

Research Article

Remote Sensing Image Recognition Based on LOG-T-SSA-LSSVM and AE-ELM Network

Chang-Jian Sun ¹ and Fang Gao ^{2,3}

¹College of Electronic Science and Engineering, Jilin University, Changchun 130012, China

²College of Computer Science and Technology, Jilin University, Changchun 130012, China

³Chang Guang Satellite Technology Co., Ltd., Changchun 130000, China

Correspondence should be addressed to Fang Gao; gaofang@163.com

Received 12 December 2021; Accepted 30 December 2021; Published 25 January 2022

Academic Editor: Ahmed Mostafa Khalil

Copyright © 2022 Chang-Jian Sun and Fang Gao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the influence of different working conditions on recognition accuracy in remote sensing image recognition, this paper adopts hierarchical strategy to construct a network. Firstly, in order to establish the classification relationship between different samples, labeled samples are used for classification. A Logistic-T-distribution-Sparrow Search Algorithm-Least Squares Support Vector Machines (LOG-T-SSA-LSSVM) classification network is proposed. LOG-T-SSA algorithm is used to optimize parameters in LSSVM to establish a better network to achieve accurate classification between sample sets and then identify according to different categories. Through UCI dataset test, the accuracy of LOG-T-SSA-LSSVM network classification is significantly improved compared with that of contrast network. The autoencoder is integrated with Extreme Learning Machine, and the autoencoder is used to realize data compression. The advantages of Extreme Learning Machine (ELM) network, such as less training parameters, fast learning speed, and strong generalization ability, are fully utilized to realize efficient and supervised recognition. Experiments verify that the autoencoder-extreme learning machine (AE-ELM) network has a good recognition effect when the sigmoid activation function is selected and the number of hidden layer neurons are 2000. Finally, after image recognition under different working conditions, it is proved that the recognition accuracy of AE-ELM based on LOG-T-SSA-LSSVM classification is significantly improved compared with traditional ELM network and Particle Swarm Optimization-Extreme Learning Machine (PSO-ELM) network.

1. Introduction

As the basis of computer vision application, object detection is one of the most widely concerned problems in real life. Generally speaking, the universal target recognition mainly has two subtasks: one is to judge the class probability of a specific target, and the other is to give the specific location of the target. Target recognition algorithm plays a very important role in daily life and has been successfully used in facial recognition [1–3], pedestrian detection [4–7], video analysis [8–10], and beacon positioning and recognition [11, 12]. With the continuous development of machine learning and its continuous application in the field of target

detection, the accuracy of target detection in common scenarios has been greatly improved. However, it is still a hot research issue for target detection in complex environments with a large number of targets and variable scales.

Existing target detection methods can be divided into two categories: methods based on manual feature construction and methods based on deep learning. The focus of the manual method is to extract the hand-made features to represent the temporal and spatial features of the video sequence [13, 14]. For instance, literature [15] proposed a new descriptor of spatial and temporal features based on optical flow information, called histograms of optical flow orientation and magnitude and entropy. Literature [16]

shows through experiments that the histogram-oriented gradient (HOG) descriptor of grid is significantly better than the existing feature set. Literature [17] proposed a scheme based on support vector machines. The flexible genre model (FGM) is proposed [18], which aims to characterize the data population at the point level and population level to detect various types of population anomalies. Although the method based on manual feature construction has achieved some achievements, the traditional target detection algorithm based on manual feature construction is not suitable for solving existing problems because of its complicated process and large amount of calculation [19]. Recently, with the continuous success of deep learning technology in various fields, object detection based on deep learning has become a research hotspot. AlexNet [20], proposed in 2012, is the first deep neural network that has made a breakthrough in large-scale image recognition. After this, deep neural networks began to be widely used in the field of computer vision. For example, VGGNET was proposed in 2014 [21]. ResNet was proposed in 2016 [22]. ResNeXt was proposed in 2017 [23]. SENet [24] was proposed in 2018. ExtremeNet was proposed in 2019 [25].

Target detection is also a hot topic in remote sensing field. However, it should be noted that methods in the field of computer vision cannot be directly applied in the field of remote sensing [26, 27], because commonly used remote sensing images and natural images are quite different. For example, remote sensing images often capture the top features of the target, while natural images capture the contour features of the target. However, as deep learning-based methods have made great achievements in the field of target detection, related extended methods have also been applied to remote sensing images. Deep learning-based target detection methods can generally be divided into two categories: region proposal-based methods, namely, two-stage detection, and regression-based methods, represented as one-stage detection.

Two-stage detection divides the detection task into two stages: (1) proposal generation and (2) proposal prediction. The first phase focuses on generating a series of candidate region proposals that might contain objects. The objective of the second phase is to classify the candidate area proposals from the first phase into object classes or backgrounds and further fine-tune the coordinates of the bounding boxes. In the two-stage algorithm, the representative method is R-CNN [28] as well as the variant method based on R-CNN, such as Faster R-CNN [29] and rotation-invariant CNN [30].

Although R-CNN and its variant methods have been successfully applied in the field of remote sensing image detection, it is undeniable that the training process is very clumsy and slow. Recently, in order to achieve real-time target detection, some researchers have begun to study the detection method based on regression, also known as one-stage detection. For example, Tang proposed Oriented_SSD (Single Shot MultiBox Detector), which improved the efficiency and accuracy of vehicle detection [31, 32]. Liu proposed SSD and its validity was verified through multiple datasets [33–35].

In this paper, a hierarchical strategy is used to construct a network for remote sensing image recognition. Firstly, in order to establish the classification relationship between different samples, labeled samples are used for classification. A LOG-T-SSA-LSSVM classification network is proposed. LOG-T-SSA algorithm is used to optimize parameters in LSSVM to establish a better network to achieve accurate classification between sample sets and then identify according to different categories. The autoencoder is integrated with Extreme Learning Machine, and the autoencoder is used to realize data compression. The advantages of ELM network, such as less training parameters, fast learning speed, and strong generalization ability, are fully utilized to realize efficient and supervised recognition.

The rest of this article is arranged as follows. The second part introduces LOG-T-SSA-LSSVM classifier. The third part introduces AE-ELM network recognizer. The fourth part constructs the recognizer combining LOG-T-SSA-LSSVM and AE-ELM. The fifth part carries on the relevant experiment verification. The last part is the conclusion and future development.

2. LOG-T-SSA-LSSVM Classifier and AE-ELM Recognizer

2.1. The LOG-T-SSA Algorithm. Sparrow Search Algorithm (SSA) is a new swarm intelligence optimization algorithm [36], which is superior to Grey Wolf Optimizer (GWO), Particle Swarm Optimization (PSO), Gravity Search Algorithm (GSA), and other algorithms. Entrants, scouts, and discoverers are mainly responsible for guiding the overall optimization direction of the population. Compared with entrants, their search scope is larger, generally accounting for 10% to 20% of the total population. The sparrows with better performance in each iteration take the role, and their specific position update formula is as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t \cdot e^{-(\alpha/iter_max)} R < ST \\ x_{i,j}^t + Q \cdot LR \geq ST \end{cases}, \quad (1)$$

where sparrow is in row i and column j in generation $t + 1$ of $x_{i,j}^{t+1}$. t represents the current iteration number. α represents random number with a range of $[0, 1]$. Warning value R is in the range of $[0, 1]$. Safe value ST is in the range of $[0.5, 1]$. Q is a random number subject to standard integer normal distribution. L is a matrix of $1 \times D$, where D is the latitude of the problem, and all elements are 1.

The entrants are nondiscoverers, and the proportion of them always remains the same. The updating formula is related to discoverers, and the formula is as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{best}^t + \beta |x_{i,j}^t - x_{best}^{t+1}| f_i > f_{g_best} \\ x_{i,j}^t + k \left(|x_{i,j}^t - x_{worst}^t| / f_i - f_{g_worst} + \xi \right) f_i = f_{g_best} \end{cases}, \quad (2)$$

where x_{best}^t represents the individual with the worst fitness value in the iteration of t generation and x_p^{t+1} represents the individual with the best fitness value in the iteration of $t + 1$

generation. A is the matrix of $1 \times D$. The elements are either 1 or -1 . $A^+ = A^T(AA^T)^{-1}$.

The scout is jointly held by the discoverer and the entrant, indicating that the dangerous individuals are aware of in the population, accounting for 10%~20% of the total population. The position update formula is as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{\text{best}}^t + \beta |x_{i,j}^t - x_{\text{best}}^{t+1}| f_i > f_{g\text{-best}} \\ x_{i,j}^t + k(|x_{i,j}^t - x_{\text{worst}}^t| / |f_i - f_{g\text{-worst}} + \xi|) f_i = f_{g\text{-best}} \end{cases}, \quad (3)$$

where x_{best}^t represents the population individual with the best fitness in generation t . β follows the standard normal distribution and controls the update step size. k belongs to $[-1, 1]$ and is a random number. f_i represents the current individual fitness value. $f_{g\text{-best}}$ represents the current global optimal fitness. $f_{g\text{-worst}}$ represents the current worst fitness value, and ξ is a constant term to prevent the denominator from being 0.

In order to improve the initial population quality of the algorithm, logistic mapping is introduced [37]. To a certain extent, logistic mapping is a time-discrete demographic model, which can fully demonstrate chaos dynamics. The expression is as follows:

$$x(t) = \mu \cdot x(t) \cdot [1 - x(t)], \quad (4)$$

where t is the number of iterative steps; $x(t) \in [0, 1]$. μ is the adjusting parameter. To ensure that the mapping range is between 0 and 1, $\mu \in [0, 4]$. $x(t)$ is the proportion of the population to the maximum possible population size at time t (i.e., the ratio of the existing population to the maximum possible population). When parameter μ is changed, the equation will show different dynamic limit behaviour. When $0 < \mu < 1$, the limit behaviour of the population tends to a fixed value of 0. When μ is between 1 and 3, the population value will approach $(\mu - 1)/\mu$. Different μ values can adjust the convergence rate. When μ is between 1 and 4, it will show periodic fluctuations. At the same time, the adaptive T-distribution is introduced to improve the update step size, and the T-distribution update formula is as follows:

$$x_i^{t+1} = x_i^t + x_i^t \cdot t(\text{iter}), \quad (5)$$

where x_i^{t+1} is the position of sparrow after mutation. x_i^t is the position of the i th individual sparrow of t generation. $t(\text{iter})$ is the T-distribution of the degree of freedom taking the number of iterations of the algorithm as the parameter. This formula makes full use of the current population information and takes the number of iterations t as the degree of freedom parameter. Cauchy-like variation with small t in the early stage has strong global search ability, and gauss variation with large t in the late stage has strong local search ability. Thus, the search ability of the algorithm is improved.

2.2. LSSVM Network. Least Squares Support Vector Machine (LSSVM) is a method to transform support vector machine into linear problem [38]. Through the sum of

squares of minimum error, the fitting object is close to the target. LSSVM changed the inequality constraint in SVM to equality constraint, and the LSSVM structural risk minimization formula is as follows:

$$\min J(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^N e_k^2, \quad (6)$$

where $\min J(w, e)$ is the objective function, w is the weight coefficient, γ is the penalty factor, and e is the error; the equation condition is as follows:

$$y_i \mathbf{w}^T \varphi(x_i) b e_i = 0, \quad (7)$$

where y_i is the corresponding output variable. $\varphi(x_i)$ is the nonlinear transformation function of input data. \mathbf{w} is the weight vector. b is the bias term. Lagrange function is constructed by the following formula:

$$\begin{aligned} L(\mathbf{w}, b, e, \alpha) &= J(\mathbf{w}, e) - \sum_i \alpha_i [(\mathbf{w}^T \varphi(x_i) + b) + e_i - y_i] \\ &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\gamma}{2} \sum_i e_i^2 - \sum_i \alpha_i [(\mathbf{w}^T \varphi(x_i) + b) + e_i - y_i], \end{aligned} \quad (8)$$

where $L(\mathbf{w}, b, e, \alpha)$ is the Lagrange expression and α is the Lagrange multiplier. According to the Karush-Kuhn-Tucker (KKT) optimization conditions, the following conditions are satisfied:

$$\begin{cases} \frac{\partial L}{\partial \mathbf{w}} = 0, \\ \frac{\partial L}{\partial b} = 0, \\ \frac{\partial L}{\partial e} = 0, \\ \frac{\partial L}{\partial \alpha} = 0. \end{cases} \quad (9)$$

2.3. Optimizing the LSSVM Network. LOG-T-SSA algorithm was used to optimize LSSVM parameters. The specific process is shown in Figure 1, and the steps are as follows:

Step 1: initialize the parameters of the sparrow search algorithm, including the initial number of sparrow population n , the proportions of finder and follower in the population, the warning value R , the safety value ST , the random value Q , and other parameters

Step 2: use logistic chaos mapping function to generate chaotic sequence, that is, individual member of sparrow population in the initial solution space position

Step 3: establish LSSVM network, and take LSSVM network classification error rate as fitness function

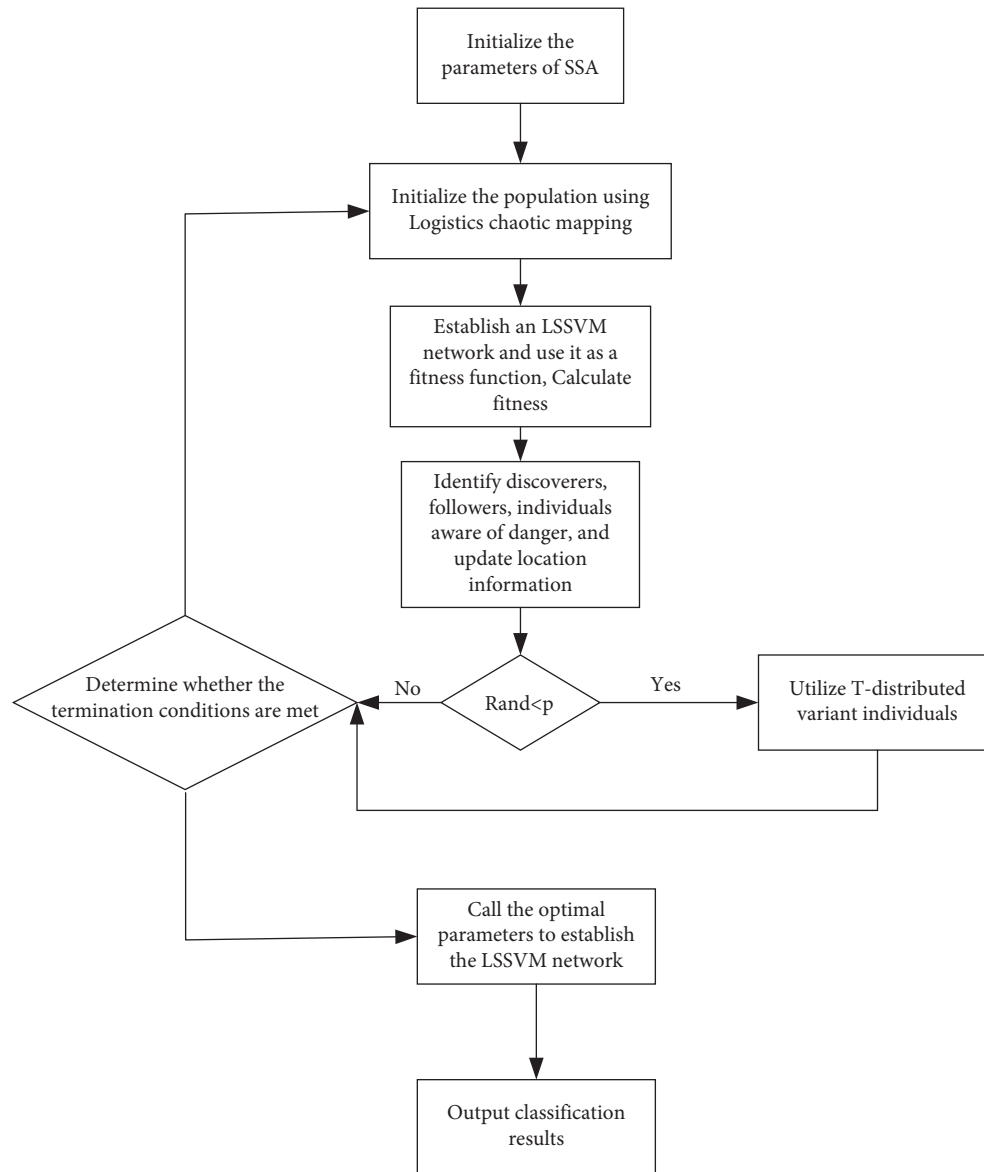


FIGURE 1: Flowchart of optimizing LSSVM by LOG-T-SSA algorithm.

Step 4: calculate the fitness value of each sparrow to determine the individual position of the optimal solution and the worst solution

Step 5: identify the finder in the population and update the location of the finder

Step 6: identify the follower and update the position of the follower

Step 7: determine the number of dangerous individuals in the population and calculate the update position

Step 8: $R \text{ and } < p$, T-distribution variation for individuals

Step 9: calculate the population fitness before and after variation and determine the optimal solution of the population

Step 10: if the maximum number of iterations is reached or the threshold is met, output the optimal

kernel parameter and penalty factor; if it is met, go back to Step 4

Step 11: LSSVM network was established by using optimal kernel parameters and penalty factors, classification was carried out, and classification results were output

2.4. Optimizing the LSSVM Network. In order to test the classification ability of LOG-T-SSA-LSSVM network, glass dataset in UCI was selected for verification. Three comparative classification networks were selected, namely, SSA-LSSVM, Tent-SSA-LSSVM, and EOBL-SSA-LSSVM. The experimental simulation environment was Windows 10, CPU: 2.80 GHz, 16 GB memory, operating environment: Matlab 2019b. The classification network parameters are shown in Table 1.

TABLE 1: Comparison algorithm parameter table.

Algorithm	Parameter setting
SSA-LSSVM	ST=0.6; PD=0.4; initial kernel parameter 20; initial penalty factor 100
Tent-SSA-LSSVM	Tent Beta=0.4; ST=0.6; PD=0.4; initial kernel parameter 20; initial penalty factor 100
EOBL-SSA-LSSVM	Learning rate 0.5; ST=0.6; PD=0.4; initial kernel parameter 20; initial penalty factor 100
LOG-T-SSA-LSSVM	Log coefficient 0.4; T-distribution degree of freedom is the number of iterations; ST=0.6; PD=0.4; initial kernel parameter 20; initial penalty factor 100

In Figure 2, accuracy is used as the evaluation standard. Under the same number of iterations, it can be seen intuitively that LOG-T-SSA algorithm has a faster convergence speed in the initial stage, indicating that logistic chaotic mapping enables the population to have a good initial distribution and the population diversity increases significantly. In the process of population renewal, compared with SSA, Tent-SSA, and EOBL-SSA, SSA had stronger optimization performance. SSA fell into local optimum within 10 generations, EOBL-SSA fell into local optimum within 40 generations, and Tent-SSA fell into optimum within 70 generations. However, LOG-T-SSA does not stop optimization until less than 90 generations, indicating that it has better search performance.

Figure 3 shows that, for glass datasets, the classification accuracy of SSA-LSSVM is 69.7%, that of Tent-SSA-LSSVM is 79.0%, and that of EOBL-SSA-LSSVM is 74.4%. By contrast, the classification accuracy of LOG-T-SSA network is as high as 93%. It shows that the improved classification network in this paper has good classification ability on the multicategory dataset.

$$\begin{cases} \dot{x} = v \cos \gamma \cos \psi, \\ \dot{y} = v \cos \gamma \sin \psi, \\ \dot{h} = v \sin \gamma, \\ \dot{v} = g(n_x - \sin \gamma), \\ \dot{\gamma} = \frac{g}{v} (n_z \cos \mu - \cos \gamma), \\ \dot{\chi} = \frac{g}{v \cos \gamma} n_z \sin \mu. \end{cases} \quad (10)$$

3. AE-ELM Network

3.1. Autoencoder. Autoencoder is an artificial neural network for unsupervised learning, which consists of three layers: input layer, output layer, and hidden layer [39]. At present, there are two main applications of autoencoders, one is data denoising, and the other is for visual and dimension reduction. Since high-dimensional data are often located in a low-dimensional manifold or nearby, the encoder nonlinearly maps the input data set to the hidden layer through the encoding process, and the data set is

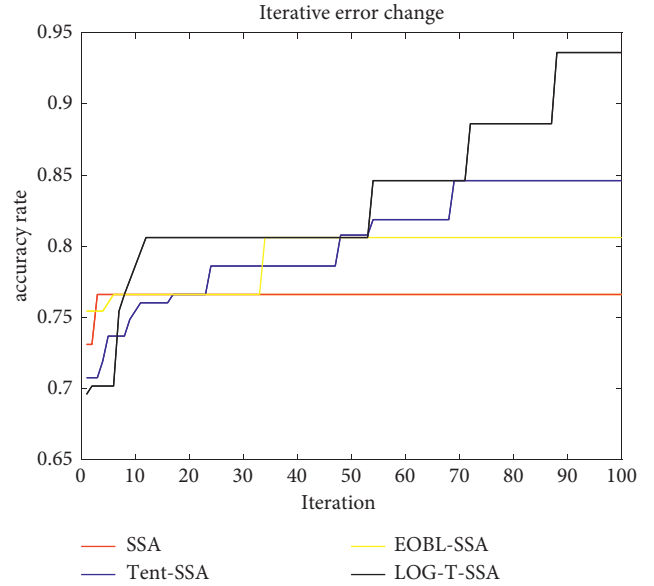


FIGURE 2: Iteration curve of accuracy.

compressed and encoded. That is, the characteristic information of the original data in another dimension space can be obtained, which is enough to reproduce the information of the input layer, so as to achieve the purpose of reducing the data dimension and improving the computing efficiency. The network structure of autoencoder is shown in Figure 4.

Autoencoder (AE) consists of encoding and decoding. The encoding process is to map input x to the hidden layer through a nonlinear activation function. The decoding process is to transform the hidden layer data h into the output value Y to reconstruct the input. Encoding process formula is

$$h = f(x) = s_f(W_1 x + b). \quad (11)$$

The decoding formula is

$$Y = g(h) = s_g(W_2 x + \tilde{b}). \quad (12)$$

The loss function is

$$J_{AE}(\theta) = \frac{1}{N} \sum_{i=1}^N L(X_i, Y_i), \quad (13)$$

where s is the activation function. W_1 and b are the encoder weight and bias, respectively. W_2 and \tilde{b} are the decoder weight and bias, respectively. $L(x_i, y_i) = 1/2 \|x_i - y_i\|^2$ is the error function.

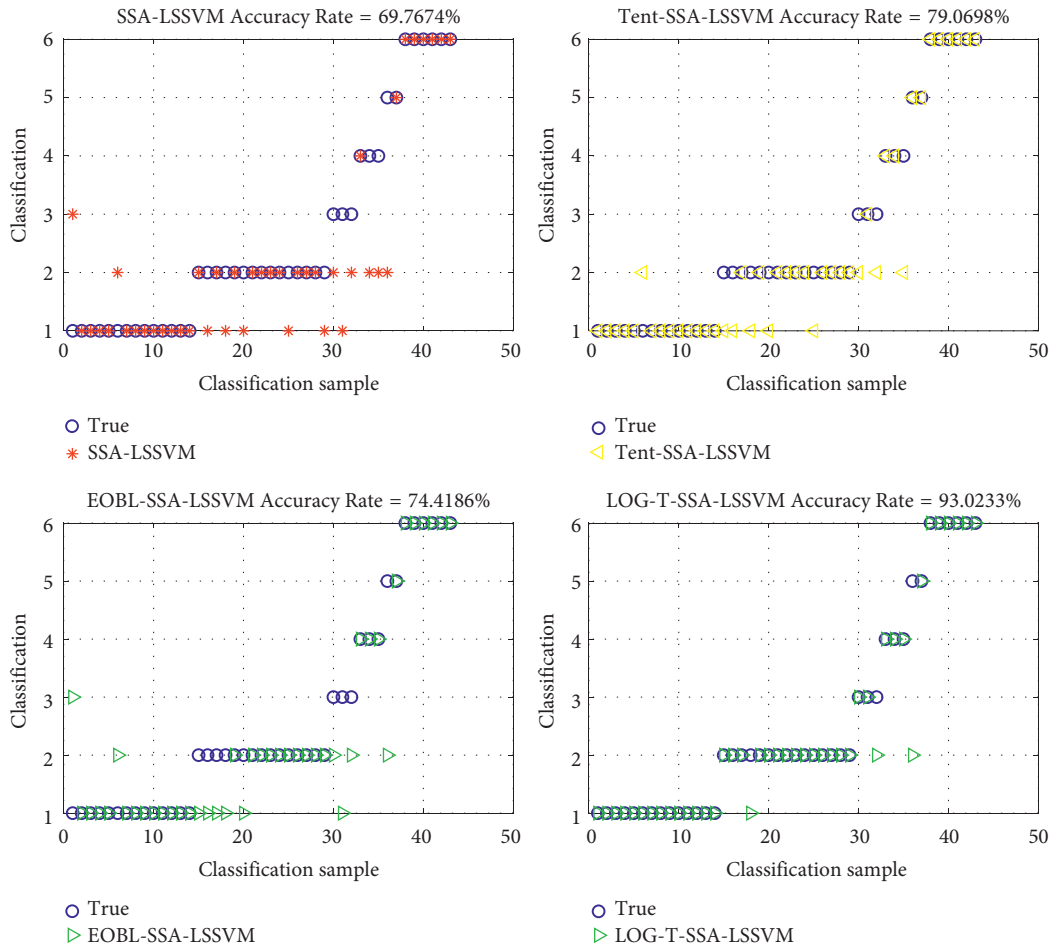


FIGURE 3: Classification results of glass datasets.

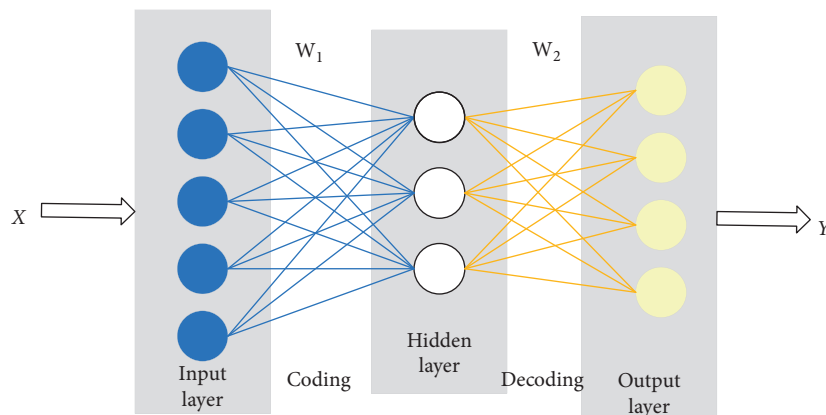


FIGURE 4: Network structure diagram of autoencoder.

3.2. *ELM Network.* Extreme Learning Machine (ELM) is a single hidden layer feedforward neural network [40]. The network structure model of ELM is shown in Figure 5. The training set has N samples $\{x, y\} = \{(x_i, y_i) | x_i \in R^{dn}, y_i \in R, i = 1, 2, \dots, N\}$. Then its model is

$$f(x_i) = \beta g(Wx_i + b), \quad (14)$$

where W is the weight vector from the input layer to the hidden layer. b is the bias vector. β is the output weight from the hidden layer to the output layer. ELM matrix expression is

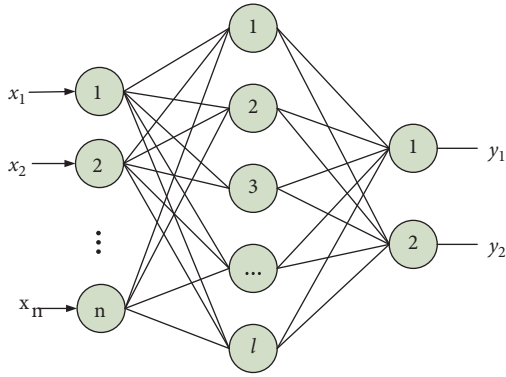


FIGURE 5: ELM network structure diagram.

$$\mathbf{H}\beta = \mathbf{Y}, \quad (15)$$

where \mathbf{H} is the output matrix of the hidden layer and \mathbf{Y} is the real matrix of the sample target output. The training process is to solve the least squares solution β ; namely,

$$\min(\mathbf{H}\beta - \mathbf{Y}). \quad (16)$$

The output weight matrix can be solved by Moore-Penrose generalized inverse formula to obtain

$$\beta = \mathbf{H}^+ \mathbf{Y}. \quad (17)$$

$$\mathbf{H}^+ = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T [41].$$

3.3. AE-ELM Network Settings. After AE data dimensionality reduction, ELM network is used for fast recognition. For ELM network, different activation functions have different recognition effects. In order to select the most suitable activation function, Root Mean Square Error was used as an evaluation index, and the optimal activation function was selected through iterative calculation of different neuron numbers.

Hardlim function, Radbas function, Sigmoid function, Sine function, and Tribas function are selected, respectively, and Inria Aerial Image Labeling dataset is used for training. Root Mean Square Error (RMSE) results are shown in Figure 6. It can be seen intuitively that Sigmoid function has the best activation effect. The Sine function has the worst effect. There is no significant difference between Hardlim, Radbas, and Tribas functions. Therefore, this article selects the Sigmoid function as the activation function.

The number of hidden layer neurons in ELM network has significant influence on the recognition result. In order to select the optimal structure, on the basis of determining the activation function, different numbers of neurons are set for recognition, and the recognition results are shown in Figure 7. It can be seen that, with the increase of the number of hidden layer neurons, the recognition effect of the network is also significantly improved. However, when the number of neurons reaches 2000, the Root Mean Square Error (RMSE) cannot be significantly reduced by increasing the number of neurons, so the number of network neurons in this paper is 2000.

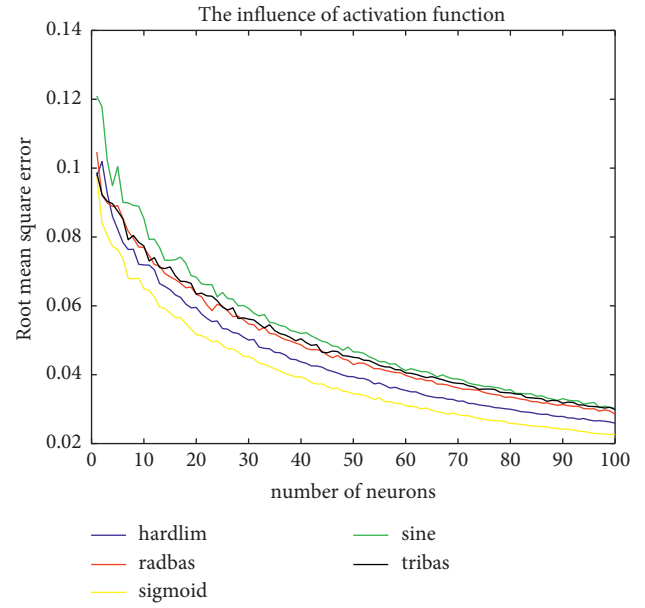


FIGURE 6: Activation function selection.

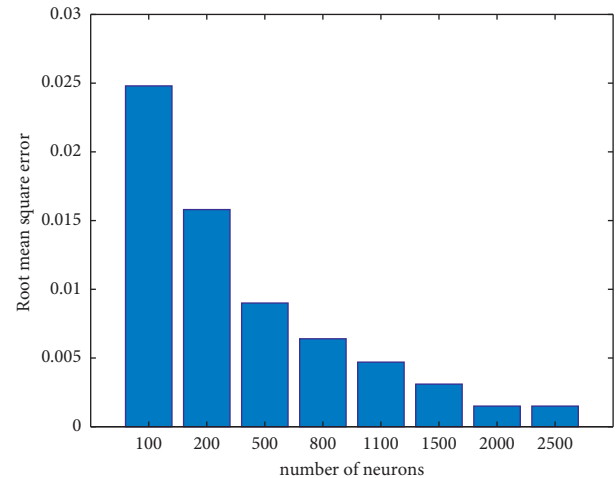


FIGURE 7: Selection of number of neurons.

4. Constructing a Recognizer Combined with LOG-T-SSA-LSSVM and AE-ELM

In the process of image shooting, angle is not uniform; it is a major problem. In this paper, the LOG-T-SSA-LSSVM network is used to fit the relationship between images and labels, and a strong classification network is established to extract effective information. On this basis, AE-ELM network is used to compress and extract data, and supervised learning method is used to establish a high accuracy recognizer. The process is shown in Figure 8, and the steps are as follows:

- Step 1: train the LOG-T-SSA-LSSVM network with labeled data
- Step 2: input images into the LOG-T-SSA-LSSVM network after training to extract effective information

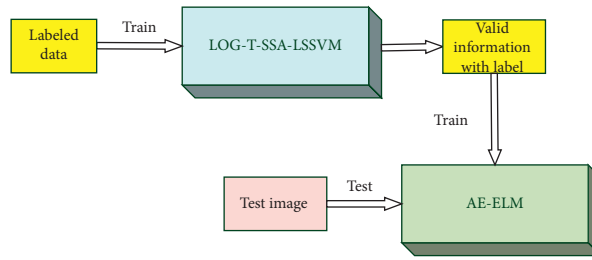


FIGURE 8: Flowchart of LOG-T-SSA-LSSVM combined with AE-ELM.

TABLE 2: Test data information.

Technical index	Parameters
Ground pixel resolution of subsatellite point	0.75 m (panchromatic)/3 m (multispectral)
Parrot sequoia	(a) Panchromatic P: 450 nm to 800 nm (b) Blue B1: 450 nm to 510 nm (c) Green B2: 510 nm to 580 nm (d) Red B3: 630 nm to 690 nm (e) Simulation near-infrared B4: 770 nm to 895 nm
Digitalizing bit	12 bits
Standard scene size (substellar point)	23 km × 23 km (6-point camera mode, default) 46 km × 46 km (3-point camera mode)
Orbit altitude	481.56 km
Positioning accuracy without control (CE90)	20 m

TABLE 3: Comparison network parameters.

Algorithm	Parameter setting
ELM	Number of neurons: 2000
PSO-ELM	$c1 = 2$; $c2 = 2$; maximum number of iterations: 1000; number of neurons: 2000
AE-ELM	Number of AE layers: 15; number of ELM neurons: 2000

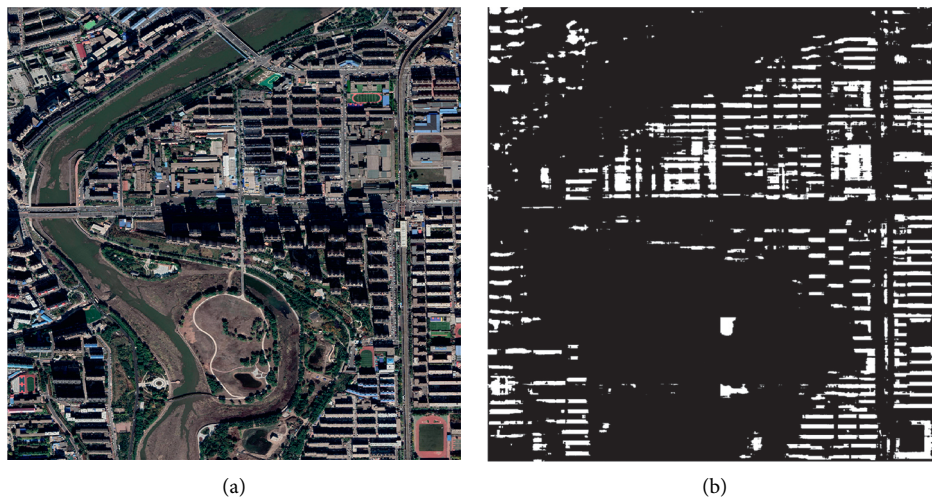


FIGURE 9: Continued.

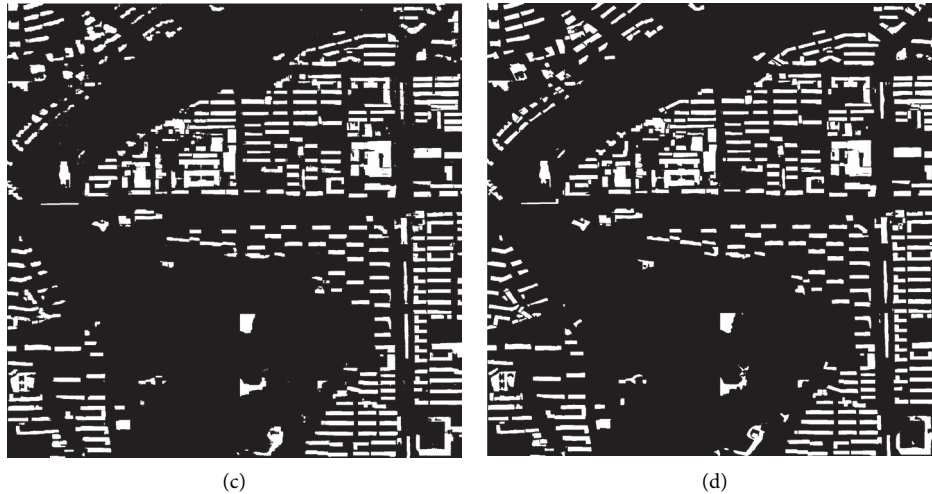


FIGURE 9: Recognition results of Case 1. (a) The original image. (b) ELM recognition result. (c) PSO-ELM recognition result. (d) AE-ELM recognition result.

TABLE 4: RMSE identification and accuracy.

	Case 1 (RMSE/accuracy)	Case 2 (RMSE/accuracy)
ELM	0.0603/92.93%	0.1178/69.58%
PSO-ELM	0.0414/97.44%	0.0782/75.28%
AE-ELM	0.0018/99.11%	0.0025/93.32%

Step 3: input effective information and labels as AE-ELM network to train AE-ELM network

Step 4: use the test image to verify the recognition accuracy

5. Experimental Verification

In order to verify the recognition accuracy of the method proposed in this paper, image sets under three different working conditions were selected. The recognition sample data were all extracted from the WIDE-amplitude star L2E class data product of JL-1. After orthofusing-drop processing, they were RGB true color 8-bit image products with a resolution of 0.75 m and taken on October 18, 2020. The shooting location is the main city of Changchun.

The JL101K satellite can obtain high-resolution panchromatic images and multispectral images. The pendulum angle of the imaging can be customized according to user requirements and is widely used in economic survey, disaster prevention and mitigation, social development research, and other fields. The main indicators are shown in Table 2.

The traditional ELM network and PSO-ELM network were selected as the comparison recognizers, and the recognition was carried out under three working conditions. The experimental simulation environment was Windows 10,

and CPU was 2.80 GHz, with 16 GB memory, and the operating environment was Matlab 2019b. Network parameters are shown in Table 3.

As can be seen from Figure 9, the ELM network recognition is fuzzy. Small individuals cannot be recognized, and they are vulnerable to the influence of edge signals. Compared with ELM network, PSO-ELM network can better identify objects through algorithm optimization, but its recognition accuracy also decreases significantly in complex geographical situations. After the image is processed by LOG-T-SSA-LSSVM classification, the output to AE-ELM network can achieve a better recognition effect. Table 4 shows that its recognition accuracy is as high as 99.11%, which has been significantly improved.

As can be seen from Figure 10, the recognition accuracy of ELM and PSO-ELM has decreased significantly, and the recognition accuracy is below 80%. The poor processing ability of images from different angles indicates that the network universality is poor without LOG-T-SSA-LSSVM first classification. Although the recognition accuracy of AE-ELM network decreased slightly, it can be seen from Table 4 that the accuracy still remains above 90%. Therefore, LOG-T-SSA-LSSVM was first used to classify the sampled images, and then images of different categories were identified, and the accuracy was significantly improved.

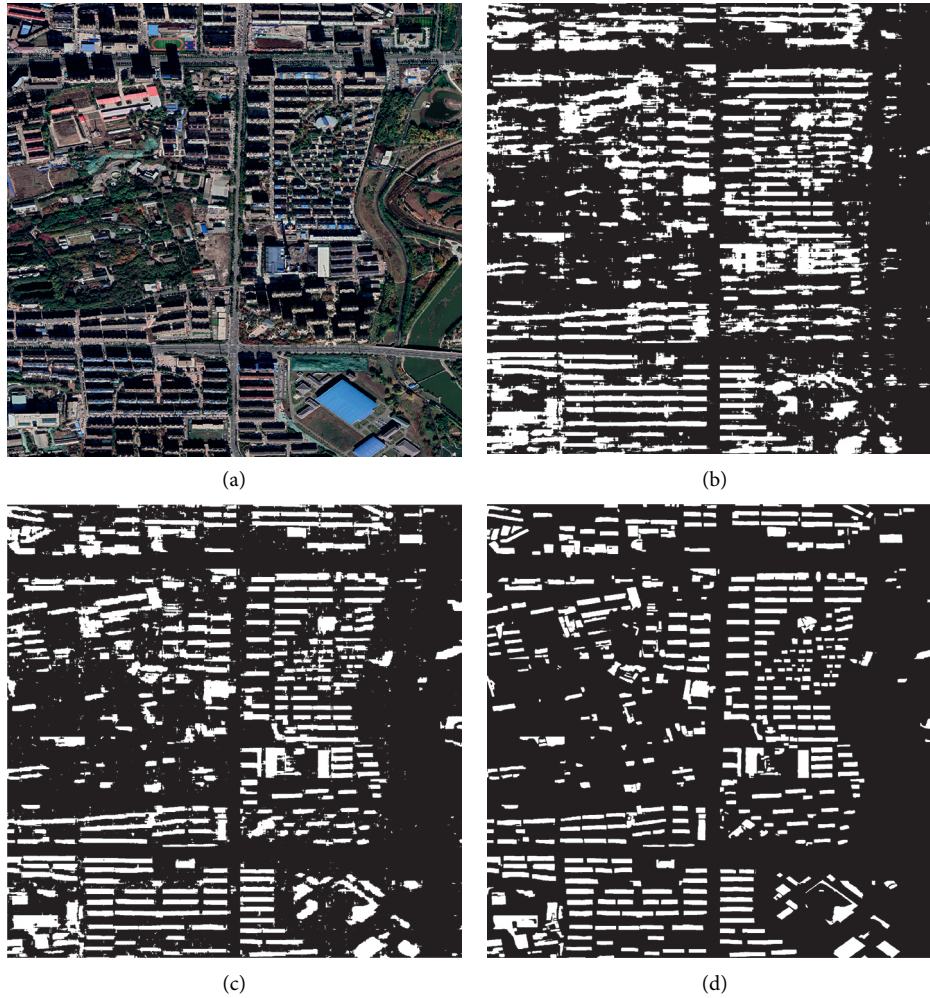


FIGURE 10: Case 2 recognition results. (a) The original image. (b) ELM recognition result. (c) PSO-ELM recognition result. (d) AE-ELM recognition result.

6. Conclusion

In this paper, a hierarchical strategy is used to construct a network for remote sensing image recognition. Firstly, in order to establish the classification relationship between different samples, labeled samples are used for classification. A LOG-T-SSA-LSSVM classification network is proposed. LOG-T-SSA algorithm is used to optimize parameters in LSSVM to establish a better network to achieve accurate classification between sample sets and then identify according to different categories. The autoencoder is integrated with Extreme Learning Machine, and the autoencoder is used to realize data compression. The advantages of ELM network, such as fewer training parameters, fast learning speed, and strong generalization ability, are fully utilized to realize efficient and supervised identification. The following conclusions are drawn after the test verification:

Through UCI dataset test, LOG-T-SSA-LSSVM network classification has significantly improved classification accuracy compared with SSA-LSSVM, Tent-SSA-LSSVM, and EOBL-SSA-LSSVM.

After image recognition under different working conditions, the recognition accuracy of AE-ELM based on LOG-T-SSA-LSSVM classification is significantly improved compared with traditional ELM network and PSO-ELM network.

The future research direction will focus on image recognition in fuzzy background.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors acknowledge funding received from the following science foundations: Key Scientific and Technological Research and Development Projects of Jilin

(20200401093GX) and Scientific and Technological Plan of Changchun (21ZGG14).

References

- [1] Y. Sun, Y. Chen, and X. Wang, "Deep learning face representation by joint identification-verification," in *Proceedings of the Conference on Neural Information Processing Systems (NIPS)*, Montreal, Canada, December 2014.
- [2] W. Y. Liu, Y. D. Wen, Z. D. Yu, M. Li, B. Raj, and L. Song, "SphereFace: deep hypersphere embedding for face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6738–6746, Honolulu, HI, USA, July 2017.
- [3] Y. Sun, D. Liang, X. Wang, and X. Tang, "DeepID3: face recognition with very deep neural networks," 2015, <https://arxiv.org/abs/1502.00873>.
- [4] H. Liu, Y. Pan, S. Li, and Y. Chen, "Synchronization for fractional-order neural networks with full/under-actuation using fractional-order sliding mode control," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 7, pp. 1219–1232, 2018.
- [5] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, "Scale-aware fast R-CNN for pedestrian detection," *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 985–996, 2018.
- [6] H. Liu, S. G. Li, G. J. Li, and H. Wang, "Robust adaptive control for fractional-order financial chaotic systems with system uncertainties and external disturbances," *Information Technology and Control*, vol. 46, no. 2, pp. 246–259, 2017.
- [7] J. Hosang, M. Omran, R. Benenson, and B. Schiele, "Taking a deeper look at pedestrians," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4073–4082, Boston, MA, USA, June 2015.
- [8] A. Angelova, A. Krizhevsky, V. Vanhoucke, A. Ogale, and D. Ferguson, *Real-Time Pedestrian Detection with Deep Network Cascades*, BMVC, Swansea, UK, 2015.
- [9] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. F. Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1725–1732, Columbus, OH, USA, June 2014.
- [10] H. Mobahi, R. Collobert, and J. Weston, "Deep learning from temporal coherence in video," in *Proceedings of the Annual International Conference on Machine Learning*, pp. 737–744, Montreal, Canada, June 2009.
- [11] H. Su, X. T. Zhu, and S. G. Gong, "Deep learning logo detection with data expansion by synthesising context," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 530–539, Santa Rosa, CA, USA, March 2017.
- [12] H. Su, S. Gong, and X. J. Zhu, "Scalable deep learning logo detection," 2018, <https://arxiv.org/abs/1803.11417>.
- [13] P. Wu, J. Liu, M. Li, S. Yujia, and S. Fang, "Fast sparse coding networks for anomaly detection in videos," *Pattern Recognition*, vol. 107, 2020.
- [14] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2720–2727, Sydney, Australia, December 2013.
- [15] R. V. H. M. Colque, C. Caetano, M. T. L. de Andrade, and W. R. Schwartz, "Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 3, pp. 673–682, 2017.
- [16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, San Diego, CA, USA, July 2005.
- [17] C. Yunqiang, Z. Xiang Sean, and T. S. Huang, "One-class SVM for learning in image retrieval," in *Proceedings of the 2001 International Conference on Image Processing (Cat. No.01CH37205)*, pp. 34–37, Thessaloniki, Greece, October 2001.
- [18] L. Xiong, B. Póczos, and J. Schneider, "Group anomaly detection using flexible genre models," in *Proceedings of the Twentyfourth International Conference on Neural Information Processing Systems*, pp. 1071–1079, Curran Associates Inc, Granada, Spain, December 2011.
- [19] A. Zimek, E. Schubert, and H.-P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," *Statistical Analysis and Data Mining*, vol. 5, no. 5, pp. 363–387, 2012.
- [20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [21] K. Simonyan and A. J. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016, <https://arxiv.org/abs/1512.03385>.
- [23] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5987–5995, Honolulu, HI, USA, July 2017.
- [24] J. Hu, L. Shen, and S. Albanie, "Squeeze-and-Excitation networks," 2017, <https://arxiv.org/abs/1709.01507>.
- [25] X. Zhou, J. Zhuo, and P. Krahenbuhl, "Bottom-up object detection by grouping extreme and center points," in *Proceedings of the Thirtysecond nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 850–859, Long Beach, CA, USA, June 2019.
- [26] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing images: a survey and a new benchmark," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 296–307, 2020.
- [27] B. Jin, L. Cruz, and N. Goncalves, "Deep facial diagnosis: deep transfer learning from face recognition to facial diagnosis," *IEEE Access*, vol. 8, pp. 123649–123661, 2020.
- [28] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 580–587, Columbus, OH, USA, June 2014.
- [29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *Neural Information Processing Systems*, vol. 39, 2015.
- [30] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7405–7415, 2016.
- [31] T. Tang, S. Zhou, Z. Deng, L. Lei, and H. Zou, "Arbitrary-oriented vehicle detection in aerial imagery with single convolutional neural networks," *Remote Sensing*, vol. 9, no. 11, 2017.
- [32] M. Zhao, A. Jha, Q. Liu et al., "Faster Mean-shift: GPU-accelerated clustering for cosine embedding-based cell

- segmentation and tracking,” *Medical Image Analysis*, vol. 71, Article ID 102048, 2021.
- [33] B. Jin, L. Cruz, and N. Goncalves, “Face depth prediction by the scene depth,” in *Proceedings of the 2021 IEEE/ACIS nineteenth International Conference on Computer and Information Science (ICIS)*, pp. 42–48, Shanghai, China, June 2021.
- [34] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. R. C. Y. Fu, and A. C. Berg, “SSD: single Shot MultiBox detector,” 2015, <https://arxiv.org/abs/1512.02325>.
- [35] W. Liu, L. Ma, and H. Chen, “Arbitrary-oriented ship detection framework in optical remote-sensing images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 6, pp. 937–941, 2018.
- [36] J. Xue and B. Shen, “A novel swarm intelligence optimization approach: sparrow search algorithm,” *Systems Science & Control Engineering*, vol. 8, no. 1, pp. 22–34, 2020.
- [37] R. M. May, “Simple mathematical models with very complicated dynamics,” *Nature*, vol. 261, no. 5560, pp. 459–467, 1976.
- [38] J. A. K. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [39] X. Ye and J. Zhao, “Multi-manifold clustering: a graph-constrained deep nonparametric method,” *Pattern Recognition*, vol. 93, pp. 215–227, 2019.
- [40] S. Ding, H. Zhao, Y. Zhang, X. Xu, and R. Nie, “Extreme learning machine: algorithm, theory and applications,” *Artificial Intelligence Review*, vol. 44, no. 1, pp. 103–115, 2015.
- [41] L. Yang, J. Zhang, X. Wang, Z Li, and Y He, “An improved ELM-based and data preprocessing integrated approach for phishing detection considering comprehensive features,” *Expert Systems with Applications*, vol. 165, 2021.