Optimization of Small Object detection based on Generative Adversarial Networks

Zhang Ruiqiang[1*], Zeng Yu[1] and Jin Xin[1]

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

Abstract. Small object detection is one of the fundamental problems in computer vision applications. Existing small object detection techniques usually focus on detecting small objects with multiple scale of features with low efficiency due to high computational cost. In this paper, we investigate small object detection problem based on generative adversarial architecture that utilizes features of small objects. We propose an Optimized Perceptual Generative Adversarial Network (OPGAN) to present more features of small objects. Specifically, the generator of OPGAN learns to present the low-resolution features of the small objects to highly resolved features similar to large objects as input image of the discriminator model. After then, the discriminator of OPGAN computes the generated feature and generates a new perceptual requirement parameter into the model to train the model iteratively. Extensive experiments on the challenging benchmark data sets demonstrate the effectiveness of OPGAN in detecting small objects.

1 Introduction

Small object detection is a one of the fundamental parts related to image understanding and computer vision applications which has been widely used in online real time object tracking [1], image segmentation [2,3], image captioning and labelling [4], people action detection [5], real and virtual scene understanding [6].

Some methods [7-10] have been focused on small object detection scenarios. [11, 12] propose methods to increase feature scale of input images to enhance resolution of small objects and generate high-resolution feature images. [13, 14, 15] develop deep learning networks to generate multi-scale features to enhance high-level small-scale features with multiple lower-level features layers.

In this paper, we optimize the parameters of GAN generator and the discriminator network by solving a min-max optimization problem. In the OPGAN model, the generator is trained to generate with largely similar features from small objects. However, the discriminative capability of the discriminator is trained and improved with super-resolved features from real large objects, and feedback the features back to the generator.

The main works of this paper include:(1) We apply an optimized GAN-alike model to solve small-scale object detection problems. (2) We present a generator model that learns the additive residual representation between large and small objects. (3) A perceptual discriminator is used to provide comprehensive supervision beneficial for detections, instead of barely differentiating fake and real.

The paper is organized as follows. We give a small object review in section 2. In section 3, we introduce the structure of GAN model. In Section 4, the proposed object tracking approach is described. Experimental studies are presented in Section 5. We present final conclusion in Section 6.

2 Related Works

Small object detection is largely applied in traffic sign detection and recognition [16-21]. Traditional techniques for this task include [22] [23]. Recently, CNN-based approaches have been utilized in object detection in traffic sign detection and classification applications. [11] applied multi-stage features to the classifier to improve traffic sign recognition precision. [15] trained the CNN with loss and obtained better test accuracy and faster stable convergence. [22] used a CNN to detect traffic signs. [21] trained two CNNs for traffic signs classification.

Some works have used small object detection techniques in pedestrian detection. In [21] and [22], the authors utilize Integral Channel Features and Aggregated Channel Features to improve precisions. Some machine learning-based methods have been proposed to improve the detection performance of pedestrian detection [12-15]. [16] adds a deformation hidden layer in the CNN to present mixture poses features. [17] optimized pedestrian detection in semantic tasks. In the training process of GAN model, [19] utilized multiple levels of features and shape information of small objects to learn the detectors.[14] firstly propose the concept of Generative Adversarial Networks. To improve the image training data sets, [5] and [6] adopted generative network to generate more types of fake images. In [22] and [23], GANs were used to obtain a feature mapping from one manifold space

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

to another space. Unsupervised representation learning based on GAN was proposed in [18]. GAN was also utilized to generate super resolution images in [21].

3 Basics of Perceptual Generative Adversarial Network

Generative Adversarial Network [22] has been largely used in object tracking scenarios to enhance the image features. GAN is formed with two parts: the generative model and discriminative model. The generative model is trained with noise data. The discriminative model takes input data from the generative model or other training data, and outputs the classification probability. The model of the GAN learning process can be formed in formula (1).

$$L=\min_{G}\max_{D}E_{x\sim p_{data(x)}}[\log D(x)]+E_{z\sim p_{noise(z)}}[\log(1-D(G(z)))] \quad (1)$$

In (1), G is the generative model; D is the discriminative model, E is a mean operator. $\boldsymbol{x}$ is input vector of D and z is input vector of D which follows distribution $P_{data}(x)$ and $P_{noise}(z)$, respectively.

The image feature mapping is generated by G network as $G(I)$. The value of the element (i, j) is denoted as $\hat{M}_{ij}$. The input image is transform into a vector as I, and the value of element (i, j, k) on image I as I_{ijk}. The dropout operation in the network is given in formula (2).

$$I_{ijk}^{0}=I_{ijk}\hat{M}_{ij} \quad (2)$$

where I_{ijk}^{0} denotes the dropout operation over image I.

4 Object Tracking Method based on GAN

In small object detection model, a deep residual network is used and the low level image is enhanced and feeded into the model.The Perceptual Generative Adversarial network is shown in Figure 1.

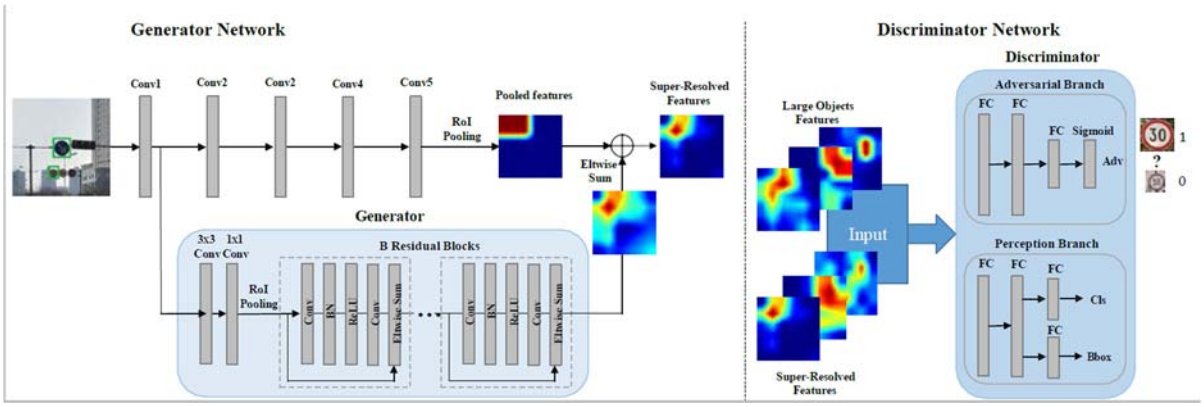

Figure 1. The network framework of the proposed OPGAN

In Figure. 1, the generative model of OPGAN framework encodes the input image into feature representation, and decode it into corresponding outputs feature representation. The discriminative model used in OPGAN is a standard convolutional neural network.

The SGD based loss function is optimized:

$$\arg\min\frac{1}{N}\sum_{i=1}^{N}\left(-\sum_{j=1}^{2}p(j)\log(q(j))\right) \quad (3)$$

where p and q are training samples and corresponding classification labels, respectively; N is training samples number.

To train the object detect model, we use the mean squared error (MSE) to measure the difference between estimated attention response feature and ground truth feature. The mean squared loss used in this paper is as formula (4).

$$L_{MSE}=\frac{1}{N}\sum_{j=1}^{N}\left(S_j-\hat{S}_j\right)^2 \quad (4)$$

$\hat{S}$ and S are the attention feature, response feature maps and corresponding ground truth feature.

MSE loss function only works on pixel-level features. We use adversarial loss function in the training process to improve the tracking performance. The improved training loss function is described in (5) :

$$L_{dis\_p}=L_{cls}(p,g)+\mathbf{1}[g\geq 1]L_{loc}(r_g,|r^*) \quad (5)$$

where L_{cls} is classification loss, and L_{loc} is the bounding-box regression loss.

To improve the converge speed, we add adversarial loss into the MSE loss model. The final loss function used in the adversarial training mode is formulated in (6):

$$L_{GAN}=L_{AL}\left(D(C,G(C)),1\right)+\lambda L_{MSE} \quad (6)$$

where λ is set as 1/20 in the experiments according to previous papers work.

5 Experimental Analysis

In this section, we present the experimental study of the proposed tracking method. We use the Tsinghua-Tencent traffic-sign benchmark for small object detection in order to compare with other method. The image data set contains 30,000 traffic-sign objects for small objects detection. We use Caltech benchmark [9] for pedestrian detection task.

Recall and accuracy results of OPGAN are shown in Table 1 compared with other techniques on traffic-sign detection tasks. From Table 1, we can find that the proposed OPGAN has a better precision over the counterparts.

Table 1. Small object detection performance comparison over Tsinghua-Tencent 100K.

Object size	Small size	Medium size	Large size
Fast R-CNN [11] (Recall)	47%	72%	76%
Fast R-CNN [11] (Accuracy)	74 %	83%	81%
Faster R-CNN [22] (Recall)	51%	85%	92%
Faster R-CNN [22] (Accuracy)	26%	67 %	81%
OPGAN (Recall)	89 %	95%	88%
OPGAN (Accuracy)	84 %	92%	92%

We evaluate the OPGAN over Caltech benchmark pedestrian data set. We compare the detection result of OPGAN with other existing methods that have worked on the Caltech data sets. The results are shown in Figure 2. From Figure 2 we can see, the OPGAN outperforms counterparts with the lowest miss rate of 9.4%.

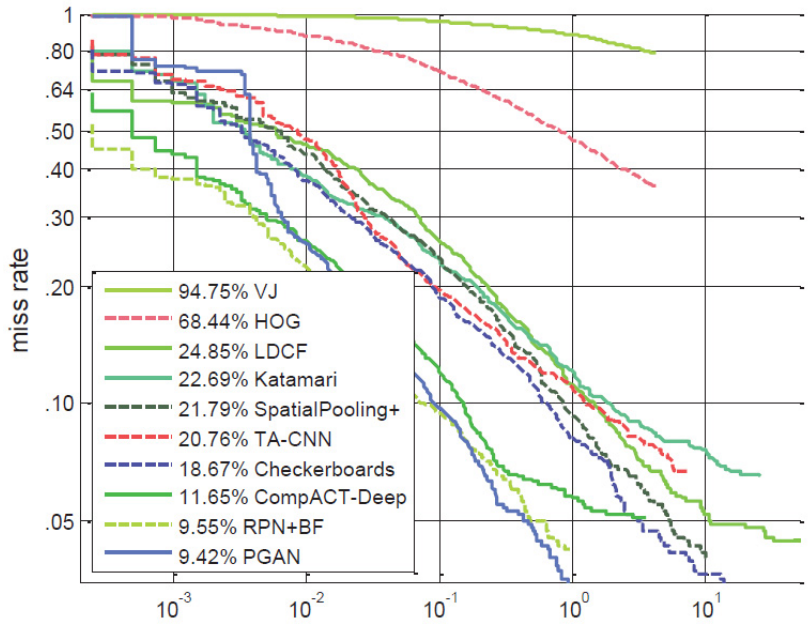

Figure 2. Detection performance comparisons of different techniques on the Caltech data set.

To validate the effectiveness of the feature extraction method in our model, we compare our method with other popular methods. The results are shown in Table 2. As we can see OPGAN outperforms other feature representation methods in both average recall and accuracy.

Table 2. Comparisons of different feature repesentation methods. (R): Recall,(A): Accuracy.

Object size	Small size data	All data
Skip Pooling (Recall)	71%	88%
Skip Pooling (Accuracy)	83 %	87%
Large Scale Images (Recall)	85.3%	92.4%
Large Scale Images (Accuracy)	82%	87 %
Multi Scale Images(Recall)	88%	92%
Multi Scale Images(Accuracy)	78%	84 %
OPGAN (Recall)	89.5 %	94%
OPGAN (Accuracy)	84.2 %	88.1%

Figure 3. Comparison between [22] and Our PGAN.

We compare OPGAN with method proposed in [22] to detect small objects in real application. From Figure 3, we can see that OPGAN can detect most small traffic signs as shown in [22].

6 Conclusion

In this paper, we investigate optimized perceptual generative adversarial network(OPGAN) to improve small object detection efficiency. The proposed OPGAN can output highly resolved features of small objects. The OPGAN generator utilizes a residual feature from lower level layers, we utilize a more efficient loss function to accelerate the training process. Experiments demonstrate the effectiveness of the proposed OPGAN in small object detection and pedestrian detection applications. In the future, we will consider to improve the computational cost of the training model by parameter tuning.

Acknowledgments

The work in this paper is fully supported by the project of the State Grid Sichuan Electric Power Company under grant No. 521947180034.

References

1. S. Bell, C. L. Zitnick, K. Bala, and R. Girshick. Inside-outside net:Detecting objects in context with skip pooling and recurrent neural networks. arXiv preprint arXiv:1512.04143, 2015. 1, 6
2. R. Benenson, M. Omran, J. Hosang, and B. Schiele. Ten years of pedestrian detection, what have we learned? In ECCV, pages 613–627, 2014. 6
3. K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531, 2014. 5
4. X. Chen, K. Kundu, Y. Zhu, A. G. Berneshawi, H. Ma, S. Fidler, and R. Urtasun. 3d object proposals for accurate object class detection. In NIPS, pages 424–432, 2015. 1
5. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, pages 886–893, 2005. 6
6. E. L. Denton, S. Chintala, R. Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In NIPS, pages 1486–1494, 2015. 2
7. P. Doll´ar, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. TPAMI, 36(8):1532–1545, 2014. 2, 5
8. P. Doll´ar, Z. Tu, P. Perona, and S. Belongie. Integral channel features. In BMVC, volume 2, page 5, 2009. 2
9. P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection:An evaluation of the state of the art. TPAMI, 34(4):743–761, 2012.1, 2, 5, 6
10. M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. 88(2):303–338, 2010. 8
11. R. Girshick. Fast r-cnn. In ICCV, pages 1440–1448, 2015. 1, 4, 5, 6,7, 8
12. R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, pages 580–587, 2014. 5
13. X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In Aistats, volume 9, pages 249–256, 2010. 5
14. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In NIPS, pages 2672–2680, 2014. 2, 3

15. M. Haloi. A novel plsa based traffic signs classification system. arXiv preprint arXiv:1503.06643, 2015. 2
16. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015. 5
17. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In ACM Multimedia, pages 675–678, 2014.5
18. H. Jiang and S.Wang. Object detection and counting with low quality videos. In Technical Report, 2016. 1
19. J. Jin, K. Fu, and C. Zhang. Traffic sign recognition with hinge loss trained convolutional neural networks. IEEE Transactions on Intelligent Transportation Systems, 15(5):1991–2000, 2014. 2
20. T. T. Le, S. T. Tran, S. Mita, and T. D. Nguyen. Real time traffic sign detection using color and shape-based features. In Asian Conference on Intelligent Information and Database Systems, pages 268–278. Springer, 2010. 2
21. C. Ledig, L. Theis, F. Husz´ar, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image superresolution using a generative adversarial network. arXiv preprint arXiv:1609.04802, 2016. 2
22. C. Li and M. Wand. Combining markov random fields and convolutional neural networks for image synthesis. arXiv preprint arXiv:1601.04589, 2016. 2
23. H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua. A convolutional neural network cascade for face detection. In CVPR, pages 5325–5334, 2015. 1