Physics Letters A 400 (2021) 127306

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

A real-time Threat Image Projection (TIP) model base on deep learning for X-ray baggage inspection

Yiru Wei, Zhiliang Zhu *, Hai Yu, Wei Zhang

Software College, Northeastern University, Shenyang, 110819, China

ARTICLE INFO

Article history:
Received 5 January 2021
Received in revised form 22 February 2021
Accepted 16 March 2021
Available online 23 March 2021
Communicated by S. Khonina

Keywords:
Threat Image Projection (TIP)
X-ray baggage inspection
Convolutional Neural Network (CNN)
Deep learning

ABSTRACT

Real-time TIP is that X-ray security machine is scanning bag while projecting X-ray image of threat object into X-ray image of bag. When TIP starts, the front part of bag is only scanned. Threat object is very likely to be projected outside of bag. Therefore, we propose a real-time TIP model. This model contains a CNN-based classifier that can predict the size type of the entire bag through the front part of bag. After predicting the size type of the bag, X-ray image of the same type of threat object is projected into X-ray image of the latter part of bag. Moreover, we propose a mapping method, which can map the real-world size of bags in optical images to the size of bag in X-ray images. In addition, our model uses a novel fusion method to project the X-ray image of threat object into the X-ray image of bag.

1. Introduction

X-ray baggage inspection aims at identifying threat objects in X-ray images, which can guarantee the safety of personnel and maintain transport security [1]. The detection of threat objects is performed by screeners who visually examine X-ray images and decide whether a bag has threat objects or not. The detection results mainly rely on the experience of screeners. Training and visual experience are important for screeners [2–4]. In order to generate a large amount of X-ray images of bags with threat objects, Threat Image Projection (TIP) was developed. TIP is that X-ray images of threat objects are projected into X-ray images of bags, by which a large number of TIP images can be obtained. Therefore, screeners can be trained with these TIP images, which can enhance their attention and alertness and thus improve threat objects detection performance.

TIP has been studied for years. It has been applied in X-ray security inspection. 2D TIP methods are developed for Computer Based Training (CBT) and conventional X-ray baggage screening systems [5–7]. It superimposes threat objects from a threat database into X-ray images of bags on a random location. A TIP framework is proposed in cargo transmission X-ray imagery [8]. The method uses the approximately multiplicative nature of X-ray images to generate a library of threat objects. These threat objects can be inserted into X-ray images of cargos. This framework can be used to train Machine Learning (ML) based automated threat detection algorithms or to train safety screeners to identify threat objects in X-ray images. In addition, TIP technology is also extended to 3D Computed Tomography (CT) screening systems [9–11]. 3D TIP methods automatically determine an appropriate place for insertion of the threat object.

These methods have achieved good performances and the generated TIP images are very similar to the real-world X-ray images. However, these earlier methods can only be applied in non-real-time situations, in which the entire bags have been scanned and thus the X-ray images of bags are entire. Then, X-ray image of threat object can be easily projected into the X-ray image of entire bag. Real-time TIP is mean that X-ray security machine is scanning bag while projecting X-ray image of threat object into X-ray image of bag. Therefore, when TIP starts, the front part of bag is only scanned and the X-ray image of threat object will be projected into X-ray image of the latter part of bag. Since the size of the entire bag is unknown, threat object is very likely to be projected outside of bag, which generates wrong TIP images. Therefore, it is crucially important to predict the size type of the entire bag in real-time TIP.

Recently, deep learning has shown promising results in many image-based tasks. Convolutional Neural Networks (CNNs) [12] are the derivatives of deep learning, which have been widely used in various applications, such as medical image analysis and applications [13–15], face detection [16], speech recognition [17], pose estimation [18] and other computer vision tasks. This leads real-time TIP to explore the deep learning methods to generate accurate TIP images.

* Corresponding author.
E-mail addresses: weiyiru0228@yeah.net (Y. Wei), zzl_neu@yeah.net (Z. Zhu).

https://doi.org/10.1016/j.physleta.2021.127306
0375-9601/

The humans can rapidly recognize the size type of the entire bag, *i.e.* large-sized, small-sized, or medium-sized, when given the front part of bag, such as a corner. Therefore, we propose a real-time TIP model to simulate the native ability of human vision. Our model contains a CNN-based classifier, which can predict the size type of the entire bag by the front part of bag. After predicting the size type of the entire bag, X-ray image of the same type of threat object is projected into X-ray image of the latter part of bag.

The existing CNNs, *e.g.*, VGG16 [19] and ResNet [20], contain a large number of parameters and thus cannot be directly applied to our real-time TIP model. Therefore, instead of using the existing CNNs, we construct a special CNN-based classifier in our model. It can take X-ray image of the front part of bag as input to predict the size type of the entire bag, *i.e.*, large, medium and small. The fundamental idea behind is based on the principle that X-ray images of threat objects are projected into X-ray image of the same type of bags, *e.g.*, large-sized threat images are projected into large-sized bags images or small-sized threat images are projected into small-sized bags images, which is favorable to generate accurate TIP images. Therefore, the CNN-based classifier can greatly improve the accuracy of real-time TIP images.

Moreover, because the CNN-based classifier needs a great deal of X-ray images of bags to train and real-world X-ray images are not public, we propose a mapping method that can map the real-world size of bags in optical images to the size of bag in X-ray images. This method can generate a large number of simulated X-ray images of bags. In addition, our real-time TIP model uses a novel fusion method to project the X-ray images of threat objects into X-ray images of bags. In general, our proposed model can bring benefits to both parties – our model can produce a large number of accurate TIP images that can greatly rich the TIP images library; with the aid of real-time TIP model, screeners can be trained to recognize threat objects in real-time TIP images.

The main contributions of this paper are as follows:

– We propose a real-time TIP model based on deep learning. Different from previous TIP methods, our model contains a CNN-based classifier, which can predict the size type of the entire bag through the front part of bag. According to the size type of bag, X-ray image of the same type of threat object is projected into the X-ray image of the latter part of bag.

– We propose a mapping method that can map the real-world sizes of bags in optical images to the sizes of bags in X-ray images. Using this method, we prepare a dataset of simulated X-ray images of bags that could be used to train the CNN-based classifier.

– Our model employs a novel fusion method that can project the X-ray images of threat objects into X-ray images of bags. The resulting TIP image is difficult to distinguish from real-world X-ray image of bag with threat object.

– Our proposed model can overcome the limitations of real-time TIP technology and achieve the state-of-the-art performance on real-world X-ray images of bags. The experimental results show that the projection precision of our proposed model can achieve 96.74% that is better than the random projection method.

2. Our proposed real-time TIP model

In this section, we will describe the proposed model in detail. In Sec. 2.1, the overall structure of the model will be introduced. In Sec. 2.2, we prepare a dataset of simulated X-ray images of bags. In Sec. 2.3, we construct a CNN-based classifier for size classification of X-ray images of the front parts of bags. In Sec. 2.4, a new fusion method will be introduced.

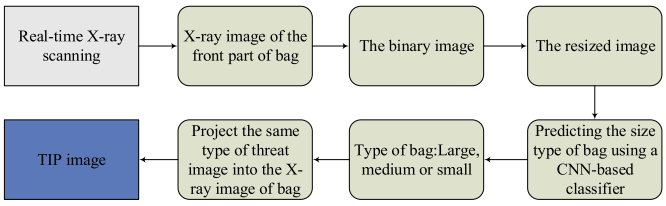

Fig. 1. Flowchart of the overall scheme of model.

Algorithm 1 Real-time TIP model.

1: Input B: the X-ray image of the front part of bag(The first 60 rows)
2: Binary image of B: D
3: Resized image to the size of 64 × 12 pixels: E
4: The size type of E is predicted by a CNN-based classifier
5: A threat object T is picked according to size type of bag B
6: X-ray image of threat object T is projected into X-ray image of bag B in real time
7: Output: S, a new X-ray image of bag with a threat object

2.1. Overview of model

The overall scheme of the model is described in Fig. 1. The CNN-based classifier takes X-ray image of the front part of bag as input rather than the whole image. According to the classification result, X-ray image of the same type of threat object is projected into X-ray image of the latter part of bag. The detailed process of real-time TIP model is summarized in Algorithm 1. To illustrate the model step-by-step, Fig. 2 shows the steps of Algorithm 1.

Since the size of the entire bag is unknown in real-time TIP, threat object is often projected the outside of bag. Fig. 3 shows the failed TIP. Our proposed model can solve this problem effectively. The main principle is that the CNN-based classifier can predict the size type of the entire bag through the front part of bag, and then the X-ray image of the same type of threat object is projected into X-ray image of the latter part of bag in real-time. Fig. 4 shows the X-ray images of three different types of threat objects are projected into X-ray images of the same type of bags.

2.2. Data preparation

A large number of X-ray images of passenger bags are not public. Furthermore, we only need to classify the size types of bags and are uninterested in the content of bags. Therefore, we propose a mapping method that can map real-world size of bags in optical images to the sizes of bags in X-ray images, by which a lot of X-ray images of bags can be produced.

After the X-ray beam goes through the bags, the X-ray residual energy reaches the X-ray flat panel detector. X-ray flat panel detector be composed of many small pixels, which is shown in Fig. 5. The residual X-ray is first transformed into visible light, then into electrons and finally into digital signals. The digital signals reach the computer and be transformed into pixel values of X-ray images. Size of small pixel reflects the size of each pixel of X-ray image. In our case, the size of small pixel is 0.16 cm. If real-world width a and height b (both in centimeters) of a bag in optical image are known, pixel width k and height d of the bag in the simulated X-ray image can be computed using Eq. (1):

$$\frac{a}{k} = \frac{b}{d} = 0.16 \tag{1}$$

In Fig. 6, the first image is an optical image of bag, and the second image is a binary image. The real-world width and height of the bag in the first image are 26 cm and 17 cm. Using Eq. (1), pixel width and height of this bag in the X-ray image are 162 and 106, respectively. The third image is the simulated X-ray image.

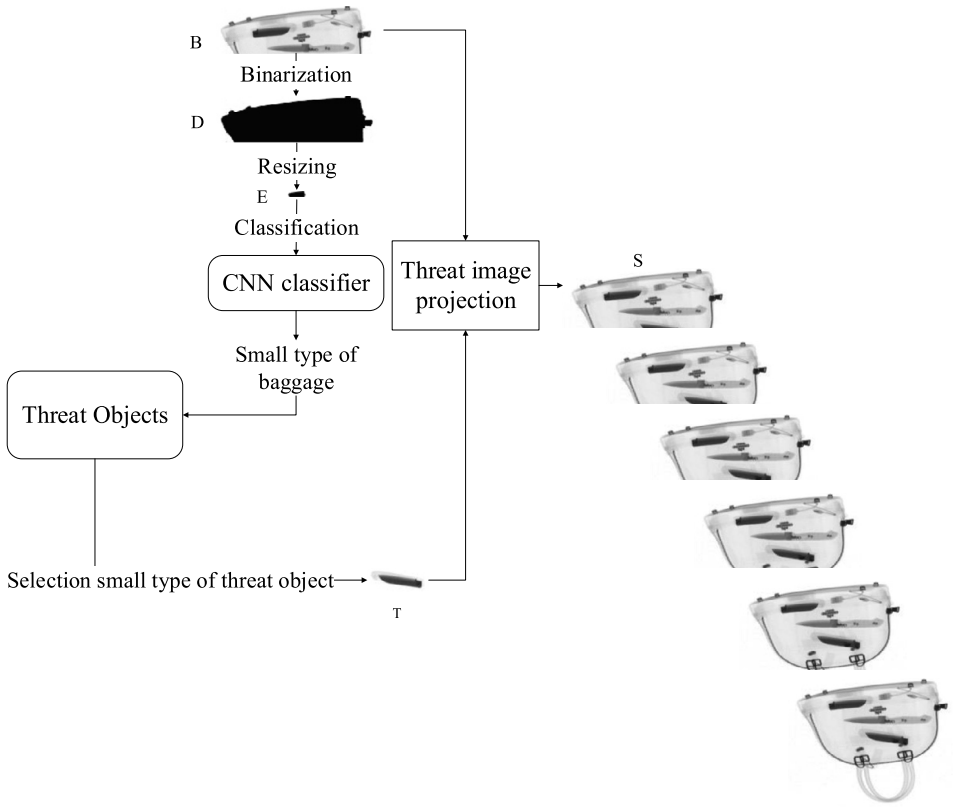

Fig. 2. Steps of Algorithm 1.

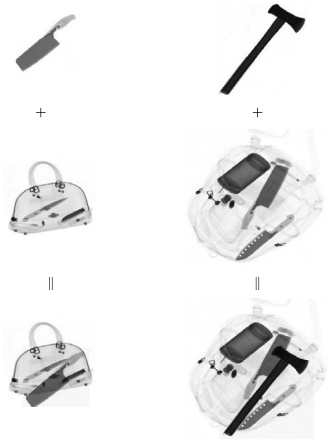

Fig. 3. Examples of failed TIP.

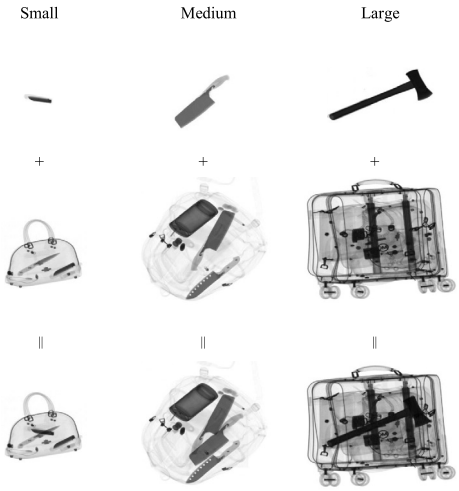

Fig. 4. The X-ray images of large, medium and small types of threat objects are projected into the X-ray images of the same types of bags.

We collected 800 representative bag images. Using Eq. (1), different size types of bags images are converted into simulated X-ray images and then are classified into three types: large, medium and small. The longest side of large-sized bags is longer than 50 cm. The longest side of medium-sized bags ranges from 30 cm to 50 cm. The longest side of small-sized bags is less than 30 cm. The simulated X-ray images of three different types of bags are shown in Fig. 7.

Our real-time TIP model contains a CNN-based classifier. The purpose of this classifier is to predict the size type of the entire bag through the front part of bag. Therefore, we use the first 60 rows pixels of bag image to simulate the front part of the bag. Through this method, the training data of the CNN-based classifier is generated. Specific operation processes are illustrated as follows:

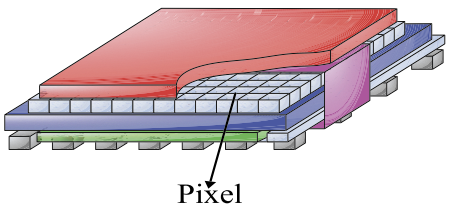

Fig. 5. Structure of an X-ray flat panel detector.

Step 1: The simulated X-ray images of bags are rotated in six directions (45°, 90°, 135°, 180°, 225°, and 270°).

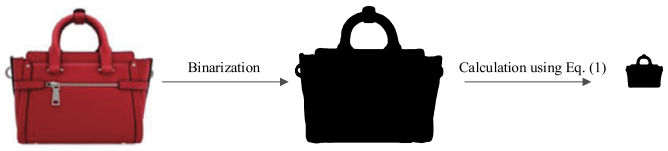

Fig. 6. Process of obtaining simulated X-ray image of bag.

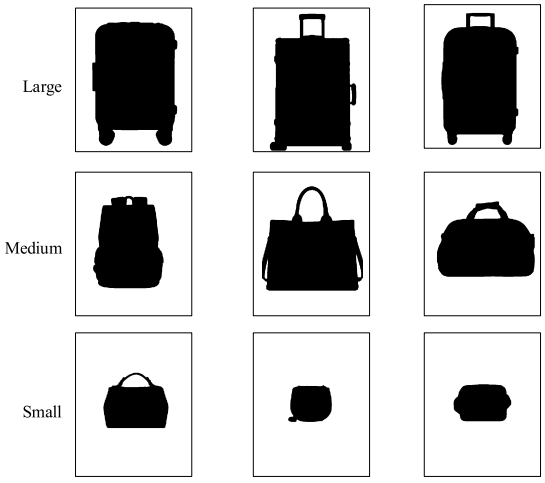

Fig. 7. The simulated X-ray images of three different types of bags.

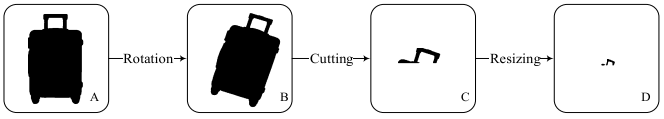

Fig. 8. The preparation process of the training data.

Step 2: The first 60 rows pixels of bags images are intercepted from the rotated images, which represent the front part of bags. In Sec. 3.4, we compare the training accuracy and validation accuracy of CNN-based classifiers that are trained by different size of bags image (the first 30, 45, 60, 75 and 80 rows). The experimental results show the classifier that be trained by the first 60 rows pixels of bags images achieves best accuracy. This parameter cannot be set too small or too large. If the front part of bag is set too small, which is bad to predict the size type of the entire bag. Conversely, the front part of bag is set too large, which is easy to generate the wrong TIP images.

Step 3: All images are resized to the size of 64 $\times$ 12 pixels.

Ultimately, we obtain 15873 simulated X-ray images of bags, including 5291 large-sized bags images, 5291 medium-sized bags images and 5291 small-sized bags images. The preparation work of the training data is completed. The whole process is shown in Fig. 8.

In Fig. 9, we randomly selected 30 images from each type of bag image; the first line shows 30 the front parts of large-sized bags images, the second displays 30 the front parts of medium-sized bags images, and the last presents 30 the front parts of small-sized bags images.

Since these are three types of bags, threat objects are also divided into three types: large, medium and small. The sizes of threat objects are smaller than the sizes of the same type of bags. The longest side of large-sized threat object ranges from 30 cm to 50 cm. The longest side of medium-sized threat object ranges from 10 cm to 30 cm. The longest side of small-sized threat object is less than 10 cm. In Fig. 10, we show four images from each type.

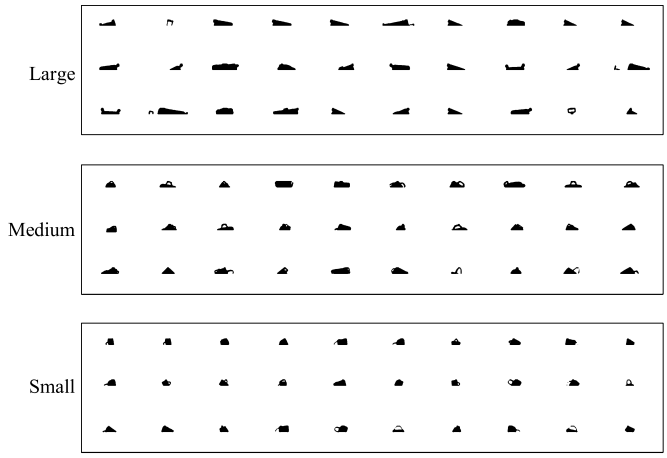

Fig. 9. The front parts of three types of bags images.

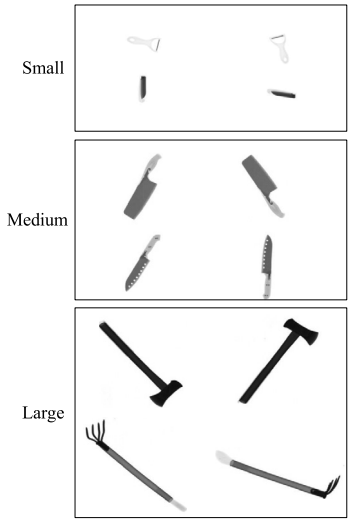

Fig. 10. Three types of threat objects.

2.3. Constructing a CNN-based classifier for size classification of the front parts of bags

The existing CNN models, such as VGG16 [19] or ResNet [20], are too complex and time-consuming for our real-time TIP. Therefore, we construct a low cost, high-performance the CNN-based classifier, which is shown in Fig. 11.

The CNN-based classifier is composed of one input layer, five convolutional layers, five batchNorm layers, one pooling layer, and one output layer. Convolutional layers perform convolutions on the images using small-sized kernels, which can extract a large number of features hidden in the images. After the features have been extracted, they are connected to the outputs. The specific parameters of each layer are shown in Table 1.

The classifier takes the front part of bag, rather than the whole bag, as input and extracts the features. The size classification of bags is a three-classification problem, where the input is a binary image, and the output is one of three labels, indicating the large, medium or small type of bag. The configurations of the CNN-based classifier are described in detail as follows.

Input layer: In this research, we used binary images of the size of 64 $\times$ 12 pixels as inputs.

Convolutional layer: In this layer, convolution operation is performed on input images or feature maps. The output feature map is computed using Eq. (2):

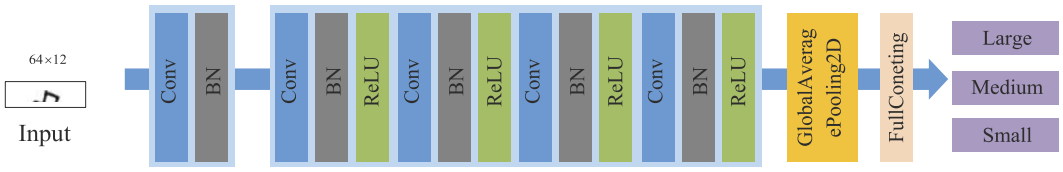

Fig. 11. Overall architecture of the proposed CNN-based classifier.

$$Z_{k+1} = (W_k * Z_k) + b_k \tag{2}$$

where W_k is a kernel, $*$ denotes the convolution operator, Z_{k+1} and Z_k are the output and input of the convolutional layer, and b_k is the bias term. *BatchNorm2d:* BatchNorm2d is used as the normalization layer of the network. It can accelerate the speed of network convergence. The principle is given by Eq. (3).

$$y = \frac{x - mean[x]}{\sqrt{var[x] + \varepsilon}} \times \gamma + \beta \tag{3}$$

where x is the input data, $mean[x]$ is the mean value, and $var[x]$ is the variance, γ is the normalized weight, β is the normalized bias, y is the normalization result. ε is a variable that can prevent zero in the denominator.

Rectified Linear Units (ReLU): The ReLU [21] is an activation function. It can change the negative value of some neurons to zero, which can prevent overfitting. Convolutional layers and fully connected layers both use ReLU as the activation function. It applies the following function to all outputs:

$$\mathrm{f}(x) = \max(0, x) \tag{4}$$

Padding: Zero padding cells are added around the input image to keep the edge information.

Stride: Stride can be used to control the size of the feature map. A shift by 'n' pixels is performed on the convolutional layer.

GlobalAveragePooling2D: It reduces the dimensionality from 3D to 1D and outputs 1 response for every feature map.

Output layer: It is a fully connected layer [22] containing 3 neurons, which can map the learned features to the sample labels. SoftMax is often used in multiple classification tasks. It normalizes the outputs of multiple neurons to the interval of (0, 1), so the output of it can be regarded as a probability to carry out multiple classification. Therefore, SoftMax is employed to output the probability of each class. The probability that the output belongs to class-i is computed as follows:

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{5}$$

where p_i denotes the output probability of class-i, z_i indicates the output of neuron-i, and n denotes the number of categories.

Loss function: The cross-entropy loss is used as the loss function to measure the distance between the output and the ground truth. The smaller the cross-entropy loss is, the closer the output and the ground truth are to each other. It is illustrated as Eq. (6):

$$\mathrm{L} = -\sum_{i=1}^{n} y_i \log(p_i) \tag{6}$$

where y_i is the ground truth, and p_i is output probability.

2.4. A new fusion method for X-ray images of threat objects and X-ray images of bags

During X-ray baggage inspection, the Beer-Lambert absorption law characterizes the intensity distribution of X-rays through bags.

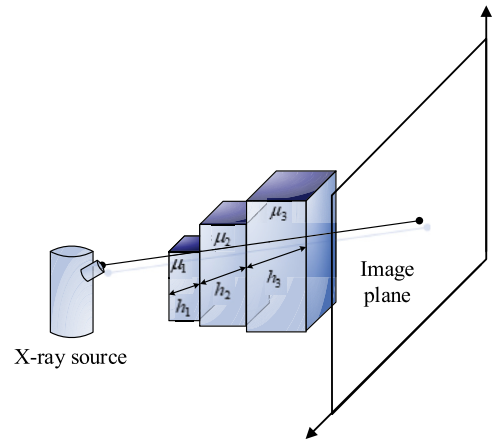

Fig. 12. X-ray beam passage through substances with different absorption coefficients.

Different substances in the bag absorb a certain amounts of energy of the X-ray beam. The residual energy reaches the X-ray flat panel detector and is transformed into digital images. In Fig. 12, after the X-ray beam passes through an example with n=3 substances, the residual energy can be expressed using Eq. (7):

$$\varphi = \varphi_0 exp(-\sum_{i=1}^{n} \mu_i h_i) \tag{7}$$

where φ_0 is the incident X-ray energy intensity, n is the number of substances in the bag, μ is the absorption coefficient, h is the thickness of the irradiated substance, and φ is the residual energy intensity.

When the incident X-ray beam passes through the threat object and the bag, respectively, the residual energy intensity can be expressed as Eq. (8) and Eq. (9):

$$\varphi_f = \varphi_0 e^{-\mu_f h_f} \tag{8}$$

$$\varphi_b = \varphi_0 e^{-\mu_b h_b} \tag{9}$$

In this case, in Eq. (8), φ_f is the residual energy intensity after passage through the threat object, μ_f is the absorption coefficient of threat object, and h_f is the thickness of the threat object. In Eq. (9), φ_b is the residual energy intensity after passage through the bag, μ_b is absorption coefficient of the bag, and h_b is the thickness of the bag.

Following Eq. (7), when an incident X-ray beam passes through the bag with a threat object, we can regard that the incident X-ray beam firstly passes through the threat object and the residual energy density subsequently passes through other materials of the bag. The residual energy intensity can be modeled by Eq. (10):

$$\varphi_t = \varphi_0 e^{-\mu_f h_f} e^{-\mu_b h_b} \tag{10}$$

It is clear that φ_t is the final residual energy intensity after the incident beam has passed through the bag with a threat object. According to Eq. (8) and Eq. (9), Eq. (10) can also be written as Eq. (11):

Table 1
Specific parameters of the proposed network model.

Layer	Kernel size	Stride	Padding	Feature map size	Activate function
Input	–	–	–	64 × 12	–
Conv2d	3	2	1	32 × 6 × 32	–
BatchNorm2d	–	–	–	32 × 6 × 32	–
Conv2d	3	1	1	32 × 6 × 64	–
BatchNorm2d	–	–	–	32 × 6 × 64	ReLU
Conv2d	3	1	1	32 × 6 × 64	–
BatchNorm2d	–	–	–	32 × 6 × 64	ReLU
Conv2d	3	2	1	16 × 3 × 128	–
BatchNorm2d	–	–	–	16 × 3 × 128	ReLU
Conv2d	3	1	1	16 × 3 × 256	–
BatchNorm2d	–	–	–	16 × 3 × 256	ReLU
GlobalAveragePooling2D	2	1	1	256	–
Fully connected	–	–	–	3	ReLU

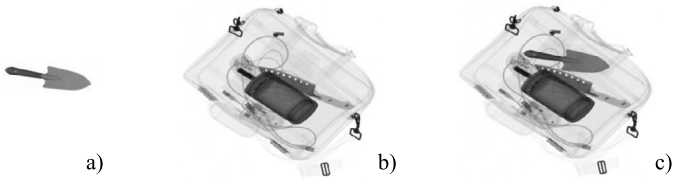

Fig. 13. The X-ray image of threat object is projected into the X-ray image of the cluttered bag: a) the X-ray image of threat object; b) the X-ray image of the cluttered bag; c) the new X-ray image of bag containing threat object.

$$\varphi_t = \varphi_f \cdot \frac{\varphi_b}{\varphi_0} \tag{11}$$

Due to X-ray energy intensity being proportional to pixel values of X-ray digital image, Eq. (11) can be transformed into (12):

$$I_t = \frac{I_f \cdot I_b}{A} \tag{12}$$

where I_f is the pixel value of the threat image, I_b is the pixel value of the bag image, and I_t is the pixel value of the bag image with the threat object. Parameter A can be obtained using a calibration approach [23]. In our case, A=15500. It is clear that Eq. (12) can be used to generate new X-ray images. In Fig. 13 a is X-ray image of the threat object, in Fig. 13 b is X-ray image of the cluttered bag, and in Fig. 13 c is the new X-ray image of bag with the threat object, generated by Eq. (12).

3. Experiments

To demonstrate the effectiveness of the real-time TIP model, we carry out a series of experiments. Firstly, we introduce the training and testing of the CNN-based classifier, and then verity its classification accuracy on real-world X-ray images of bags. Secondly, we show the overall process of our proposed real-time TIP model. Finally, we evaluate our model and compare it with the random projection method.

3.1. Implementation details

The experiments were run on ubuntu 16.04. The experimental codes were implemented in Python 0.4.1. We used one NVIDIA Titan-V GPU (with 12GB memory) for training and testing of the CNN-based classifier. During training, mini-batch is set to 16. The base learning rate is set to 0.001 and the moment is set to 0.9 for adjusting the learning rate. Since the CNN-based classifier converges after 50 epochs, the total number of iterations is set to 50 epochs. In order to optimize the network, we employ Adaptive Moment Estimation (Adam) [24] as optimizer.

Table 2
Description of dataset.

	Large	Medium	Small	Total
Training set	3386	3387	3385	10158
Validation set	847	845	848	2540
Testing set	1058	1059	1058	3175

3.2. Dataset

In Sec. 2.2, we prepare 15873 simulated X-ray images of the front parts of bags, of which the proportion of large-sized bags images, medium-sized bags images and small-sized bags images is 1 : 1 : 1. These images are first arranged in an unordered order and then are divided into training set, validation set and testing set in the approximate proportion of 8 : 1 : 1. The description of the dataset is shown in Table 2.

3.3. Evaluation metrics

We evaluate the CNN-based classifier by classification accuracy. It can be illustrated as follows:

$$\text{Accuracy} = \frac{TC}{N} \tag{13}$$

where TC denotes the number of the bags images that are correctly classified, N denotes the total number of bags images, and Accuracy presents the classification accuracy.

We evaluate our proposed model by projection precision. It can be defined as follows:

$$\text{Precision} = \frac{TP}{N} \tag{14}$$

where TP stands for the number of the correct TIP images, N denotes the total number of bags images, and Precision presents the projection precision.

3.4. Training and testing of the CNN-based classifier

We use different size of bags image (the first 30, 45, 60, 75 and 80 rows pixels of bags images) to train the CNN-based classifier and compare their classification accuracy. The comparison results are shown in Table 3, from which we can see that the classifier is trained by the first 60 rows pixels of bags images achieves the best result. Meanwhile, Fig. 14 shows the training and validation processes of it. The accuracy curves of training data and validation data are consistent and rise steadily and the loss curves decline steadily. When the training reaches 50 epochs, the loss curve of the validation data tends to flatten, which indicates that the loss has reached the minimum value. Therefore, the CNN-based classifier

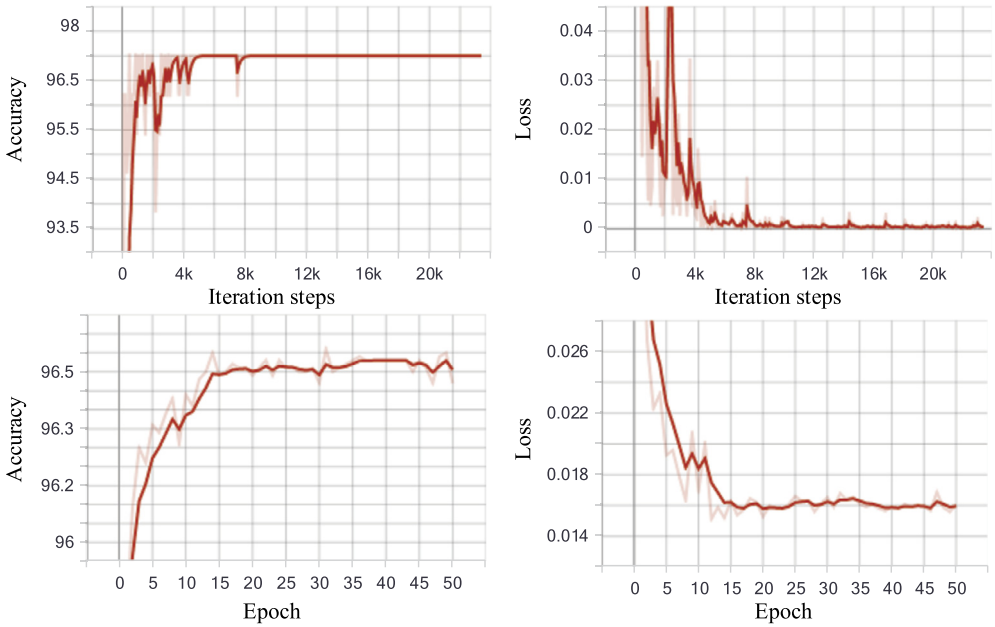

Fig. 14. Result curves: the first line is the accuracy and loss values of training phase; the second line is the accuracy and loss values of validation phase.

Table 3
Accuracy comparisons of the CNN-based classifiers trained with different size of bags images. The classifier_N denotes the CNN-based classifier trained by the first N rows pixels of bags images.

	Training accuracy	Validation accuracy
The Classifier_30	66.26%	63.06%
The Classifier_45	78.91%	76.98%
The Classifier_60	96.87%	96.2%
The Classifier_75	91.46%	90.67%
The Classifier_90	83.57%	82.72%

Table 4
Classification accuracy of the CNN-based classifier on testing set.

	Large	Medium	Small	Total
Accuracy(first part)	332/340	327/341	330/341	989/1022
Accuracy(second part)	358/371	351/371	357/370	1066/1112
Accuracy(third part)	334/347	330/347	331/347	995/1041
Accuracy(average)	96.79%	95.18%	96.21%	96.06%

trained by the first 60 rows pixels of bags images is suitable for our classifier.

After training, the classifier is tested on the testing data. The testing data is divided into three parts. The first part includes 1022 images (340 large, 341 medium and 341 small). The second includes 1112 images (371 large, 371 medium and 370 small). The third includes 1041 images (347 large, 347 medium and 347 small). The classification accuracy is shown in Table 4. Average classification accuracy can reach 96.06%.

3.5. Classification of real-world X-ray images of bags using our CNN-based classifier

In this section, our proposed classifier is tested on real-world X-ray images of bags. We prepare 600 X-ray images of bags, including 202 large-sized bags images, 200 medium-sized bags images and 198 small-sized bags images. All images must be initialized and then fed into the classifier. The initialization process is shown as follows:

Step 1: The minimum enclosing rectangle image is extracted using edge detection.

Step 2: Binarization.

Table 5
Classification accuracy of real-world X-ray images of bags using our CNN-based classifier.

	Large	Medium	Small	Total
	195/202	190/200	190/198	575/600
Accuracy	96.53%	95%	95.96%	96%

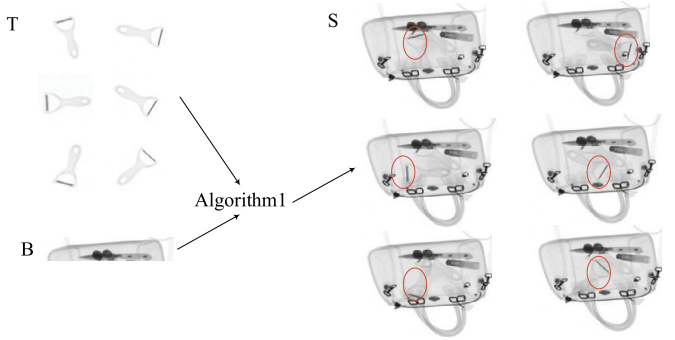

Fig. 15. Process of projecting small-sized threat images (T) in different directions into the X-ray image of the latter part of small-sized bag (B); the TIP images (S) include an original threat object and a projected threat object (note the red circles in the images). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Step 3: The first 60 rows pixels of bags images are intercepted, which represents the front part of bags images.

Step 4: All images be resized to the size of 64 × 12 pixels.

The classification results are shown in Table 5. The overall classification accuracy can reach 96%; in particular, accuracy for large-sized bags reaches 96.53%, that for medium-sized bags reaches 95%, and that for small-sized bags reaches 95.96%. Thus, our proposed CNN-based classifier is effective.

3.6. Real-time TIP

The CNN-based classifier firstly predicts the size type of the entire bag and then X-ray image of the corresponding type of threat object is projected into X-ray image of the latter part of bag. In this section, we show real-time TIP using Algorithm 1. Moreover,

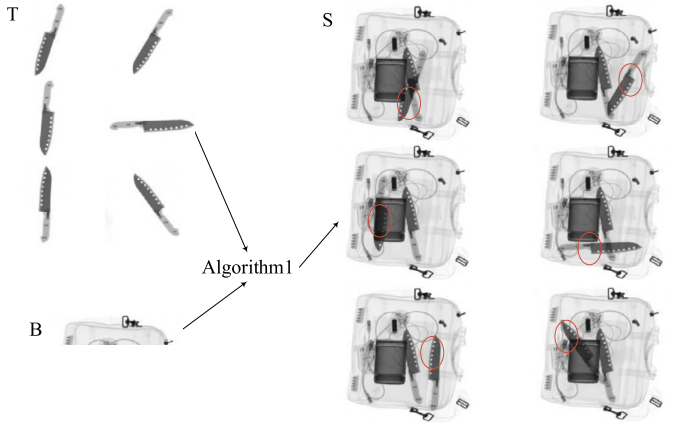

Fig. 16. Process of projecting medium-sized threat images (T) in different directions into the X-ray image of the latter part of medium-sized bag (B); the TIP images (S) include an original threat object and a projected threat object (note the red circles in the images).

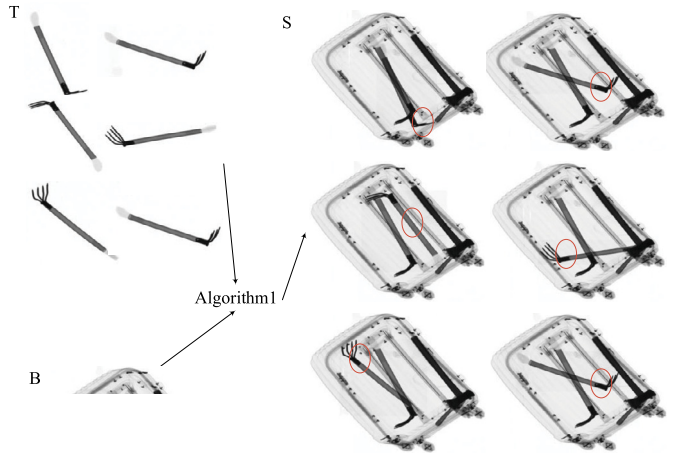

Fig. 17. Process of projecting large-sized threat images (T) in different directions into the X-ray image of the latter part of large-sized bag (B); the TIP images (S) include an original threat object and a projected threat object (note the red circles in the images).

in order to demonstrate the similarity between the TIP images and the original X-ray of images of bags, original X-ray images of bags contain an original threat object, and the TIP images contain an original threat object and a projected threat object. The results are shown in Fig. 15, 16 and 17. The experimental results demonstrate that our model can project threat objects into bags accurately. Additionally, both threat objects – original and projected – are so similar that it is difficult to distinguish which one is the original and which one is the projected threat object.

3.7. Comparative experiment

Since our proposed model is first one to address the problem of accuracy low of real-time TIP, there is no other methods to compare. Therefore, in order to demonstrate the effectiveness of our proposed model, we compare our model with the random projection method. The random projection method is that threat images are randomly selected and projected into X-ray images of the latter part of bags during X-ray baggage scanning. The entire threat objects are projected into bags, which indicates the success of TIP operation and can generate correct TIP images.

In this experiment, both methods are tested on 706 different size types of bags (230 large, 236 medium and 240 small). The obtained results (precision) are summarized in Table 6. Results show that our model can reach a very high precision, outperforming the random projection method. Therefore, our proposed model can significantly enhance the accuracy of real-time TIP images and effectively solve the problem of real-time TIP failure.

Table 6
Comparisons of projection precision between our real-time TIP model and random projection method. The first row (Precision(large)) denotes the projection precisions of the two method that both are tested on large-sized bags images; the second row (Precision(medium)) denotes the projection precisions of the two method that both are tested on medium-sized bags images; the third row (Precision(small)) denotes the projection precisions of the two method that both are tested on small-sized bags images.

Method	Random projection method	Ours
Precision(large)	94.78% (218/230)	96.08% (221/230)
Precision(medium)	39.86% (94/236)	97.46% (230/236)
Precision(small)	34.58% (83/240)	96.67% (232/240)
Precision(average)	55.95%	96.74%

4. Conclusion

In this paper, we propose a real-time TIP model based on deep learning for X-ray baggage inspection. Different with previous TIP methods which project threat image into the X-ray image of bag that has been scanned, our model can address the limitation of real-time TIP. Our model takes a CNN-based classifier to predict the size type of the entire bag through the front part of bag and then projects X-ray image of the same size type of threat object into X-ray image of the latter part of bag during X-ray scanning bag. Moreover, a mapping method is proposed that can map the real-world sizes of bags in optical images to the sizes of bags in X-ray images. Using this method, a large number of simulated X-ray images of bags are produced, which are used to train the CNN-based classifier. In addition, our model employs a novel fusion method to project X-ray image of threat object into X-ray image of bag. Experimental results demonstrate the effectiveness of our real-time TIP model. Therefore, our model can be used to train screeners to recognize threat objects in real-time TIP images and can be applied to automated detection of threat objects research in the future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant Nos. 61977014, 61902056, 61603082).

References

[1] Y.F.A. Gaus, N. Bhowmik, T.P. Breckon, On the use of deep learning for the detection of firearms in X-ray baggage security imagery, in: 2019 IEEE International Symposium on Technologies for Homeland Security (HST), 2019, pp. 1–7.
[2] S. Michel, N. Hättenschwiler, M. Zeballos, A. Schwaninger, Comparing e-learning and blended learning for threat detection in airport security X-ray screening, in: 2017 International Carnahan Conference on Security Technology (ICCST), 2017, pp. 1–6.
[3] Y. Sterchi, N. Hättenschwiler, S. Michel, A. Schwaninger, Relevance of visual inspection strategy and knowledge about everyday objects for X-ray baggage screening, in: 2017 International Carnahan Conference on Security Technology (ICCST), 2017, pp. 1–6.
[4] N. Hättenschwiler, S. Merks, A. Schwaninger, Airport security X-ray screening of hold baggage: 2D versus 3D imaging and evaluation of an on-screen alarm resolution protocol, in: 2018 International Carnahan Conference on Security Technology (ICCST), 2018, pp. 1–5.
[5] R. Riz á Porta, Y. Sterchi, A. Schwaninger, Examining threat image projection artifacts and related issues: a rating study, in: 2018 International Carnahan Conference on Security Technology (ICCST), 2018, pp. 1–4.

[6] V. Cutler, S. Paddock, Use of threat image projection (TIP) to enhance security performance, in: 43rd Annual 2009 International Carnahan Conference on Security Technology, 2009, pp. 46–51.
[7] S.M. Steiner-Koller, A. Bolfing, A. Schwaninger, Assessment of X-ray image interpretation competency of aviation security screeners, in: 43rd Annual 2009 International Carnahan Conference on Security Technology, 2009, pp. 20–27.
[8] T.W. Rogers, N. Jaccard, E.D. Protonotarios, J. Ollier, E.J. Morton, L.D. Griffin, Threat image projection (tip) into X-ray images of cargo containers for training humans and machines, in: 2016 IEEE International Carnahan Conference on Security Technology (ICCST), 2016, pp. 1–7.
[9] N. Megherbi, T.P. Breckon, G.T. Flitton, A. Mouton, Fully automatic 3D threat image projection: application to densely cluttered 3D computed tomography baggage images, in: 2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA), 2012, pp. 153–159.
[10] N. Megherbi, T.P. Breckon, G.T. Flitton, Radon transform based automatic metal artefacts generation for 3D threat image projection, in: Optics and Photonics for Counterterrorism, Crime Fighting and Defence IX; Optical Materials and Biomaterials in Security and Defence Systems Technology X, vol. 8901, 2013, pp. 94–102.
[11] K.K. Kishan, K.V.M. Prashanth, Techniques for detecting and tracking of baggages in airports, in: 2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT), 2017, pp. 333–338.
[12] A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 84–90 (6) (2017).
[13] P. Chriskos, C.A. Frantzidis, P.T. Gkivogkli, P.D. Bamidis, C. Kourtidou-Papadeli, Automatic sleep staging employing convolutional neural networks and cortical connectivity images, IEEE Trans. Neural Netw. Learn. Syst. 31 (1) (2020) 113–123, https://doi.org/10.1109/TNNLS.2019.2899781.
[14] H. Kang, Accelerator-aware pruning for convolutional neural networks, IEEE Trans. Circuits Syst. Video Technol. 30 (7) (2020) 2093–2103, https://doi.org/10.1109/TCSVT.2019.2911674.
[15] H. Chen, C. Wu, B. Du, L. Zhang, L. Wang, Change detection in multisource VHF images via deep Siamese convolutional multiple-layers recurrent neural network, IEEE Trans. Geosci. Remote Sens. 58 (4) (2020) 2848–2864, https://doi.org/10.1109/TGRS.2019.2956756.
[16] J. Deng, J. Guo, S. Zafeiriou, Single-stage joint face detection and alignment, in: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 1836–1839.
[17] M. Yousefi, J.H.L. Hansen, Block-based high performance CNN architectures for frame-level overlapping speech detection, IEEE/ACM Trans. Audio Speech Lang. Process. 29 (2021) 28–40, https://doi.org/10.1109/TASLP.2020.3036237.
[18] X. Du, T. Kurmann, P. Chang, M. Allan, S. Ourselin, R. Sznitman, J.D. Kelly, D. Stoyanov, Articulated multi-instrument 2-D pose estimation using fully convolutional networks, IEEE Trans. Med. Imaging 37 (5) (2018) 1276–1287, https://doi.org/10.1109/TMI.2017.2787672.
[19] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556, 2014.
[20] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[21] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines vinod nair, in: Proceedings of the 27th International Conference on Machine Learning (ICML), 2010, pp. 21–24.
[22] S. Haykin, Neural Networks: A Comprehensive Foundation, 3rd edition, Pearson, 2007.
[23] D. Mery, Computer Vision for X-Ray Testing, Springer, 2015.
[24] D.P. Kingma, J. Ba Adam, A method for stochastic optimization, arXiv:1412.6980, 2014.