ABSTRACT  Due to the low detection accuracy of YOLOv3 target detection method, this paper proposes an improved target detection method of YOLOv3 remote sensing image. Firstly, the feature extraction network DarkNet53 is strengthened to improve the ability of feature extraction; Secondly, the original Leaky ReLU activation function is replaced by the Mish activation function, thus improving the generalization of the method in this paper; Finally, the Learning rate and BatchSize parameters are modified to prevent overfitting. The remote sensing image datasets of RSOD and TGRS-HRRSD are used in this paper. The Average Precision (AP) results of the method on the RSOD datasets in this paper show that the mAP value is 5.33 percent higher than that of the previous YOLOV3 method. The log average miss-rate (LAMR) results show that the LAMR value is 0.1100 lower than that of the previous YOLOv3 method. The mAP results of the method on the TGRS-HRRSD datasets show that the MAP value is 1.29 percent higher than that of the previous YOLOv3 method, and the LAMR results show that the LAMR value is 0.0338 lower than that of the previous YOLOv3 method.

INDEX TERMS  Target detection, remote sensing image, feature extraction network, mish activation function, RSOD datasets, TGRS-HRRSD datasets.

I. INTRODUCTION

Nowadays, the research in the direction of computer vision plays a pivotal role in promoting robot technology, unmanned intelligent transportation, military field of monitoring and security, and aerospace field, etc. Therefore, the research on detection of small and medium targets in remote sensing images has a very far-reaching significance [1]–[6]. In recent years, due to the rapid development of satellite field at home and abroad, remote sensing satellite with high resolution and aerospace field can provide a large number of remote sensing image datasets, which opens up a very good prospect for target detection in remote sensing image under computer vision. Traditional target detection technology requires manual setting of feature images [7]–[9], and the robustness of such target detection algorithms is poor, so it is difficult to solve the problems faced by the current target detection field. Therefore, Convolutional Neural Networks (CNN) should be used as the feature extractor for the research on detection of small and medium targets in remote sensing images. Among them, the classic ones are VggNet [10], [11], ResNet [12] and AlexNet [13], [14].

Nowadays, the acquisition of remote sensing image datasets mainly depends on remote sensing satellite platform. Since remote sensing image is a data set with small target and complex background, and the acquired remote sensing image will face noise, weather, illumination, occlusion and background interference, therefore, the research on the direction of small target detection in remote sensing image is faced with huge problems, but also brings greater challenges. To solve this problem, convolutional neural network is used to deal with target detection in remote sensing images. Although most of the detection algorithms have a relatively good effect on the common datasets and good generalization, the detection accuracy is still very low in the small target datasets of remote sensing images. Lin et al. [15] proposed FPN(Feature Pyramid Networks), which firstly integrates the up-sampling of small-scale feature maps with large-scale feature maps to improve the accuracy of small target detection. The Perceptual Generative Adversarial Network (Perceptual GAN)
proposed by Li et al. [16] improves the accuracy of small target detection by reducing the representation gap between small target and large target. The Regional Proposal Network (RPN) proposed by Cheng et al. [17] is combined with SSD algorithm [18], integrating the advantages of single-stage algorithm and two-stage algorithm, and adding the feature pyramid structure. The accuracy of small target detection is improved by using the information of multi-convolutional layer fusion feature and high-level feature. Zhang et al. [19] proposed a multi-scale feature fusion algorithm, designed a feature fusion module to integrate the position information of the shallow feature map and the semantic information of the deep feature layer, designed a module to remove redundant information, and further extracted features through convolution operation, thus improving the detection accuracy of small targets. Zhang et al. [20] improved the YOLOv4 [21] algorithm by strengthening the CSP feature extraction network of the original algorithm to increase its feature extraction capability. They also replaced the original activation function Leaky ReLU with Mish activation function and added the space pyramid pooling module, finally improving the accuracy of small target detection. Hou et al. [22] improved the trunk feature extraction network of YOLOv4, and used DenseNet, a dense connection network in the convolutional neural network, to enhance the feature extraction of aircraft targets. In addition, K-means algorithm was used in the datasets to obtain the number and size of the best priori box, which improved the detection accuracy of small targets.

This paper made an improvement on the algorithm of YOLOv3 [23]. YOLOv3 algorithm uses the main feature extraction network DarkNet53 [24], [25]. Firstly, on the feature extraction network, the input size of the image is modified from 416*416*3 to 800*800*3, which improves the DarkNet53 feature extraction network and strengthens the ability of extracting small and medium targets from remote sensing images. Secondly, in order to prevent overfitting, the learning rate was reduced and the batch size was increased. Finally, Zhang et al. [20] used Mish activation function to replace the original activation function Leaky ReLU for reference, thus improving the generalization of the algorithm. Finally, the experimental results on RSOD and TGRS-HRRSD remote sensing image datasets show that the proposed algorithm has higher advantages in small target detection of remote sensing images.

II. YOLOv3 TARGET DETECTION MODEL

A. CONVOLUTIONAL NEURAL NETWORK

Yann LeCun of New York University proposed Convolutional Neural Network in 1998, which is essentially a multi-layer perceptron, mainly composed of convolutional layer, pooled layer and fully connected layer. Each layer has multiple feature maps, each feature map extracts and inputs a feature through a convolution kernel, and each feature map has multiple neurons. It is characterized by local connection and weight sharing. On the one hand, the network is easy to be optimized because the number of weights is reduced. On the other hand, the complexity of the model is reduced, thus making the network model not easy to over fit. The advantage of the convolutional neural network is that the image can be directly input as the network, avoiding the complicated process of feature extraction and data reconstruction in the traditional recognition algorithm, which has a great advantage in the processing of remote sensing image. The network model can automatically extract the image features including color, texture, shape and image topology, especially in the recognition of displacement, scaling and other forms of distortion invariance application has good robustness and efficiency, has been widely used in pattern classification, target detection and target recognition.

B. BACKBONE FEATURE EXTRACTION NETWORK

Figure 1 shows the structure diagram of the main feature extraction network of YOLOv3. Compared with the YOLOv2 [26] algorithm, the difference is that the YOLOv3 algorithm greatly improves the detection accuracy of the target detection algorithm. The YOLOv3 algorithm uses DarkNet53 as the main feature extraction network. An important feature of the backbone feature extraction network is the used of Residual Network (ResNet) [27], [28]. First of all, the residual convolution in DarkNet53 carried out a convolution with a convolution kernel size of 3*3 and a step size of 2, which would compress the width and height of the inputted feature layer to obtain a feature layer. Secondly, a 1*1 convolution and a 3*3 convolution are carried out on the feature layer, and then the result is added to the feature layer to form the residual structure. Finally, the network is deepened by 1*1 convolution, 3*3 convolution and the superposition of residual edges. The characteristics of the residual network are easy to optimize, and the accuracy can be improved by adding a certain depth. However, the internal residual block adopts a jump connection, which alleviates the gradient disappearance problem caused by increasing the depth in the deep neural network. Each convolution part of darkNet53 uses the unique darkNetConv2D structure. There is L2 regularization during each convolution, and the Batch Normalization (BN) and activation function Leaky ReLU [29] after the convolution are completed. Whereas normal ReLU sets all negative values to zero, Leaky ReLU gives all negative values a non-zero slope. The mathematical formula is shown in formula (1):

\[
y_l = \begin{cases} 
  x_l & \text{if } x_l \geq 0 \\
  \frac{x_l}{a_l} & \text{if } x_l < 0 
\end{cases}
\]

C. YOLOv3 MODEL STRUCTURE

The YOLOv3 model structure uses three feature layers for border prediction, and its process is shown in Figure 2. The size in the original DarkNet53 was trained on the image classification training set, so the size of the input image is 256*256. The following figure is drawn based on the YOLOv3 800 model, so the size of the input is 800*800. The
III. IMPROVED YOLOv3 MODEL

A. BACKBONE FEATURE EXTRACTION NETWORK

First of all, as shown in Figure 3, the input image size of darkNet53, the backbone feature extraction network, was modified from $416^2416^3$ to $800^2800^3$, and the residuals were convolved once, where the size of the convolution kernel was $3^3$, to a convolution with a convolution block of $(1^3 + 3^1 + 3^3)$ and the step size was 2. This will also compress the width and height of the input feature layer, thus obtaining a feature layer. Secondly, a $1^1$ convolution and a $3^3$ convolution are modified to a convolution block of $(1^3 + 3^1 + 3^3)$ for the feature layer, and then the result is added to the feature layer to form the residual structure. Finally, the network is continuously deepened by $1^1$ convolution, $(1^3 + 3^1 + 3^3)$ convolution and the superposition of residual edges. Each convolution part of darkNet53 uses the unique darkNetConv2D structure, with L2 regularization during each convolution, and Batch Normalization (BN) and Mish activation function after the convolution is completed.

B. MISH ACTIVATION FUNCTION

Mish activation function is a self-regularization non-monotonic neural activation function proposed by He et al. [30]. Firstly, the Mish activation function has no upper bound and lower bound. No upper bound is a required feature of any activation function, because it avoids the gradient saturation that leads to the sharp decline of training speed, so it speeds up the training process. No lower bound is helpful to achieve strong regularization effect. Secondly, the Mish activation function is a non-monotonic function, which helps to maintain a small negative value, so as to stabilize the network gradient flow. Finally, Mish activation function has infinite order continuity and smoothness. It is a smooth function, which has better generalization ability.
and effective optimization ability of results, and can improve the quality of results. Smoothing activation function allows better information to go deep into the neural network, thus obtaining better accuracy and generalization. The expression for the Mish activation function is shown in formula (2):

\[
\text{Mish} = x \ast \tanh(\ln (1 + e^x))
\]

The Mish activation function curve is shown in Figure 4, and because the maximum value of this function is unlimited, saturation is avoided due to the limit of its upper limit. The activation function can be differentiated everywhere in the defined domain, which can better transfer the gradient flow, and the smooth activation function allows the shallow information to better penetrate into the neural network, so that the algorithm has a higher accuracy and recall rate.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, in order to verify the effectiveness of the improved YOLOv3 algorithm, the proposed method is used for the RSOD data annotated by Wuhan University and the TGRS-HRRSD data set annotated by the Xi'an Institute of Optical Precision Machinery, Chinese Academy of Sciences. The comparison algorithms used in the experiment include the target detection algorithm based on Efficientnet-YOLOv3, the target detection algorithm based on RetinaNet [31], [32], the target detection algorithm based on SSD single-stage, the target detection algorithm based on YOLOv4 and the target detection algorithm based on the original YOLOv3. In order to objectively analyze the performance of the target detection algorithm in this paper, the mean Average Precision (mAP) and logarithmic Average Miss Rate (LAMR) values are used as the evaluation indexes. The larger the mAP value, and the smaller the LAMR value, the better the performance of the model.

A. EXPERIMENTAL ENVIRONMENT CONFIGURATION

The experimental environment platform built in this paper: computer configuration i5-8250 CPU, 8GRAM, 64-bit Windows10 operating system, and the server is configured as a GPU queue configuration for BSCC-N22: each machine is configured with 8 pieces of NVIDIA Tesla V100-SXM 232GB explicit memory of the GPU, Each GPU card is assigned 8 CPU cores and 36GB memory, i.e., the ratio of GPU, CPU, and memory is 1 GPU card, eight-core CPU and 36GB memory. The training process is shown in Figure 5.

B. EXPERIMENTAL RESULTS ON RSOD DATA SET

1) RSOD DATA SET

As shown in Figure 6, the RSOD data set is an open data set which is applied to the small target detection in remote sensing images. The datasets include four categories: Figure 6 (a) aircraft, Figure 6 (b) oil tank, Figure 6 (c) Overbridge and Figure 6 (d) playground. The format of the datasets is PASCAL VOC and the datasets consists of 4 files, each of which is an object. There are 4,993 aircraft
in 446 images, 191 amusement parks in 189 images, 180 overpasses in 176 images and 1,586 fuel tanks in 165 images.

2) mAP VALUE WAS USED AS EVALUATION INDEX

For the small target detection algorithm of remote sensing image, the detection accuracy of the algorithm is very important. This time, the average accuracy (mAP) is selected as the evaluation index. The AP value actually refers to the area under the curve drawn by using the combination of different precision and recall points. Different precision and recall can be obtained by taking different
confidence levels. When the confidence levels are dense enough, a lot of precision and recall can be obtained, mAP is the average of AP values of all classes. As shown in Figure 7, which is the experimental results of mAP value in RSOD remote sensing image datasets. In detail, Figure 7 (a) is obtained by YOLOv3 method, Figure 7 (b) is obtained by
efficientnet-YOLOv3 method, Figure 7 (c) is obtained by retinanet method, Figure 7 (d) is obtained by SSD method, Figure 7 (e) is obtained by YOLOv4 method and Figure 7 (f) is obtained by the method of this paper.

As shown in Figure 7, the experimental results mAP of numerical indicators obtained by all the methods in this paper on the RSOD remote sensing image datasets. The abscissa in Figure 7 is the AP value of a single class. There are four classes tested in this experiment. The ordinate indicates all the classes corresponding to this target detection. The top of each subgraph is the mAP value of each algorithm. From the mAP value at the top of each graph, we can see that the
results of mAP obtained by the method of this paper are better than those of the five algorithms. Especially, based on the method in Figure 7 (a) YOLOv3, the mAP value is increased by 5.33%; Compared with Figure 7 (b) Efficientnet-YOLOv3 method, the mAP value of this method is increased by 5.17%; Compared with Figure 7 (c) RetinaNet method, the mAP value of this method is increased by 6.74%; Compared with the SSD method in Figure 7 (d), the mAP value of this method is increased by 8.56%; Compared with the method in Figure 7 (e) YOLOv4, the mAP value of this method is increased by 7.45%, which further shows that the method proposed in this paper has achieved good results in the process of small target detection in remote sensing images.

3) LAMR VALUE WAS USED AS EVALUATION INDEX
The relationship curve between miss rate (MR) and false positive per image (FPPI) is used to evaluate the small target detection algorithm of remote sensing images. In this paper, the logarithmic mean of MR is used as the evaluation standard of data when the logarithm of FPPI is in interval [0.01,100], which is called LAMR for short.

As shown in Figure 8, which are the experimental results of LAMR values, in detail, Figure 8 (a) is the result obtained by YOLOv3 method, Figure 8 (b) is from Efficientnet-YOLOv3 method, Figure 8 (c), Figure 8 (d), Figure 8 (e) and Figure 8 (f) are from RetinaNet method, SSD method, YOLOv4 method and the method of this paper respectively.

As shown in Figure 8, the experimental results of the numerical index LAMR obtained by all the methods in this paper on the RSOD remote sensing image datasets are shown. The abscissa in Figure 8 shows the MR value of a single class, and the ordinate indicates all the classes corresponding to this target detection. There are four classes in this experiment.

LAMR refers to the logarithm average undetected rate. The smaller the experimental result of each class, the better the algorithm performance. Among them, the LAMR value is reduced by 0.1100 on the basis of Figure 8 (a) YOLOv3 method. Compared with Figure 8 (b) Efficientnet-YOLOv3 method, the LAMR value of this method is reduced by 0.1525; Compared with the RetinaNet method in Figure 8 (c), the LAMR value of this method is reduced by 0.1250; Compared with the SSD method in Figure 8 (d), the LAMR value of this method is reduced by 0.2075; Compared with the method in Figure 8 (e) YOLOv4, the LAMR value of this method is reduced by 0.1325, which shows that the method proposed in this paper has achieved good results in the process of target detection.

C. EXPERIMENTAL RESULTS OF TGRS-HRRSD DATASETS
1) TGRS-HRRSD DATASETS
As shown in Figure 9, the TGRS-HRRSD data set is a data set produced by the optical image analysis and learning center of Xi’an Institute of Optics and precision machinery, Chinese Academy of Sciences for the study of high-resolution remote sensing image target detection. The format of the data set is Pascal VOC. The datasets includes: Figure 9 (a) aircraft, figure 9 (b) bridge, figure 9 (c) intersection, figure 9 (d) ship, figure 9 (E) vehicle, figure 9 (f) port, figure 9 (g) playground, figure 9 (H) storage tank, figure 9 (I) basketball...
Figure 10. Experimental results of mAP value of TGRS-HRRSD data.

Figure 10. Experimental results of mAP value of TGRS-HRRSD data.

2) mAP VALUE WAS USED AS EVALUATION INDEX
For the small target detection algorithm of remote sensing image, whether the algorithm is effective or not has a great relationship with the detection accuracy of the algorithm. This time, the average accuracy (mAP) is selected as the evaluation index. The AP value actually refers to the area under the curve drawn by using the combination of different precision and recall points. Different precision and recall can be obtained by taking different confidence levels. When the confidence levels are dense enough, a lot of precision and recall.
recall can be obtained, mAP is the average of AP values of all classes. As shown in Figure 10 is the experimental result of mAP value in TGRS-HRRSD remote sensing image datasets. Among them, Figure 10 (a) is YOLOv3 method, Figure 10 (b) is EfficientNet-YOLOv3 method, Figure 10 (c) is RetinaNet method, Figure 10 (d) is SSD method, Figure 10 (e) is YOLOv4 method and Figure 10 (f) is the method of this paper.

As shown in Figure 10, the experimental results mAP of numerical indicators obtained by all methods in this paper on TGRS-HRRSD remote sensing image datasets. The abscissa in Figure 10 is the AP value of a single class. There are 13 classes tested in this experiment. The ordinate indicates all the classes corresponding to this target detection. The top of each sub graph is the mAP value of each algorithm. From the mAP value at the top of each graph, we can see that the map numerical results obtained by this method are excellent compared with the other five algorithms. Especially, based on the method in Figure 10 (a) YOLOv3, the mAP value is increased by 1.29%; Compared with the results in Figure 10 (b) by
Efficientnet-YOLOv3 method, the mAP value of this method is increased by 2.54%; Compared with the RetinaNet method shown in Figure 10 (c), the mAP value of this method is increased by 4.83%; Compared with the SSD method shown in Figure 10 (d), the mAP value of this method is increased by 11.01%; Compared with the result of YOLOv4 method shown in Figure 10 (e), the mAP value of this method is increased by 12.52%, which further shows that the method proposed in this paper has achieved good results in small target detection of remote sensing images.

3) LAMR VALUE WAS USED AS EVALUATION INDEX
Small target detection algorithms in remote sensing images are generally evaluated by the relationship between miss rate (MR) and average false positive per image (FPPI). In this paper, the logarithmic mean value of MR when the logarithm of FPPI is in the interval [0.01, 100] is used as the evaluation standard of data, which is called LAMR for short. As shown in Figure 11, which is the experimental results of LAMR value in TGRS-HRRSD remote sensing image datasets. The result of Figure 11 (a) is obtained by YOLOv3, and the result of Figure 11 (b), Figure 11 (c), Figure 11 (d), Figure 11 (e) are obtained by Efficientnet-YOLOv3, RetinaNet, SSD method, YOLOv4 method and the method of this paper.

As shown in Figure 11, the experimental results of numerical index LAMR obtained by all the methods in this paper on TGRS-HRRSD remote sensing image datasets are shown. The abscissa in Figure 11 is the MR value of a single class, and the ordinate is all the classes corresponding to this target detection. There are 13 classes in this experiment. LAMR refers to the logarithm average miss detection rate. The smaller the experimental results of each class, the better the algorithm performance. Among them, the LAMR value is reduced by 0.0338 based on the YOLOv3 method of Figure 11 (a), and the LAMR value of this method is reduced by 0.0731 compared with the Efficientnet-YOLOv3 method shown in Figure 11 (b); Compared with the RetinaNet method shown in Figure 11 (c), the LAMR value of this method is reduced by 0.1023; Compared with the SSD method in Figure 11 (d), the LAMR value of this method is reduced by 0.2115; Compared with the YOLOv4 method shown in Figure 11 (e), the LAMR value of this method is reduced by 0.2032, which again shows that the method proposed in this paper has achieved good results in the process of target detection.

V. CONCLUSION
This paper proposes an improved method of object detection in remote sensing image based on YOLOv3. Compared with the original method, the improved method improves the mAP value of small object detection in remote sensing image without adding any model parameters, reduces the LAMR value of small object detection in remote sensing image, and improves the detection accuracy. It also reduces the miss detection rate and enhances the stability of the algorithm. The method proposed in this paper has achieved good results in both RSOD and TGRS-HRRSD remote sensing image datasets, but there are two shortcomings in the small target detection of remote sensing image. Firstly, the method in this paper is modified on the basis of YOLOv3 method, and the feature extraction network of DarkNet53 is strengthened, which makes the complexity of the backbone extraction network model increase; Secondly, the generalization of the proposed method is very weak in the face of small target detection in remote sensing images. We hope that in the further work, we will consider to optimize the model to reduce the model parameters, remove the redundancy of the model, adjust the training parameters of the model, and improve the performance of the small target detection method facing the remote sensing image.

REFERENCES
[1] Y. Q. Wang, L. Ma, and Y. Tian, “State-of-the-art of ship detection and recognition in optical remotely sensed imagery,” Acta Automatica Sinica, vol. 39, no. 9, pp. 1029–1039, 2011.
[2] L. Yang, F. Zhengye, and Z. Fengbin, “Review on high resolution remote sensing image classification and recognition,” J. Geo-Inf. Sci., vol. 17, no. 9, pp. 1080–1091, 2015.
[3] A. Farooq, J. Hu, and X. Jia, “Efficient object proposals extraction for target detection in VHR remote sensing images,” in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Jul. 2017, pp. 3337–3340, doi: 10.1109/IGARSS.2017.8127712.
[4] X. Wang, Y. Ban, H. Guo, and L. Hong, “Deep learning model for target detection in remote sensing images fusing multilevel features,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., Jul. 2019, pp. 250–253, doi: 10.1109/IGARSS.2019.8898759.
[5] L. Zong-ling, W. Yu-yuan, R. Yang, C. Bo-wen, H. Liang, C. Shuai, L. Zhen, and Y. Jia-feng, “Remote sensing ship target detection and recognition system based on machine learning,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., Jul. 2019, pp. 1272–1275, doi: 10.1109/IGARSS.2019.8898599.
[6] G. You and Y. Zhu, “Target detection method of remote sensing image based on deep learning,” in Proc. Cross Strait Radio Sci. Wireless Technol. Conf. (CSRWTSC), Dec. 2020, pp. 1–2, doi: 10.1109/CSRWTSC50769.2020.9372469.
[7] G.-X. Wang, X.-T. Huang, and Z.-M. Zhou, “UWB SAR change detection of target in foliage based on local statistic distribution change analysis,” J. Electron. Inf. Technol., vol. 36, no. 1, pp. 49–54, Feb 2011.
[8] W. Wu, Research on Knowledge-Based Target Recognition and Tracking Techniques, Harbin, China: Harbin Institute of Technology, 2017.
[9] J. X. Cao and A. G. Song, “Research on texture image segmentation method based on Markov random field,” J. Instrum., vol. 36, no. 4, pp. 776–786, 2015.
[10] U. Muhammad, W. Wang, and S. P. Chatha, “Pre-trained VGGNet architecture for remote-sensing image scene classification,” in Proc. 24th Int. Conf. Pattern Recognit. (ICPR), Aug. 2018, pp. 1622–1627, doi: 10.1109/ICPR.2018.8545591.
[11] H. Jun, L. Shuai, S. Jinming, L. Yue, W. Jingwei, and J. Peng, “Facial expression recognition based on VGGNet convolutional neural network,” in Proc. Chin. Autom. Cong. (CAC), 2018, pp. 4146–4151, doi: 10.1109/CAC.2018.8623238.
[12] S. Targ, D. Almeida, and K. Lyman, “Resnet in resnet: Generalizing residual architectures,” 2016, arXiv:1603.08029.[Online], Available: http://arxiv.org/abs/1603.08029.
[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. NIPS, vol. 1, 2012, pp. 1097–1105.
[14] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size,” 2016, arXiv:1602.07360.[Online], Available: http://arxiv.org/abs/1602.07360.
[15] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2117–2125.
KAIJUN WU received the Ph.D. degree from Lanzhou Jiaotong University, China, in 2017. He is currently a Professor with the School of Electronics and Information Engineering, Lanzhou Jiaotong University. His research interests include video target detection and video anomaly detection.

CHENSHUAI BAI is currently pursuing the master’s degree with the School of Electronics and Information Engineering, Lanzhou Jiaotong University. His research interests include video target detection and video anomaly detection.

DICONG WANG is currently pursuing the Ph.D. degree with the School of Electronic Information Engineering, Lanzhou Jiaotong University, and the Department of Intelligence and Computing, Tianjin University. His research interests include video target detection and video anomaly detection.

ZHENGNAN LIU is currently pursuing the master’s degree with the School of Electronics and Information Engineering, Lanzhou Jiaotong University. Her research interest includes the intelligent optimization of differential evolution algorithm.

TAO HUANG is currently pursuing the master’s degree with the School of Electronics and Information Engineering, Lanzhou Jiaotong University. His research interest includes video anomaly detection.