorithm.

In the next section, we review related work on the dual-energy X-ray material classifier and the contraband detection based on dual-energy X-ray data. In Section 3, we discuss the contraband detection technology based on the XMC R-CNN model, including the detection model design and the algorithm implementation. In Section 4, we introduce the training and test results of the model. We summarize the work of this paper in Section 5.

2. Related Work

We now summarize the related work on the dual-energy X-ray material classifier and the contraband detection based on dual-energy X-ray data.

2.1. X-Ray Material Classifier. The dual-energy method of material classifier has been widely used in the X-ray security inspection systems by measuring the difference of attenuation coefficient of different materials for high- and low-energy X-ray in [39–46]. The physical principle of the dual-energy X-ray radiography is based on the exponential law of photon radiation attenuation. When an X-ray beam passes through an object, the detector signal can be described by the Beer–Lambert law:

$$\mathbf{I} = \int \mathbf{I}_0(E) e^{-\mu(E,Z)\tau} dE, \quad (1)$$

where E is the photon energy, $I_0(E)$ is the initial energy intensity of photons emitted from the source, $\mu(E, Z)$ is the attenuation coefficient of a material with atomic number Z for impinging photons with energy E , and τ is the material thickness.

In previous work for the material classifier in the X-ray baggage images, the log-ratio R is defined as the vertical axis, and $1/H_i$ is defined as the horizontal axis by Ogorodnikov and Petrunin in [47]:

$$\mathbf{R} = \frac{\log(\mathbf{H}_i)}{\log(\mathbf{L}_o)}, \quad (2)$$

where Hi is the high-energy transparency and Lo is the low-energy transparency. By formulas (1) and (2), for a given material with atomic number Z , R is the unique value of the material and does not depend on the thickness. Therefore, R can be used to discriminate materials.

The α -curve method computes the features of the X-ray material in [48–50]:

$$\begin{cases} \alpha_1 = -\log(Hi), \\ \alpha_2 = -\log(Hi) + \log(Lo), \end{cases} \quad (3)$$

where α_2 is defined as the vertical axis and α_1 is defined as the horizontal axis. The two materials can be easily differentiated if their α -curves are largely separated. On the contrary, the materials cannot be easily differentiated if their α -curves are too close.

2.2. Contraband Detection. Most contraband detection algorithms based on deep learning use the RGB images and do not really use the X-ray high- and low-energy data. Therefore, the material information is not fully used by the detection algorithm. The contraband detection within X-ray baggage imagers has not been well explored by machine vision community due to the lack of publicly available X-ray image datasets. Since the contraband detection within X-ray baggage imagers is a challenging problem, the detection models that use the X-ray high- and low-energy data are significantly more limited in the literature.

The dual-energy images which provide material information about the objects are used for the contraband detection in [21, 51]. The authors show that using multiple local color and texture features improves classification and detection performance. They first present an extensive evaluation of standard local features for object detection on a large X-ray image dataset in a structured learning framework. Then, they propose two dense sampling methods as a key-point detector for textureless objects and extend the SPIN (the image generation process can be visualized as a sheet spinning about the normal of a point) color descriptor to utilize the material information. Finally, they propose a multiview branch-and-bound search algorithm for multi-view object detection.

The two-channel dual-energy networks and the four-channel dual-energy networks are described in [37]. For each input, the high- and low-energy images are transformed into the feature space of a given method. The final convolutional layer feeds into the fully connected (FC) layers. The networks used in this work consist of 16 convolutional layers and 3 fully connected layers. The author experimented with the four-channel dual-energy networks, and the four-channel dual-energy networks with $\{-\log H, \Delta, -\log L, \Sigma\}$ perform the best performance metrics, where $\Sigma = H + L$ and $\Delta = H - L$ are used.

3. Detection Model Design

In this paper, the Faster R-CNN model is combined with the characteristics of the X-ray baggage image, and the XMC R-CNN model is designed for the automatic detection of the

typical guns and the typical knives in the X-ray baggage image. The framework of the system is Caffe, and the backbone model of the system is VGG-16 [8].

3.1. XMCR-CNN Detection Model. The XMC R-CNN model is composed of two modules. The first module is X-ray material classifier (XMC) that strips organic and inorganic, and the second module is the Faster R-CNN detector that uses the proposed regions. The XMC R-CNN algorithm model mainly includes the following steps, as shown in Figure 2.

3.1.1. X-Ray Material Classifier. High-energy data and low-energy data are input, and different material values of organic, mixture, and inorganic are output through the material classification algorithm.

3.1.2. Organic Stripping and Inorganic Stripping. High-energy data, low-energy data, and material values are input, and the material value of the organic, the gray value of the organic, the material value of the inorganic, and the gray value of the inorganic are output through the organic stripping and inorganic stripping algorithm.

3.1.3. Convolutional Layer Features. A convolutional layer features adopts the Simonyan and Zisserman model [8] (VGG-16), which has 13 shareable convolutional layers. The organic, mixture, and inorganic images are input, which sizes are different. The feature maps of the input image are output, which will be used for the region proposal network (RPN) layers and the full connection layers.

3.1.4. Region Proposal Networks. A region proposal network [14] is mainly used to generate region proposals. Firstly, anchor boxes of multiple scales and aspect ratios are designed. By the SoftMax layer, it can be determined that anchor boxes belong to foreground or background. Then, the anchor boxes are modified by binding box regression and a more accurate region proposal is obtained.

3.1.5. Region of Interest Pooling. The RoI (Region of Interest) pooling layer in Faster R-CNN is adopted. The feature map from the last layer of VGG-16 and the region proposals from RPN are input, and the fixed size 7×7 proposal feature map is output.

3.1.6. Contraband Detection. The proposal feature maps are input. Through the full convolutional layer and the SoftMax layer, we can get the classification of region proposal. At the same time, we can get the locations of the detection boxes by the binding box region. Finally, the classifications and the locations of the typical guns and the typical knives are output.

The detailed implementation of Step (3), (4), (5), and (6) can be seen in the Faster R-CNN [14] and VGG-16 [8]. The

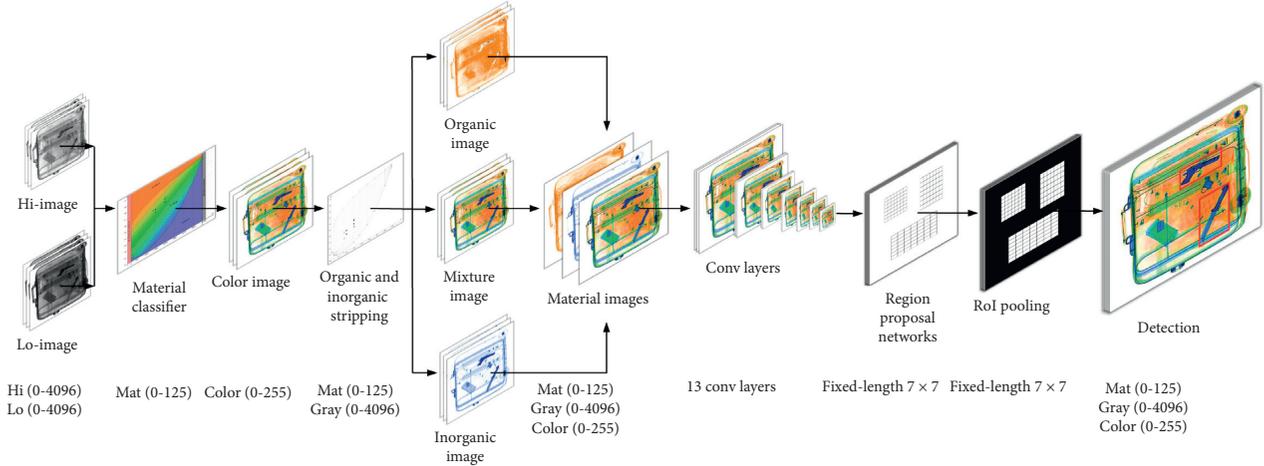


FIGURE 2: XMC R-CNN model.

X-ray material classifier algorithm of Step (1) and the organic stripping and inorganic stripping algorithm of Step (2) will be described in the following sections.

3.2. X-Ray Material Classifier. In this paper, the organic glass (chemical formula: $C_5H_8O_2$) is selected to represent the organic material in Table 1, the aluminum (model: 2A12) is selected to represent the mixture material in Table 2, and the carbon steel (model: 45[#]) is selected to represent the inorganic material in Table 3.

Three kinds of materials are selected from different thickness ranges, and data are collected; then, the curves are drawn as shown in Figure 3. The horizontal axis is the high-energy value (Hi) of the dual-energy X-ray, and the vertical axis is the material value (Mat) of the dual-energy X-ray. The log-ratio R is defined as the material value. 122 curves are evenly inserted between the three curves, which, respectively, represent the different materials of organic, mixture, and inorganic. Through these 125 curves, the material values of any point in the space are calculated, and the material table of the dual-energy X-ray is generated.

The whole material space is divided into five regions: low-gray unrecognizable region where the image shows red, high-gray unrecognizable region where the image shows gray, organic region where the image shows orange, mixture region where the image shows green, and inorganic region where the image shows blue, as shown in Figure 3.

The material value of the low-gray unrecognizable region is defined as 125, the material value of the high-gray unrecognizable region is defined as 0, the material value of the organic region is defined as 1, the material value of the inorganic region is defined as 124, and the material value of the mixture region needs to be calculated, which is the transition region from the inorganic region to the organic region. Three material curves in the mixture region have been obtained. According to the X-ray material classifier model in Figure 3, the low- and high-energy data of other 122 curves with different thickness are calculated. Through the materials of 125 curves and 100 data sampling points of different material thickness, the mixture region can be

divided into 12276 grid cells. As shown in Figure 3, each grid cell is composed of four points (A , B , C , and D). Using the linear interpolation algorithm, we can calculate the material value of any point in the quadrilateral (A , B , C , and D). Finally, the material values of all points in the whole material space can be calculated and exported to the material table.

3.3. Organic Stripping and Inorganic Stripping. By means of physical model designing, experimental data sampling, and the using of the grid cell method and linear interpolation algorithm, the organic or the inorganic can be screened separately from the overlapped X-ray baggage image. The organic glass (chemical formula: $C_5H_8O_2$) is selected to represent the organic material, the carbon steel (model: 45[#]) is selected to represent the inorganic material, and the overlapped simulants of the organic glass and the carbon steel are selected to represent the mixture material, whose thickness is different.

The effective atomic number [52] is given by

$$Z_{eff} = \left(\frac{\sum a_i * Z_i^{3.5}}{\sum a_i} \right)^{1/3.5}, \quad (4)$$

$$a_i = \frac{m_i * Z_i}{A_i}, \quad (5)$$

where Z_{eff} is the effective atomic number, m_i is the mass percentage of the element, Z_i is the atomic number, and A_i is the atomic weight. According to formulas (1) and (2), the effective atomic number Z_{fe} of the carbon steel is equal to 25.91, and the effective atomic number Z_{gl} of the organic glass is equal to 6.56. The material of the effective atomic number Z_n (7~25) can be obtained by overlapping the organic glass and the carbon steel with different thicknesses.

As shown in Figure 4, the horizontal axis is the high-energy value (Hi) of the dual-energy X-ray and the vertical axis is the material value (Mat) of the dual-energy X-ray. The log-ratio R is defined as the material value. The orange curve is the organic glass curve, which represents the organic material. The blue curve is the carbon steel curve, which

TABLE 1: Element content data of organic glass.

Element	Atomic number	Atomic weight	Percentage
C	6	12.01	59.98
H	1	1.008	8.05
O	8	16.00	31.97

TABLE 2: Element content data of aluminum.

Element	Atomic number	Atomic weight	Percentage
Cu	29	63.55	4.35
Mg	12	24.31	1.5
Mn	25	54.94	0.6
Al	13	26.98	93.55

TABLE 3: Element content data of carbon steel.

Element	Atomic number	Atomic weight	Percentage (%)
C	6	12.01	0.46
Si	14	28.09	0.27
Mn	25	54.94	0.65
P	15	30.97	0.04
S	16	32.07	0.40
Cr	24	52.00	0.25
Ni	28	58.69	0.25
Fe	26	55.85	97.68

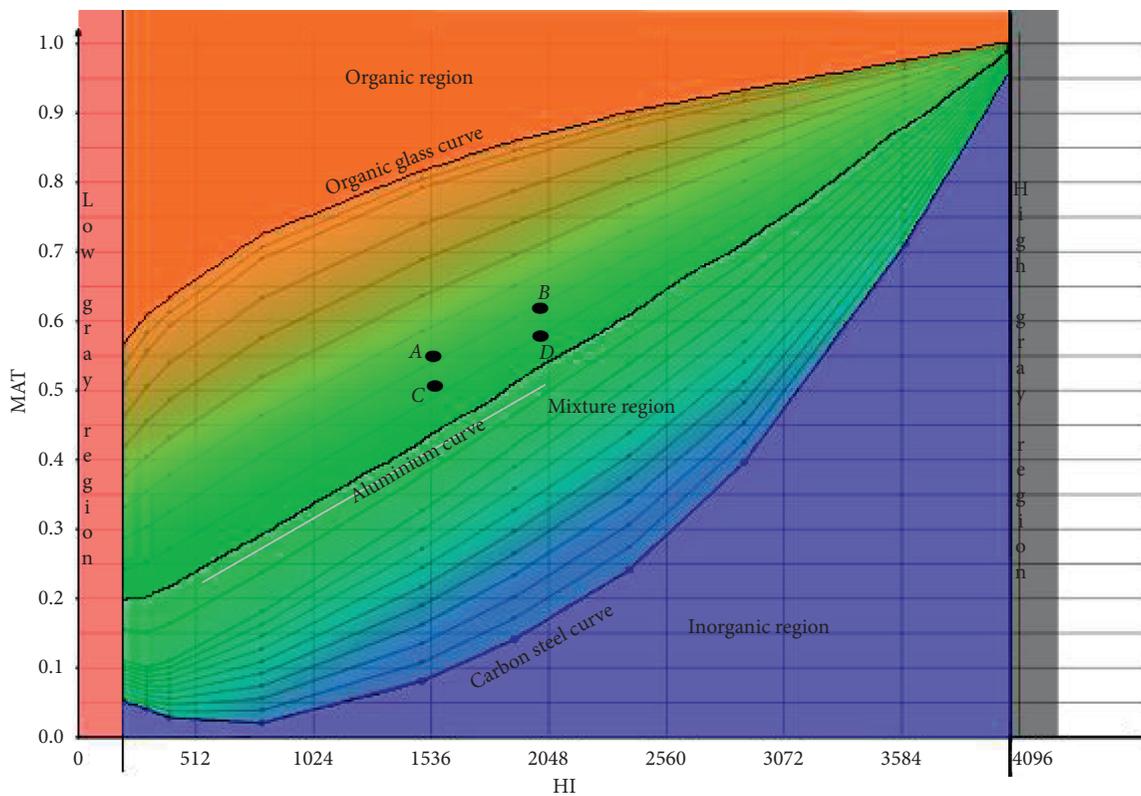


FIGURE 3: X-ray material classifier model.

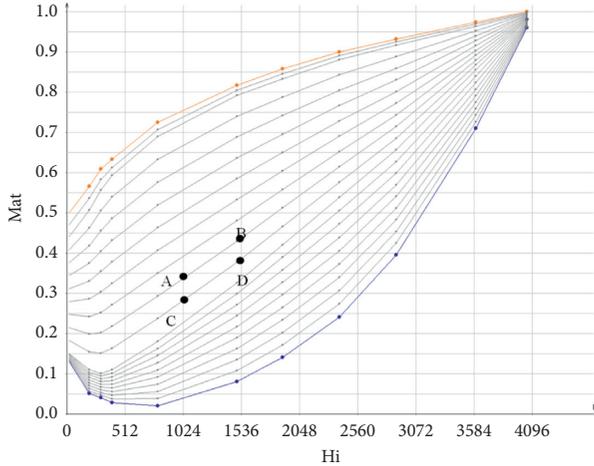


FIGURE 4: Organic tripping and inorganic stripping model.

represents the inorganic material. The curves between the orange curve and the blue curve are the different thicknesses overlapped curves of the organic glass and the carbon steel, which represents the mixture material.

By overlaying the organic glass ($Z_{gl} = 6.5$) and the carbon steel ($Z_{fe} = 26$), the effective atomic number Z_n (7~25) is obtained. For example, if we want to calculate the material with $Z_n = 7$, take $Z_{eff} = 7$ into formulas (1) and (2), so we can get the mass ratio of the organic glass and the carbon steel ($m_{gl}:m_{fe} = 0.997572383:0.002427617$). The length and width of the organic glass and the carbon steel models are the same in the test process, we know that the density of the carbon steel is 7.82 g/cm^3 , and the density of the organic glass is 1.18 g/cm^3 , so that the volume ratio of the organic glass and the carbon steel can be calculated and then converted to the thickness ratio ($T_{gl}:T_{fe} = 0.999632928:0.000367072$). Similarly, we can calculate the material with $Z_n = 8-25$, and the corresponding thickness of the organic glass and the carbon steel is shown in Table 4.

According to Table 4, the material of the effective atomic number $Z_n = 7-25$ can be obtained by overlapping the organic glass and the carbon steel with different thicknesses. 9 sampling points are designed for each material, and 171 sampling points are obtained in total. Six groups of data were collected at each sampling point, and a total of 1026 sampling data points were obtained.

For example, the thickness of the organic glass is 32.7 mm, and the thickness of the carbon steel is 0.012 mm. We know that the density of the carbon steel is 7.82 g/cm^3 , and the density of the organic glass is 1.18 g/cm^3 . In this way, we can calculate the mass ratio of the organic glass and the carbon steel ($m_{gl}:m_{fe} = 38.568:0.09384$), and bring the m_{gl} and m_{fe} into formulas (1) and (2), so that we can get $Z_{eff} = 7.00039831652385 \approx 7$ (i.e., lines 7-6 in Table 5). Similarly, it can be calculated that each point in Table 5 meets the corresponding effective atomic number. See Table 5 for the experimental data of the design sampling point, where Z_n is the effective atomic number and the serial number of the sampling point, T_{gl} is the thickness of the organic glass, T_{fe} is the thickness of the carbon steel, G_m is

the high- and low-energy value of the mixture material, G_{gl} is the high- and low-energy value of the organic glass, and G_{fe} is the high- and low-energy value of the carbon steel.

An organic stripping and inorganic stripping table is generated from the data in Table 5. As shown in Figure 4, there are four points: A, B, C, and D. Through Table 5, we can get A (G_m, G_{gb}, G_{fe}), B (G_m, G_{gb}, G_{fe}), C (G_m, G_{gb}, G_{fe}), and D (G_m, G_{gb}, G_{fe}). Through the bilinear interpolation algorithm, we can calculate the (G_m, G_{gb}, G_{fe}) value of any point in the quadrilateral (A, B, C, and D), so that we can calculate the (G_m, G_{gb}, G_{fe}) value of any point in the coordinate system. Finally, we can realize the organic stripping and inorganic stripping of dual-energy X-ray security images.

In Figure 5(a), a knife is covered with some objects. In the original image, the outline of the knife is not clear, and its color is shown as orange, so it cannot be judged as a dangerous metal. In Figure 5(b), the knife is clearly visible, and it is shown as a metal color (blue), because the organic stripping algorithm is used. The outline of this knife is clear in the organic stripping image, and it is easy to be detected by the XMC R-CNN model.

In Figure 6(a), an explosive is hidden under several steel plates. In the original image, it is difficult for the security inspector to determine that there is a dangerous object in this area. In Figure 6(b), the explosive is clearly visible, and it is shown as an explosive color (orange), because the inorganic stripping algorithm is used. This is a good basis for detecting more kinds of contraband in X-ray baggage images in the future.

4. Evaluation

Training and testing are performed via the use of Caffe [53], a deep learning tool designed and developed by the Berkley Vision and Learning Center. The framework of the system is Caffe, the backbone model of the system is VGG-16, and the algorithm model of the system is XMC R-CNN.

4.1. Dataset Design. The experimental samples of different material guns are shown in Table 6, and the experimental samples of different type knives are shown in Table 7.

4.2. Training and Testing. The thirty pieces of baggage of different sizes and materials are filled with different objects, whose six pieces of baggage are complex, eighteen pieces of baggage are medium complex, and six pieces of baggage are simple. Each baggage is divided into three layers with the top facing up, and each layer is divided into nine areas. As shown in Figure 7, 1-9, 11-19, and 21-29 are used to represent the nine areas of each layer, and the baggage internal space is divided into 27 areas for placing contraband samples.

Four directions of each luggage are as shown in Figure 8. The top of the luggage is upward to enter the X-ray equipment channel, the top of the luggage is upward to enter the X-ray equipment channel at an angle of about 45 degrees, the bottom of the luggage is upward to enter the X-ray equipment channel, and the bottom of luggage is upward to

TABLE 4: Effective atomic number ratio data.

Effective atomic number	Organic glass thickness percentage	Carbon steel thickness percentage
6.5	1	0
7	0.999632928	0.000367072
8	0.998548186	0.001451814
9	0.997047753	0.002952247
10	0.995034526	0.004965474
.	.	.
.	.	.
.	.	.
22	0.896018167	0.103981833
23	0.863513626	0.136486374
24	0.817831277	0.182168723
25	0.749959651	0.250040349
26	0	1

TABLE 5: Sampling experimental data.

Z-n	T_{gl}	T_{fe}	G_m		G_{gl}		G_{fe}	
			Hi	Lo	Hi	Lo	Hi	Lo
7-1	10.9	0.004	2622	2308	2687	2485	3960	3732
7-2	13.6	0.005	2009	1657	2081	1839	3913	3575
7-3	16.3	0.006	1727	1374	1833	1590	3880	3478
7-4	21.8	0.008	1270	951	1349	1116	3823	3342
7-5	27.2	0.01	1137	826	1220	993	3797	3256
7-6	32.7	0.012	837	579	907	708	3754	3151
7-7	43.6	0.016	635	415	698	526	3651	2952
7-8	49.0	0.018	313	184	358	247	3532	2712
7-9	54.5	0.02	163	86	194	126	3387	2440
.
.
.
10-1	11.77	0.06	3142	2850	3213	3073	3996	3773
10-2	19.96	0.1	2631	2255	2724	2524	3932	3602
10-3	35.79	0.18	1863	1447	1998	1759	3801	3260
10-4	44.95	0.22	1523	1129	1674	1432	3738	3126
10-5	58.14	0.3	1129	773	1270	1041	3613	2866
10-6	79.93	0.4	708	440	827	638	3460	2569
10-7	99.96	0.5	458	264	553	403	3335	2361
10-8	119.9	0.58	300	161	375	261	3202	2156
10-9	159.6	0.78	144	71	178	115	2978	1848
.
.
.
25-1	0.77	1	2605	1435	4026	4007	2658	1468
25-2	1.63	2.12	1773	731	3951	3916	1828	753
25-3	2.08	2.70	1458	537	3923	3881	1520	560
25-4	2.66	2.47	1245	421	3889	3834	1305	442
25-5	4.00	3.92	713	197	3796	3729	764	209
25-6	6.37	6.38	301	70	3613	3522	334	78
25-7	9.34	12.12	190	44	3520	3417	213	48
25-8	10.79	14.0	112	26	3384	3254	134	31
25-9	11.72	15.2	59	16	3223	3078	69	17

enter the X-ray equipment channel at an angle of about 45 degrees.

The gun samples in Table 6 and the knife samples in Table 7 are placed in 27 space areas within each luggage, and the sample data are collected in four different

directions as shown in Figure 8. The total number of collected samples is $(10 + 20) \times 30 \times 27 \times 4 = 97200$. As shown in Figure 9, the 80000 samples are selected as the training dataset, and the 17200 samples are selected as the test dataset.

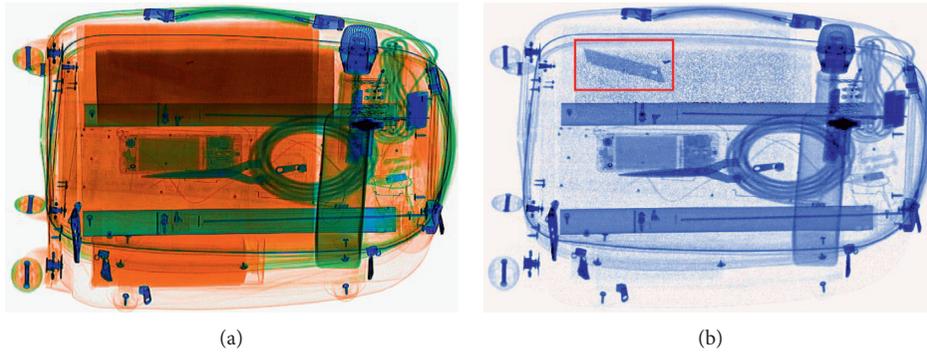


FIGURE 5: (a) Original X-ray image; (b) organic stripping image.



FIGURE 6: (a) Original X-ray image; (b) inorganic stripping image.

TABLE 6: Experimental samples of the gun.

No.	Photo	Name	Material
G-1		Pistol	Metal
G-2		Pistol	Plastic
G-3		Pistol	Wood
G-4		Pistol	Metal
G-5		Revolver	Metal
G-6		Revolver	Plastic
G-7		Revolver	Wood
G-8		Revolver	Metal
G-9		Rifle	Metal
G-10		Rifle	Plastic

TABLE 7: Experimental samples of the knife.

No.	Photo	Name	Material
K-1		Dagger	Metal
K-2		Dagger	Metal and plastic
K-3		Dagger	Metal
K-4		Double edged knife	Metal
K-5		Double edged knife	Metal
K-6		Jackknife	Metal
K-7		Jackknife	Metal
K-8		Keychain knife	Metal
K-9		Watermelon knife	Metal
K-10		Tools	Metal
K-11		Fruit knife	Metal
K-12		Fruit knife	Metal
K-13		Utility knife	Metal
K-14		Utility knife	Metal and plastic
K-15		Blade	Metal
K-16		Kitchen knife	Metal
K-17		Scissors	Metal and plastic
K-18		Scissors	Metal and plastic
K-19		Sword	Metal
K-20		Sword	Metal

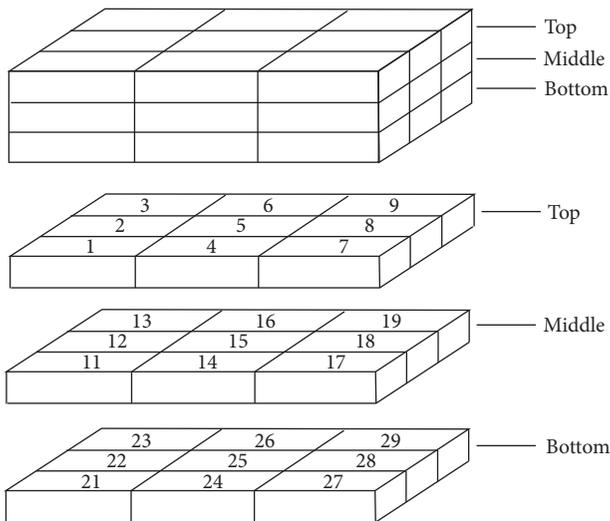


FIGURE 7: Interior space division of the baggage.

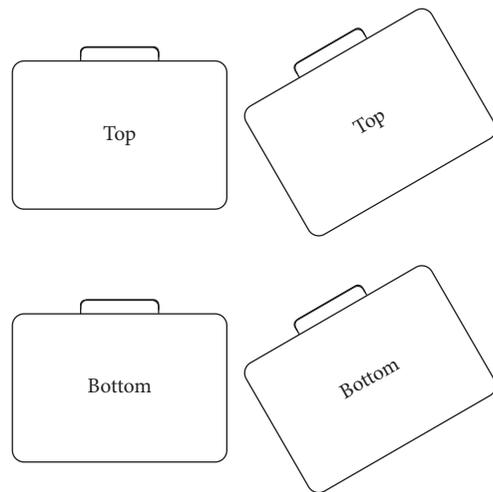


FIGURE 8: Sketch map of luggage direction.

4.3. *Test Results.* At present, the accuracy and speed of various classical models are tested through the ILSVRC (ImageNet Large-Scale Visual Recognition Challenge) dataset or the PASCAL VOC (Pattern Analysis, Statistical Modelling, and Computational Learning, and Visual Object

Classes) dataset. Because the characteristics of the images in these datasets are different from the characteristics of the X-ray baggage images, the performance and indicators of different models on the X-ray baggage image datasets are different from the ILSVRC dataset and the PASCAL VOC dataset. According to the test results of various classical

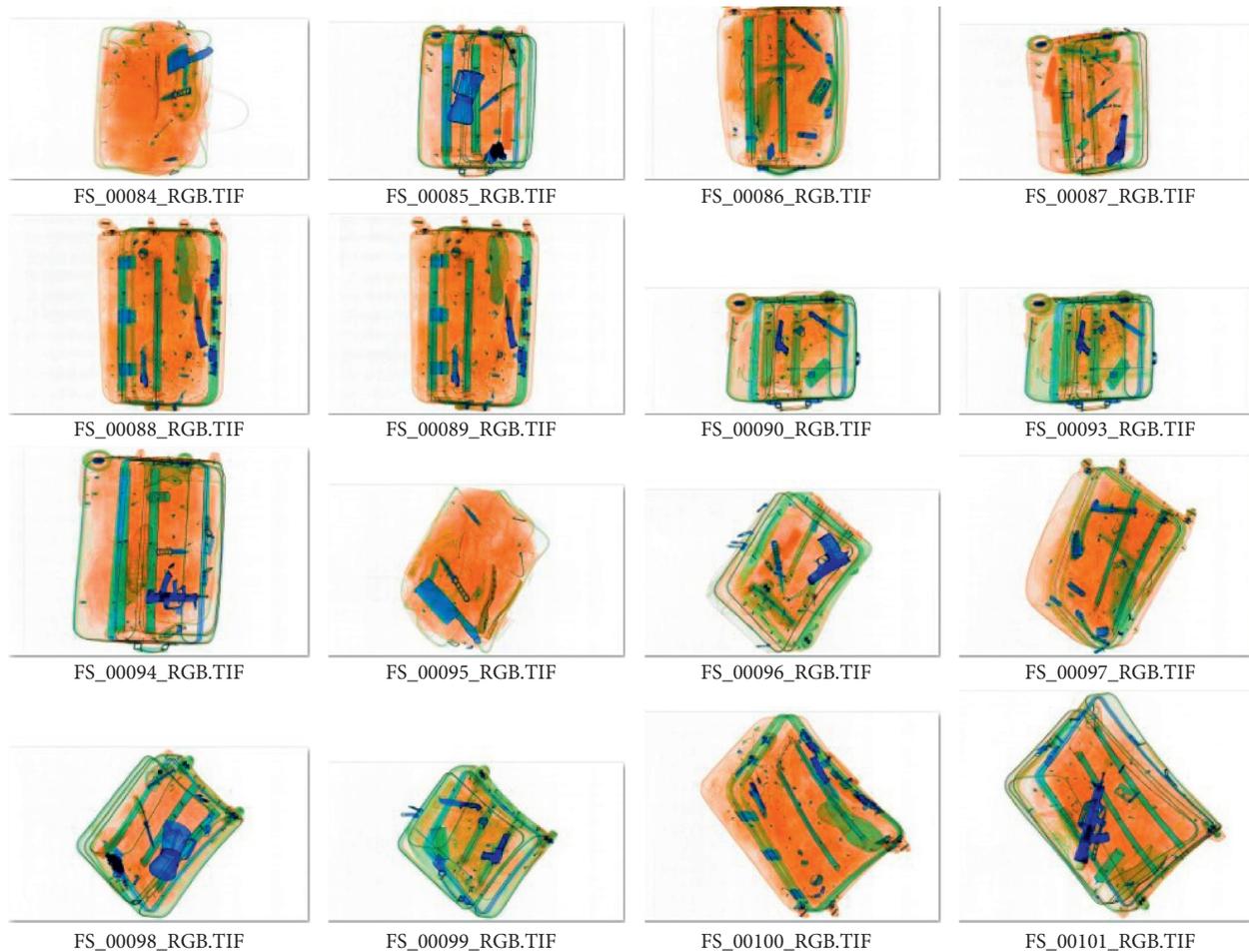


FIGURE 9: Contraband images.

models on the ILSVRC dataset and the PASCAL VOC dataset, then two classic models are selected for comparative testing on the X-ray baggage images dataset.

By comparing the speed and accuracy of various classical models in [12–18], the basic conclusion is that the accuracy of the Faster R-CNN model is higher, while the processing speed of the SSD model and the YOLO model are faster. Therefore, the Faster R-CNN model and the SSD model are selected for comparative testing on the X-ray baggage image dataset. The experimental environment is built as follows. The configuration of the high-performance graphics workstation is shown in Table 8. The operating system is Windows 10×64. The CUDA (Compute Unified Device Architecture) version is CUDA 10.0, and the CUDNN (CUDA Deep Neural Network) version is CUDNN 10.1, and the Python version is Python 3.7.

The Faster R-CNN model and the SSD model are tested on the X-ray baggage image dataset. The Caffe deep learning framework is used in the Faster R-CNN model, and the Torch deep learning framework is used in the SSD model, and both network models are the VGG-16 model. IoU (intersection-over-union) is a concept used in the target detection, which is the overlap rate between the generated candidate bound and the ground truth bound. Here, IoU is

equal to 0.5, which is used to test the X-ray baggage image dataset.

The performance of the proposed method and the prior work is evaluated by comparing the following indicators: true Positive (TP%), a positive sample is predicted to be positive, and it can be called true accuracy. True negative (TN%), a negative sample is predicted to be negative, and it can be called false accuracy. False positive (FP%), a negative sample is predicted to be positive, and it can be called false alarm rate. False negative (FN%), a positive sample is predicted to be negative, and it can be called miss rate. Precision ($PRE = TP / (TP + FP)$), which is the proportion of true positive samples in the predicted positive examples. Recall ($REC = TP / (TP + FN)$), which is the proportion of true positive samples in the actual positive examples. Accuracy ($ACC = (TP + TN) / (TP + FN + FP + TN)$), which is the proportion of true positive samples and true negative samples in the all samples. The PR curve is a precision-recall curve, and AP is the area under the PR curve; mAP (mean average precision) is the average value of each category of AP. AP measures the quality of the model in each category, and mAP measures the quality of the model in all categories.

The Faster R-CNN model is that merging RPN and Fast R-CNN into a single network by sharing their convolutional

TABLE 8: High-performance graphic workstation configuration.

Specifications	
Processor	Intel Xeon® Processor E5-2623 v3 (10M Cache, 3.00 GHz)
Memory	16 GB (2 × 8GB) DDR4 2133 MHz with ECC REG
GPU card	Nvidia Tesla Titan X
System disk	120 GB Intel® SSD Pro 2500 Series 2.5 SATA 6.0 Gb/s Solid State Drive
Storage	4 TB Seagate Enterprise Class 3.5 SATA 6.0 Gb/s

TABLE 9: Test results of the X-ray baggage image dataset.

Model	mAP (VOC)	mAP (X-ray)	FPS (VOC)	FPS (X-ray)
Faster R-CNN	73.2	62.5	7	4
SSD512	76.8	52.1	19	10

TABLE 10: Test results of contraband detection.

	TP (%)	TN (%)	FP (%)	FN (%)	PRE	REC	ACC
Guns	97.8	96.9	3.1	2.2	91.3	97.8	97.1
Knives	95.8	91.3	8.8	4.2	87.9	95.8	93.1

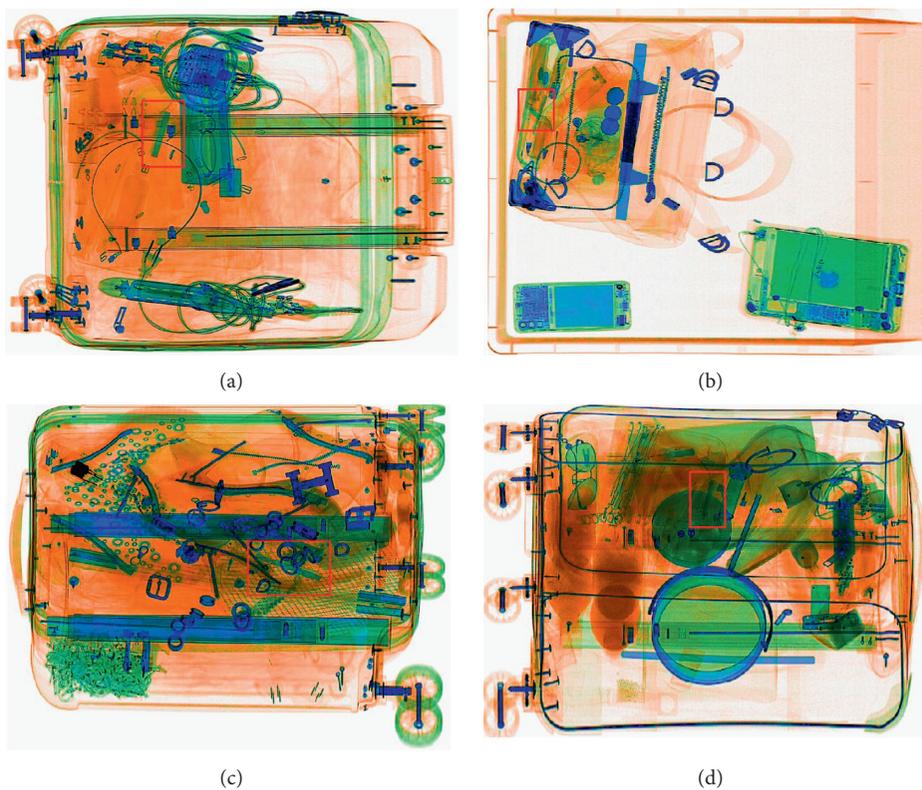


FIGURE 10: Test images.

features. The framework of the system is Caffe, and the backbone model of the system is VGG-16. The SSD model is a one-stage object detection algorithm. The framework of the system is Torch, and the backbone model of the system is VGG-16. Table 9 shows the test results of the PASCAL VOC dataset and the X-ray baggage image dataset. The mAP (VOC) and the FPS (VOC) are obtained on the PASCAL

VOC dataset, and the mAP (X-ray) and the FPS (X-ray) are obtained on the X-ray baggage image dataset. The accuracy of the Caffe + VGG-16 + Faster R-CNN model is lower on the PASCAL VOC dataset, but the accuracy of the Caffe + VGG-16 + Faster R-CNN model is higher on the X-ray baggage images dataset. The processing speed of the Torch + VGG-16 + SSD512 model is faster on the X-ray

baggage image dataset and the PASCAL VOC dataset. The processing speed of the Caffe + VGG-16 + Faster R-CNN model and the Torch + VGG-16 + SSD512 model is faster on the PASCAL VOC dataset than the X-ray baggage images dataset, because the image size of the PASCAL VOC dataset is 500×375 , and the image size of the X-ray baggage images dataset is 1024×700 .

Considering the accuracy of the Faster R-CNN model is higher than the SSD model and the processing speed of the Faster R-CNN model can also meet the current application needs, which it takes 2 seconds to collect an X-ray image. Therefore, the XMC R-CNN model was designed based on the Faster R-CNN model combined with the characteristics of the X-ray baggage images for the automatic detection of the typical guns and the typical knives in the X-ray baggage images.

Table 10 shows the test results of the XMC R-CNN model. The detection rates (TP%) of guns and knives are 96.5% and 95.8%, the miss rates (FN%) of guns and knives are 2.2% and 4.2%, and the accuracies (ACC) of guns and knives are 97.1% and 93.1%. It can be seen from the data that the X-ray image features of the guns are obvious, and all types of guns include barrels, butts, triggers, and other components, and these features are obviously different from the features of the common objects in the passenger baggage, so the detection rate of the gun is relatively high and the miss alarm rate of the gun is relatively low. On the contrary, the X-ray image features of the knives are weak, and the features of different types of knives are significantly different. In addition, these features are similar to the features of objects in the passenger baggage, such as baggage handles, baggage locks, umbrellas, and electronic equipment. Therefore, the detection rate of the knife is relatively low and the miss alarm rate of the knife is relatively high. At present, the maximum processing speed of the XMC R-CNN model is 250 milliseconds per image, and the processing speed can meet the requirement of collecting an X-ray image for 2 seconds. In Figure 10, the knives in these images are covered with some objects. The knives are not detected in the X-ray baggage image dataset without the material classifier, and these knives are detected in the XMC R-CNN model with the material classifier.

5. Conclusion and Future Work

In this work, the XMC R-CNN model is explored in the tasks of classification and detection within X-ray baggage images. The main contribution of the XMC R-CNN model is used to solve the problem of contraband detection in overlapped X-ray baggage images by the X-ray material classifier algorithm and organic stripping and inorganic stripping algorithm, and the detection rate and the miss rate that meet the requirements of screening on-site are achieved by the deep learning method. The detection rate is greater than 95%, and the miss rate is less than 5%. In some applications, it has exceeded the level of security inspectors.

The automatic detection technology of the contraband based on the XMC R-CNN model is applied to the X-ray baggage security image. According to user needs, the safe

X-ray baggage images can be automatically filtered in some specific fields, which reduces the number of X-ray baggage images that security inspectors need to screen. The efficiency of security inspection is improved and the labor intensity of security inspection is reduced. In addition, the security inspector can screen the X-ray baggage image according to the box of automatic detection, which can improve the effect of security inspection.

Future work will consider exploring the threat image projection (TIP) research in order to increase the training dataset, which can improve the accuracy of the automatic detection algorithm. On the other hand, future work will consider developing the XMC R-CNN model based on the safe objects to reduce the false alarm rate. These safe objects have the marked features and are common in passenger baggage, such as baggage handles, baggage locks, umbrellas, watches, power banks, and electronic equipment. The detection rate of the safe object is improved, and the false alarm rate of the contraband is reduced.

Data Availability

The datasets used in this work were provided by the FISCAN Systems. The dataset involving security cannot be shared.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The datasets used in this work were provided by the FISCAN Systems, and this work was partially funded by the First Research Institute of the Ministry of Public Security Project AMRSS (FRI-AX2003). This work was supported in part by the FISCAN Systems.

References

- [1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: a large-scale hierarchical image database," in *Proceedings of the Computer Vision and Pattern Recognition*, Miami, FL, USA, June 2009.
- [2] H. Dong, *Object Detection by Deep Learning: Core Technologies and Practices*, Machine Press, vol. 1, no. 1–8, Beijing, China, 2020.
- [3] M. Abadi, P. Barham, and J. Chen, "Tensorflow: a system for large-scale machine learning," in *Proceedings of the Operating Systems Design and Implementation*, vol. 16, pp. 265–283, Savannah, GA, USA, November 2016.
- [4] J. Li, *Get Started and Best Practices with TensorFlow for Deep Learning*, Vol. 1, Machine Press, Beijing, China, 2018.
- [5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proceedings of the International Conference on Neural Information Processing Systems*, Curran Associates Inc, Lake Tahoe, NV, USA, pp. 1097–1105, December 2012.
- [7] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proceedings of the Computers*

- Vision-European Conference on Computer Vision 2014*, pp. 818–833, Springer, Zurich, Switzerland, September 2014.
- [8] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proceedings of the International Conference on Learning Representations*, San Diego, CA, USA, May 2015.
- [9] C. Szegedy, W. Liu, and Y. Jia, “Going deeper with convolutions,” in *Proceedings of the Computer Vision and Pattern Recognition*, IEEE, Boston, MA, USA, pp. 1–9, June 2015.
- [10] K. He, X. Zhang, and S. Ren, “Deep residual learning for image recognition,” in *Proceedings of the Computer Vision and Pattern Recognition*, IEEE, Las Vegas, NV, USA, pp. 770–778, June 2016.
- [11] Z. Li, C. Peng, G. Yu, X. Zhang, and Y. Deng, “DetNet: a backbone network for object detection,” 2018, <https://arxiv.org/abs/1804.06215>.
- [12] R. Girshick, J. Donahue, Trevor Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the Computer Vision and Pattern Recognition*, IEEE, Columbus, OH, USA, pp. 580–587, June 2014.
- [13] R. Girshick, “Fast R-CNN,” 2015, <https://arxiv.org/abs/1504.08083>.
- [14] S. Ren, K. He, and R. Girshick, “Faster R-CNN: towards real-time object detection with region proposal networks,” in *Proceedings of the Neural Information Processing Systems*, Montreal, Canada, December 2015.
- [15] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in *Proceedings of the International Conference on Computer Vision*, IEEE, Venice, Italy, pp. 2980–2988, October 2017.
- [16] W. Liu, D. Anguelov, and D. Erhan, “SSD: single shot multibox detector,” 2016, <https://arxiv.org/abs/1512.02325>.
- [17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 2016.
- [18] J. Dai, Y. Li, and K. He, “R-FCN: object detection via region-based fully convolutional networks,” arXiv preprint arXiv:1605.06409, 2016.
- [19] S. Akçay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon, “Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2203–2215, 2018.
- [20] T. Franzel, U. Schmidt, and S. Roth, “Object detection in multi-view X-ray images,” in *Proceedings of the Joint DAGM (German Association for Pattern Recognition) and OAGM Symposium*, pp. 144–154, Graz, Austria, August 2012.
- [21] M. Bastan, W. Byeon, and T. M. Breuel, “Object recognition in multi-view dual energy X-ray images,” in *Proceedings of the British Machine Vision Conference*, Bristol, UK, September 2013.
- [22] M. Domingo and R. Vladimir, “Automated object recognition using multiple X-ray views,” 2014.
- [23] D. Mery, “Inspection of complex objects using multiple-X-ray views,” *IEEE Transactions on Mechatronics*, vol. 20, no. 1, pp. 338–347, 2014.
- [24] D. Mery and V. Riffo, “Automated object recognition in baggage screening using multiple X-ray views,” 2013.
- [25] D. Mery, “Automated X-ray object recognition using an efficient search algorithm in multiple views,” 2013.
- [26] D. Mery, G. Mondragon, V. Riffo, and I. Zuccar, “Detection of regular objects in baggage using multiple X-ray views,” *Insight-Non-destructive Testing and Condition Monitoring*, vol. 55, no. 1, pp. 16–20, 2013.
- [27] D. Mery, “Automated detection in complex objects using a tracking algorithm in multiple X-ray views,” in *Proceedings of the IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum*, pp. 41–48, Washington, DC, USA, June 2011.
- [28] G. T. Flitton, T. P. Breckon, and N. M. Bouallagu, “Object recognition using 3D SIFT in complex CT volumes,” 2010.
- [29] A. Mouton, “An evaluation of image denoising techniques applied to CT baggage screening image,” 2013.
- [30] N. Megherbi, “A comparison of classification approaches for threat detection in CT based baggage screening,” 2012.
- [31] G. Flitton, T. P. Breckon, and N. Megherbi, “A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery,” *Pattern Recognition*, vol. 46, no. 9, pp. 2420–2436, 2013.
- [32] M. Toews and W. M. Wells, “Efficient and robust model-to-image alignment using 3D scale-invariant features,” *Medical Image Analysis*, vol. 17, no. 3, pp. 271–282, 2013.
- [33] A. Mouton, “A novel intensity limiting approach to metal artefact reduction in 3D CT baggage image,” 2012.
- [34] Q. Hao, “Efficient 2D-to-3D correspondence filtering for scalable 3d object recognition,” 2013.
- [35] T.-H. Yu, O. J. Woodford, and R. Cipolla, “A performance evaluation of volumetric 3D interest point detectors,” *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 180–197, 2013.
- [36] S. Akçay, M. E. Kundegorski, M. Devereux, and T. P. Breckon, “Transfer learning using convolutional neural networks for object classification within X-ray baggage security image,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, IEEE, Beijing, China, June 2016.
- [37] T. W. Rogers, N. Jaccard, and L. D. Griffin, “A deep learning framework for the automated inspection of complex dual-energy x-ray cargo image,” in *Proceedings of the SPIE*, San Diego, CA, USA, July 2017.
- [38] N. Jaccard, T. W. Rogers, E. J. Morton, and L. D. Griffin, “Detection of concealed cars in complex cargo X-ray image using deep learning,” 2016, <https://arxiv.org/abs/1606.08078>.
- [39] <http://www.smithsdetection.com>.
- [40] <http://www.l-3com.com>.
- [41] <http://www.rapiscansystems.com>.
- [42] <http://www.as-e.com>.
- [43] <http://www.nuctech.com>.
- [44] <http://www.fri.com.cn>.
- [45] <http://www.gesecurity.com>.
- [46] <http://www.vividusa.com>.
- [47] S. Ogorodnikov and V. Petrunin, “Processing of interlaced images in 4–10 MeV dual energy customs system for material recognition,” *Physical Review Special Topics-Accelerators and Beams*, vol. 5, no. 10, Article ID 104701, 2002.
- [48] L. Li, R. Li, S. Zhang, T. Zhao, and Z. Chen, “A dynamic material discrimination algorithm for dual MV energy X-ray digital radiography,” *Applied Radiation and Isotopes*, vol. 114, pp. 188–195, 2016.
- [49] X. W. Wang, J. M. Li, and K. J. Kang, “Material discrimination by high-energy X-ray dual-energy imaging,” *High Energy Physics & Nuclear Physics*, vol. 31, no. 11, pp. 1076–1081, 2007.
- [50] V. Novikov, S. Ogorodnikov, and V. Petrunin, “Dual energy method of material recognition in high energy introscopy

- systems,” *Questions of Atomic Science and Technology [translated from Russian]*, no. 4, pp. 93–95, 1999.
- [51] M. Bastan, “Multi-view object detection in dual energy X-ray images,” *Machine Vision and Application*, vol. 26, pp. 1045–1060, 2015.
- [52] F. E. Rachard, “Development and evaluation of simulants for X-ray based explosive detection systems,” in *Proceeding of Second Explosives Detection Technology Symposium & Aviation Security Technology Conference*, Atlantic City, NJ, USA, November 1996.
- [53] Y. Jia, E. Shelhamer, J. Donahue et al., “Convolutional architecture for fast feature embedding,” 2014, <https://arxiv.org/abs/1408.5093>.