GAN-based data augmentation of prohibited item X-ray images in security inspection*

ZHU Yue (朱越)[1], **ZHANG Hai-gang** (张海刚)[2], **AN Jiu-yuan** (安久远)[1], and **YANG Jin-feng** (杨金锋)[2]**

1. *Tianjin Key Lab for Advanced Signal Processing, Civil Aviation University of China, Tianjin 300300, China*
2. *Institute of Applied Articial Intelligence of the Guangdong-Hong Kong-Macao Greater Bay Area, Shenzhen Polytechnic, Shenzhen 518055, China*

(Received 11 July 2019; Revised 28 August 2019)

Convolutional neural networks (CNNs) based methods for automatic discriminant of prohibited items in X-ray images attract attention increasingly. However, it is difficult to train a reliable CNN model using the available X-ray security image databases, since they are not enough in sample quantity and diversity. Recently, generative adversarial network (GAN) has been widely used in image generation and regarded as a power model for data augmentation. In this paper, we propose a data augmentation method for X-ray prohibited item images based on GAN. First, the network structure and loss function of the self-attention generative adversarial network (SAGAN) are improved to generate the realistic X-ray prohibited item images. Then, the images generated by our model are evaluated using GAN-train and GAN-test. Experimental results of GAN-train and GAN-test are 99.91% and 98.82% respectively. It implies that our model can enlarge the X-ray prohibited item image database effectively.

Document code: A **Article ID:** 1673-1905(2020)03-0225-5
DOI https://doi.org/10.1007/s11801-020-9116-z

X-ray baggage inspection is widely used to maintain public transportation security[1]. But the reliability and efficiency of traditional X-ray baggage inspection is undesirable. It also brings great working pressure to security inspection operators. It is crucial to establishing a model of detecting and recognizing prohibited items automatically for ensuring the safety of passengers. In the latest years, convolutional neural networks (CNNs) have shown strong performance on image classification and object detection[2], and it has also been applied in X-ray baggage security inspection tasks. In Ref.[3], the CNN model is used for X-ray prohibited item image classification. Xu et al[4] proposed an attention-based CNN model for detecting the prohibited items in X-ray images.

Currently, there are two available security inspection X-ray databases including GDX-ray[5] and SIXray[6]. The GDX-ray is grayscale image database, which is different from the pseudo color image in current security inspection. The SIXray database contains more than one million images, but only 8 929 images contain prohibited items. These databases are difficult to meet requirements of CNN training in sample quantity and diversity. It is also difficult to collect enough images manually for building a database by X-ray machine. Therefore, a reasonable alternative is to automatically generate new training samples using data augmentation.

Generative adversarial networks (GANs) have achieved considerable success in image generation. Many derived GAN models have been proposed to generate images with high quality[7-9]. Self-attention generative adversarial network (SAGAN)[10] and BigGAN[11] have improved the quality and diversity of the generated images obviously. The increasingly photorealistic sample quality of generated image models demonstrate their feasibility in data augmentation such as biomedical image data augmentation[12,13]. Recently, the GAN method is also used to enlarge the X-ray prohibited item image database. Ref.[14] first applied GAN to enlarge the X-ray prohibited items image database. This method can be used to generate different kinds of X-ray prohibited item images. However, the generated images are not ideal in visual quality and diversity. Therefore, we focus on generating more realistic X-ray prohibited item images with better quality and diversity. In this work, we are only interested in X-ray prohibited items, and it is another work to synthesize the X-ray prohibited items into X-ray images.

In this paper, a GAN-based method is proposed to enlarge the database for X-ray prohibited item images. First, a preliminary X-ray prohibited item image database is introduced. Then, the improved SAGAN model is

* This work has been supported by the National Natural Science Foundation of China (No.61806208), the Fundamental Research Funds for the Central Universities (No.3122018S008), and the Tianjin Education Committee Research Project (No.2018KJ246).
** E-mail: jfyang@szpt.edu.cn

proposed to match the X-ray prohibited item image database and generate the realistic new images. The generative images are evaluated using GAN-train and GAN-test[16]. Experimental results show that the images generated by our model can be used for data augmentation.

In this paper, the X-ray prohibited item image database were collected by an X-ray machine. In order to enable GAN model to better learn common feature of X-ray prohibited items, the foregrounds of prohibited items are extracted from the collected X-ray images according to the image preprocessing method in Ref.[14]. In this way, the interference caused by the background can be avoided when images are generated. The database consists of 12 categories prohibited items including handgun, fruit knives, lighters, blade, wrenches, screwdrivers, power bank, scissors, hammers, forks, and liquids material (shown in Fig.1). Each type of prohibited items involves about 200—400 images. The resolution of each images is 256×256.

Fig.1 Database image

The X-ray prohibited item image database has several characteristics which are the important guidance for improving GAN model. 1) The database is small in size. 2) The color of X-ray images provides useful information since different colors represent different materials during X-ray imaging. 3) The different poses of the same prohibited item can be determined by the contour and texture of the images. The contour and texture information of images is crucial to generating X-ray prohibited item images.

The color, contour and texture of image are called global image features. To generate high-quality and diverse X-ray prohibited item images, the generation model needs to learn the global image feature well in the small sample database.

SAGAN[10] introduces the self-attention mechanism[17] to improve the ability of both Generator and Discriminator to model global structure. The model is widely used to generate realistic images. Because of the small database, SAGAN model can not work well in generating the X-ray prohibited item images. Therefore, we improve the network structure and loss function of the SAGAN model to generate the realistic images.

From the section above, we can see that the global images features are very important for X-ray prohibited item images. So we improve the network structure to learn the global information of the images on the small database. The improved GAN model is illustrated in Fig.2.

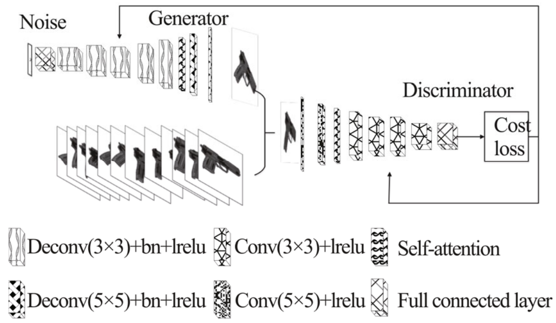

Fig.2 Improved SAGAN architecture

In order to facilitate the GAN training on the small database, we use the convolution and deconvolution structures as the Discriminator and Generator. We deepen the convolutional network structure so that the convolutional networks can learn the long-range correlation of X-ray prohibited item images. The Discriminator contains six convolutional layers and one fully connected layer. The Generator consists of seven deconvolutional layers and one fully connected layer. We use two sizes of convolution kernels connected in series for Discriminator and Generator. We remove the Batch Normalization of the discriminator layer and only retain the Batch Normalization in the generator. When WGAN-GP and Spectral Normalization are combined, the model is easy to over-fit, so we omit Spectral Normalization in the models, finding that they perform well without it.

Compared with WGAN-GP, we find that the hinge loss function[18] is not suitable for the X-ray prohibited item image database. So the WGAN-GP loss function is applied in SAGAN model. Compared to the two-sided penalty used by the original WGAN-GP, we empirically find that one-sided penalty is more suitable for our X-ray prohibited item database. The loss function is defined as

$$L=E[D(G(z))]-E(D(x))+\lambda GP, \tag{1}$$

$$GP = E[\max(0,(\|\nabla_{\hat{x}}D(\hat{x})\|_2-1))^2] \ , \tag{2}$$

$$\hat{x} = \varepsilon x + (1-\varepsilon)G(z) \ , \tag{3}$$

where G and D represents Generator and Discriminator of GAN model, respectively. The input z is the random uniform noise vector. The λ is the penalty coefficient.

Gradient penalty object $\hat{x}$ is uniformly sampled from the straight line between the generated data and real data. If the model train images with the same category, the interpolated $\hat{x}$ can be more reasonable and the generated images more realistic.

Evaluating the generated images by drawing on subjective visual does not work well when the difference is not obviously. We need quantitative measures to evaluate the performance of GANs and the quality and diversity of generated images. In this section, we present FID score[15], GAN-train and GAN-test[16] to evaluate the generated images.

Currently, Inception score[19] and Frechet Inception

Distance (FID) score are widely used to evaluate the generated images. Inception score does not consider original samples, so it cannot measure the approximation between the generated samples and the real samples. FID has been shown to be more consistent with human evaluation in assessing the realism and variation of the generated samples. The lower FID score means the better performance. We use FID to compare different GAN models and select the optimal model.

Although FID can measure the quality and diversity between the generated sample and the real sample, the FID measure cannot separate image quality from image diversity. Thus, we use the GAN-train and GAN-test measures to evaluate the quality and diversity of the generated images.

The illustration of GAN-train and GAN-test is shown in Fig.3. The Inception V3 network is used as the classifier. GAN-train outputs the accuracy of a classifier trained on the generated images and tested on real images. GAN-train evaluates the diversity and realism of the generated images by GANs. GAN-test outputs the accuracy of a classifier trained on the real images and tested on generated images. GAN-test evaluate how realistic the images generated by GANs.

Similarly, we can quantify the amount of feature information into the classification accuracy. When the accuracy of GAN-train is higher than the accuracy of GAN-test, it means that the generated sample contains more feature information than the original sample.

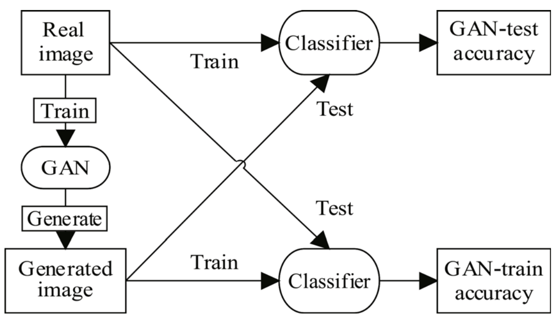

Fig.3 Illustration of GAN-train and GAN-test

In this section, we will introduce our experiments and show the experimental results. All of the following experiments are based on our database.

Some images with different visual quality are generated based on six GAN models (showing in Fig.4). The first model is the original deep convolutional generative adversarial network (DCGAN)[8]. We find the images generated by DCGAN are with noisy and no texture information like real images. The second model is DCGAN with Hinge loss function[18]. The images have small improvement in quality comparing with the images generated by DCGAN. The third model is SAGAN[10]. The images generated by this model have almost no noise, but the shape of the X-ray prohibited items is distorted. The fourth and fifth model are WGAN-GP and WGAN-LP[9]. The image quality has been improved, but the image edges of some prohibited items are blurred. The sixth is our model. Compared with other models, the visual quality of images generated by our model have been improved obviously.

Fig.4 The images generated by different models

Tab.1 presents the FID scores of the six GAN models. The loss function of DCGAN is the same as the original GAN. The loss functions of Hinge model and SAGAN model are Hinge. The loss function of WGAN-GP and WGAN-LP are the two-sided penalty and the one-sided penalty respectively. Compared with the previous five models, the images generated by WGAN-LP get the lowest FID score. It means the WGAN-GP loss function with one-sided penalty is more suitable for prohibited item images database. Compared WGAN-LP with our model, we find the self-attention mechanism can effectively improve the quality of the images.

Tab.1 FID scores for five GAN models

Model	Handgun	Fruit knife
DCGAN	174	208
Hinge	161	177
SAGAN	121	164
WGAN-GP	70	127
WGAN-LP	67	114
Our model	57	82

As shown in Fig.5, some images of different X-ray prohibited items are generated by our model. The experimental parameters are set as Tab.2. G-lr and D-lr denote the learning rate of Generator and Discriminator. The quality of the generated images is very close to the real images shown in Fig.1. These 12 classes of generated images have photorealistic quality.

From the result above, we know our model is better than other models. Thus, it is not necessary to compare the five models in this experiment. In this section the images generated by our model and CTGAN which is used in Ref.[14] are evaluated using GAN-train and GAN-test. We generate 147 434 images including 12 classes using our model. There are 10 thousand images generated by CTGAN including 9 classes. The classification results are shown in Tab.3.

Fig.5 Some generated image samples

Tab.2 Hyperparameter setting

Epoch	Batch size	G-lr	D-lr
10 000	36	0.000 1	0.000 4

Tab.3 Results of GAN-train and GAN-test

Image evaluation	GAN-train	GAN-test
Our	99.91%	98.82%
CTGAN	91.85%	99.12%

The high accuracy of GAN-train indicates that the quality and diversity of the generated images is similar to real images. The GAN-test with a high value denotes that the generated images are a realistic approximation of the distribution of real images.

Tab.4 reports the accuracy of prohibited item images generated by the two models. Compared GAN-train of our model with CTGAN, we can presume the classification network can learn the more feature distribution in the image generated by our model and identify correctly the real image better. It means the images generated by our model contain almost all the feature distributions of the real images. GAN-test accuracy of our model is reduced compared with GAN-train accuracy. The classification network with real images feature information cannot identify the feature information of image generated by our model well. This means that the image generated by our model contains more feature information than the real images. However, the accuracy of CTGAN is contrary to our model. It means that the images generated by CTGAN contains less feature information than the real images. Therefore, the images generated by our model are suitable for enlarging X-ray prohibited item database.

Tab.4 Classification accuracy of different items

Item	GAN-train		GAN-test	
	Our	CTGAN	Our	CTGAN
Handgun	100.0%	89.41%	100.0%	100%
Fruit knife	100.0%	92.85%	97.61%	98.6%
Blade	100.0%		95.83%	
Screw-driver	100.0%	94.27%	98.97%	99.6%
Scissors	98.87%		99.76%	
Liquid	100.0%	97.92%	99.83%	100%
Pliers	100.0%	84 .88%	99.98%	97.5%
Lighter	100.0%	97.92%	98.27%	100%
Wrench	99.59%	85.23%	99.62%	99.6%
Power bank	100.0%	94.61%	99.97%	99.9%
Hammer	100.0%		99.25%	
Fork	100.0%	89.29%	95.70%	98.6

In this paper, we propose a data augmentation method based on GAN model for X-ray prohibited item images. We improve the SAGAN to generate the realistic X-ray prohibited item images. Compared with other models, our model is more suitable for generating the X-ray prohibited item images. To validate the conclusion, the images generated by our model and CTGAN are evaluated using GAN-train and GAN-test. Evaluation results indicate that the images generated by our model contain more feature information. Therefore, our work achieves data augmentation for X-ray prohibited item images effectively.

References

[1] Akcay S, Kundegorski M E, Devereux M and Breckon Y P, Transfer Learning Using Convolutional Neural Networks for Object Classification within X-Ray Baggage Security Imagery, IEEE International Conference on Image Processing, 1057 (2016).

[2] R. Girshick, Fast R-CNN, IEEE International Conference on Computer Vision, 1440 (2015).

[3] Akcay S, Kundegorski M E, Willcocks C G and Breckon Y P, IEEE Transactions on Information Forensics and Security **13**, 2203 (2018).

[4] Xu Mao-shu, Zhang Hai-gang and Yang Jin-feng, Prohibited Item Detection in Airport X-Ray Security Images via Attention Mechanism Based CNN, Chinese Conference on Pattern Recognition and Computer Vision, 429 (2018).

[5] Mery D, Svec E, Arias M, Riffo V, Saavedra J M and Banerjee S, IEEE Transactions on Systems, Man, and Cybernetics **47**, 682 (2017).

[6] Miao C, Xie L, Wan F, Su C, Liu H, Jiao J anf Ye Q, SIXray: A Large-scale Security Inspection X-ray Benchmark for Prohibited Item Discovery, Computer Vision and Pattern Recognition, 2119 (2019).

[7] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B,

Warde-Farley D, Ozair S, Courville A and Bengio Y, Generative Adversarial Nets, the 27th International Conference on Neural Information Processing Systems, 2672 (2014).

[8] Radford A, Metz L and Chintala S, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, arXiv:1511.06434, (2015).

[9] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V and Courville A C, Improved Training of Wasserstein GANs, arXiv:1704.00028, (2017).

[10] Zhang H, Goodfellow I, Metaxas D and Odena A, Self-Attention Generative Adversarial Networks, arXiv:1805.08318, (2018).

[11] Brock A, Donahue J and Simonyan K, Large Scale GAN Training for High Fidelity Natural Image Synthesis, arXiv: 1809.11096, (2018).

[12] Antoniou A, Amos S and Harrison E, Data Augmentation Generative Adversarial Networks, arXiv: 1711.04340, (2017).

[13] Frid-Adar M, Klang E, Amitai M, Goldberger J and Greenspan H, Synthetic Data Augmentation Using GAN for Improved Liver Lesion Classification, 15th International Symposium on Biomedical Imaging, 289 (2018).

[14] Zhao Z H, Zhang H G and Yang J F, A GAN-Based Image Generation Method for X-Ray Security Prohibited Items. Pattern Recognition and Computer Vision, 420 (2018).

[15] Heuse H, Ramsauer H, Unterthiner T, Nessler B and Hochreiter S, GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, Neural Information Processing Systems, 6629 (2017).

[16] K Shmelkov, C Schmid and K Alahari, How Good is My GAN? European Conference on Computer Vision, 213 (2018).

[17] Wang X, Girshick R, Gupta A and He Kai-ming, Non-local Neural Networks, IEEE Conference on Computer Vision and Pattern Recognition, 7794 (2018).

[18] Lim J H and Ye J C, Geometric GAN, arXiv: 1705.02894, (2017).

[19] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A and Chen X, Improved Techniques for Training GANs, Neural Information Processing Systems, 2234 (2016).