Improved Object Detection using Data Enhancement method based on Generative Adversarial Nets

Zhang Ruiqiang 1*, Liu Jiajia 1, Zeng Yu 1, Pan Kejia 1, and Huang Lin 1

1 Information and Communication Company of State Grid Sichuan Electric Power Company, Chengdu, Sichuan Province, 610041, China

*Corresponding author’s e-mail: rqzhang@yeah.net

Abstract. In deep learning-based object detection, especially in face detection, small target and small face has always been a practical and common difficult problem due to its low resolution, blurred image, less information and more noise. In some applications, sensing image data is hard to collect, leading to limited object detection performance. In this paper, we investigate using a generative adversarial network model to augment data for object detection in images. We use generative adversarial network to generate the diverse objects based on the current image data. An improved generative adversarial network is added in the network and a new loss function is applied during the training process to generate diverse and high-quality training images. Experiments show that images generated by generative adversarial network have higher quality than counterparts.

1. Introduction

Small object detection is one of the most fundamental parts images understanding and computer vision. Small object detection is technically the basis of many other computer vision applications, such as real time object tracking [1], image object segmentation [2,3], image captioning and labelling [4], person action capturing [5], virtual scene understanding [6]. If the images themselves are in a low resolution, then small objects may only be represented by a few pixels. With so limited information from the images, even humans would not perform well enough. In addition, most of the computer vision architectures scale the input size to a fixed size which is usually very small. This may reduce the image size further. Moreover, many computer vision models use pooling layers which may remove features from the images. This would be an essential problem if some features of the small objects are removed after pooling. While detection on small objects may not always be necessary, it could be typically crucial to some real-world problems. For example, the vehicle perception models for self-driving requires extremely accurate detection behaviours on small objects. Any failures of traffic lights detection can cause severe problems.

To alleviate the inefficiency caused by small training images problem, data enhancement techniques have been proposed to increase the images numbers and diversity. These methods can be classified into three catalogues [7]. The first kind is spatial geometric transformation methods which uses rotation, flip, zoom, crop and shift to generate diverse kinds of images. The second type of method considers pixels transformation such as colour jittering, noise added and random erasing to generate more types of training images to feed into the machine learning model. The third way of image enhancement is to use over-sampling technique [8], sample pairing [9] to generate images. However, these image data enhancement methods depend on many hand-crafted rules, which is difficult to get for commonly unexperienced use and the images produced are also of limited diversity.
In this paper, a novel generative adversarial network based (GAN) model is proposed to generate high-quality and diverse images. The main works of this paper include: (1) We apply an optimized GAN model to argument small-scale object detection problems. (2) We use data distribution feature in the generate model to improve the clarity of the generated images. (3) R-CNN based samples selection technique is used to sample features from objects to obtain object diversity. (4) Experimental studies show the effectiveness of the proposed method.

The rest of the paper is organized as follows. In Section 2, related work of existing small object detection algorithms and argumentation methods is described. Section 3 introduces the GAN framework. In Section 4, we introduce the object tracking network and optimization. Experimental studies are presented in Section 5. In Section 6, we conclude the paper.

2. Related Works

Object detection is widely used in image understanding and video surveillance applications. The object detection methods based on deep convolutional neural networks (DCNNs)[11-13] have improved image detection accuracy greatly over traditional methods. In some applications, training images are hard to obtain [13] and small sized. Thus, different image enhancement techniques are used to improve the accuracy of object detection methods. [14] used Faster R-CNN model and covered image pixels with random values to enhance image generation ability. [15] used image mirroring, rotation, Gaussian blur and Gaussian noise to enhance the training image. [16] enhance the train image data by randomly dropping some objects in the training images which can produce richer variation image data set than method in [17]. In [18], authors adopted colour jittering and geometric alteration to generate the diverse kinds of training images. [19] simulates camera shooting and illumination synthesis to generate different marine uneven illuminating images for marine organisms’ detection. [20] generates images by changing physical characters to improve object detection.

3. Basics of Generative Adversarial Network

The GAN model was firstly proposed in [21]. Generally, GAN consists of two models: a generator G and a discriminator D. The optimization objective function of GAN is defined as follows:

$$L = \min \max_E \sum_{x \sim P_{data}} \left[ \log D(x) \right] + \sum_{z \sim P_{noise}} \left[ \log (1 - D(G(z))) \right]$$

where $G$ denotes the generative model; $D$ represents the discriminative model, $E$ is the mean operation. $z$ is the noise following Gaussian distribution $P_{noise}(z)$.

Figure 1 shows the discriminative network used in this paper.

![Figure 1. The basic discriminative network structure of GAN network.](image)

An improvement of GAN is to replace the convolutional layers with residual functions which produce a residual generative network (RGAN) by adding the output feature map into the input layer.
4. Network improvement and optimization

We optimize the network firstly by replacing the generator G resolution network with a 32 residual blocks of size $96 \times 96$, and add discriminator network D to form a complete GAN network. Actually, to compare with other methods, we use the Wasserstein GAN (WGAN) [22] model and add a gradient penalty parameter in the model. The loss function of the discriminator used in the model is formulated in (2):

$$
\mathcal{L}_D = \sum_{x_{HR}} -D_{\Phi}(x_{HR}) - \sum_{x_{SR}} D_{\Phi}(x_{SR}) + \lambda \sum_{x} \left( \|
abla D_{\Phi}(x)\|_2 - 1 \right)^2
$$

where $x_{HR}$ is the resolution of the image, $x_{SR} = G_{sr}(x_{LR})$ is the generated super resolution image, $x_{LR}$ is the low resolution of the image, $G_{sr}$ is the super-resolution of the image, $D_{\Phi}$ is the discriminator model, $P_{r}$ is real image data distribution, $P_{g}$ is the generated image data distribution; $r_{D\Phi}$ is gradient parameter of the discriminator. $\bar{x}$ is a random vector sampled from $x_{HR}$ and $x_{SR}$.

The next proposed improvement uses an HR image as input to generate a low resolution image as shown in Figure. 2.

The loss function used is as follows:

$$
\mathcal{L}_{SR-GAN} = \mathcal{L}^{1,1}(x_{HR}, G_{hr}(x_{LR})) + \mathcal{L}^{1,1}(x_{LR}, G_{lr}(x_{HR})) + \mathcal{L}^{MSE}(x_{HR}, G_{hr}(G_{lr}(x_{HR}))) + \mathcal{L}^{MSE}(x_{LR}, G_{lr}(G_{hr}(x_{LR})))
$$

In the loss function (3), we try to integrate the super resolution image into a the discriminator model. GAN model computes loss function between the super-resolved images generated by $G_{hr}$ and high-resolution images. If the super resolution image is similar to the high resolution image, the low resolution image is generated by the high resolution image.

5. Experimental Study

To validate the performance of the proposed GAN model, we compare our model with two popular GAN methods (WGAN [24], LSGAN [25]) and run experiments to evaluate the performance of these GAN models. To compare with other methods, we set parameters similar to experiments in those works (the epochs of all GAN model is set to 500). The model is implemented with Python 3.5 and Pytorch library.
Figure 3 shows the images generated by the GAN model over the YOLO network. We can see the images generated are better than LR, EDSR methods. From Table 1, we can see that, compared to the result obtained from high resolution images, the GAN based models can detect all vehicles from the image, but only part of the vehicles were detected from the super resolution images generated by the bicubic interpolation algorithm and EDSR model. Among the tested models, the Yolo based model provided the most accurate results as shown in Table 1.

The proposed image enhancement method is also compared with traditional data augmentation techniques and other GAN models. We first use the GAN model to enhance the training images, then the images were rotated, flipped and transformed to get more versatile forms. Table 2 shows the detection results.

### Table 1. Experiments over different IoUs for all data enhancement methods.

| Method             | IoU(0.05) | IoU(0.25) | IoU(0.25) |
|--------------------|-----------|-----------|-----------|
| High resolution    | 97.35%    | 94.56%    | 83.43%    |
| Bicubic            | 23.67%    | 18.2%     | 10.12%    |
| SRGAN              | 63.65%    | 59.55%    | 44.56%    |
| SRGAN-Yolo         | 77.80%    | 72.35%    | 56.05%    |

### Table 2. Experiment over different data augmentation methods (APship, APstor, APbase denote the ship, store, baseball).

| datasets          | APship   | APstor   | APbase   |
|-------------------|----------|----------|----------|
| original          | 73.64%   | 79.95%   | 97.94%   |
| original + tradition| 72.21%  | 90.15%   | 97.34%   |
| original + WGAN    | 79.28%   | 80.34%   | 98.69%   |
| original + LSGAN   | 78.79%   | 80.03%   | 97.46%   |
| original + SRGAN   | 79.90%   | 79.97%   | 97.99%   |
original + SRGAN + tradition 78.72%  89.54%  97.99%

From Table 2 we can see, the average precision value increases. The reason is that the training images become fluent after rotating, flipping and transforming. With more training images, the detection model can be well-trained and object detection accuracy is greatly improved. With the combination method, images are expanded greatly and diversity of images is enough to reduce imbalance distribution of positive and negative image samples in the training process. With both image size and diversity both improved, the accuracy of object detection can be largely improved.

6. Conclusion
In this paper, we investigate a GAN model to enhance image datasets to improve accuracy of object detection in small size image sets. We use the GAN model to generate different level of images. Furthermore, the diversity of the images are obtained with some mapping. With the generated new images, the learning model is well-trained and more object features are extracted. Experimental studies with some popular methods showed that the proposed method can improve the average detection precision by 4.48%. In the future, we will work on parameter tuning of the model, to reduce training computing time and memory usage.

Acknowledgments
This paper is supported by the project of the State Grid Sichuan Electric Power Company under grant No. 521947180034.

References
[1] K. Kang, H. Li, J. Yan, X. Zeng, B. Yang, T. Xiao, C. Zhang, Z.Wang, R.Wang, X.Wang, W. Ouyang, T-CNN: tubelets with convolutional neural networks for object detection from videos, IEEE Trans. Circ. Syst. Video Tech. 28 (10) (2018) 2896–2907.
[2] J. Dai, K. He, J. Sun, Instance-aware semantic segmentation via multi-task network cascades, Computer Vision and Pattern Recognition 2016, pp. 3150–3158.
[3] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2020) 386–397.
[4] Qi Wu, Chunhua Shen, Peng Wang, Anthony R. Dick, Anton van den Hengel, Image captioning and visual question answering based on attributes and external knowledge, IEEE Trans. Pattern Anal. Mach. Intell. 40 (6) (2018) 1367–1381.
[5] S. Herath, M. Harandi, F. Porikli, Going deeper into action recognition: a survey, Image Vis. Comput. 60 (2017) 4–21.
[6] B. Zhou, Z. Hang, X. Puig, T. Xiao, S. Fidler, A. Barriuso, A. Torralba, Semantic understanding of scenes through the ADE20K dataset, Int. J. Comput. Vis. 127 (3) (2016) 302–321.
[7] Z. Cui, M. Zhang, Z. Cao, C. Cao, Image data augmentation for SAR sensor via generative adversarial nets, IEEE Access 7 (2019) 42255–42268.
[8] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, Smote: synthetic minority oversampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[9] H. Inoue, Data augmentation by pairing samples for images classification, arXiv: 1801.02929 (2018).
[10] S. Ren, K. He, R. Girshick, J. Sun, Faster r-CNN: Towards real-time object detection with region proposal networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2015, pp. 91–99.
[11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, SSD: Single shot multibox detector, in: Proceedings of the European Conference on Computer Vision, Springer, 2016, pp. 21–37.
[12] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.

[13] X.-Y. Zhang, S. Wang, X. Yun, Bidirectional active learning: a two-way exploration into unlabeled and labeled data set, IEEE Trans. Neural Netw. Learn. Syst. 26 (12) (2015) 3034–3044.

[14] Y. Ren, C. Zhu, S. Xiao, Deformable faster R-CNN with aggregating multi-layer features for partially occluded object detection in optical remote sensing images, Remote Sensing 10 (9) (2018) 1470.

[15] Y. Li, W. Hu, H. Dong, X. Zhang, Building damage detection from post-event aerial imagery using single shot multibox detector, Appl. Sci. 9 (6) (2019) 1128.

[16] H. Qu, L. Zhang, X. Wu, X. He, X. Hu, X. Wen, Multiscale object detection in infrared streetscape images based on deep learning and instance level data augmentation, Appl. Sci. 9 (3) (2019) 565.

[17] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, Random erasing data augmentation, arXiv: 1708.04896 (2017).

[18] R. Wang, J. Xu, T.X. Han, Object instance detection with pruned alexnet and extended training data, Signal Process. Image Commun. 70 (2019) 145–156.

[19] H. Huang, H. Zhou, X. Yang, L. Zhang, L. Qi, A.-Y. Zang, Faster R-CNN for marine organisms detection and recognition using data augmentation, Neurocomputing 337 (2019) 372–384.

[20] P. Wei, J. Ball, D. Anderson, Fusion of an ensemble of augmented image detectors for robust object detection, Sensors 18 (3) (2018) 894.

[21] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, X. Bing, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Proceedings of the International Conference on Neural Information Processing Systems, 2014.

[22] Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein gans. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 5767–5777.

[23] Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.

[24] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: Proceedings of the International Conference on Machine Learning, 2017, pp. 214–223.

[25] X. Mao, Q. Li, H. Xie, R.Y. Lau, Z. Wang, S. Paul Smolley, Least squares generative adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2794–2802.