A Semi-Supervised Railway Foreign Object Detection Method Based on GAN

Yanqi Chen
School of Automation, Southeast University and Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Nanjing China
18741932221@163.com

Shuzhen Tong
School of Automation, Southeast University and Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Nanjing China
114464169@qq.com

Xiaobo Lu*
School of Automation, Southeast University and Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Nanjing China

Yun Wei
Beijing Mass Transit Railway Operation Corporation Limited, Beijing China
luckyboy0309@163.com

ABSTRACT

The rapid development of deep learning provides new technical means for railway foreign object detection. However, in practical applications, the datasets of railways with foreign objects are scarce. In order to solve this problem, by improving the loss function and anomaly image evaluation standard, this paper proposes a new semi-supervised anomaly detection method based on GAN (Generative Adversarial Networks). Experiments show that our method can achieve railway foreign object detection without anomaly prior knowledge. Regarding anomaly recognition, a 0.058 AUC (Area Under Curve) and a 6% classification accuracy relative improvement for the railway dataset used in this paper are obtained.

CCS CONCEPTS

• **Computing methodologies**; • **Artificial intelligence**; • **Computer vision**; • **Computer vision problems**;

KEYWORDS

Railway, Anomaly detection, GAN, Semi-supervised

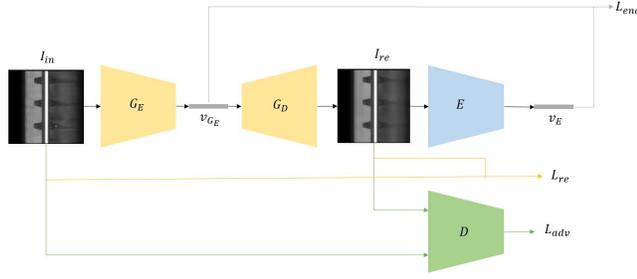*Corresponding author: xblu2013@126.com

1 INTRODUCTION

With the development of high-speed rail technology, railway mileage has increased rapidly. Railway anomalies can cause damage to rails and trains, which brings serious hidden dangers and threats to the safety of railway. Therefore, railway safety issues have become increasingly prominent. The task of railway anomaly detection becomes more and more important. At present, railway anomaly detection tasks can be divided into rail defect detection, fastener defect detection, sleeper defect detection, and railway foreign object detection. The purpose of this paper is the detection of railway foreign objects.

So far, some researchers have successfully proposed several methods based on visual images to collision avoidance for railway foreign objects detection [1]. However, this task still faces difficulties on (1) detecting a foreign object in complex railway scenes, (2) a low success rate of foreign object detection in dynamic camera scenes, (3) the lack of datasets with foreign objects. To solve the above problems, Passarella et al. [2] utilized a distance sensor-based method to detect foreign objects on tracks. However, this method relies too much on the sensor's accuracy and is easily affected by the surrounding environment. Mockel et al. [3] proposed a method for detecting foreign objects in tracks based on Lidar and computer vision technology. However, this method requires more computing resources and equipment. In addition, the actual effect is not good in long-distance detection. Gasparini et al. [4] adopted a GAN-based railway foreign objects detection method to solve the lack of anomaly images. But this method is easily affected by the railway environment because the railway images are obtained via a drone. For a general understanding of Neural Networks in general and GANs in particular, readers are referred to <Neural Networks and Deep Learning>.

This paper proposes a semi-supervised anomaly detection method based on GAN, which detects railway foreign objects on the railway inspection dataset. First of all, in the training stage, we take normal railway images as the input so that the model can learn and generate normal images. Then, in the inference stage, the normal samples' images are similar to the input. Images generated by abnormal samples are far from the input. After then, anomaly scores are calculated according to the differences between generated images and

Figure 1: Overall Network Architecture, Where the Yellow Is G, the Blue Is E, the Green Is D.

Figure 2: Sub-Network G Structure, Where the Left Blue Part Is G_E, and the Right Orange Part Is G_D.

input images. Finally, images with a score greater than a threshold will be determined as an anomaly. Compared with GANomaly [5] on the railway inspection dataset, this paper's method improves the AUC by 0.058, and the classification accuracy increases by 6%.

2 RELATED WORK

Currently, the most common anomaly detection methods are based on image reconstruction, which adopts sparse coding algorithm [6] and deep autoencoder algorithm [7] to learn feature distribution of normal samples and reconstruct them. The biggest challenge of anomaly detection is the lack of anomaly samples in practical applications. The proposal of the GAN brought an unsupervised machine learning algorithm [8]. AnoGAN firstly adopted GAN for anomaly detection [9]. This method trained the model without anomaly samples. GANomaly proposed a GAN-based anomaly detection method, which exploited the difference between hidden space features to infer anomaly samples [5].

3 METHOD

3.1 A GAN-Based Anomaly Detection Model

This paper proposes a semi-supervised railway foreign objects detection method based on GAN. The model is divided into three sub-networks: G(i.e., G_E, G_D), D, and E, as shown in Figure 1

**Sub-network G**: The sub-network G is a combination of an encoder and a decoder, as shown in Figure 2. With inputting an image I_{in}, $I_{in} \in \mathbb{R}^{w \times h \times c}$, it is firstly forwarded through the encoder G_E to obtain the compressed vector v_{G_E}, $v_{G_E} \in \mathbb{R}^z$. v_{G_E} is a vector with the smallest dimension but the most comprehensive feature representation of I_{in} in high-level feature space. Then, v_{G_E} is input to the decoder G_D in turn. Finally, I_{re} is reconstructed by v_{G_E} through G_D.

Sub-network D: The sub-network D acts as a discriminator similar to G_E but removes the last convolutional layer of G_E. Therefore, D utilizes the original penultimate layer as a feature extraction layer. A sigmoid activation function is added at the end, as shown in Figure 3 (a).

Sub-network E: The sub-network E is exactly the same as G_E but different in parameter updating. E re-encodes an image reconstructed by G to obtain a compressed vector v_E, $v_E \in \mathbb{R}^z$. v_E is also a vector with the smallest dimension but the most comprehensive
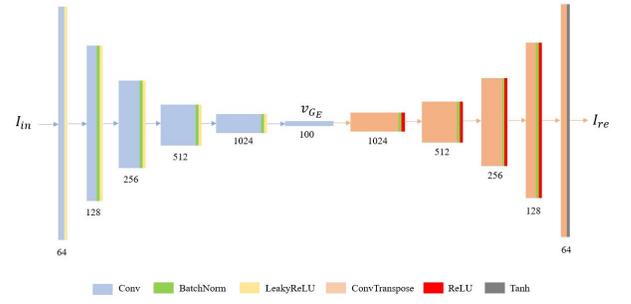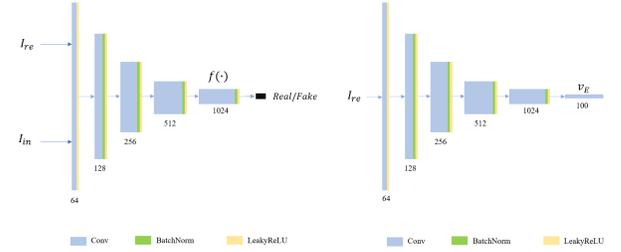
Figure 3: (a) Left: Sub-Network D Structure (b) Right: Sub-Network E Structure.

feature representation of I_{re} in high-level feature space, as shown in Figure 3 (b).

3.2 Training

In the training stage, the model learns feature representation and data distribution of normal images because it takes normal images I_{in} as the input. In the inference stage, with inputting a normal image I_{in}, a reconstructed image I_{re} and a vector v_E that is very similar to the input will be obtained. With inputting an anomaly image $\hat{I}_{in}$, the model parameters are not suitable for reconstructing anomaly images because the model failed to capture anomaly image information. Hence, an image $\hat{I}_{re}$ lacks anomaly features, which is quite different from $\hat{I}_{in}$. Meanwhile, the difference between the $\hat{v}_E$ and $\hat{v}_{G_E}$ is also large.

3.2.1 *Adversarial Loss.* Salimans et al. [10] proposed a GAN loss function based on feature matching. Feature matching emphasizes that feature representation of a discriminator's latent space is used to replace a true or false output to update the model, which avoids the problem of unstable GAN training process. Therefore, we combine 2-norm loss based on feature matching with traditional GAN adversarial loss as the model's adversarial loss function, as shown in formula (1) and formula (2).

For D

$$\mathcal{L}_{adv} = \mathbb{E}_{x \sim pX} ||f(x) - \mathbb{E}_{x \sim pX} f(G(x))||_2 + (\mathbb{E}_{x \sim pX}[logD(x)] + \mathbb{E}_{x \sim pX}[log(1 - D(G(x)))])$$

(1)

For G

$$\mathcal{L}_{adv} = \mathbb{E}_{x \sim pX}||f(x) - \mathbb{E}_{x \sim pX}f(G(x))||_2 + \\ \mathbb{E}_{x \sim pX}[\log(1 - D(G(x)))] \quad (2)$$

Mao et al. [11] proposed Least Squares GANs (LSGANs) based on least-square loss. Through LSGANs, only when a generator pulls generated images far from the decision boundary to the vicinity of the decision boundary, a least-square loss is small. Therefore, generated images based on least-square loss will be more realistic. We replace a cross-entropy loss with a least-square loss as the model's adversarial loss function, as shown in formula (3) and formula (4).

For D

$$\mathcal{L}_{adv} = \mathbb{E}_{x \sim pX}||f(x) - \mathbb{E}_{x \sim pX}f(G(x))||_2 + \\ (\mathbb{E}_{x \sim pX}[(D(x) - 1)^2] + \mathbb{E}_{x \sim pX}[(D(G(x)))^2]) \quad (3)$$

For G

$$\mathcal{L}_{adv} = \mathbb{E}_{x \sim pX}||f(x) - \mathbb{E}_{x \sim pX}f(G(x))||_2 + \\ \mathbb{E}_{x \sim pX}[(D(G(x)) - 1)^2] \quad (4)$$

3.2.2 Reconstruction Loss. A single adversarial loss function does not contain semantic information of image-level so that the necessary information to update a generator is not comprehensive. Therefore, we utilize distance information between the real input images I_{in} and reconstructed images I_{re} to reconstruct more realistic images. The reconstruction loss function is shown in formula (5).

$$\mathcal{L}_{re} = \mathbb{E}_{x \sim pX}||I_{in} - I_{re}||_1 \quad (5)$$

3.2.3 Encoder Loss. These two types of loss functions can make GAN networks generate real reconstructed images. In addition, a condition is added to constrain a generator in this paper. The difference between v_{G_E} and v_E in high-level abstract space is adopted to infer abnormalities. This additional abstraction level can significantly improve abilities to resist noise interference and learn a more robust anomaly detection model. The encoder loss function is shown in formula (6).

$$\mathcal{L}_{enc} = \mathbb{E}_{x \sim pX}||v_{G_E} - v_E||_2 \quad (6)$$

We add the weighted sum of the above three types of loss functions, as shown in formula (7).

$$\mathcal{L} = \omega_{adv}\mathcal{L}_{adv} + \omega_{re}\mathcal{L}_{re} + \omega_{enc}\mathcal{L}_{enc} \quad (7)$$

3.3 Inference

In the inference stage, we combine the reconstruction loss $\mathcal{L}_{re}$ and encoder loss $\mathcal{L}_{enc}$ as a criterion for identifying anomaly images. A reconstruction loss focuses more on local differences. In contrast, an encoder loss focuses more on global differences. The evaluation criterion as shown in formula (8).

$$A(x) = \alpha * ||x - G(x)||_1 + (1 - \alpha) * ||G_E(x) - E(G(x))||_2 \quad (8)$$

With inputting a normal image, the $A(x)$ is small. For an anomaly image, the $A(x)$ is large. So, we set a threshold θ, when $A(x) < \theta$, the input is judged to be normal; when $A(x) > \theta$, the input is judged to be abnormal. We normalized the calculated anomaly scores, as shown in formula (9).

$$\hat{A}(x) = (A(x) - minA(x))/(maxA(x) - minA(x)) \quad (9)$$
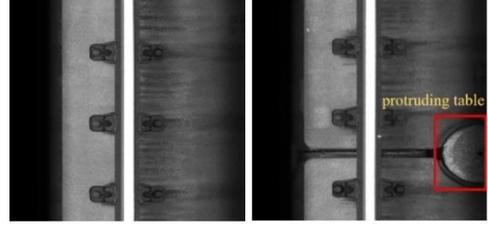
Figure 4: Railway Inspection Dataset. Left- (a) A Normal Railway Image; Right- (b) An Anomaly Railway Image.

4 EXPERIMENTS

In the railway anomaly detection, we expect to detect debris, bottles and peels and other foreign objects to avoid damage to the railway. This paper regards railway images with a protruding table as abnormal samples and railway images with only rails as normal samples due to the lack of images with foreign objects in the railway inspection dataset, as shown in Figure 4. There are 2543 normal images in the training dataset. And in the testing dataset, there are 628 normal images and 125 anomaly images. Our method sets the input image size to 128×128, epoch to 200, learning rate to 2e-4, batch-size to 16, weights of loss functions to ω_{adv} =1, ω_{re} =50, ω_{enc} =1, and the weight of reconstruction loss to α =0.5.

4.1 Results

This paper uses the area under the curve (AUC) of Receiver Operating Characteristic (ROC) and classification accuracy rate to evaluate the performance of our model [9] because anomaly detection is often treated as a classification task.

This paper conducted comparative experiments to obtain the ROC curve, as shown in Figure 5. The AUC and classification accuracy rate of different methods are presented in Table 1. The ROC curve in Figure 5 (a) is the result of GANomaly. Figure 5 (b) and Figure 5 (c) are ROC curves of our method, where the adversarial loss function is based on cross-entropy loss and least-square loss, respectively. The generated reconstruction images of anomaly samples are not perfect because the anomaly feature representation is not obtained in the training stage. This paper compared input images and anomaly reconstructed images, as shown in Figure 6

4.2 Analysis

Because our method adopts an adversarial loss function that combines 2-norm loss based on feature matching and adversarial loss of traditional GAN, the model avoids overtraining. Besides, an anomaly image evaluation standard based on reconstruction loss and encoder loss is adopted to infer whether the image is an anomaly sample from a local and global perspective. Through the above experiments, our method is stronger than GANomaly in the AUC and classification accuracy, as shown in Figure 5 and Table 1. Apart from these, this paper also considers that the adversarial loss function based on least-square can pull images far from the decision boundary to the vicinity of the decision boundary. This can punish a generator and generate more realistic railway images, as shown in Figure 6. For anomaly images, GANomaly will reconstruct the black cracks, as shown in Figure 6 (b). However, in our method, the
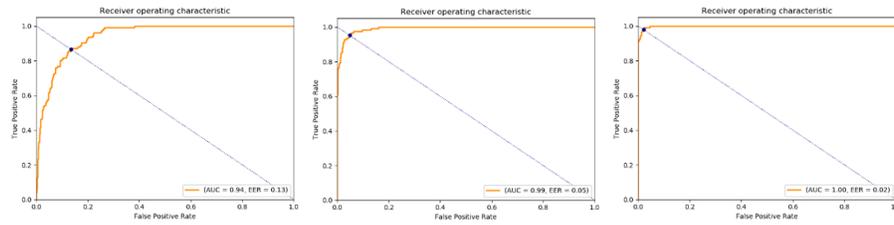
Figure 5: ROC Curve of Different Methods, (a) GANomaly, (b) Our Method with Cross-Entropy Adversarial Loss, (c) Our Method with Least-Square Adversarial Loss.

Table 1: AUC and Classification Accuracy of Different Methods

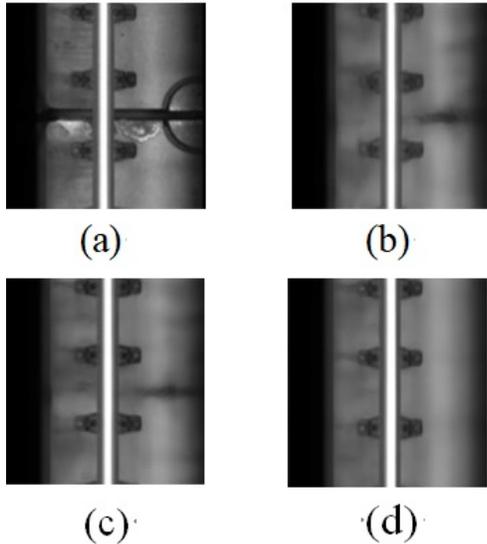| Method	AUC	Classification accuracy (%)
GANomaly	0.940	92
Our method(cross-entropy)	0.993	96
Our method(least-square)	0.998	98

Figure 6: Anomaly Railway Image and Reconstructed Railway Images of Different Methods. (a) Input, (b) GANomaly; (c) Our Method with Cross-Entropy Adversarial Loss; (d) Our Method with Least-Square Adversarial Loss.

black cracks of the reconstructed image based on cross-entropy adversarial loss are lighter and thinner, as shown in Figure 6 (c). In the reconstructed image based on least-square adversarial loss, there is nearly no black crack, as shown in Figure 6 (d). Therefore, the difference between the generated railway image without crack and the input railway image with a protruding table is more obvious, and it is more accurate in anomaly identification.

5 CONCLUSION

This paper proposed a semi-supervised anomaly detection method based on GAN to detect foreign objects in a railway environment. Our method has two contributions: (1) propose an adversarial loss function based on the combination of feature matching and least-square loss, (2) adopt an anomaly image evaluation standard based on local features and global features. On the railway inspection dataset, our method's AUC and classification accuracy are higher than GANomaly by 0.058 and 6% respectively, which shows effective results without prior knowledge of abnormalities. However, it only limits to finds out anomaly images excluding the anomaly location. In future work, we will explore ways to locate abnormalities on this basis.

ACKNOWLEDGMENTS

REFERENCES

[1] Maire F (2007, September). Vision based anti-collision system for rail track maintenance vehicles. In 2007 IEEE Conference on Advanced Video and Signal Based Surveillance (pp. 170-175). IEEE
[2] Passarella R, Tutuko B, & Prasetyo A P (2011). Design concept of train obstacle detection system in Indonesia. IJRRAS, 9(3), 453-460.
[3] Mockel S, Scherer F, & Schuster P F (2003, June). Multi-sensor obstacle detection on railway tracks. In IEEE IV2003 Intelligent Vehicles Symposium. Proceedings (Cat. No. 03TH8683) (pp. 42-46). IEEE.
[4] Gasparini R, Pini S, Borghi G, Scaglione G, Calderara S, Fedeli E, & Cucchiara R (2020, September). Anomaly detection for vision-based railway inspection. In European Dependable Computing Conference (pp. 56-67). Springer, Cham.
[5] Akcay S, Atapour-Abarghouei A, & Breckon T P (2018, December). GANomaly: Semi-supervised anomaly detection via adversarial training. In Asian conference on computer vision (pp. 622-637). Springer, Cham.
[6] Zhao B, Fei-Fei L, & Xing E P (2011, June). Online detection of unusual events in videos via dynamic sparse coding. In CVPR 2011 (pp. 3313-3320). IEEE
[7] Hasan M, Choi J, Neumann J, Roy-Chowdhury A K, & Davis L S (2016). Learning temporal regularity in video sequences. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 733-742).
[8] Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S & Bengio Y (2014). Generative adversarial networks. arXiv preprint arXiv:1406.2661.

[9] Schlegl T, Seeböck P, Waldstein S M, Schmidt-Erfurth U, & Langs G (2017, June). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In International conference on information processing in medical imaging (pp. 146-157). Springer, Cham.

[10] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, & Chen X (2016). Improved techniques for training gans. arXiv preprint arXiv:1606.03498.

[11] Mao X, Li Q, Xie H, Lau R Y, Wang Z, & Paul Smolley S (2017). Least squares generative adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2794-2802).