Research Article

A Data Augmentation Method for Prohibited Item X-Ray Pseudocolor Images in X-Ray Security Inspection Based on Wasserstein Generative Adversarial Network and Spatial-and-Channel Attention Block

Dongming Liu,1,2 Jianchang Liu,1,2 Peixin Yuan,3 and Feng Yu1,2

1College of Information Science and Engineering, Northeastern University, Shenyang 110819, China
2State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China
3School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China

Correspondence should be addressed to Jianchang Liu; liujianchang@mail.neu.edu.cn

Received 23 January 2022; Accepted 26 February 2022; Published 18 March 2022

Abstract

For public security and crime prevention, the detection of prohibited items in X-ray security inspection based on deep learning has attracted widespread attention. However, the pseudocolor image dataset is scarce due to security, which brings an enormous challenge to the detection of prohibited items in X-ray security inspection. In this paper, a data augmentation method for prohibited item X-ray pseudocolor images in X-ray security inspection is proposed. Firstly, we design a framework of our method to achieve the dataset augmentation using the datasets with and without prohibited items. Secondly, in the framework, we design a spatial-and-channel attention block and a new base block to compose our X-ray Wasserstein generative adversarial network model with gradient penalty. The model directly generates high-quality dual-energy X-ray data instead of pseudocolor images. Thirdly, we design a composite strategy to composite the generated and real dual-energy X-ray data with background data into a new X-ray pseudocolor image, which can simulate the real overlapping relationship among items. Finally, two object detection models with and without our data augmentation method are applied to verify the effectiveness of our method. The experimental results demonstrate that our method can achieve the data augmentation for prohibited item X-ray pseudocolor images in X-ray security inspection effectively.

1. Introduction

With the frequent population flow, people carry more and more items in the baggage and the types of prohibited items have become increasingly abundant. In order to ensure public security, the research studies on the detection method of prohibited items are significant. Currently, X-ray inspection technology has been widely used in the security inspection of public places, reducing criminal behavior effectively. In the process of security inspection, security inspectors need to determine whether there are prohibited items in baggage through the X-ray images. In some special situations, such as the traffic rush hours, the frequency of baggage passing is greatly increased, which requires security inspectors to complete the inspection in a very short time. In addition, there is still no unified standard for the training of security inspectors, and the accuracy of inspection depends on the experience and working status of security inspectors. For these reasons, the accuracy of manual detection methods cannot be guaranteed [1]. Accordingly, a fast and effective automatic detection method of prohibited items for X-ray security inspection is significant.

The detection method of prohibited items has been developed for many years. In the early stage, the automatic detection method of prohibited items usually used different feature extraction algorithms to extract the features of prohibited items and then classified the extracted features. These feature extraction methods included scale-invariant feature transform (SIFT), Haar-like features (Haar), bag of visual words (BOW), histogram of oriented gradients...
(HOG), and so on [2]. Based on these feature extraction methods, researchers proposed some detection methods of prohibited items. In the case of the single-energy X-ray image, Turcsany et al. [3] proposed a method based on BOW for detecting firearms using a dual-view approach. In [4], a method using visual vocabulary and an occurrence structure was proposed to detect three different prohibited items. In addition, an approach called adaptive sparse representation (XASR+) [5] was proposed to recognize items automatically in cases with less constrained conditions including contrast variations, pose variations, image size variations, and focal distance variations. In the case of the dual-energy X-ray image, Riffo and Mery [6] proposed an active X-ray testing framework that is able to find an adequate view of the object item to detect razor blades in different cases. The applicability of BOW methods in X-ray image classification and retrieval was discussed in [7]. In practical applications, the performance of these automatic detection methods based on manual features cannot meet the requirements due to the wide variety of items, occlusion, noise, clutter, and other reasons.

In recent years, the convolutional neural networks (CNNs) have been widely used in image analysis and processing. Methods based on deep learning have achieved great success in many computer vision tasks [8–10]. A CNN-based classification algorithm was proposed in [9], after which CNN-based classifiers have received wide attention in the field of computer vision. These methods based on the CNN have achieved satisfactory results in image classification [11], object detection [12], target segmentation [13], etc. For the object detection methods based on the CNN, the methods are mainly divided into two classifications. One is the two-stage method, such as R-CNN [12], Fast R-CNN [14], Faster R-CNN [15], R-FCN [16], and FPN [17]. These methods first generate a set of candidate region suggestions and then classify, filter, and refine the candidate region suggestions to achieve object detection. The other is the one-stage method, such as YOLO [18–21] and SSD [22]. These methods predict the classification and bounding box directly from a single convolutional network. Such methods have a faster detection speed, but the accuracy is lower than that of the two-stage methods. Both of them have achieved success in the object detection of natural optical images and have been applied in various fields [23–25].

Compared with natural optical images, X-ray pseudocolor images are quite different in several aspects. Natural optical images are formed by the light reflection, while X-ray pseudocolor images are formed by irradiating the objects with X-ray, which loses a lot of information about the surface of objects [26]. In addition, X-ray pseudocolor images consist of shadows from overlapping transparent layers. The transparency of the pseudocolor image is determined by the material density along the X-ray path and different materials will appear in different colors. The overlap between objects also makes the same object appear in different colors, and high-density objects (e.g., thick metal) can obscure other overlapping objects [27]. These phenomena make the research on object detection of X-ray images difficult. In [28], Mery et al. used transfer learning to classify three kinds of prohibited items based on the X-ray grayscale images, and the experimental results showed that the method is effective. In [29], Akcay et al. compared a BoVW approach with a CNN approach, and the experiments showed that the methods based on the CNN outperform the BoVW methods. In [30], the researchers proposed a method using deep convolutional neural networks to detect objects in X-ray security inspection. The method adopted a specific data enhancement technique, feature enhancement blocks, and multiscale fusion regions of interest. Most of the previous research studies for X-ray security inspection used transfer learning based on the ImageNet dataset. However, the direct adoption of the pretrained networks limits the adjustment of the network structure for X-ray pseudocolor images and may reduce the detection performance.

In addition, the method based on deep learning requires a large amount of data. Nevertheless, the X-ray pseudocolor image dataset is scarce and the color definition of X-ray pseudocolor images varies from company to company, which presents an enormous challenge to the object detection method of prohibited items for X-ray security inspection. There are two methods to solve this problem. One is to collect sufficient X-ray images containing the prohibited items with various poses and scales, which requires huge costs. The other is data augmentation. Using data augmentation methods can improve the generalization ability of the methods based on deep learning. Data augmentation methods include rotation, translation, scaling, etc. However, these methods have limited performance gains for the methods based on deep learning. In the last few years, the generative adversarial network (GAN) has achieved considerable success in image generation [31]. To improve the quality of the generated images, some models based on GAN have been proposed, such as the CGAN [32], the Cycle GAN [33], and the WGAN-GP [34]. Recently, some data augmentation methods based on the GAN have been used for X-ray image datasets. In [35], Yang et al. proposed a method of prohibited item X-ray pseudocolor image generation using the GAN. Their generated X-ray pseudocolor images only contain one single prohibited item and the image quality and diversity are not ideal. In [36], Zhu et al. proposed a method based on Cycle GAN to transform the item natural images into X-ray pseudocolor images. These methods used the methods based on GAN to directly generate prohibited item X-ray pseudocolor images and then replaced the target area of the real background X-ray pseudocolor image with the generated images. It is worth noting that these methods cannot show the real overlapping relationship among items. This also resulted in the fact that the final composite X-ray pseudocolor images are not ideal.

To generate more realistic and higher-quality X-ray pseudocolor images, we propose an effective data augmentation method for prohibited item X-ray pseudocolor images in X-ray security inspection (DA-PIX). The main contributions of the method are as follows. Firstly, we design the framework of the DA-PIX to achieve the dataset augmentation using the datasets with and without prohibited items. Secondly, based on the WGAN-GP, we design a
spatial-and-channel attention block (SCAB) and a new base block to compose our X-ray Wasserstein generative adversarial network model (SCAB-XWGAN-GP). The model directly generates high-quality dual-energy X-ray data instead of X-ray pseudocolor images. Thirdly, in order to generate more realistic X-ray pseudocolor images, we design a composite strategy based on the absorption law of X-ray to compose the generated and real dual-energy X-ray data with background data into a new more realistic X-ray pseudocolor image, which can simulate the real overlapping relationship among items.

The rest of the paper is organized as follows. Section 2 describes the dataset creation. In Section 3, the proposed method is described. Section 4 presents the experiments and results. Finally, Section 5 concludes the paper and discusses some directions for future work.

2. Dataset Creation

The dual-energy X-ray method has been widely used in X-ray security inspection systems. The method realizes the material classification of the detected objects by measuring the difference of attenuation coefficient of different materials under high-energy and low-energy X-ray. The low-energy and high-energy data are converted into an X-ray pseudocolor image by a lookup table to facilitate the interpretation of the detected objects, and the lookup table can be obtained through calibration [37]. In order to clearly distinguish the material of the detected object, orange represents the organic matter, green represents the mixture matter, and blue represents the inorganic matter in the X-ray pseudocolor image. Figure 1 shows the imaging process of the X-ray pseudocolor image.

The dual-energy X-ray security inspection equipment was provided by Shenyang DT Inspection Equipment Co., Ltd., China. The X-ray tube voltage is 140 kV. The X-ray tube current is 0.75 mA. The value range of the dual-energy X-ray data is normalized to 0–15200. The size of each data is 600 × 600 × 2. Our X-ray dataset is divided into two parts. One is the X-ray prohibited item dataset (XD-P). The XD-P is collected in the simulated situation. To establish the XD-P, different baggage with the random prohibited items is packed and sent into the security inspection equipment. The XD-P consists of 8,000 dual-energy X-ray data and corresponding X-ray pseudocolor images. The prohibited items include brass knuckles, firecrackers, guns, hammers, knives, lighters, metal bottles, pliers, and scissors. The other is the X-ray security item dataset (XD-S). The data in the XD-S are the basic data for the data augmentation. The data in XD-S were collected from the real subway security inspection. The XD-S consists of 10,000 dual-energy X-ray data and corresponding X-ray pseudocolor images without prohibited items. Some X-ray pseudocolor images of the dual-energy X-ray data used in this paper are shown in Figure 2.

3. DA-PIX Method

The dataset plays an important role in the object detection methods based on deep learning. Due to the security, the dataset is scarce. In order to augment the dataset and improve the robustness of the method, some methods based on the GAN are proposed. However, these methods are limited for our dataset because these methods cannot show the real overlapping relationship between the generated prohibited items and the background. This also results in the fact that the final composite X-ray pseudocolor images are not ideal. Accordingly, we propose an effective data augmentation method DA-PIX for prohibited item X-ray pseudocolor images in X-ray security inspection. Firstly, we design the framework of the DA-PIX. Secondly, we design the SCAB-XWGAN-GP model that directly generates high-quality dual-energy X-ray data instead of X-ray pseudocolor images. Thirdly, we design a composite strategy to compose the generated and real dual-energy X-ray data with background data into a new more realistic X-ray pseudocolor image.

3.1. The Framework of DA-PIX. To obtain more ideal composite X-ray pseudocolor images and improve the robustness of the detection method, we design the DA-PIX. The framework of the DA-PIX is shown in Figure 3. First, the SCAB-XWGAN-GP model is trained using a training dataset extracted manually from the XD-P. Second, the real dual-energy X-ray data of prohibited items and the generated dual-energy X-ray data of prohibited items using the SCAB-XWGAN-GP model are combined into a dataset for the composite strategy. Third, using the secure background dual-energy X-ray data from the XD-S and the foreground dual-energy X-ray data of prohibited items from the dataset for the composite dual-energy X-ray data, new composite dual-energy X-ray data can be obtained through the composite strategy. Then, the composite dual-energy X-ray data can be converted into an X-ray pseudocolor image by a lookup table. The composite images can more realistically simulate the actual situation. Finally, we can obtain the final augmented dataset.

3.2. SCAB-XWGAN-GP Model. Some methods based on GAN are applied to generate prohibited item X-ray pseudocolor images. A typical GAN model is composed of a generator and a discriminator. The architecture of the GAN model is shown in Figure 4. The generator is trained to generate the new data from random noise. The discriminator is trained to distinguish between the real data and the generated data. However, these methods mainly focus on directly generating the X-ray pseudocolor images of prohibited items and cannot show the real overlapping relationship among items. Hence, we need to generate high-quality dual-energy X-ray data instead of X-ray pseudocolor images, which allows our next composite strategy to compose more realistic and higher-quality X-ray pseudocolor images. For the methods based on GAN, most models use upsampling + convolution or transposed convolution. Meanwhile, the X-ray prohibited items dataset is sparse and the training of the methods based on the GAN is challenging, which would result in the methods overfitting or collapse. In order to generate higher-quality prohibited item dual-energy
X-ray data, we design the SCAB-XWGAN-GP model. In our model, we use the WGAN-GP as the basis and the dual-energy X-ray data as the generated object. Meanwhile, we design a new base block to compose the generator. The new base block can effectively improve the quality of the generated images and reduce the training difficulty of the discriminator and the generator.

The structure of the base block is shown in Figure 5. In the base block, we employ the convolutional layer and the PixelShuffle in the subpixel convolution [38] to achieve...
upsampling. The PixelShuffle can rearrange the elements of a $H \times W \times C \cdot r^2$ feature map to a feature map of shape $r \cdot H \cdot r \cdot W \times C$. After that, we use a residual block containing convolutional layers with $3 \times 3$ kernels to optimize the quality and detail of the generated data. In the process, in order to improve the fitting ability of the block and reduce the training difficulty, we use the ParametricRelu (Prelu) [39] as the activation function.

In addition, to generate more realistic dual-energy X-ray data, we design the SCAB based on the spatial-and-channel squeeze-and-excitation [40] block to fit our SCAB-XWGAN-GP model. The SCAB can make the details of data generated by the generator more realistic. Moreover, it can make the discriminator ignore less meaningful information and focus on more meaningful information. The architecture of the SCAB is shown in Figure 6. Given an input $X \in R^{H \times W \times C}$, we can get the calibrated output $\hat{X} \in R^{H \times W \times C}$. The SCAB is divided into the spatial and channel branches.

For the channel branch, $X$ can be expressed as $X = [x_1, x_2, \ldots, x_i, \ldots, x_C]$, where $x_i \in R^{H \times W}$. First, the global average pooling and the global max pooling are used to generate two $1 \times 1 \times C$ feature maps $S_a = [S_{a1}, S_{a2}, \ldots, S_{aC}]$ and $S_m = [S_{m1}, S_{m2}, \ldots, S_{mC}]$ to express $X$ in general. Second, the channel-wise dependencies $\hat{S}$ can be obtained using fully connected (FC) layers and nonlinearity layers. The process can be expressed as

$$\hat{S} = \varphi(W_{a2}\sigma(W_{a1}S_a) + W_{m2}\sigma(W_{m1}S_m)),$$

where $\varphi$ is the sigmoid function, $\sigma$ represents the Relu activation function, $W_{a1}, W_{m1} \in R^{C \times C}$, $W_{a2}, W_{m2} \in R^{C \times C \times r}$ are the weights of the fully connected layers, and $r$ is a ratio parameter. Third, the recalibrated feature map of the channels $\hat{X}_c$ can be obtained by

$$\hat{X}_c = \text{Scale}(\hat{S}, X) = [\hat{s}_1x_1, \hat{s}_2x_2, \ldots, \hat{s}_Cx_C],$$

where Scale($\cdot$) refers to channel-wise multiplication between the scalar $\hat{S}$ and the input $X$.

For the spatial branch, $X$ can be expressed as $X = [x_{i1}, x_{i2}, \ldots, x_{i,j}, \ldots, x_{i,HW}]$, where $x_{i,j} \in R^{1 \times 1 \times C}$. The spatial-wise dependencies $\hat{O}$ can be extracted using $1 \times 1$
convolution layer and the nonlinearity layers. The process can be expressed as

\[ O = \varphi(\sigma(W, X)), \]

where \( W \in \mathbb{R}^{1 \times 1 \times C \times 1} \) is the weight of the \( 1 \times 1 \) convolution layer. Then, the recalibrated feature map of the spatial can be obtained by

\[ \tilde{X}_s = \text{Scale}(\bar{S}, X) = [\bar{s}_{1,1}x_{1,1}, \bar{s}_{1,2}x_{1,2}, \ldots, \bar{s}_{H\times W}\times x_{H\times W}]. \]

After obtaining the recalibrated feature maps, the final calibrated output recalibrated feature map \( \tilde{X} \) can be obtained by

\[ \tilde{X} = X + \tilde{X}_c + \tilde{X}_s. \]

Based on the base block and the SCAB, we design the SCAB-XWGAN-GP model. The architecture of the SCAB-XWGAN-GP model is shown in Figure 7. In the generator, six base blocks and one SCAB are concatenated to generate \( 256 \times 256 \) data and the upsampling multiplier for each base block is 2. It is worth noting that the PixelShuffle requires four times the number of channels on its upper layer, which leads to a dramatic increase in computation and the number of parameters. Therefore, we limit the maximum number of channels of the base block to 256. It also means that the number of convolutional kernels per convolutional layer is at most 1024. Finally, we can obtain the generated dual-energy X-ray data through a convolution + tanh block. In the discriminator, the real and fake data are used as input data, and then the feature extraction is performed by six
convolution+Prelu blocks and one SCAB. Finally, the predicted results are output through the fully connected layer.

For our SCAB-XWGAN-GP model, we use the same training method and the loss function as the WGAN-GP. The loss function $L$ of the SCAB-XWGAN-GP model can be defined as

$$L = E[D(G(z))] - E[D(x)] + \lambda \cdot E[\nabla_x D(\tilde{x})_2 - 1]^2,$$

where $G$ is the generator, $D$ is the discriminator, $z$ is the random noise vector, $\lambda$ is the penalty coefficient, and $\tilde{x}$ is the gradient penalty object sampled from the sample space of the generated data and the real data uniformly, $\tilde{x} = \epsilon z + (1 - \epsilon)G(z)$, where $\epsilon \in [0, 1]$.

3.3. The Composite Strategy. Through manual collection and the SCAB-XWGAN-GP model, we can obtain a large amount of dual-energy X-ray data for the prohibited items. For compositing the ideal X-ray pseudocolor image, we design the composite strategy based on the absorption law of X-ray. Based on the exponential law of photon radiation attenuation, the energy flux density after passage through the object with the thickness of $d$ can be expressed as

$$\varphi = \varphi_0 e^{-\sigma d},$$

where $\varphi_0$ is the incident energy flux density and $\sigma$ is the attenuation coefficient. If X-ray radiation passes through $n$ different objects, the energy flux density after passage through the objects can be expressed as

$$\varphi = \varphi_0 e^{-\sum_{i=1}^{n} \sigma_i d_i}. \quad (8)$$

The gray value of a pixel can be linearly modeled as [37]

$$D = A \cdot \varphi + B,$$ \quad (9)

where $A$ and $B$ are constant parameters of the model. Since the composite strategies for the high and low energies are same, we take single energy as an example. Define the gray value of the foreground object ($D_{\text{fore}}$) and the gray value of background object ($D_{\text{back}}$) as

$$D_{\text{fore}} = A \cdot \varphi_{\text{fore}} + B = A \cdot \varphi_0 e^{-\sigma_{\text{fore}} d_{\text{fore}}} + B,$$ \quad (10)

$$D_{\text{back}} = A \cdot \varphi_{\text{back}} + B = A \cdot \varphi_0 e^{-\sigma_{\text{back}} d_{\text{back}}} + B,$$ \quad (11)

where $\sigma_{\text{fore}}$ and $\sigma_{\text{back}}$ represent the attenuation coefficient of $D_{\text{fore}}$ and $D_{\text{back}}$, and $d_{\text{fore}}$ and $d_{\text{back}}$ represent the thickness of $D_{\text{fore}}$ and $D_{\text{back}}$. From (8), the energy flux density of the composite data $\varphi_c$ can be modeled as

$$\varphi_c = \varphi_0 e^{-\sigma_{\text{fore}} d_{\text{fore}}} e^{-\sigma_{\text{back}} d_{\text{back}}}.$$ \quad (12)

The gray value of the composite data $D_c$ can be expressed as

$$D_c = A \cdot \varphi_c + B = A \cdot \varphi_0 e^{-\sigma_{\text{fore}} d_{\text{fore}}} e^{-\sigma_{\text{back}} d_{\text{back}}} + B.$$ \quad (13)

From (10), (11), and (13), we can obtain

![Figure 7: The architecture of the SCAB-XWGAN-GP model.](image-url)
\[ \frac{D_c - B}{K} = \frac{D_{\text{fore}} - B}{K} \cdot \frac{D_{\text{back}} - B}{K} \]  

where \( K \) is the constant and it can be obtained using a calibration approach \([37]\). From (14), we can obtain the gray value of the composite data as follows:

\[ D_c = \frac{(D_{\text{fore}} - B) \cdot (D_{\text{back}} - B)}{K} + B. \]  

We can obtain a composite X-ray pseudocolor image using equation (15) and a lookup table. The details of the composite process are shown in Figure 8. Two random dual-energy X-ray data \( D_{\text{fore}} \) and \( D_{\text{back}} \) are selected (\( D_{\text{back}} \) belongs to the XD-S and \( D_{\text{fore}} \) belongs to the dataset for the composite images). Then, we randomly select the composite position \( P \) and generate the composite matrix \( \text{Mark} \). \( \text{Mark} \) is a two-valued matrix of zeros and ones, which is obtained from \( D_{\text{fore}} \) by binarization. Finally, the composite image can be obtained by

\[ I_c = \text{Lookup}(D_{\text{back}} \cdot (1 - \text{Mark}) + C(D_{\text{back}}, D_{\text{fore}}, P, \text{Mark})), \]  

where \( \text{Lookup}() \) represents the conversion of dual-energy X-ray data to X-ray pseudocolor images and \( C() \) indicates the composite operation of \( D_{\text{back}} \) and \( D_{\text{fore}} \) at position \( P \) using (15).

To demonstrate the superiority of our composite strategy, we show the original X-ray pseudocolor image and the composite X-ray pseudocolor images using different methods in Figure 9. The traditional method uses the foreground to replace the target area of the background. From Figure 9, we can find that the composite X-ray pseudocolor image using our composite strategy can simulate the real overlapping relationship between the prohibited item and the background compared to the traditional method, which makes the composite X-ray pseudocolor image more realistic.

In addition, our composite strategy can not only realize the composite of the single prohibited item and the safety images but also realize the composite of the multiple prohibited items and safety images or images containing prohibited items. Some composite X-ray pseudocolor images are shown in Figure 10. From Figure 10, we can find that the composite X-ray pseudocolor images using our method can simulate the real situation for the single prohibited item and multiple prohibited items. The composite X-ray pseudocolor images using our composite strategy can increase the diversity of our dataset.

Utilizing our composite strategy, we can randomly select the real and generated foreground objects to combine with the data in XD-S into an augmented dataset. By controlling the selection of the foreground objects and the composite position, we can easily control the position of objects in the augmented dataset and achieve a balance in the number of each object. Using the method, we can obtain a nearly infinite augmented dataset.

### 4. Experiments and Results

In this section, we first test our SCAB-XWGAN-GP model. Then, to verify the effectiveness of our DA-PIX, we design a comparative experiment to evaluate the performance of our method. The experiments are run on a GPU system with the following specifications: Intel Core i9-10900k CPU, 64 GB RAM, and NVIDIA GeForce GTX 3090 GPU.

#### 4.1 Evaluation Criteria

For evaluating the generated X-ray pseudocolor images, the Fréchet inception distance (FID) \([41]\) is used. The FID is a comprehensive metric that has been shown to be more consistent with human assessments in assessing the realism and variability of the generated samples. We can obtain the FID by calculating the Wasserstein-2 distance between the generated X-ray pseudocolor images and the real X-ray pseudocolor images in the feature space of the Inception-v3 network. Lower FID value means that the distance between the generated data and the real data distribution is closer and the model works better.

For evaluating the performance of the object detection model with and without our DA-PIX, average precision (AP) and mean average precision (mAP) are used. AP can be calculated as

\[ \text{AP} = \frac{1}{11} \sum_{Re \in \{0.0,0.1,0.2,\ldots,1\}} \max_k (PrRe), \]  

where \( Pr \) is the precision (\( Pr = TP/TP + FP \)), \( Re \) is the recall (\( Re = TP/TP + FN \)), \( TP \) is the number of true-positive samples, \( FP \) is the number of false-positive samples, \( FN \) is the number of false-negative samples, and \( \text{ Recall } Re \) is the measured precision at recall \( Re \). Subsequently, mAP can be defined as

\[ \text{mAP} = \frac{1}{K} \sum_{k=1}^{K} \text{AP}_k. \]  

#### 4.2 Results and Discussion

In order to better verify the effectiveness of our SCAB-XWGAN-GP model, we trained the DCGAN model and the WGAN-GP model to compare them with our model. For the training dataset used by the generative models, first we extracted the object data from the labeled XD-P by labeling the extracted data manually. Second, we kept the target data unchanged and replaced the background with 15200 by manual labeling. Third, we filled these data into a square with the maximum value of the width and height of the data and resized them to \( 256 \times 256 \times 2 \) using the bilinear interpolation. Finally, these dual-energy X-ray data formed the training dataset used by the generative models. The adaptive moment estimation (Adam) optimization algorithm was used. The batch size was 32, the penalty coefficient \( \lambda \) was 10, and the learning rate of the generator and discriminator was 0.0002 and 0.0001, respectively. The update ratio of generator and discriminator was 5 for our model and the WGAN-GP model. For the DCGAN model, the update ratio of generator and discriminator was 1. It is worth noting that the generated data by these models are dual-energy X-ray data, and these data are converted into the X-ray pseudocolor images by a lookup
The background dual-energy data

The composition position

The composition matrix

The foreground dual-energy data

Figure 8: The composite process.

The composite dual-energy data

Lookup table

The composite image

Figure 9: Continued.

(a) 

(b)
Figure 9: The original X-ray pseudocolor image and the composite X-ray pseudocolor images using different methods. (a) The original X-ray pseudocolor image. (b) The composite X-ray pseudocolor image using our composite strategy. (c) The composite X-ray pseudocolor image using the traditional method.

Figure 10: Continued.
Some X-ray pseudocolor images generated by these models are shown in Figure 11.

From Figure 11, it can be found that the generated X-ray pseudocolor images by the DCGAN have a lot of noise, the images appear to have severe distortion, and the detailed information is blurred. Meanwhile, for the brass knuckle and plier, the DCGAN showed pattern collapse. For the WGAN-GP, the quality of the generated X-ray pseudocolor images has improved, but there is still noise and distortion in the images, the detailed information is still blurred, and the visual quality is poor. Compared with the methods, although there is still a little noise and minor deformation in the generated X-ray pseudocolor images, the quality of the X-ray pseudocolor images generated using our SCAB-XWGAN-GP is significantly improved.

In addition, to objectively evaluate our method, we compared the methods quantitatively using the FID score. The FID scores of these models are shown in Table 1. From Table 1, we can find that our method has the lowest FID score. This means that our model can better approximate the distribution of the real dual-energy X-ray data of the prohibited items.

Then, to verify the effectiveness of our DA-PIX, we trained the object detection models with and without our DA-PIX. We employed the YOLOV4-tiny model [21] as the object detection model. We randomly divided the XD-P into the training dataset and the testing dataset. The training dataset contained 7200 X-ray pseudocolor images and the test dataset contained 800 X-ray pseudocolor images. Our DA-PIX was used to augment the training dataset. In this process, first, we manually cropped the generated dual-energy X-ray data and resized them to their real size. Second, we combined these dual-energy X-ray data with the real dual-energy X-ray data extracted from the XD-S to form the dataset for the composite strategy. Third, we used the composite strategy to composite the new X-ray pseudocolor images with the data in the XD-S as the backgrounds. These composite X-ray pseudocolor images contained 2–7 prohibited items in each image. The augmented training dataset contained 17,200 X-ray pseudocolor images. After augmenting the dataset, we trained the YOLOV4-tiny model on the training dataset and the augmented training dataset, respectively. The stochastic gradient descent with momentum (SGDM) was used. The batch size was 16, the initial
The learning rate with warm-up was 0.001, and the momentum was 0.9. The detection results are shown in Table 2.

From Table 2, we can find that the mAP of the YOLOV4-tiny model trained by the augmented training dataset is higher than that of the YOLOV4-tiny model trained by the original training dataset, with an increase of 9.25%.

Moreover, the AP of each prohibited item has been enhanced using our method. It means that our DA-PIX method can effectively improve the performance of the detection model for prohibited item X-ray pseudocolor images in X-ray security inspection. In addition, the YOLOV4-tiny model is a lightweight detection model, and

Table 1: The FID scores of different models.

| Items        | WGAN-GP | DCGAN | SCAB-XWGAN-GP |
|--------------|---------|-------|---------------|
| Brass knuckle| 185.3   | 143.7 | 103.2         |
| Firecracker  | 16.3    | 18.7  | 13.8          |
| Gun          | 73.1    | 121.7 | 18.47         |
| Hammer       | 32.6    | 41.8  | 19.2          |
| Knife        | 25.8    | 32.3  | 9.4           |
| Lighter      | 106.9   | 108.6 | 78.3          |
| Metal bottle | 45.2    | 48.6  | 41.4          |
| Plier        | 43.8    | 45.6  | 27.2          |
| Scissors     | 44.5    | 48.4  | 22.4          |
we believe that the mAP will be further improved under more complex detection models.

5. Conclusion

In this paper, an effective data augmentation method for prohibited item X-ray pseudocolor images in X-ray security inspection is proposed. The innovation is mainly reflected in three major aspects. First, we design the framework of our DA-PIX to achieve dataset augmentation using the datasets with and without prohibited items. Second, in order to generate high-quality dual-energy X-ray data, a SCAB-XWGAN-GP model is designed based on the SCAB and a new base block. Third, a composite strategy based on the absorption law of X-ray is proposed to composite a new X-ray pseudocolor image, which can simulate the real overlapping relationship among items. Two YOLOV4-tiny models with and without our DA-PIX are trained to verify the effectiveness of our method. The experimental results demonstrate that our DA-PIX can effectively improve the mAP of the YOLOV4-tiny model. Therefore, our DA-PIX is effective for the data augmentation of prohibited item X-ray pseudocolor images in X-ray security inspection.

In future work, we will focus on designing effective detection models for the detection of prohibited item in X-ray security inspection.

Data Availability

The dataset used to support the findings of this study was supplied by Shenyang DT Inspection Equipment Co., Ltd., China, under license and the dataset involving security cannot be shared.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (no. 61773106) and Major Scientific and Technological Projects of the Ministry (no. JB2016GD034).

Table 2: The detection results.

| Items                  | YOLOV4-tiny + the original training dataset | YOLOV4-tiny + the augmented training dataset |
|------------------------|--------------------------------------------|------------------------------------------------|
| Brass knuckle          | 88.75%                                     | 95.76%                                        |
| Firecracker            | 86.63%                                     | 92.88%                                        |
| Gun                    | 86.80%                                     | 93.84%                                        |
| Hammer                 | 67.81%                                     | 80.59%                                        |
| Knife                  | 74.20%                                     | 80.58%                                        |
| Lighter                | 86.69%                                     | 94.81%                                        |
| Mental bottle          | 93.17%                                     | 96.23%                                        |
| Plier                  | 89.77%                                     | 90.11%                                        |
| Scissors               | 72.79%                                     | 87.06%                                        |
| mAP                    | 82.96%                                     | 90.21%                                        |

References

[1] S. Michel, J. Ruiter, M. Hogervorst, S. Koller, and A. Schwaninger, “Computer-based training increases efficiency in x-ray image interpretation by aviation security screeners,” in Proceedings of the 41st Annual IEEE International Carnahan Conference on Security Technology, pp. 201–206, Ottawa, ON, Canada, 8–11 Oct. 2007.

[2] V. Riffo, S. Flores, and D. Mery, “Threat objects detection in x-ray images using an active vision approach,” Journal of Nondestructive Evaluation, vol. 36, no. 44, 2017.

[3] D. Turcsany, A. Mouton, and T. P. Breckon, “Improving feature-based object recognition for x-ray baggage security screening using primed visual words,” in Proceedings of the 2013 IEEE International Conference on Industrial Technology (ICIT), pp. 1140–1145, Cape Town, South Africa, 25–28 Feb. 2013.

[4] V. Riffo and D. Mery, ”Automated detection of threat objects using adapted implicit shape model,” IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 46, no. 4, pp. 472–482, 2016.

[5] D. Mery, E. Svec, and M. Arias, “Object recognition in x-ray testing using adaptive sparse representations,” Journal of Nondestructive Evaluation, vol. 35, no. 3, pp. 1–9, 2016.

[6] V. Riffo and D. Mery, “Active x-ray testing of complex objects,” Insight - Non-Destructive Testing and Condition Monitoring, vol. 54, no. 1, pp. 28–35, 2012.

[7] M. Bastan, M. R. Yousefi, and T. M. Breuel, “Visual words on baggage x-ray images,” in Proceedings of the Computer Analysis of Images and Patterns, pp. 360–368, Seville, Spain, August 2011.

[8] A. S. Elkorany and Z. F. Elsharkawy, “Covidetection-net: A tailored covid-19 detection from chest radiography images using deep learning,” Optik, vol. 231, 2021.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[10] O. Russakovsky, J. Deng, H. Su et al., “Imagenet large scale visual recognition challenge,” International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.

[11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, Las Vegas, NV, USA, 27-30 June 2016.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Region-based convolutional networks for accurate object detection and segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 1, pp. 142–158, 2016.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic
segmentation,” in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, 23-28 June 2014.

[14] R. Girshick, “Fast r-cnn,” in *Proceedings of the 2015 IEEE International Conference on Computer Vision*, pp. 1440–1448, Santiago, Chile, 7–13 Dec. 2015.

[15] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[16] J. Dai, Y. Li, K. He, and J. Sun, “R-fcn: object detection via regionbased fully convolutional networks,” in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 379–387, Barcelona Spain, December 2016.

[17] T. Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, ”Feature pyramid networks for object detection,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 936–944, Honolulu, HI, USA, 21-26 July 2017.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: unified, real-time object detection,” in *Proceedings of the 2016 IEEE conference on computer vision and pattern recognition*, pp. 779–788, Las Vegas, NV, USA, 27-30 June 2016.

[19] J. Redmon and A. Farhadi, “Yolo9000: better, faster, stronger,” in *Proceedings of the 2017 IEEE conference on computer vision and pattern recognition*, pp. 6517–6525, Honolulu, HI, USA, 21-26 July 2017.

[20] J. Redmon and A. Farhadi, “Yolov3: an incremental improvement,” 2018, https://arxiv.org/abs/1804.02767.

[21] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, “Yolov4: optimal speed and accuracy of object detection,” 2020, https://arxiv.org/abs/2004.10934.

[22] W. Liu, D. Anguelov, D. Erhan et al., “Ssd: single shot multibox detector,” in *Proceedings of the European Conference on Computer Vision*, Amsterdam, The Netherlands, October 2016.

[23] Z. Zhao, K. Hao, X. Ma et al., “SAI-YOLO: a lightweight network for real-time detection of driver mask-wearing specification on resource-constrained devices,” *Computational Intelligence and Neuroscience*, vol. 2021, p. 15, Article ID 4529107, 2021.

[24] Z. Wu, X. Wang, and C. Chen, “Research on lightweight infrared pedestrian detection model architecture for embedded platform,” *Security and Communication Networks*, vol. 2021, p. 7, Article ID 1549772, 2021.

[25] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, “Fighting against covid-19: a novel deep learning model based on yolo-v2 with resnet-50 for medical face mask detection,” *Sustainable Cities and Society*, vol. 65, 2021.

[26] Z. Chen, Y. Zheng, B. Abidi, D. L. Page, and M. Abidi, “A combinational approach to the fusion, de-noising and enhancement of dual-energy x-ray luggage images,” in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05) - Workshops*, San Diego, CA, USA, 21-23 Sept. 2005.

[27] V. Rebuffel and J.-M. Dinten, “Dual-energy x-ray imaging: benefits and limits,” *Insight - Non-Destructive Testing and Condition Monitoring*, vol. 49, no. 10, pp. 589–594, 2007.

[28] D. Mery, E. Svec, M. Arias, V. Riffó, J. M. Saavedra, and S. Banerjee, “Modern computer vision techniques for x-ray testing in baggage inspection,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 4, pp. 682–692, 2017.

[29] 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 imagery,” in *Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP)*, pp. 1057–1061, Phoenix, AZ, USA, 25-28 Sept. 2016.

[30] B. Gu, R. Ge, Y. Chen, L. Luo, and G. Coatrieux, “Automatic and robust object detection in x-ray baggage inspection using deep convolutional neural networks,” *IEEE Transactions on Industrial Electronics*, vol. 68, no. 10, pp. 10248–10257, 2021.

[31] C. Ledig, L. Theis, F. Huszár et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114, Honolulu, HI, USA, 21-26 July 2017.

[32] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” 2014, https://arxiv.org/abs/1411.1784.

[33] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, Venice, Italy, 22-29 Oct. 2017.

[34] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, “Improved training of wasserstein gans,” 2017, https://arxiv.org/abs/1704.00028.

[35] J. Yang, Z. Zhao, H. Zhang, and Y. Shi, “Data augmentation for x-ray prohibited item images using generative adversarial networks,” *IEEE Access*, vol. 7, pp. 28894–28902, 2019.

[36] Y. Zhu, Y. Zhang, H. Zhang, J. Yang, and Z. Zhao, “Data augmentation of x-ray images in baggage inspection based on generative adversarial networks,” *IEEE Access*, vol. 8, pp. 86536–86544, 2020.

[37] D. Mery, *Computer Vision for X-ray Testing*, Springer International Publishing, Heidelberg, Germany, 2015.

[38] W. Shi, J. Caballero, F. Huszár et al., “Real-Time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1874–1883, Las Vegas, NV, USA, 27-30 June 2016.

[39] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: surpassing human-level performance on ImageNet classification,” in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034, Santiago, Chile, 7-13 Dec. 2015.

[40] A. G. Roy, N. Navab, and C. Wachinger, “Concurrent spatial and channel ‘Squeeze & excitation’ in fully convolutional networks,” in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, Venice, Italy, 22-29 Oct. 2017.

[41] H. Heuse, H. Ramsauer, B. Unterthiner, B. Nessler, and S. Hochreiter, “GANs trained by a two time-scale update rule converge to a local Nash equilibrium,” in *Proceedings of the International Conference on Neural Information Processing Systems*, pp. 6629–6640, Montreal Canada, December 2017.