Received January 18, 2019, accepted February 5, 2019, date of publication February 7, 2019, date of current version March 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2898028

Multi-Scale Convolutional Neural Networks for Space Infrared Point Objects Discrimination

QIUQUN DENG, HUANZHANG LU, HUAMIN TAO, MOUFA HU, AND FEI ZHAO
National Key Laboratory of Science and Technology on ATR, College of Electronic Science, National University of Defense Technology, Changsha 410073, China

Corresponding author: Qiuqun Deng (dengqiuqun75@163.com)

This work was supported in part by the National Key Laboratory of Science and Technology on ATR, National University of Defense Technology.

ABSTRACT Object discrimination plays an important role in an infrared (IR) imaging system. However, at a long observing distance, the presence of detector noise and the absence of robust features make space objects' discrimination difficult to tackle with. In this paper, a multi-scale convolutional neural network (MCNN) is proposed for feature learning and classification. It consists of three parts: transformation, local convolution, and full convolution. Different from previous objects' classification methods, the MCNN can automatically extract features of objects at multi-timescales and multi-frequencies. Low-level features are combined with high-level features to simultaneously capture long-term tendency and short-term fluctuations of the time sequences of IR radiation intensity. Training data are generated from IR radiation models considering micro-motion dynamics and inherent properties of space point objects under different scenarios. The simulation results indicate that our method not only promotes the performance but is also robust to the detector noise. The classification accuracy can reach 96% at a strong noise level (signal-to-noise ratio is 10 dB) in a simulation scenario.

INDEX TERMS Convolutional neural network, space point objects, infrared radiation, discrimination, multi-scale.

I. INTRODUCTION

Objects discrimination using IR sensor is a key technology for space tracking systems and surveillance systems [1] and is significant to space security. In the case of long observing distance, objects emerge as small dots on the IR image plane lacking shape and attitude information. And when objects are free from air resistance and gravity, they follow the same trajectory during flight. Those objects have similar IR radiation intensity, which poses a great challenge to the discrimination system. Recently, object discrimination based on time sequence of object IR radiation intensity has attracted much attention, which provides a feasible way for the problem solving. Thus, the problem of IR point objects discrimination can be converted into the classification of time sequences of IR signatures.

Existing space IR objects discrimination methods [2]–[7] can be mainly classified into two categories, i.e., either hand-crafted models or learning based approaches. Most hand-crafted models focus on feature extraction which is of vital importance for object classification. Silberman [2] extracted several statistical features of input signatures, e.g., mean, variance, to build classifier for ballistic targets. The analysis of objects infrared signatures showed that different objects may possess different temperatures and cool at varying rates. Wang and Yang [4] proposed probabilistic neural network for exo-atmospheric target discrimination using the temporal evolutional characteristics of temperature and emissivity-area products as inputs of the neural network. Temperature feature can be extracted through the radiation ratio of two different wavelengths. Besides the temperature feature, the micro-motion features and geometrical shape also serve as important features. Exo-atmospheric objects are always in rotational or vibrational motion, referred to as micro-motion, until they re-enter the atmosphere [5]. Micro-motion of object leads to periodic fluctuations of the time sequence of IR signature [7]. Although the shape of object cannot be resolved by IR imaging, different shape of micro-motion object can induce different time sequence of IR signature [8], [9]. Wu *et al.* [10] and Liu *et al.* [11] estimated the micro-motion and shape parameters based on the model they proposed, respectively. However, these approaches are limited by the assumptions of specific parameters, when the conditions change, the performance of these hand-crafted features will obviously decrease.

The associate editor coordinating the review of this manuscript and approving it for publication was Gang Mei.

The feature extraction of space infrared point object remains challenging.

Deep learning techniques have recently achieved impressive results in a variety of domains [12]–[19], which provide a new perspective for the problem mentioned above. Instead of extracting hand-crafted features, learning based approaches can adopt deep neural network for feature learning from the raw data directly. As one of deep neural networks, convolutional neural network (CNN) has been successfully applied to object detection [17], visual recognition [18], [20] and time series classification [21]–[28]. However, for space IR point objects, the dissimilarity between classes is small. In typical CNN, the first convolution layer may lose some important information in bottom layers of the CNN, limiting its performance in multiple feature extracting. To overcome this shortcoming, CNN combining Multi-scale representation [21], [29]–[31] is explored in this work. Since the IR signatures are generated by dynamical systems which are mainly caused by motion, including orbit motion and micro-motion, and are susceptible to detector noise. The challenge of the application of CNNs is to explore the framework that is appropriate for learning both fast variables and slow variables and is robust to noise. Considering the properties of IR signature, we propose a multi-scale convolutional neural network (MCNN) to automatically learn suitable feature in both time and frequency domains for objects classification.

Our contributions can be summarized as follows:

(1) The infrared radiation intensity sequence model of space point objects is established. By analyzing the characteristics of motion, temperature and shape projection of space object, a direct mathematical model of micro-motion observation and the mathematical expression of projection area series are derived. On this basis, time sequence of IR signature can be easily obtained.

(2) We propose a multi-scale CNN structure for discrimination of space IR point objects, where the multi-timescale and multi-frequency information are used and serve as input to the network, enabling more richer feature learning.

(3) Extensive evaluations are performed and validate that the proposed framework can significantly improve the performance of space IR point objects discrimination under limited data and strong noise scenarios.

The rest of this paper is organized as follows. Section 2 introduces IR signature model of space point objects. Section 3 presents the detailed methodology of our work. We present results with the simulated IR radiation datasets to demonstrate the effectiveness of our classifier in Section 4. And conclusion is drawn in Section 5.

II. IR SIGNATURES MODELING

Due to consuming time, high expense, limited number of experiments depending only on measurement, it is inconvenient to analyze IR signature of space point objects under various complex conditions. Thus, the simulation is an alternative method [6], which provides an effective research facility to analyze IR signature with more flexibility and lower costs. There are a number of models used for simulating IR signature measurements in literature [8]–[11]. The majority of previous work are based on simplified assumptions. However, the space object IR signature detected by motion platform depends on a wide range of complicated factors, including the object surface temperature, emissivity of material, geometry, motion attitude and detector characteristics [5]. To bridge the gap between simulation and real scenarios, we provides a relative comprehensive consideration for space IR objects applications. Our IR object signature simulation is performed by IR radiation intensity model, projection area model, and time sequence model.

A. IR RADIATION INTENSITY MODEL

According to Planck's Law of Radiation, an object's radiation is mainly depended on its temperature distribution. To calculate the temperature, we divide the objects' surfaces into small pieces using finite element method [9]. Compared to the sun and the earth, the other sources contribute little radiation to the object. Therefore, the main factors that affect an object's temperature are as follows:

1) External heat source radiation $q1$,
2) Internal heat source radiation $q2$,
3) Heat exchange between adjacent nodes $q3$,
4) Radiation from other nodes$q4$,
5) Node's own heat radiation $q5$,
6) Node's temperature variation $q6$.

In the finite element method, the above heat radiations factors all have mature calculation formulas and details are available in the references [9]. Based on law of energy conservation, the node heat equivalence function can be understood as the following:

$$q1 + q2 + q3 + q4 = q5 + q6 \tag{1}$$

The temperature of every node at a given time can be calculated according to the initial temperatures supplied in Table 1. Thus, for an object at absolute temperature T, the power received by the sensor in the wave band $\lambda_1 \sim \lambda_2$ is

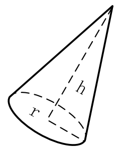
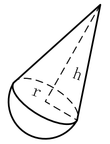
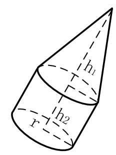
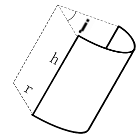
$$P_T(\lambda_1 \sim \lambda_2) = \frac{\pi D^2}{4} \varepsilon \tau A_{proj} \cdot \int_{\lambda_1}^{\lambda_2} M_T(\lambda) \mathrm{d}\lambda \tag{2}$$

where D is the optical aperture diameter, R is the distance between the sensor and the object, τ is the optical transmittance of the system, ε is the emissivity of the object, $M_T(\lambda)$ is the spectral radiant exitance of blackbody. According to Planck's law, $M_T(\lambda)$ can be represented as

$$M_T(\lambda) = 2hc^2/\lambda^5[\exp(hc/(k\lambda T)) - 1]^{-1} \tag{3}$$

where c denotes the velocity of light in vacuum, k denotes the Boltzmann entropy constant, and h denotes the Planck constant. A_{proj} is the projection area along the line of sight (LOS) of the detector. During the flight of space object outside the atmosphere, the detector is approaching the object continuously, the projection area is varying due to the micromotion. So the A_{proj}, T and R are the keys to the changing of IR

TABLE 1. Simulation parameters of four classes and IR detector.

Objects type	O1	O2	O3	O4
3-D models				
Shape parameters	r = 0.25~0.35m h = 0.75~1.25m	r = 0.25~0.35m h = 0.75~1.25m	r = 0.25~0.35m h1 = 0.25~0.55m h2 = 0.50~0.70m	r = 0.20~0.40m h = 0.30~0.70m φ = 0.5π~0.7π
Micro-motion mode	Coning	Coning	Tumbling	Tumbling
micro-motion parameters	θ=0.1π~0.2π ω_s=3.0π~6.0π rad/s α_c=0.0π β_c=0.25π~0.35π ω_c=0.25π~0.3π rad/s	θ=0.15π~0.35π ω_s=2.0π~5.0π rad/s α_c=0.0π β_c=0.2π~0.4π ω_c=0.3π~0.35π rad/s	θ=0.0π~0.25π α_t=0.0π β_t=0.1π~0.3π ω_t=0.3π~0.35π rad/s	θ=0.15π~0.25π α_t=0.0π β_t=0.05π~0.2π ω_t=0.25π~0.3π rad/s
Initial temperature	320K	320K	320K	600K
Coating material $\alpha_V/\varepsilon_{IR}$	0.87/0.7	0.27/0.50	0.43/0.36	0.54/0.20
Specific Heat	0.904	1.15	0.732	0.812
IR detector parameters	Wave band: 8~12 μm; Observation time: T = 20s; Sample frequency: f = 20Hz;			

signature while the other parameters are invariant. More crucially, the object projection area A_{proj} is a complex variable that closely related to the shape and the motion attitude.

B. PROJECTION AREA MODEL

Generally, the shape model of space objects detected by IR detector include irregular fragments and symmetrical likes flat-base cone, ball-base cone, cylinder, cone-cylinder, sphere etc. In this work, our classification mainly considers four representative categories: flat-base cone, ball-base cone, cone-cylinder and arc-shaped debris. We define them as object 1 (O1), object 2 (O2), object 3 (O3) and object 4 (O4), respectively. We choose them because they are similar in shape. Once the algorithm can accurately discriminate them, it can also discriminate those objects that have larger difference in shape.

To calculate the projection area, it is necessary to deduce the micro-motion process of the objects as it varies with micro-motion parameters. The geometry of the sensor and a

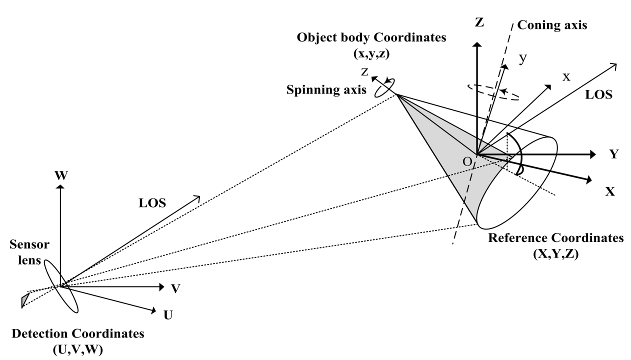


FIGURE 1. Geometry of the sensor and a coning object.

coning object is illustrated in Fig. 1. Coning is a rigid body rotation about an axis that intersects with an object body coordinate, which usually combined with spinning. The sensor coordinate system is *(U, V, W)*, the object body coordinate is *(x, y, z)*, the reference coordinate *(X, Y, Z)* is parallel to the sensor coordinate and its origin locates in object's mass center. α,β are the azimuth and elevation angle of the sensor LOS in reference coordinate.

It is supposed that the convex object surface is assembled by N small triangular patches, and the area and normal vector of patches are a_i and $\mathbf{n}_i (i = 1, 2, \ldots, N)$ respectively. $\mathbf{R}_{init}$ is initial rotation matrix, described by Euler angle (ϕ, θ, ψ). the matrix of Euler angle can be expressed as

$$\mathbf{R}_{init} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \quad (4)$$

Viewed in the reference coordinate system, when an object rotates about a rotation axis $\overrightarrow{oz}$ whose azimuth and elevation angle in the reference coordinates *(X, Y, Z)* are α_R and β_R with the angular velocity ω_R, the point $\vec{p}_1$ at time $t_0 = 0$ described in the reference coordinate would move to a new position $\vec{p}_2$ at time t by rotation matrix $\mathbf{R}_t$. According to Rodrigues formula [7], the rotation matrix $\mathbf{R}_t$ can derived as

$$\mathbf{R}_t = \mathbf{I} + \hat{\mathbf{e}} \sin\omega_R t + \hat{\mathbf{e}}^2 (1 - \cos\omega_R t) \quad (5)$$

where $\hat{e}$ is a skew symmetric matrix.

$$\hat{e} = \begin{bmatrix} 0 & -\sin\beta_R & \sin\alpha_R \cos\beta_R \\ \sin\beta_R & 0 & -\cos\alpha_R \cos\beta_R \\ -\sin\alpha_R \cos\beta_R & \cos\alpha_R \cos\beta_R & 0 \end{bmatrix} \quad (6)$$

So at time t, the normal vector $\vec{n}_i = [x_0, y_0, z_0]$ of any an object patch will rotate from its initial vector in the object local coordinate system to a new vector $\vec{n}_i^{new}$ in the reference coordinate system by

$$\mathbf{n}_i^{new} = \mathbf{R}_t \cdot \mathbf{R}_{init} \cdot \mathbf{n}_i \tag{7}$$

Therefore, according to Eq. (5)-(7), the object's projection area along the LOS at time t can be expressed as

$$A_{\mathrm{proj}} = \sum_{i=1}^{N} a_i \cdot \max[\cos(\mathbf{n}', \mathbf{n}_i^{new}), 0] \tag{8}$$

C. TIME SEQUENCE MODEL

In subsection A and B, the main factors affecting the varying of the IR radiation have been discussed in detail. In this subsection, the time sequence of IR radiation is established based on the IR signature model.

The detection distance is computed based on the ellipse trajectory theory. The flight trajectory of object and detector is illustrated in Fig. 2. We set the start position of geographical coordinates of objects is (130° E, 80° N, 150 km) and the end position is (80° E, 40° N, 150 km). The peak distance of trajectory from the ground is 469.3km. The detector observes the objects with speed of 6 km/s from start position (95° E, 35° N, 300 km). The data are gathered from 300 s to 320 s. A series of IR signature sequences are generated by sampling randomly from a distribution of physical attributes and dynamic states as listed in Table 1.

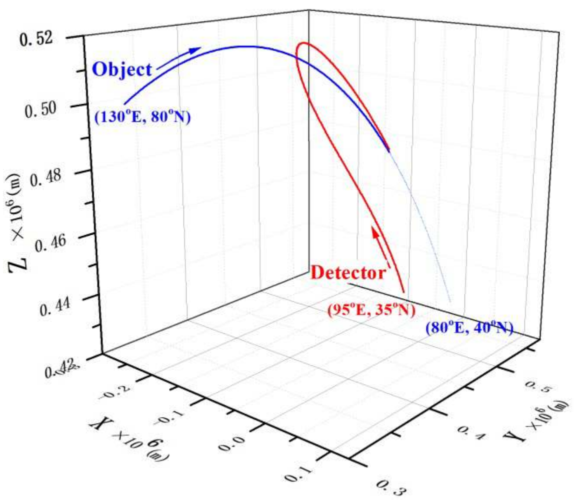


FIGURE 2. The flight trajectory of object and detector.

Based on the modeling method stated above and simulation parameters listed in Table 1, four types of objects were simulated. Fig. 3 shows the idealized radiation intensity of objects varies over time. We can see that the IR signatures have not only the long-term variation but also short-term fluctuation characteristics. They all have periodic fluctuations due to micromotion. The O1 has relatively small periodic variation amplitude while the O2 and O3 have larger periodic variation

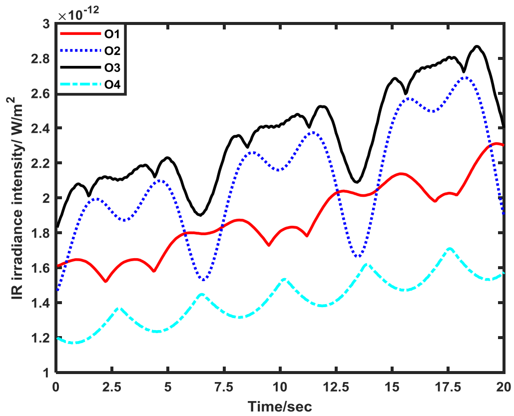


FIGURE 3. IR radiation intensity sequences under ideal condition.

amplitude. The IR radiation sequences of the O1, O2 and O3 are unsmooth and somewhat jagged while O4 is relatively uniform. There are complicated factors induced this variation, including the object surface temperature, emissivity of material, geometry, motion attitude and observing angle. It cannot be discriminated simply depending on several specific features.

It should be noted that the description of the IR radiation intensity sequence model of the target is inevitably idealized. In fact, the temperature variation on the surface of the object is not completely inconsistent, and the IR sensor sensitivity, response rate, etc. will change slightly. For these reasons, in the IR radiation intensity sequence simulation, these factors are usually described as Gaussian additive white noise to improve the authenticity of the data description. Fig. 4 shows the IR radiation sequences of four classes at noise level of SNR =5 dB. The SNR is defined as the ratio of signal power P_s to the noise power P_n and is computed by the formula: $SNR = 10\log_{10}(P_s/P_n)$. It poses a great difficulty for discrimination as four classes affected by strong noise are very similar.

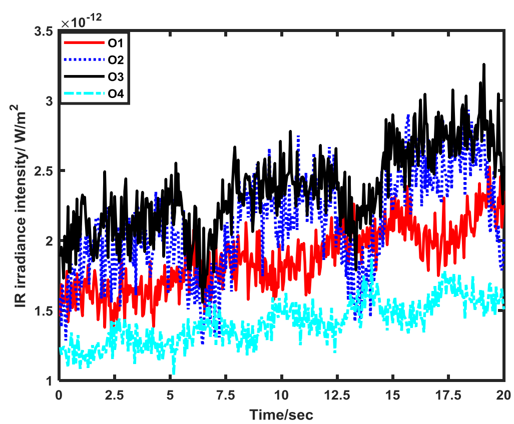


FIGURE 4. IR radiation intensity sequences with noise.

III. PROPOSED CLASSIFICATION FRAMEWORK

Given a time sequence of IR signature, our goal is to learn a set of discriminative features using the CNNs, which captures the essence of different objects. The challenge is the insufficiency of feature extraction for objects classification. In this work, we propose a MCNN framework to extract multi-scale features from IR signatures. As shown in Fig. 5, the framework contains three sequential parts, namely, transformation part, local convolution part and full convolution part. The transformation part includes identity mapping, down-sampling transformations in time domain, and spectral transformations in the frequency domain. The local convolution has three CNN in parallel which learn features from different transformations of input. The full convolution part concatenates all extracted features from local convolution part and apply two more convolutional layers and max pooling layers, one fully connected layer and a softmax layer for final classification.

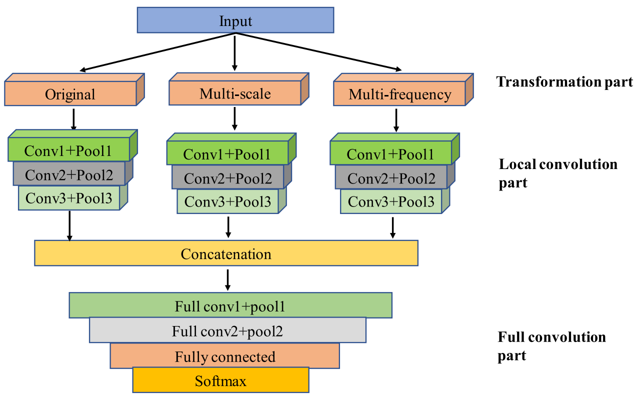


FIGURE 5. Proposed MCNN framework.

A. TRANSFORMATION OF INPUT SPACE

The transformation of input IR signature is inspired by Cui et al. [21]. In fact, our IR signature classification problem can be viewed as a special time series classification problem. There are complicated factors induced the fluctuate of time sequence of IR signature both in short range and long range. Moreover, the time sequence of IR signature is often distorted by detector noise. It remains challenge to learn an appropriate feature representation from raw data. In this work, we make a try to address this problem through diversity transformation of input space before feature extracting. The transformation includes down-sampling in time domain and smoothing in frequency domain. By means of down-sampling and smoothing with different window sizes, we can get multiple time sequences of IR signature with different time scales and frequencies. We denote a time sequence of IR signature as $T = \{t_1, t_2, \cdots, t_n\}$ and the down-sampling rate is m, then the new time series T^m can be expressed as

$$T^m = \{t_{1+m*i}\}, \quad i = 0, 1, \cdots, \frac{n-1}{m} \tag{9}$$

Moving average is applied to generate multiple time sequences of IR signature with different degrees of smoothness. Defining the window size as s, then the new time series T^s can be expressed as

$$T^s = \frac{x_i + x_{i+1} + \ldots + x_{i+s-1}}{s}, \quad i = 0, 1, \cdots, n-s+1 \tag{10}$$

By doing multi-scale transformation, both short-term features and long-term features can be utilized at the same time. While traditional CNN learns features from low level to high level, which may lose some important features for classification. Moving average with different window sizes plays a role to reduce the noise level, which is equivalent to a low frequency filter. After transformation, the input is divided into three branches, namely, original branch, multi-timescale branch and multi-frequency branch.

B. CONFIGRATION OF THE PROPOSED MCNN

The local convolution block has three channels sharing the same CNN architecture, while the full convolution block adopts a different structure. The details of layouts of each network are described in Table 2.

TABLE 2. Architectures of the proposed MCNN.

Type	Kernel size	Feature maps	Strides
Conv-local	3	12	1
Maxpool-local	2	12	2
Conv-local	3	12	1
Maxpool-local	2	12	2
Conv-local	3	24	1
Maxpool-local	2	24	2
Conv-full	3	12	1
Maxpool-full	2	12	2
Conv-full	3	24	1
Maxpool-full	2	24	2
Fully-connect		204/396/ 792	
Softmax		4	

First, we design the local convolution block with three identical CNN structure, which separately extracts features from the output of transformation part. The architecture of CNN in each channel contains three convolution layers, three max pooling layers and one fully connected layer. The convolution operation between an input feature map x and a convolutional kernel W is defined by

$$\mathrm{h}(x) = f(x * W + b) \tag{11}$$

where $*$ denotes the convolution operator, f is the activation function for each layer that adds nonlinearity to the feature vector. We use ReLu for layer activation, which is defined by

$$f(x) = \max(0, x) \tag{12}$$

Following the convolutional layer, a max-pooling layer is applied to reduce feature maps' size as well as the number of following layers' parameters to reduce redundancy and improve computation efficiency. After convolution and max pooling, a fully connected layer is introduced for the later feature extraction block.

Then, we concatenate the output of each fully connected layer from three branches and feed them into the next block as input.

Finally, we design the full convolution block, which is composed of two convolution layers, two max pooling layers, one fully connected layer, and a softmax classifier. ReLu is used for layer activation and dropout is applied to avoid overfitting. A softmax function is used to restrict outputs in the ranges (0,1), which is defined by

$$y_j(h(x)) = \frac{\exp(h_j)}{\sum_i \exp(h_i)} \tag{13}$$

Cross entropy is used as loss function and can be expressed as

$$L(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log P(y^i|x^i;\theta) \tag{14}$$

where vector θ are the parameters of network, $\{(x^i, y^i), i = 1, 2, \ldots n\}$ are the set of labeled training set, y^i corresponding to the true label of the sample x^i. Then the MCNN is trained by adaptive moment estimation (Adam) [32] by minimizing the cross-entropy loss between the outputs and the labeled data. The parameters updating rule is

$$\theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{\hat{v}_t} + \varepsilon}\hat{m}_t \tag{15}$$

$$\hat{m}_t = m_t/(1-\beta_1^t)$$
$$\hat{v}_t = v_t/(1-\beta_2^t) \tag{16}$$

where α is the learning rate, t is the iteration step, $\hat{m}_t$, $\hat{v}_t$ are the first moment estimation and second moment estimation respectively. The parameters β_1,β_2 are exponential decay rate and the default value of ε is 10^{-8} generally. Adam is an algorithm for first-order gradient-based optimization of stochastic objective functions which is well suited for problems with noisy or sparse gradients. Empirical results show that Adam algorithm performs well in practice and has great advantages compare to other stochastic optimization algorithms.

IV. SIMULATION RESULTS

In this section, the classification framework is validated using simulated data which are generated from physics-based models as mentioned in section 3. Comparison with traditional methods are conducted to evaluate the performance of our proposed approach on different noise level and data length (L). Our method is implemented in Python with the TensorFlow wrapper and run on a PC with 2.7 GHz CPU and 8 GB 1600 MHz DDR3 memory.

A. DATA DESCRIPTION

To meet the requirement of real situation, we simulated data with different length and noise level. Firstly, we transform the simulated 20 second objects IR signature data into three parts: the first 5 second, the first 10 second and 20 second, respectively. We make this division according to the micro-motion periodicities of simulated signatures which are set in the range of 6 second to 8 second, as displayed in table 1. Then, we add White Gaussian noise to the data with different signal-to-noise ratios (SNR =5 dB, 10 dB, 15 dB, 20 dB, 25 dB, 30 dB) to validate the performance of our proposed method. In all the experiments, we use 60 percent of data for training, 20 percent of data for validation and 20 percent of data for test. To achieve the best performance, a set of experiments are conducted to choose the suitable amount of dataset. The classification results are the average of 10 runs. Fig. 6 is the accuracy curve under different amount of training samples where the horizontal axis represents for the dataset size. It can be seen that when the amount of training samples reach about 1600, the accuracy gradually converges.

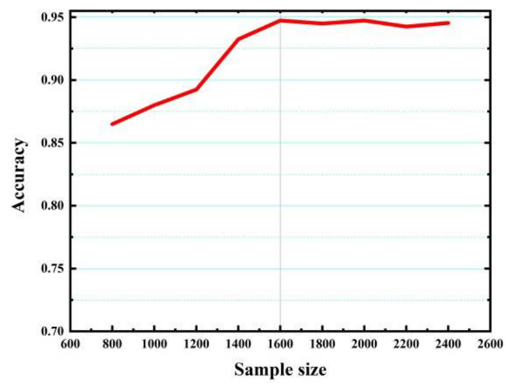


FIGURE 6. Classification accuracy using different number of training samples.

B. EFFECT UNDER DIFFERENT SCENARIOS

To verify the effectiveness of our proposed method to the space point IR objects discrimination. We conduct simulations under different scenarios. Table 3 displays the classification performance of MCNN with different data length under a range of noise level. The results show that the accuracy increases with the input data length L and SNR. This is because the higher noise level can distort the signature, which directly worsen the classification accuracy. On the

TABLE 3. Classification results under different scenarios.

SNR	Accuracy		
	L = 5s	L = 10s	L = 20s
5 dB	77.00%	84.50%	92.75%
10 dB	79.25%	90.50%	96.00%
15 dB	84.25%	93.50%	95.50%
20 dB	86.75%	94.75%	96.50%
25 dB	89.50%	94.25%	96.25%
30 dB	88.75%	95.75%	96.50%

other hand, for periodic IR signature, it would be a challenge to the classifier when the objects information is less than one cycle. However, the accuracy would not increase too much with the increment of the input data length when the input contains more than one cycle information. The classification accuracy of our proposed method can reaches 96% at strong noise level (SNR = 10 dB), demonstrating its robustness to noise. And when the IR signatures are less polluted by noise (SNR = 25 dB), it can achieves 89.5% with limited input information (L = 5 s).

C. EFFECTIVENESS OF THE MULTI-SCALE CONVOLUTION

To validate the effectiveness of multi-scale transformation, we perform simulation under three different scenarios, namely, Scenario I, Scenario II and Scenario III. Scenario I represents the time sequence of IR signature with 5 seconds (the length of sequence is 100) at the noise level of 30 dB. Scenario II represents the time sequence of IR signature with 10 seconds (the length of sequence is 200) at the noise level of 20 dB. Scenario III represents the time sequence of IR signature with 20 seconds (the length of sequence is 400) at the noise level of 10 dB. In each simulation, we remove the multi-frequency branch to test the effect of multi-timescale transformation on the classification accuracy and remove the multi-timescale branch to test the effect of multi-frequency transformation on the classification accuracy. We denote the down-sampling window size, step size, number and the data length after down-sampling as W_1, S_1, N_1, L_1, respectively, and denote moving average window size, step size, number and the data length after moving average as W_2, S_2, N_2, L_2, respectively. In the column of down-sampling, we only perform multi-timescale transformation. In the column of moving average, we only perform multi-frequency transformation.

An evaluation of our proposed method with varying down-sampling and moving average parameters under different scenarios is shown in Table 4-6. The results show that the different down-sampling parameters and moving average parameters affect the classification accuracy at different level. The parameters N_1 and N_2 play a major influence to the accuracy. From the last row of Table 4-6, we can see that the worst performance is obtained when applying single scale transformation. The parameters owning best performance are shown in bold. Considering three scenarios comprehensively, we select (2,1,4) as down-sampling parameters and (4,3,3) as moving average parameters. It should be noted that the data length is varied with different parameters setting. The performance of adopting multi-timescale branch alone or multi-frequency branch alone is worse than the combination of the two simultaneously. In Scenario I, the accuracy of proposed method with three branches can reach 88.75% (shown in Table 3) while that is 88.50% (shown in Table 4) with two branches. In Scenario II, the accuracy of three branches is 94.75% (shown in Table 3) while that is 93.25% and 93.75% (shown in Table 5), respectively, with two branches. In Scenario III, the accuracy of three branches is 96.00% (shown in Table 3) while that is 95.25% (shown in Table 6) with two branches. It demonstrates the effectiveness of application of multi-scale transformation.

TABLE 4. Classification accuracy (%) with varying down-sampling parameters (W_1, S_1, N_1)-L_1 and moving average parameters (W_2, S_2, N_2)-L_2 under Scenario I.

Down-Sampling	Accuracy	Moving average	Accuracy
(2,1,6)-161	85.25%	(4,2,4)-376	88.50%
(2,1,5)-146	85.25%	(4,3,4)-370	87.00%
(2,1,4)-129	**88.50%**	(2,3,3)-288	87.75%
(2,1,3)-109	86.25%	(4,2,3)-285	87.50%
(2,1,2)-84	86.25%	(3,3,3)-285	87.50%
(3,1,4)-96	87.75%	**(4,3,3)-282**	**88.50%**
(4,1,4)-77	87.50%	(6,3,3)-276	87.75%
(2,2,2)-75	87.50%	(4,3,2)-191	87.50%
(2,2,1)-50	82.50%	(4,3,1)-97	84.25%

TABLE 5. Classification accuracy (%) with varying down-sampling parameters (W_1, S_1, N_1)-L_1 and moving average parameters (W_2, S_2, N_2)-L_2 under Scenario II.

Down-Sampling	Accuracy	Moving average	Accuracy
(2,1,6)-320	**93.25%**	(4,2,4)-776	93.25%
(2,1,5)-291	92.25%	(4,3,4)-770	93.50%
(2,1,4)-257	93.00%	(2,3,3)-588	92.75%
(2,1,3)-217	91.25%	(4,2,3)-585	92.75%
(2,1,2)-167	91.25%	(3,3,3)-585	93.50%
(3,1,4)-191	91.50%	**(4,3,3)-582**	**93.75%**
(4,1,4)-153	92.50%	(6,3,3)-576	93.50%
(2,2,2)-150	91.00%	(4,3,2)-391	92.25%
(2,2,1)-100	91.00%	(4,3,1)-197	90.75%

TABLE 6. Classification accuracy (%) with varying down-sampling parameters (W_1, S_1, N_1)-L_1 and moving average parameters (W_2, S_2, N_2)-L_2 under Scenario III.

Down-Sampling	Accuracy	Moving average	Accuracy
(2,1,6)-639	94.75%	(4,2,4)-1576	94.75%
(2,1,5)-581	94.50%	(4,3,4)-1570	94.00%
(2,1,4)-514	**95.25%**	(2,3,3)-1188	94.75%
(2,1,3)-434	94.50%	(4,2,3)-1185	94.50%
(2,1,2)-334	94.50%	(3,3,3)-1185	94.50%
(3,1,4)-381	95.00%	(4,3,3)-1182	95.00%
(4,1,4)-305	95.00%	**(6,3,3)-1176**	**95.25%**
(2,2,2)-300	94.75%	(4,3,2)-791	94.00%
(2,2,1)-200	93.00%	(4,3,1)-497	93.75%

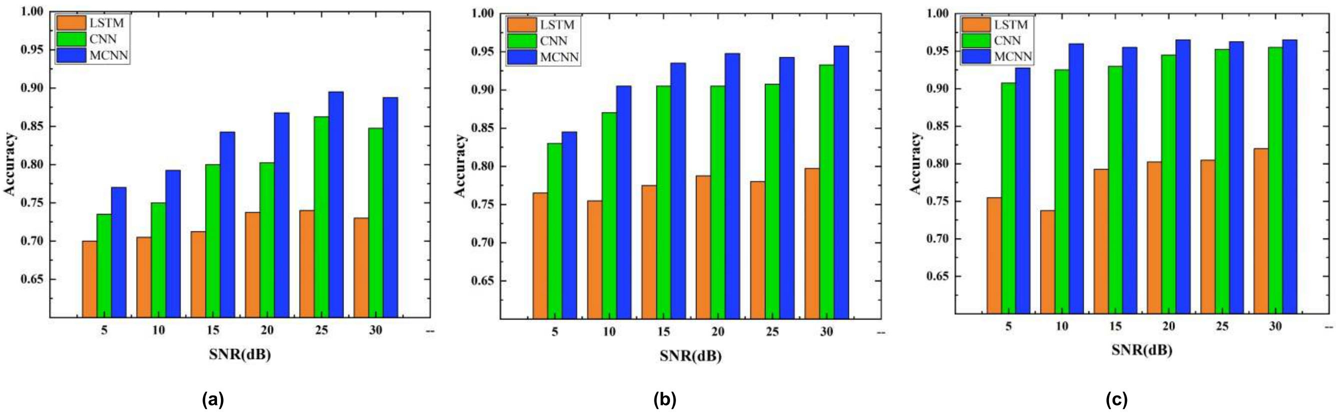


FIGURE 7. Classification performance comparison on different data length. (a) L = 5s. (b) L = 10s. (c) L = 20s.

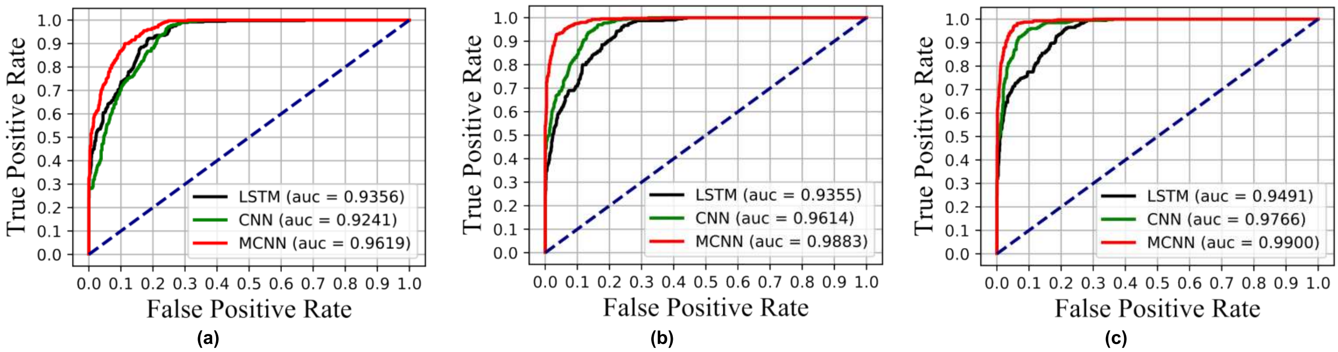


FIGURE 8. ROC curves of three methods on different data length. (a) L = 5s. (b) L = 10s. (c) L = 20s.

D. COMPARISON WITH OTHER METHODS

We evaluate the performance of the proposed framework with two classical baseline methods: Long Short-Term Memory (LSTM) and standard CNN. LSTM is a canonical recurrent network which has superior performance in sequence modeling tasks [33]–[35]. To show the benefit of using the proposed multi-scale transformations and local convolution, we test standard convolutional neural network with the same number of parameters as in MCNN. Four groups of comparisons are conducted in the following.

1) COMPARISON OF ACCURACY ON SIMULATED DATA

We use the Adam optimizer with learning rate 0.001. The cross validation is applied to select the optimal model parameters. Fig. 7 displays the comparison results of three classifiers at the data length of 5s, 10s and 20s, respectively. The results illustrate that the classification accuracies of three methods increase with the input length L and SNR. From the comparisons of three methods, we can see that the MCNN outperforms the other two networks not only in the conditions of different noise level but also in the scenes of different input length. In addition, the proposed framework still shows prominent performance even in lower noise level and with insufficient objects information. While the LSTM classifier perform not well. The comparisons of CNN and MCNN demonstrate the effectiveness of multi-scale transformations and local convolution of our proposed framework. The reason is that our method can extract features at different time scales exploring richer feature space. Instead of increasing different filter size in the same convolution layer, the multi-scale transformation of input time series can obtain different local receptive field with a same filter size. It benefits the learning of overall trends and subtle changes of time sequences, both of which are crucial to the classification. On the other hand, the transformation part can be viewed as a data augmentation technique, which significantly improves the networks' generalization capabilities, thus make the proposed model more robust to the variations of the input data compared to the standard CNN. LSTM can learn temporal dependencies in sequence and is suitable for short term sequence prediction problems. However, it still has difficulty with long term dependencies. For classification tasks, many existing works have shown that CNNs possess superior performance than LSTM. Our results are consistent with it.

2) COMPARISON OF ROC CURVES

To analyze the effectiveness of our method under different criteria, the receiver operating characteristic (ROC) curves and the areas under the ROC curves (AUC) of the above three methods at the data length of 5s, 10s and 20s are given in Fig. 8. ROC is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied and is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings [36]. AUC is a common indicator used to characterize the merits of the classification or prediction model, and the larger the AUC value, the better the performance of the model. It is noticed that the ROC curve of our method is above the other two methods and our method has the highest AUC.

3) COMPARISON OF ACCURACY ON REAL DATA

To further verify the classification performance of MCNN, we conduct experiments on real-world data sets originating from the ''UCR Time Series Classification Archive'' [37], which is a public datasets for time series classification and includes time series sets from different application domains. Since our IR radiation intensity data is essentially a time series. It is reasonable to use this datasets verifying MCNN's performance on real data. Table 7 provides the detailed information about each dataset we used.

TABLE 7. **Summary of data sets, including the number of classes, the time series length and the size of train/ validate/test set.**

Data set	Class	Length	Train/Validation/Test
Synthetic control	6	60	300/150/150
Swedish leaf	15	128	500/313/312
Face All	14	131	560/845/845
ECG200	2	96	100/50/50
50 words	50	270	450/228/227
Wafer	2	152	1000/3082/3082
OSU leaf	6	427	200/121/121

Table 8 shows the classification accuracy of three methods on real-world time series. The best results are shown in bold. It can be seen that MCNN obtains the best classification performance. CNN achieves the same performance with MCNN on two groups of data sets, which are '' Synthetic control'' and '' Wafer''. In fact, the classification accuracy of three methods are close on referred two data sets. However, MCNN shows superior performance on other data sets compared to CNN and LSTM as a whole. We believe that the main reason is that MCNN can learning more discriminative features through multi-scale convolutions and can easily converge to global optimal solutions while CNN may get local optimal solutions in some scenarios. The results on real data are consistent with the results on simulated data, which demonstrate the robustness of our proposed methods.

4) COMPARISON OF EFFICIENCY

Considering the efficiency of our proposed method, we provide the comparison of total parameters and run time of three methods in Table 9. It should be noted that the run time in the Table 9 refers to the time running on the same test

TABLE 8. **Classification accuracy comparison of three methods on real-world time series.**

Data set	MCNN	CNN	LSTM
Synthetic control	**1.000**	**1.000**	0.9800
Swedish leaf	**0.9423**	0.9262	0.6233
Face All	**0.7846**	0.7633	0.6525
ECG200	**0.9000**	0.8600	0.8000
50 words	**0.7648**	0.7384	0.4889
Wafer	**0.9967**	**0.9967**	0.9849
OSU leaf	**0.6690**	0.6527	0.4666

sets, which are the average of 10 runs. We can see that the LSTM has lowest model parameters but the longest run time. This is due to the fact that LSTM processes the input one by one. It takes more time to training and testing. While the CNNs can process the input in batch which are more efficient than LSTM. The increase of multi-channels inevitably has a slight influence on the efficiency of the proposed MCNN, which requires more parameters and run time than standard CNN. However, for the space object discrimination task, the accuracy serves as a key role. The MCNN framework achieves significantly improved performance despite with a minor increase in run time. On the whole, the design of our framework is relative lightweight.

TABLE 9. **Comparison of three methods on parameters and run time.**

Model	Total parameters	Run time (s)
LSTM	1020	0.709
CNN	3532	0.091
MCNN	6720	0.214

E. DISCUSSION

As MCNN method has significantly improved the performance of space point IR objects discrimination under different scenarios, the structure of MCNN has been analyzed. There are two features of MCNN that make MCNN really suitable for solving IR objects discrimination problem.

(1) Multi-scale convolution: time sequences of IR signatures have both long-term tendency and short-term fluctuations due to the combination of micro-motion and orbit motion. For our model, we transform the raw data into different time scales and frequency scales, which avoids losing important information in the bottom layers of MCNN and provides local features and global features at the same time after convolution. Meanwhile, the multi-frequency transformation makes the framework more robust to noise, which are essential for IR objects discrimination at long observing distance. Furthermore, by down-sampling the input time sequence instead of increasing the filter size, it can greatly

reduce the number of parameters in the local convolutional layer.

(2) Tolerance of shape changes and displacement: time sequences of IR signatures have many local period features and these features are distributed on the time axis, where every feature will appear to the center around a particular time which varies in a limited range. In other words, there are a lot of shape changes and displacements due to errors caused by motion platform and observing line of sight. To deal with the problem of variability in MCNN, max-pooling layers are inserted into the network structure. Generally, the activations of max-pooling layers are divided into some bands and there are a smaller number of bands can be obtained which provide a lower resolution feature. Those features contain more useful information and are more robust to shape changes and displacement.

Nevertheless, there are some limitations to this work. First, the training of deep neural networks is time consuming since model parameters are determined by lots of experiments and the model still can be further optimized; Second, the proposed method lacks actual flight data for test. But we believe that once trained this framework can be applied to real-data classification examples. Hence, in the future work, we plan to study and extend our framework for IR objects classification on more data sets and parameter settings.

V. CONCLUSION

In this paper, we established an IR radiation intensity sequence model of space point objects and proposed a MCNN framework for discrimination. The performance is evaluated with extensive simulations and experiments. The results show that using multi-scale convolution significantly improves the classification performance, especially in conditions with strong noise and limited information. The classification accuracy can reach 96% at strong noise level (SNR=10 dB) in simulation scenario. It overcomes the shortcoming of previous works that they only learn features with single time scale and are sensitive to detector noise. The proposed framework shows promise as a tool for space point objects discrimination. For future work, we will further optimize our framework using more data sets and applying to other similar classification problems.

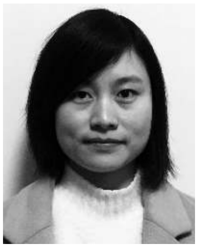
ACKNOWLEDGMENT

The authors would like to thank the editors and anonymous reviewers for their insightful comments and suggestions, which could significantly help in improving the quality of this paper.

REFERENCES

[1] H. X. Zhou, "*Infrared Detectors*, 2nd ed. Beijing, China: China Machine Press, 2014.
[2] G. L. Silberman "Parametric classification techniques for theater ballistic missile defense," in *Johns Hopkins APL Tech. Dig.*, vol. 3, 1998, pp. 322–339.
[3] J. Liu, S. Chen, H. Lu, and B. Zhao, "Ballistic targets micro-motion and geometrical shape parameters estimation from sparse decomposition representation of infrared signatures," *Appl. Opt.*, vol. 56, no. 4, pp. 1276–1285, 2017.
[4] J. Wang and C. Yang, "Exo-atmospheric target discrimination using probabilistic neural network," *Chin. Opt. Lett.*, vol. 9, no. 7, 2011, Art. no. 070101.
[5] L. Tao, "The study on ballistic midcourse targets detection and recognition based on the space-based infrared system," M.S. thesis, Nat. Univ. Defense Technol., Changsha, China, 2011.
[6] Liu, Z. "Research on techniques of detection and discrimination of point target in IR image," M.S. thesis, Nat.Univ. Defense Technol., Changsha, China, 2005.
[7] C. Resch, "Exo-atmospheric discrimination of thrust termination debris and missile segments," in *Johns Hopkins APL Tech. Dig.*, vol. 19, no. 3, 1998, pp. 315–321.
[8] L. Huang, X. Li, and J. Liu, "IR radiative properties modeling and feature extraction method on ballistic target," *Proc. SPIE*, vol. 9631, Jul. 2015, Art. no. 96310J.
[9] J. Zhang *et al.*, "Numerical analysis of infrared characteristics of exoatmospheric ballistic target," *Infr. Laser Eng.*, vol. 5, p. 007, Aug. 2008.
[10] Y. Wu, H. Lu, F. Zhao, and Z. Zhang, "Estimating shape and micro-motion parameter of rotationally symmetric space objects from the infrared signature," *Sensors*, vol. 16, no. 10, p. 1722, Oct. 2016.
[11] J. Liu *et al.*, "Micromotion dynamics and geometrical shape parameters estimation of exoatmospheric infrared targets," *Opt. Eng.*, vol. 55, no. 11, 2016, Art. no. 113103.
[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[13] G. E. Hinton, "Learning multiple layers of representation," *Trends Cognit. Sci.*, vol. 11, no. 10, pp. 428–434, Oct. 2007.
[14] Y. Bengio, A. Courville, and P. Vicent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[15] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *Comput. Sci.*, to be published.
[16] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[18] T. Y. Lin, A. Roychowdhury, and S. Maji, "Bilinear convolutional neural networks for fine-grained visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1309–1322, Jun. 2017.
[19] A. Hassan and A. Mahmood, "Convolutional recurrent deep learning model for sentence classification," *IEEE Access*, vol. 6, pp. 13949–13957, 2018.
[20] Q. Deng, H. Lu, M. Hu, and B. Zhao, "Exo-atmospheric infrared objects classification using recurrence-plots-based convolutional neural networks," *Appl. Opt.*, vol. 58, no. 1, pp. 164–171, Jan. 2019.
[21] Z. Cui, W. Chen, and Y. Chen. (2016). "Multi-scale convolutional neural networks for time series classification." [Online]. Available: https://arxiv.org/abs/1603.06995
[22] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, "Exploiting multi-channels deep convolutional neural networks for multivariate time series classification," *Frontiers Comput. Sci.*, vol. 10, no. 1, pp. 96–112, Feb. 2016.
[23] H. I. Fawaz *et al.* (2018). "Deep learning for time series classification: A review." [Online]. Available: https://arxiv.org/abs/1809.04356
[24] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *J. Syst. Eng. Electron.*, vol. 28, no. 1, pp. 162–169, Feb. 2017.
[25] J. C. B. Gamboa. (2017). "Deep learning for time-series analysis." [Online]. Available: https://arxiv.org/abs/1701.01887
[26] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2017.
[27] F. Karim *et al.* (2018). "Multivariate LSTM-FCNs for time series classification." [Online]. Available: https://arxiv.org/abs/1801.04503
[28] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, Jun. 2014.
[29] Z. Cai *et al.* (2016). "A unified multi-scale deep convolutional neural network for fast object detection." [Online]. Available: https://arxiv.org/abs/1607.07155

[30] D. Zeng and M. Zhu, "Background subtraction using multiscale fully convolutional network," *IEEE Access*, vol. 6, pp. 16010–16021, 2018.
[31] D. Yoo, S. Park, J. Y. Lee, and I. S. Kweon, "Multi-scale pyramid pooling for deep convolutional representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 71–80.
[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Comput. Sci.*, to be published.
[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[34] J. Chung *et al.* (2014). "Empirical evaluation of gated recurrent neural networks on sequence modeling." [Onine]. Available: https://arxiv.org/abs/1412.3555
[35] S. Bai, J. Z. Kolter, and V. Koltun. (2018). "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling." [Online]. Available: https://arxiv.org/abs/1803.01271
[36] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
[37] A. H. Dau.(2018). *The UCR Time Series Classification Archive*. [Online]. Available: https://www.cs.ucr.edu/ eamonn/time_series_data_2018/



QIUQUN DENG was born in 1991. She received the M.S. degree from the College of Electronic Science, National University of Defense Technology, Changsha, China, in 2015, where she is currently pursuing the Ph.D. degree with the National Key Laboratory of Science and Technology on ATR.

Her research interests include optical guidance, target recognition, and machine learning.

HUANZHANG LU was born in 1963. He received the M.S. and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 1990 and 1994, respectively, where he is currently a Professor with the National Key Laboratory of Science and Technology on ATR.

His research interests include automatic target recognition, precision, and real-time systems.

HUAMIN TAO was born in 1972. She received the M.S. degree from the National University of Defense Technology, Changsha, China, in 1997, where she is currently an Associate Professor with the National Key Laboratory of Science and Technology on ATR.

Her research interests include signal processing and real-time system design.

MOUFA HU was born in 1979. He received the M.S. and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 2003 and 2008, respectively, where he is currently an Associate Researcher with the National Key Laboratory of Science and Technology on ATR.

His research interests include image process and real-time process systems.

FEI ZHAO was born in 1983. He received the M.S. and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 2007 and 2012, respectively, where he is currently a Lecturer with the National Key Laboratory of Science and Technology on ATR.

His research interests include machine learning and real-time systems.