A GAN-Based Image Generation Method for X-Ray Security Prohibited Items

Zihao Zhao, Haigang Zhang, and Jinfeng Yang[✉]

Tianjin Key Lab for Advanced Signal Processing,
Civil Aviation University of China, Tianjin, China
jfyang@cauc.edu.cn

Abstract. Recognizing prohibited items intelligently is significant for automatic X-ray baggage security screening. In this field, Convolutional Neural Network (CNN) based methods are more attractive in X-ray image contents analysis. Since training a reliable CNN model for prohibited item detection traditionally requires large amounts of data, we propose a method of X-ray prohibited item image generation using recently presented Generative Adversarial Networks (GANs). First, a novel pose-based classification method of items is presented to classify and label the training images. Then, the CT-GAN model is applied to generate many realistic images. To increase the diversity, we improve the CGAN model. Finally, a simple CNN model is employed to verify whether or not the generated images belong to the same item class as the training images.

Keywords: Generative Adversarial Network
X-ray prohibited item images · Image generation
Feature transformation

1 Introduction

X-ray security baggage screening is widely used to ensure transport security [1]. But the accuracy of manual detection have not been desirable for a long time. The prohibited items are very difficult to detect when they are placed closely in baggage and occluded by other objects [2]. Furthermore, operators are usually allowed only a limited working time to recognize the prohibited items in baggage. A reliable automatic detection system for X-ray baggage images can significantly speed the screening process up and improve the accuracy of detection [3]. Recently, the deep learning based approaches have drawn more and more attentions in image contents analysis. They probably perform well on prohibited item detection. Unfortunately, the dataset of X-ray prohibited item images used in training human inspectors could not meet the requirements of network training. In addition, it is also difficult to collect enough X-ray images containing prohibited items with pose and scale variety in practice.

It is traditional to address the problem via using data augmentation of collected images, such as translation, rotation, and scale. But little additional information can be gained by these ways [4]. Besides data augmentation, training

the network on a pre-trained model slightly improve the performance of image processing algorithm. The Generative Adversarial Network [5] has enjoyed considerable success in data generation. It can be used to generate realistic images according to the recent development of GAN in network architecture and training process [6–9]. WGAN-GP [10] is a popular model for image generation, while PGGAN [11] and SNGAN [12] can generate images with high resolution and rich diversity.

But for the task of generating X-ray prohibited item images, existing GAN-based approaches are not trainable since the amount of training images is not enough. In addition, the items in baggage are placed randomly and packed tightly, so the X-ray prohibited items generally present various visual angles. Figure 1 shows some images of handguns. The guns in images have many poses, and the backgrounds are greatly varied. These factors are unfavorable for GAN to learn the common features of all guns.

Fig. 1. X-ray handgun images

In this paper, we propose an image generation method of X-ray security prohibited items using GAN-based approach. We take dealing with the handgun images as an instance since the detection of handgun is a classical subject. First, we introduce a pose-based classification method of handguns. Then, we facilitate the network training by adding pose labels for the collected images and extracting the object foreground with KNN-matting [13]. Next, CT-GAN [14] model is used for image generation. In order to increase the diversity of images, such as pose, scale and position, we improve the CGAN model [15]. Finally, a simple CNN model is used to verify whether or not the generated images and real images belong to the same item class. Only the images with a correct matching result given by CNN model can be used as new samples of dataset.

The rest of paper is organized as follows. In Sect. 2, we present an image preprocess method. Section 3 introduces the CT-GAN model and the improved CGAN model, Sect. 4 details the experiments and shows some generated images. In Sect. 5, we perform a verification experiment. Finally, Sect. 6 summarizes this paper.

2   Image Preprocessing

Most GAN models for image generation need a large training dataset, such as ImageNet and LSUN. The absence of training images and the pose variety of prohibited items increase the difficulty of network training. If these images are directly fed into GAN model for unsupervised learning, the network is hard to learn their common features. As shown in Fig. 2, the generated images have unreasonable shapes of handguns. To solve this problem, we remove the background and add labels for images before training the GAN model.

Fig. 2. Generated images without preprocessing

2.1   Image Classifying and Labeling

A space rectangular coordinate system is constructed as shown in Fig. 3, and its origin corresponds to the geometrical center of the handgun. Different poses of handguns can be regarded as how many angles the gun rotated around three axes in the coordinate system. And we can classify the handgun images according to the angles of rotation.
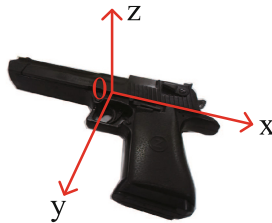

Fig. 3. Construction of space rectangular coordinate system

Rotation around z-axis changes the direction of guns, while rotation around x-axis and y-axis changes the angle. The result of classification is illustrated in Fig. 4. We set the standard position where the handgun turns the muzzle to left. The images can be divided into two classes according to the direction of muzzle. The rotations around z-axis can be roughly divided into 4 classes,

include $0° \pm 45°$, $90° \pm 45°$, $-90° \pm 45°$ and $180° \pm 45°$. The rotations around x-axis and y-axis can be divided into two classes, $0° \sim 45°$ and $-45° \sim 0°$. The geometrical view of handguns in actual security screening that corresponding to the rotation more than $\pm 45°$ is unusual, so it is not considered. When the rotation angle is more than $\pm 90°$, it repeats with the mirror position. Therefore, the handgun images can be divided into $32(2 \times 4 \times 2 \times 2)$ point classes.
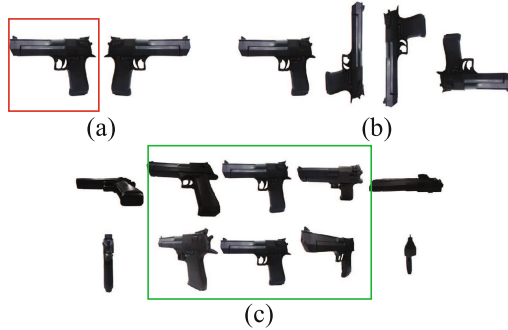

Fig. 4. The classification result of handgun images. (a) Standard and mirror position, and the red box is the standard position. (b) Classes of direction. (c) Classes of angle, the image in the green box is what this paper considers. (Color figure online)

2.2 Foreground Extracting

X-ray prohibited item images always have complex background. It is hard for network to extract common feature of background when the size of training data is not big enough. Furthermore, object foreground is much more important than background. So, matting method is here used to extract foreground of the X-ray prohibited item images, where original image, background image and trimap are required. The trimap only contains foreground, background and unknown pixel. The image foreground is extracted by Eq. (1),

$$I = \alpha F + (1 - \alpha)B, \tag{1}$$

where I is any pixel in the image, F is foreground pixel, B is background pixel, and α is fusion coefficient among 0 and 1. For certain background, $\alpha=0$, for certain foreground, $\alpha=1$. The α matrix can be obtained by KNN-matting [13]. The process for extracting foreground of handgun in X-ray images is shown in Fig. 5. Matting result shows that this method can remove the complex background and leave the foreground of interest in image.

3 Image Generative Model

The generated X-ray prohibited item images must be increased greatly in quantity and diversity. This can be achieved by two steps. First, many new images are
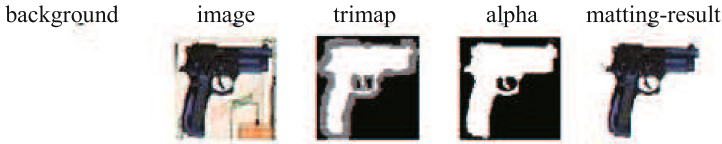
background image trimap alpha matting-result

Fig. 5. Image foreground extraction process. From left to right are the background image, original image, trimap, α matrix, and X-ray image that only has object foreground.

generated based on CT-GAN. Then, the CGAN model is improved for effectively re-adjusting the poses and scales of the generated item images. The flowchart of image generation is shown in Fig. 6.
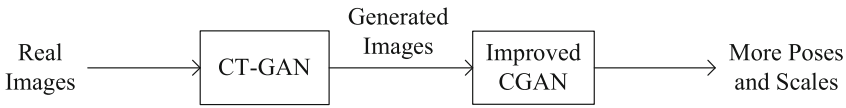

Fig. 6. Image generation flowchart

3.1 CT-GAN

CT-GAN is proposed based on the improvements of WGAN-GP. Compared with WGAN-GP, it performs better on small datasets and improves the stability of training. Here, CT-GAN is used to generate many images of X-ray prohibited items with high quality. It should be mentioned that we make some modifications to the loss function compared with Reference [14]. The loss function is defined as Eq. (2),

$$L = D(G(z)) - D(x) + \lambda_1 GP \mid_{x'} + \lambda_2 CT \mid_{x_1,x_2}, \tag{2}$$

the gradient penalty (GP) and consistency regularization (CT) are defined as Eqs. (3) and (4),

$$GP \mid_{x'} = E_{x'}[(\parallel \nabla_{x'} D(x') \parallel_2 - 1)^2], \tag{3}$$

$$CT \mid_{x_1,x_2} = E_{x \sim P_r}[max(0, d(D(x_1), D(x_2)) - M')], \tag{4}$$

where x' is uniformly sampled from the straight line between the generated data and real data. Both x_1, x_2 are real data. M' is a constant. The basic architecture of generator G is a deconvolutional neural network. The input is random Gaussian noise vector while the output is a generated image. The basic architecture of discriminator D is a convolutional neural network. Selecting suitable values of λ_1 and λ_2 can optimize the quality of generated images.

3.2   Improved CGAN

Many new images could be generated by CT-GAN, but they vary little compared with the real images. We improve the CGAN model [15] to increase the diversity of the generated images, including poses, position and scales. This model is different from the traditional GAN models, where the input of generator G is random noise. It uses an original image A and a target image B (there are different prohibited item poses in A and B) as the real data. The aim of G is to transform image A to image B'. So, image A and image B' are the fake data. Several training image pairs, $A - B$, are used to train the network. Finally, G can generate a new image based on image A without corresponding image B.
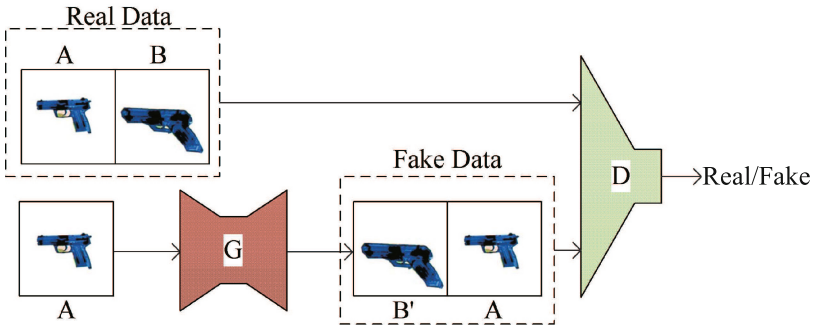

Fig. 7. The architecture of improved CGAN

The architecture of improved CGAN is shown in Fig. 7. The handguns in image A and image B are different in pose and scale. The architecture of D is still a convolution neural network, and the architecture of G adopts the structure of encoder-decoder. The images can be generated better by adding the gradient penalty. The loss function is defined as Eq. (5),

$$L = D(x, G(x)) - D(x, y) + \lambda GP, \tag{5}$$

4   Experiments and Results

In this section, the experimental details are discussed. Most X-ray prohibited item images used here are collected from Google, and a part of images is taken by a X-ray machine. This Section shows the results of various handgun images generated by CT-GAN and improved CGAN. In addition, some images of other prohibited items are also generated using the proposed method.

4.1   Generating Many Images Based on CT-GAN

CT-GAN is used to generated many new images. The dataset consists of more than 500 X-ray handgun images. All the images are resized to 96×64 pixels. The batch size is set to 64. Our model is trained for 1500 epochs with a learning rate 0.0001. The best generated image samples can be obtained when the training frequency of D is same with that of G.

(a)

(b)

(c)

(d)

Fig. 8. Some generated image samples. (a) Some real X-ray images. (b) Images generated by DCGAN. (c) Images generated by WGAN-GP. (d) Images generated by CT-GAN.

Images with different visual quality are generated based on CT-GAN and several other GAN models (shown in Fig. 8). The images that generated by DCGAN model are poor in quality. As for WGAN-GP, the resolution of most images have been improved, but some images still have ghost shapes of handguns. Compared with these models, the quality of images generated by CT-GAN have been improved obviously. Many handgun images with different poses are generated by CT-GAN, here some image samples are shown in Fig. 9.

4.2   Generating Images to Increase the Diversity by Improved CGAN

Firstly, we build 50 pairs of training image samples $A - B$. The handgun of B is different to that of A in pose, position and scale. Then, the improved

Fig. 9. Image samples generated by CT-GAN

CGAN model is trained for 500 epochs based on this dataset with a learning rate 0.0001. The new images generated (shown in Fig. 10) by the proposed method are different from rotating the images directly. There are more changes between the generated images and real images.

4.3   More Prohibited Item Image Generation

In order to test the generalization ability of the proposed method, we also generate some images of other prohibited items respectively, such as wrench, pliers, blade, lighter, kitchen knife, screwdriver, fruit knife and hammer. All the experiments performed on a dataset of 100–200 images. Some generated images are shown in Fig. 11.

The images generated here using our method only contain foreground. The complete X-ray images can be obtained by fusing the generated item images with existing background images through some rules. Here we have more interests on the foreground of images.

5   Verification

Most images generated by CT-GAN and the improved GAN are realistic. However, a part of images have poor quality because of the instability of training. Before using the generated images as new samples of dataset, it is necessary to verify whether or not the generated images belong to the same item class as the original images.

It can be verified by a simple CNN model that include three convolutional layers and three full connected layers. Both the training images and testing images are real X-ray security images, and they account for 75% and 25% respectively. The dataset has ten classes, include handgun, wrench, pliers, blade, lighter, kitchen knife, screwdriver, fruit knife, hammer and other items. Each class has

Fig. 10. The generated images based on improved CGAN. (a) Original handgun images. (b), (c) Two different generated image samples that have different handgun pose, position and scale.
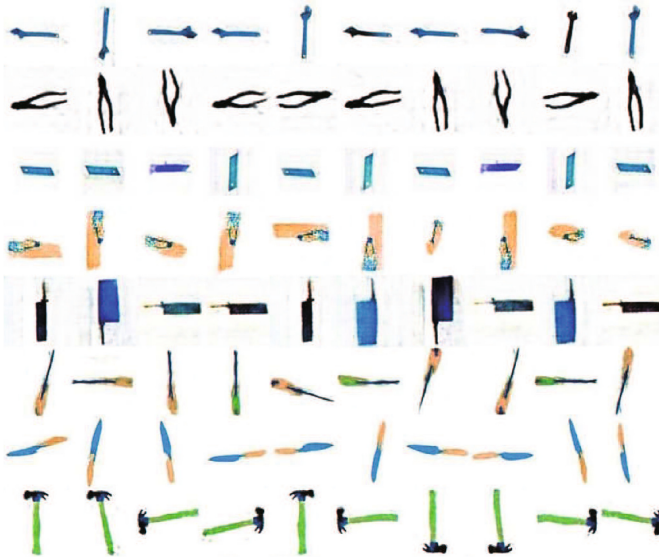

Fig. 11. Generated images of eight prohibited items. From top to bottom, the generated images are respectively wrench, pliers, blade, lighter, kitchen knife, screwdriver, fruit knife and hammer.

200 images, and different images have different item poses. Batch size is set to 64. After 25 epochs of training, the accuracy of classification on training dataset is 99.84% while the accuracy on testing dataset is 99.22%.

One hundred generated images are select randomly from each prohibited item class. Table 1 reports the count of images with correct matching labels. We can find that most images are classified correctly by CNN model.

Table 1. Matching results of CNN model

Prohibited item	Number
Handgun	100
Wrench	100
Pliers	100
Blade	87
Lighter	100
Kitchen knife	92
Screwdriver	91
Fruit knife	95
Hammer	100

6   Conclusions

In this paper, a GAN-based method was proposed to generate images of X-ray prohibited items. After image classifying and foreground extracting, many new images with various poses were generated by the CT-GAN model and the improved CGAN model. We also verified that most generated images belong to the same class with real images. Our work can increase the X-ray prohibited item image dataset effectively in both quantity and diversity.

References

1. Akcay, S., Kundegorski, M.E., Devereux, M., et al.: Transfer learning using convolutional neural networks for object classification within x-ray baggage security imagery. In: International Conference on Image Processing, pp. 1057–1061 (2016)
2. Mery, D., Svec, E., Arias, M.: Modern computer vision techniques for x-ray testing in baggage inspection. IEEE Trans. Syst. Man Cybern. **47**(4), 682–692 (2017)
3. Turcsany, D., Mouton, A., Breckon, T.P.: Improving feature-based object recognition for x-ray baggage security screening using primed visual words. In: International Conference on Industrial Technology, pp. 1140–1145 (2013)

4. Frid-Adar, M., Diamant, I., Klang, E., et al.: GAN-Based Synthetic Medical Image Augmentation for Increased CNN Performance in Liver Lesion Classification. arXiv preprint arXiv:1803.01229 (2018)
5. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. In: International Conference on Neural Information Processing Systems, pp. 2672–2680 (2014)
6. Mirza, M., Osindero, S.: Conditional Generative Adversarial Nets. Computer Science (2014)
7. Radford, A., Metz, L., Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. Computer Science (2015)
8. Salimans, T., Goodfellow, I.J., Zaremba, W., et al.: Improved techniques for training GANs. In: International Conference on Neural Information Processing Systems, pp. 2226–2234 (2016)
9. Gurumurthy, S., Sarvadevabhatla, R.K., Babu, R.V.: DeLiGAN: generative adversarial networks for diverse and limited data. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4941–4949 (2017)
10. Gulrajani, I., Ahmed, F., Arjovsky, M., et al.: Improved training of Wasserstein GANs. In: International Conference on Neural Information Processing Systems, pp. 5769–5779 (2017)
11. Karras, T., Aila, T., Laine, S., et al.: Progressive Growing of GANs for Improved Quality, Stability, and Variation. arXiv preprint arXiv:1710.10196 (2017)
12. Miyato, T., Kataoka, T., Koyama, M., et al.: Spectral Normalization for Generative Adversarial Networks. arXiv preprint arXiv:1802.05957 (2018)
13. Chen, Q., Li, D., Tang, C.: KNN matting. IEEE Trans. Pattern Anal. Mach. Intell. **35**(9), 2175–2188 (2013)
14. Wei, X., Gong, B., Liu, Z., et al.: Improving the Improved Training of Wasserstein GANs: A Consistency Term and Its Dual Effect. arXiv preprint arXiv:1803.01541 (2018)
15. Isola, P., Zhu, J., Zhou, T., et al.: Image-to-image translation with conditional adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 5967–5976 (2017)