

Deep Multimodel Cascade Method Based on CNN and Random Forest for Pharmaceutical Particle Detection

Hui Zhang^{ID}, Member, IEEE, Miao Zhao^{ID}, Li Liu^{ID}, Hang Zhong^{ID}, Zhicong Liang^{ID}, Yimin Yang^{ID}, Senior Member, IEEE, Xianen Zhou^{ID}, Q. M. Jonathan Wu^{ID}, Senior Member, IEEE, and Yaonan Wang^{ID}

Abstract—The quality detection of pharmaceutical liquid products is inevitable and crucial in drug manufacture because drugs contaminated with foreign particles are definitely not to be used. However, with the current detection methods, it is still a challenge to detect and identify the small moving particles using an imaging system. In this article, a deep multimodel cascade method combining single-frame image and multiframe images processing method to detect and identify foreign particles is proposed. The proposed method consists of three stages. First, a Faster R-CNN convolutional neural network is adopted to detect and localize the multiple suspected foreign particles of each single-frame image. Then, the k -means clustering algorithm is used to cluster the trail of that detected multiple suspected foreign particles in the eight sequential images to obtain the moving object trajectory. Finally, trajectory features are extracted and the random forest (RF) classifier is used to distinguish noises and foreign particles according to the motion feature of the moving object trajectory. Experimental results demonstrate that the proposed multitask stepwise method improves the accuracy of foreign particles detection and reduces the rate of omission in the case of strong noise, which proves the effectiveness of this method.

Index Terms—Faster R-CNN, k -means, moving foreign particles visual detection, pharmaceutical detection, random forest (RF), trajectory space feature vector.

Manuscript received June 14, 2019; revised November 18, 2019; accepted January 28, 2020. Date of publication February 13, 2020; date of current version August 11, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61971071 and Grant 61701047, in part by the National Key Research and Development Program of China under Grant 2018YFB1308200, in part by the Hunan Key Laboratory of Intelligent Robot Technology in Electronic Manufacturing under Grant IRT2018009, in part by the Hunan Key Project of Research and Development Plan under Grant 2018GK2022, and in part by the Changsha Science and Technology Project under Grant kq1907087. The Associate Editor coordinating the review process was Shutao Li. (*Corresponding authors:* Hui Zhang; Li Liu.)

Hui Zhang is with the School of Robotics, Hunan University, Changsha 410082, China, and also with the College of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410012, China (e-mail: zhanghuiby@126.com).

Miao Zhao and Zhicong Liang are with the College of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410012, China (e-mail: zhao114@qq.com; liangzhicongcslg@163.com).

Li Liu, Hang Zhong, Xianen Zhou, and Yaonan Wang are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: liuli@hnu.edu.cn; zhonghang@hnu.edu.cn; zhouxianen@hnu.edu.cn; yaonan@hnu.edu.cn).

Yimin Yang is with the Department of Computer Science, Lakehead University, Thunder Bay, ON P7B5E1, Canada (e-mail: yyang48@lakeheadu.ca).

Q. M. Jonathan Wu is with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B3P4, Canada (e-mail: jwu@uwindsor.ca).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIM.2020.2973843

I. INTRODUCTION

THE detection of foreign matters in liquid drugs is an important guarantee for the safety and reliability of products. Due to various factors in the production process, a variety of visible foreign particles will be mixed into the pharmaceutical liquid products. The patients who use the drugs that were contaminated with foreign particles may cause allergic reactions and even death [1]. However, low contrast ratio and dynamic change of foreign particles are great challenges to the existing detection methods.

The detection technology of foreign particles in pharmaceutical products has been developed in recent decades. Earlier, the detection methods include the manual visual detection system (VDS), semiautomatic detection [2], and so on. However, such detection methods are inefficient and subjective, which may cause second contamination. The methods for detecting foreign matters in pharmaceutical products using nondestructive testing techniques (NDTs) mainly include the visual cameras [3], optoelectronic method [4], and so on. In recent years, visual detection [5], [6] technology is developing rapidly and has been applied in a number of industrial detections. For example, visual detection is commonly used for detection in fabric, semiconductor wafers, railheads [7], steel plate defect inspection [8], concrete materials crack detection [9], and so on. However, due to the low contrast, low energy, random noise, deformation, and other adverse factors of tiny foreign particles, some traditional machine vision detection methods are difficult to detect effectively.

Currently, due to the advantages of high efficiency and reliability, computer vision based on the convolutional neural network (CNN) technology has been gradually applied to multiple defect detection. Kang *et al.* [10] designed an insulator surface defect detection system using a deep CNN that consists of two stages, that is, a Faster R-CNN network for object detection and the deep multitask neural network for defect detection. Chang and Wang [11] proposed to apply the CNN for detecting the bubble defect, and it can detect the bubble defects and reduce the false alarm. Gibert *et al.* [12] proposed to multitask neural network architecture for the material classification and railway track elements' detection. Liu *et al.* [13] used a Faster R-CNN network to localize the isoelectric line and achieved pretty good results. In [14], a fine-tuned CNN was employed to extract part-level features of the cropped images

for rail images' classification. In [15], CNN was proposed to detect crack patches and false positives (FPs) were effectively discarded through naive Bayes decisions. In [16], the R-CNN model is used to locate the vehicle and the corresponding parts, and the corresponding features are aggregated to train the SVM classifier. In addition, the application of CNN defect detection also includes some road cracks and concrete cracks [17], [18]. Although there are almost no relevant methods of deep learning directly used for the detection of foreign particles in pharmaceutical liquid products. Inspired by those works, deep-learning technology can also be used in the detection of small foreign particles in pharmaceutical liquid products. In fact, the speed/accuracy tradeoffs of different convolutional object detectors are different [19]. Considering the difficulty of small object detection, a Faster R-CNN [20] network is used in this article to detect and localize suspected liquid particles.

Actually, the strong noise makes it difficult to determine whether the detected object is a foreign particle or not. Hence, it is an important step to use the features of the moving object trajectory to ensure the correct distinction between foreign particles and noise. In this article, in order to obtain the trajectory of the moving object, we considered the moving object trajectory clustering algorithms [21]. As a matter of fact, not only are the motion modes of different foreign particles significantly different [22] but the trajectories of foreign particles and noises are also quite different [23]. Clustering is to group data into clusters, making the data in one group more similar than those of other groups [21], [24]. Han *et al.* [25] divided the clustering algorithm of point data into the five main categories. Partition-based methods are relatively simple, and only one parameter k ($k \leq n$, n is the number of data points in the data set) is required to set the number of the final partition of the data [21]. The representative algorithms include k -means and k -medoids. The core idea of k -means is to randomly find k clustering centers and then iterate repeatedly according to the within-cluster sum of squares to divide all the points until the clustering center point no longer changes or meets the set conditions [26]. However, the randomized initial clustering center is not only slower, but the final clustering result is very likely to be affected by the outlier and cause errors. Therefore, initializing the cluster center in k -means++ [27] can alleviate that effect.

After obtaining the moving object trajectories, this article proposes a classification method based on the trajectory features. The trajectories are composed of a series of points arranged in chronological order. Abundant knowledge and information can be obtained from the research and analysis of trajectories. In [28], through mining the vehicle trajectories information, not only its state and movement characteristics can be obtained but also abnormality detection can be completed. Trajectory mining is an important embodiment of its application in urban transportation assistance [29] and route recommendation [30]. Trajectory data classification is an important part of trajectory mining and has been applied in many aspects. Krumm and Horvitz [31] detected whether the target device is moving or stationary by using 802.11 infinite electrical signals. Sohn *et al.* [32] classified the motion of the

target into three states: stationary, walking, and driving based on the trajectory data. Zhu *et al.* [33] inferred whether the taxi carries passengers through the GPS trajectory of the taxi. Lin *et al.* [34] proposed a hierarchical inference model to achieve target recognition and important location discovery through target trajectory classification. It is difficult to distinguish the trajectories formed by foreign particles directly. Also, there are missing values and outlier, which may affect the correct classification of the trajectory. Therefore, it is considered to extract the features of the trajectories and then classify them. However, it is difficult to determine the relationship between features when classifying according to the features. Fortunately, the decision tree algorithm [35] can complete classification based on features without determining the relationship between the features. However, single decision tree classification is not robust enough, and when the learnable lower bound of the decision tree (base learner) is known, an ensemble classifier with a certain accuracy can be obtained by a large number of ensemble base learner [36]. Krogh and Vedelsby [37] proposed that the generalization error of ensemble learning systems is not only related to the generalization error of base learner but also to the degree of correlation of individual base learner. Based on the earlier analysis, we used random forest (RF) [38] as the classifier for trajectory classification. RF is a comprehensive classifier system composed of multiple random decision trees, and the classification results implementation combines classifiers by averaging their probabilistic prediction.

In this article, a new solution to the detection of foreign particles in pharmaceutical liquid products is provided. The main contributions of this article are as follows.

- 1) CNN is used to detect suspected foreign particles without classification so that the effect and robustness of detection results reach a high level.
- 2) Features used for trajectory classification are constructed so that the object classification is transferred to data domain processing.
- 3) RF ensemble learning algorithm is used to classify trajectories in the feature space to avoid the influence of point trace missing values and the complex analysis of the relationships between the features.

This article is organized as follows. Section II overviews the foreign particles detection system and data set analysis. The detection and localization of suspected foreign particles, clustering analysis, and classification of trajectory are described in Section III. The experimental results and analysis are summarized in Section IV. In Section V, a reasonable discussion is necessary. Finally, Section VI presents the conclusion and future work.

II. SYSTEM OVERVIEW

A. System Configuration

Particles lie on the bottom of the container or float in the liquid. They cannot be distinguished from dust or scratches on the outside. Therefore, in order to achieve the automatic detection of foreign particles in the liquid, it is necessary to make them rotate. A piece of special equipment is designed to

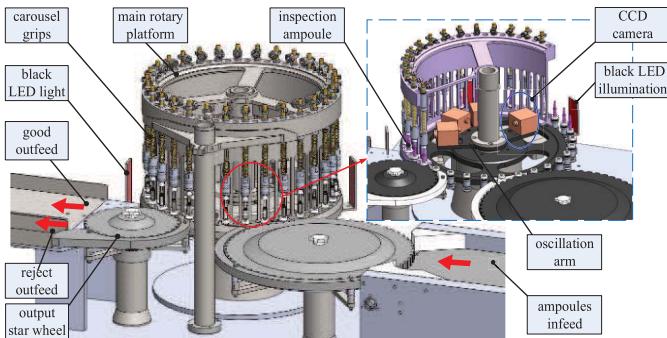


Fig. 1. Prototype of pharmaceutical detection system.

simulate the process of manual visual detection. The containers to be detected enter the detection platform from the infeed channel, and the carousel grips in the detection platform clamp the mouth of the container; then, the containers are driven by the servo rotary mechanism in the main rotary platform to make the centrifugal rotation. In this way, the liquid inside the container, including the particles, starts to rotate. After containers stop rotating for a period of time and the liquid level is relatively stable, the CCD camera immediately takes the images of the corresponding containers to obtain a series of images. Taking photographs after a period of time is to eliminate bubbles as much as possible. After acquiring the required sequence images, the machine classifies the products well according to the visual detection results. The qualified drugs can be successfully exported from a good outfeed output channel, while the defective ones are exported from reject outfeed. The overall structure also includes some other parts, and its structure is shown in Fig. 1.

B. Pharmaceutical Images Data Set

One of the concerns of this article is the accurate detection of small foreign particles in the liquid. Due to the influence of randomness of multitype foreign particles movements, the relative positional change with CCD camera, and the influence of reflection, the appearance features of the foreign particles have various uncertainties; in addition, the adverse factors, such as low contrast and low energy of the foreign particles, will also affect the detection and recognition effect. The moving objects detected in the liquid are roughly divided into two categories: one is foreign particles such as glass, rubber, hair, and fiber, and the other is the random noise caused by the surface defect of the glass bottle. Fig. 2(a) shows the partial foreign particles and interference bubbles. The experiment uses eight sequential images as a group to detect. Fig. 2(b) shows the actual image, and the image detection area is cropped based on the moving region of detection objects. All samples are RGB images of 770×600 and totally 100 groups of 800 samples.

C. Framework of the Proposed Detection and Classification Algorithm

As shown in Fig. 3, this article proposes a deep multimodel cascade learning method for foreign particles' visual detection in a strong noise environment. Deep multimodel cascading learning refers to the object detection model Faster R-CNN for suspected foreign particles detection in stage A. Also, data

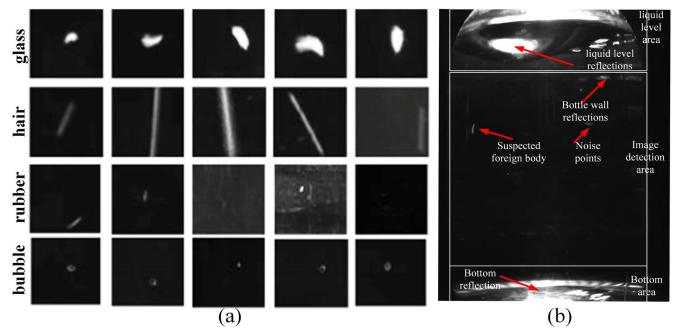


Fig. 2. (a) Typical defects of different foreign particles. (b) Actual image. The objects in (a) are the enlarged images.

clustering analysis acquires the moving object trajectories in stage B. In stage C, noise and foreign particles are classified according to the trained RF classifier. In our cascade model, the supervised learning of the method satisfies the detection of foreign particles under the interference of illumination inequality, mutual occlusion, shape change, and so on, and achieves a good recall rate. Then, the trajectory obtained by k-means clustering makes the spatiotemporal features of the foreign particles be vectorized. Finally, an RF classifier is used to distinguish the foreign particles and noises in feature space, which is fast and avoids complex analysis.

III. DETECTION AND CLASSIFICATION METHODS

A. Detection and Location Model Faster R-CNN

As discussed previously, we first need to detect and locate suspected foreign particles. In order to reduce the missed detection of foreign particles in a complex environment, the foreign particles and random noise will be detected together by Faster R-CNN [20], [39]. The Faster R-CNN consists of two CNNs: the regional proposal network (RPN) and the Fast R-CNN [40] network, as shown in part A of Fig. 3. These two deep neural networks share Vgg16 [41] feature extraction layers, which improve the detection efficiency.

The RPN can input images of any size and directly generate region proposals and their object scoring that may contain foreign particles to be detected. The RPN slides through a 3×3 spatial sliding window on the feature map of the shared convolutional layer while predicting k ($k = 9$) anchors of three scales and three aspect ratios at each sliding position. For a convolutional feature map of size $W \times H$, there is a total of $W \times H \times k$ anchors. Then, according to the score of each anchor from high to low, the top 300 anchors are selected as the region proposals and are fed to the crop and resize the pooling layer [39]. At the same time, the feature maps of the shared convolutional layer will also be fed to the crop and resize the pooling layer. The crop and resize operator, which crops and resizes feature maps to 14×14 , and then max-pools them to 7×7 to match the input size of fcs [39]. Finally, the softmax probabilities and bounding-box regression offsets are output by the subsequent fully connected layers. After training, the two networks form a unified Faster

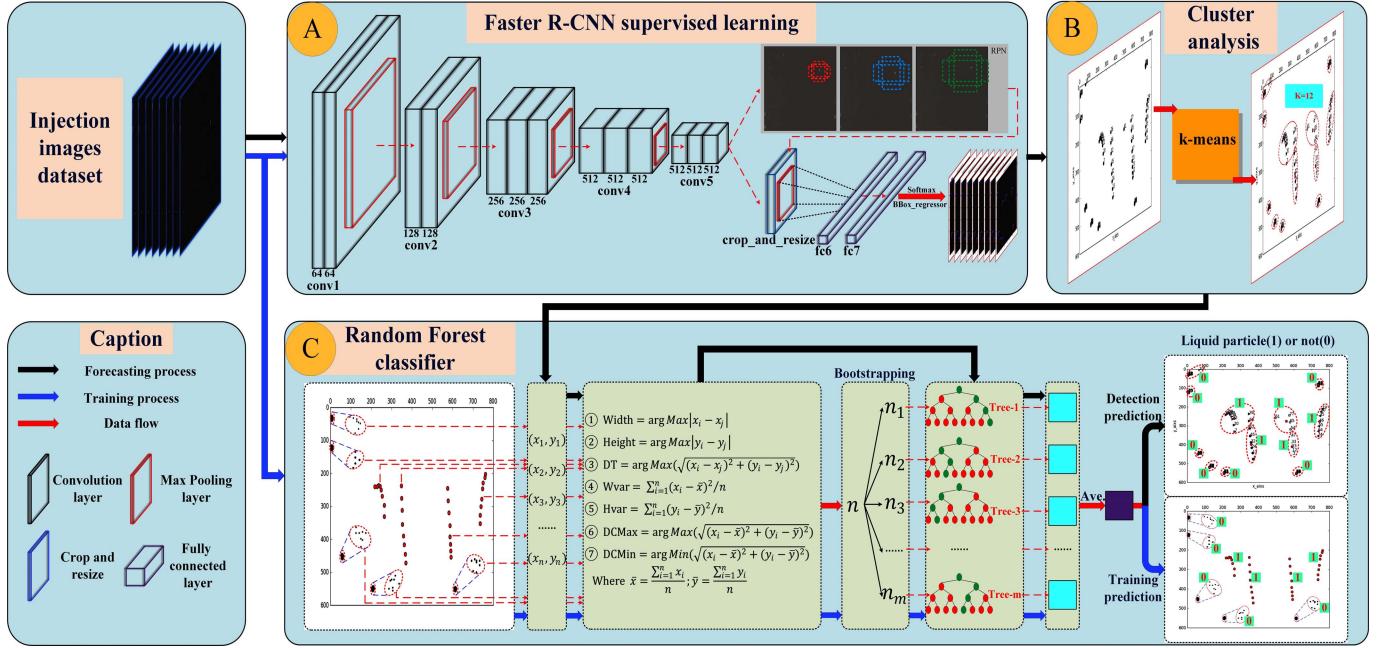


Fig. 3. Overview of the proposed particles detection and classification algorithm. A–C: stages A–C, respectively. The details are in Section III.

R-CNN network that is used to detect and localize the suspected foreign particles.

B. Cluster Processing k-Means

After detecting and locating suspected foreign particles in the complete eight sequential images based on Faster R-CNN, in order to classify suspected foreign particles according to the spatiotemporal features, this article uses a k -means algorithm that selects the optimal k -value based on the within-cluster sum of squared error (SSE) and the silhouette coefficient (SC) to obtain the trajectories quickly and accurately. In fact, the suspected foreign particles detected by the Faster R-CNN object detection algorithm include two types, one is a visible foreign particle that has a regular range of motion, and the other is random noise with irregular motion in a small range. Therefore, in order to eliminate the interference of random noise on foreign particles detection, it is necessary to design a good classification algorithm.

1) *K-Means Clustering Algorithm*: K -means is a traditional clustering method based on feature similarity in the original data set, which belongs to the unsupervised learning method. The trajectory of the foreign particles has continuous smoothness, this article proposes to use k -means to cluster the detected suspected foreign particles to obtain its movement trajectories. Toward the detected location points $D = \{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n\}$, where $\mathcal{X}_1 = (x_1, y_1)$, $\mathcal{X}_2 = (x_2, y_2), \dots, \mathcal{X}_n = (x_n, y_n)$. k -means minimizes the SSE for cluster partitioning $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$.

$$\text{SSE} = \sum_{i=1}^k \sum_{\mathcal{X} \in C_i} \|\mathcal{X} - \mu_i\|_2^2 \quad (1)$$

where $\mu_i = (1/|C_i|) \sum_{\mathcal{X} \in C_i} \mathcal{X}$ is the mean vector of cluster C_i . In fact, SSE is not a standardized metric. In general,

we think that the value of SSE should be as low as possible so that zero is the best, but when SSE is equal to zero, k is generally greater than the real clustering number. To do this, we need to use a certain method to comprehensively measure.

2) *SSE and SC*: In order to obtain the value of the parameter k more accurately, the SC is calculated while calculating the SSE. In data mining and analysis, it is generally believed that as the number of clusters k increases, subsets will be divided more precisely and the degree of aggregation of each cluster will gradually increase so that SSE will naturally decrease gradually. When k is less than the number of real clusters, the increase of k will significantly increase the degree of aggregation of each cluster, therefore the SSE will decrease greatly. However, when k reaches the real number of clusters and then increases k , the growth rate of aggregation degree of each cluster will decrease rapidly, thus the decrease rate of SSE will decrease sharply and finally, the decrease rate of SSE will become gentle with the continuous increase of k . Considering the ground truth labels are not known when we do cluster analysis, so we must use the model itself to evaluate k and select better k values. Based on this, the SC is adopted here to evaluate the clustering model with different parameters k . The SC for a single sample is defined as

$$\text{SC} = \frac{b - a}{\max(a, b)} = \begin{cases} 1 - \frac{a}{b}, & a < b \\ 0, & a = b \\ \frac{b}{a} - 1, & a > b. \end{cases} \quad (2)$$

The SC is defined for each sample and is composed of two scores. Where a is the mean distance between a sample and all other points in the same cluster. The smaller a is, the more samples should be clustered into this cluster, and a is called the intracluster dissimilarity of samples. And b is the mean distance between a sample and all other points in the next

nearest cluster. Define b as the intercluster unfamiliarity of the sample. The larger b is, the less the sample belongs to other clusters. It can be seen from (2) that the SC ranges between $[-1, 1]$. SC close to 1 indicates that the sample clustering is reasonable, and SC close to -1 indicates that the sample should be classified into another cluster. The SC around zero indicates overlapping clusters. It is worth noting that the SC for a set of samples is given as the mean of the SC for each point. In general, when choosing the parameter k , this article considers SSE and SC as effective metrics of its clustering model under different k .

3) *K-Means++ Initialization*: In order to converge as quickly as possible and avoid local minima, initialization of cluster centers using k -means++ [27]. This will cause the initialization of the cluster centers to stay away from each other and, generally, get more reliable results than random initialization. In fact, k -means++ [27] randomly selects the first cluster center $C_1(\mathcal{X}_c)$ and then calculates the shortest distance between each point and the current cluster center C_1

$$D(\mathcal{X}_j) = \sqrt{\arg \min (\|\mathcal{X}_c - \mathcal{X}_j\|_2^2)} \quad (3)$$

where \mathcal{X}_j is a point in the data. In order to select the next cluster center, select the point corresponding to the maximum value according to the calculation probability

$$P(\mathcal{X}_j) = \arg \max \left(\frac{D(\mathcal{X}_j)^2}{\sum_{\mathcal{X} \in D} D(\mathcal{X})^2} \right) \quad (4)$$

where $\mathcal{X} \in D$ is the whole data set. According to (3) and (4), the calculation is repeated until K initial cluster centers are selected. The subsequent calculations are consistent with normal k -means.

C. Feature Extraction and RF

1) *Feature Extraction*: In order to overcome the problem that foreign particles morphological features are difficult to extract, this article distinguishes the random noise and foreign particles according to the spatial features of motion trajectories. We define the motion trajectory feature vector of a moving object as

$$\mathcal{V} = [\text{Width}, \text{Height}, \text{Distance},$$

$$\text{Wvar}, \text{Hvar}, \text{DCMax}, \text{DCMin}]. \quad (5)$$

The attribute elements of “ \mathcal{V} ” are defined as follows:

$$\text{Width} = \arg \max |x_i - x_j| \quad (6)$$

$$\text{Height} = \arg \max |y_i - y_j| \quad (7)$$

$$\text{Distance} = \arg \max \left(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \right) \quad (8)$$

$$\text{Wvar} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (9)$$

$$\text{Hvar} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} \quad (10)$$

$$\text{DCMax} = \arg \max \left(\sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \right) \quad (11)$$

$$\text{DCMin} = \arg \min \left(\sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \right) \quad (12)$$

where $\bar{x} = (\sum_{i=1}^n x_i)/n$, $\bar{y} = (\sum_{i=1}^n y_i)/n$. x_i and y_i are the horizontal and vertical coordinates of the moving object point traces. The “ n ” is the number of points in the trajectory. “Width” represents the maximum distance of the trajectory point abscissa. “Height” represents the maximum distance of the trajectory point ordinate. “Distance” represents the maximum distance between the trajectory points. “Wvar” represents the variance of the abscissa of the trajectory point. “Hvar” represents the variance of the ordinate of the trajectory point. “DCMax” and “DCMin” represent the maximum and minimum distances between the trajectory point and the mean point (\bar{x}, \bar{y}) , respectively.

2) *Trajectory RF Classifier*: The training for trajectory classification is mainly divided into three steps, as shown in stage C of Fig. 3. First, the construction of the sample set is to extract the point traces from the labeled object of the training data set. Manually label and use them as positive and negative samples, respectively. It should be noted that the label used in this article is positive sample “1” and negative sample “0,” respectively, representing foreign matters and noise. The second is feature extraction, which extracts the feature parameters of the object from the trajectory according to (5). Finally, the RF classifier model is trained by the feature vector of the sample. Specifically, after obtaining the training sample set, the sample set after features extraction according to formula (5) can be expressed as

$$D = \{\mathcal{V}_i, p_i\} \quad i = 1, 2, \dots, L \quad (13)$$

where \mathcal{V}_i is the eigenvector calculated from the i th trajectory according to (5). p_i is the label of the i th trajectory, the positive sample is 1, and the negative sample is 0. L is the number of samples. Due to the small data set and difficulty in effectively dividing the training and test sets, RF uses bootstrapping to sample the training set. According to bootstrapping, several sampling sets $D^t = \{\mathcal{V}^t, p^t\} (t = 1, 2, \dots, m; l = 1, 2, \dots, L)$ can be obtained, where D^t is the t th bootstrap sample and m is the number of bootstrap samples. Each bootstrap sample $S = D^t$ will be used to train a decision tree (base learner). It is worth mentioning that the features we extracted are continuous variables, which were discretized using bipartition before being used in decision trees. For any attribute, $a \in A$, and a has L different values in S . Sort these values from smallest to largest as $\{a^1, a^2, \dots, a^L\}$. Thus, a candidate partition point set $T_a = \{(a^i + a^{i+1})/2 | 1 \leq i \leq L-1\}$ of $L-1$ elements is obtained. For any partition point of $t \in T_a$, S is divided into subsets S_t^- and S_t^+ , where S_t^- contains samples whose value is not greater than t on attribute a and S_t^+ contains those samples whose value is greater than t on attribute a . When selecting the best attribute for splitting, there are two main ways to split the decision tree. The information gain is defined as follows:

$$\begin{aligned} \text{Gain}(S, a) &= \max_{t \in T_a} \text{Gain}(S, a, t) \\ &= \max_{t \in T_a} \text{Ent}(S) - \sum_{\lambda \in \{-, +\}} \frac{|S_t^\lambda|}{|S|} \text{Ent}(S_t^\lambda) \end{aligned} \quad (14)$$

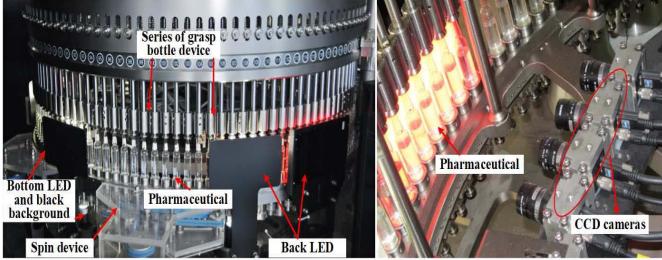


Fig. 4. Pharmaceutical VDS based on multicameras.

where

$$\text{Ent}(S) = - \sum_{k=1}^2 p_k \log_2 p_k \quad (15)$$

where p_k is the proportion of the k th sample in S . Equation (15) is actually information entropy, and the smaller the value of $\text{Ent}(S)$, the higher the purity of S . S_t^λ is a sample when S is divided according to the partition point t , and $|S_t^\lambda|, |S|$ represents the number of samples. Another partitioning is made according to the Gini index, using the same symbol as before

$$\begin{aligned} \text{Gini}(S, a) &= \min_{t \in T_a} \text{Gini}(S, a, t) \\ &= \min_{t \in T_a} \sum_{\lambda \in \{-, +\}} \frac{|S_t^\lambda|}{|S|} \text{Gini}(S_t^\lambda) \end{aligned} \quad (16)$$

where

$$\text{Gini}(s) = 1 - \sum_{k=1}^2 p_k^2. \quad (17)$$

It is worth mentioning that in order to further increase the diversity of the base learner, the RF adds random attribute selection processing in the decision tree optimal attribute selection stage.

IV. EXPERIMENT AND ANALYSIS

In actual production, a pharmaceutical VDS is shown in Fig. 4. The VDS mainly consisted of several vision stations with PointGrey cameras. The cameras are placed on the motion platform. The detection vision systems are mounted on an oscillation arm to capture the image sequence continuously. The experimental environment is as follows: Ubuntu16.04, Python 3.5, deep-learning framework Tensorflow, other algorithms framework Scikit-learn [42], Intel Core i7-6700, and GTX 1080 GPU with 8-GB memory.

A. Suspected Foreign Particles Detection by Faster R-CNN

In this section, we evaluate the performance of the detection and localizing algorithm based on the Faster R-CNN. The pharmaceutical images data set was randomly divided into two parts: 520 images for training and 280 images for testing. The data set includes the images with only foreign particles, images with only noise, and images with both foreign particles and noise.

TABLE I
MAP AND AVERAGE RECALLS (AVE-REC) TEST RESULTS
AT DIFFERENT OVERLAP THRESHOLDS

Mnemonic	Test thresholds	mAP	ave-Rec
L1	0.01	0.535	0.826
L2	0.1	0.754	0.874
L3	0.2	0.803	0.890
L4	0.3	0.801	0.890
L5	0.4	0.777	0.891
L6	0.5	0.686	0.856
L7	0.6	0.489	0.793
L8	0.7	0.400	0.723
L9	0.8	0.279	0.652
L10	0.9	0.051	0.375

To make Faster R-CNN equipped with good capacity on our data set, the truth is that we want Faster R-CNN to detect as many suspected objects as possible, that is, the average recall rate is maximized under the guaranteed mean average precision (mAP). Overlap threshold of regions of interest (ROIs) and ground-truth boxes (gt-boxes) relates to the division of the foreground (object) and background, which will affect the detection of small objects, and nonmaximum suppression (NMS) of RPN proposals will affect the recall rate of detection. The performance with different overlap threshold of ROIs and ground-truth boxes (gt-boxes) and NMS of RPN proposals will be discussed next.

1) *Test on Overlap Threshold of ROI and Gt-Boxes:* In this experiment, we design ten overlap thresholds (range from L1 to L10) in the process of model training, with a learning rate of 0.001, the training step of 20k (20000), and the batch size of 128. The data format is the standard format of VOC2007. The data partitioning is that the training set and the validation set together account for 65% of the total data set, and the test set takes up 35%.

Table I lists the test results under different overlap thresholds. It can be seen from Table I that when the overlap threshold is lower than 0.4, the average recall rate of the detection model gradually increases as the overlap threshold increases. When the overlap threshold is 0.2, 0.3, and 0.4, the average recall rate is basically equal and reaches the highest range of the test. When the overlap threshold is greater than 0.4, the average recall rate decreases significantly with the increase in the overlap threshold and finally drops to 0.375. The mAP reaches a maximum of 0.803 when the overlap threshold is 0.2. However, when the overlap threshold is greater than 0.2, the mAP decreases significantly from 0.4 as the overlap threshold increases.

Specifically, a box plot of recall is shown in Fig. 5(a) for the overall explanation. It needs to be noted that recall here is the continuous variable from zero to the maximum obtained by the VOC07 metric in the test set, so recall here is not used for outlier analysis. In the range of L1–L5, as the overlap threshold increases, the rectangle gradually becomes shorter, and both the median line and the mean increase, which indicates that the detection model from L1 to L5 has been improving, and at L3–L5, the corresponding better experimental results are achieved. Fig. 5(b) shows a close view of the L3–L5, and the mean and rectangle of L3–L5 are similar to each other, but the median line of L5 is closer to the upper

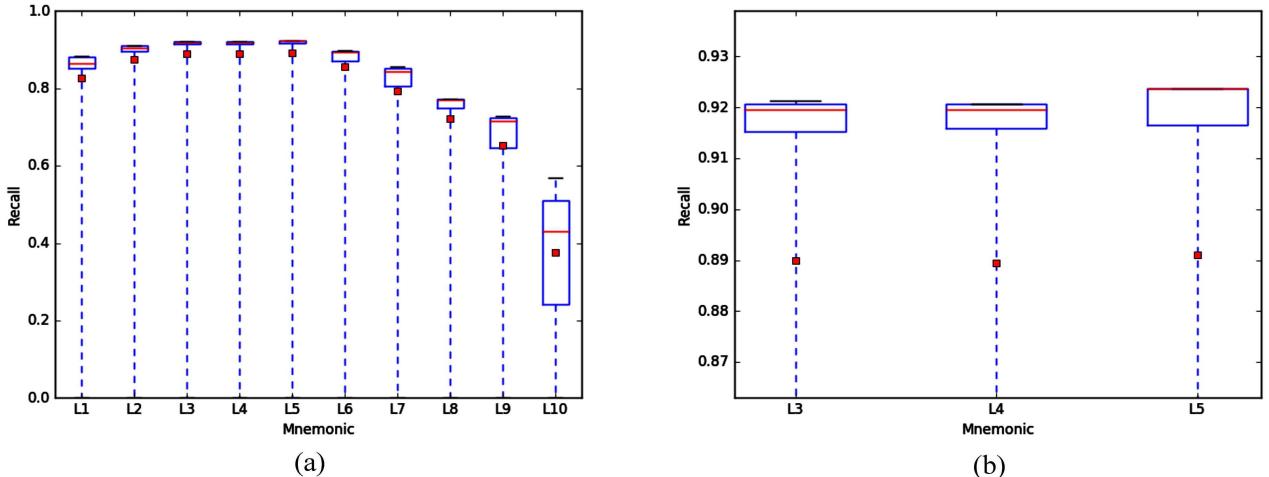


Fig. 5. (a) Box plot of recall at test threshold: L1–L10 refer to overlap thresholds, namely 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. (b) Close view of the L3–L5 in (a).

TABLE II
MAP AND AVERAGE RECALLS (AVE-REC) TEST RESULTS AT DIFFERENT NMS THRESHOLDS

Mnemonic	NMS thresholds	mAP	ave-Rec
R1	0.01	0.712	0.881
R2	0.1	0.755	0.886
R3	0.2	0.763	0.883
R4	0.3	0.791	0.891
R5	0.4	0.788	0.882
R6	0.5	0.806	0.894
R7	0.6	0.815	0.886
R8	0.7	0.812	0.896
R9	0.8	0.804	0.877
R10	0.9	0.812	0.888

quartile, indicating that recall corresponding to the threshold of L5 is better. However, the mAP of L5 is 2.6% and 2.4% lower than L3 and L4, respectively. From L6 to L10, as the overlap threshold increases, the rectangle becomes longer, and the median line and the mean decrease, indicating that the recall in the model is deteriorating. At L10, the maximum value of recall does not exceed 0.6, indicating that the model has been difficult to effectively detect the objects.

To conclude, the mAP of L3 is slightly higher than L4, and the foreground object or background is divided by 0.2 as the threshold during training.

2) *Test on NMS Threshold Used on RPN Proposals:* In this experiment, we also design ten NMS thresholds (range from R1 to R10) used on RPN proposals in the process of model training, with a learning rate of 0.001, the training step of 20k (20 000), and the batch size of 128. Also, the overlap threshold of ROI and gt-boxes is 0.2. The data set is the same as previously described.

Table II lists the test results under different NMS thresholds. It can be seen from Table II that when the NMS threshold increases, the average recall rate of the detection model fluctuates. A maximum of 0.896 was reached when the NMS threshold was equal to 0.7. In fact, it can be observed from Table II that the average recall rate fluctuates less during the increase in NMS, even if the difference between the minimum

and maximum values is only 1.9%. The same mAP change does not have global monotonicity, reaching a maximum of 0.815 when the NMS threshold is equal to 0.6.

Similarly, a box plot of recall is shown in Fig. 6(a) for the overall analysis. It needs to be noted that recall here is the continuous variable from zero to the maximum obtained by the VOC07 metric in the test set, so recall here is not used for outlier analysis. As can be seen from Fig. 6(a), R1–R10 have a slight fluctuation in the rectangular length, the median line, and the mean as the NMS threshold increases, which indicates that the recall of the detection model from R1 to R10 is variable. Fig. 6(b) shows a close view of R1–R10. It can be observed in Fig. 6(b) that R8 not only has the largest mean but its median line is closest to the upper quartile. Comparing R8 with R7, although the mAP of R8 is 0.3% lower than R7, the average recall rate is 1% higher. Also, it can be clearly seen from Fig. 6(b) that R8 is better than R7.

To conclude, the NMS threshold is more appropriate considering R8; therefore, 0.7 is used as an NMS threshold used on RPN proposals during training.

B. Obtain Trajectories by Clustering Point Traces

In this section, we test the performance of the k -means that got the trajectories. In the experiment, we directly used position trace data that have been obtained by the Faster R-CNN. In this article, the detected trace data of the suspected foreign particles are the center of the detected bounding boxes.

Some experimental results are shown in Fig. 7. Fig. 7(a1) shows the 56 point data detected by the corresponding eight-frame sequence images through Faster R-CNN. Corresponding to Fig. 7(a1), the results of evaluating different values of k are shown as in Fig. 7(a2). Obviously, when $k = 7$, the SC reaches a maximum of 0.974, and the SSE has also dropped to a minimum range. Therefore, the clustering result of Fig. 7(a1) based on $k = 7$ is shown as in Fig. 7(a3). Similarly, a total of 74 suspected foreign particles are detected in Fig. 7(b1), and according to Fig. 7(b2), when $k = 11$, the SC obtains a maximum value of 0.922 and the SSE is also in a small range. According to $k = 11$, the corresponding clustering result is

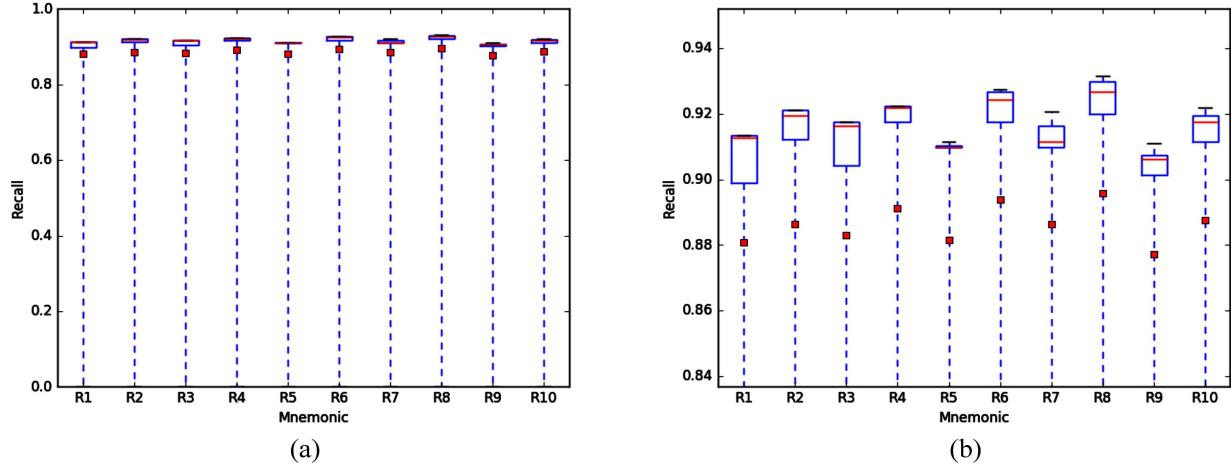


Fig. 6. (a) Box plot of recall at test threshold: R1–R10 refer to NMS thresholds, namely 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. (b) Close view of (a).

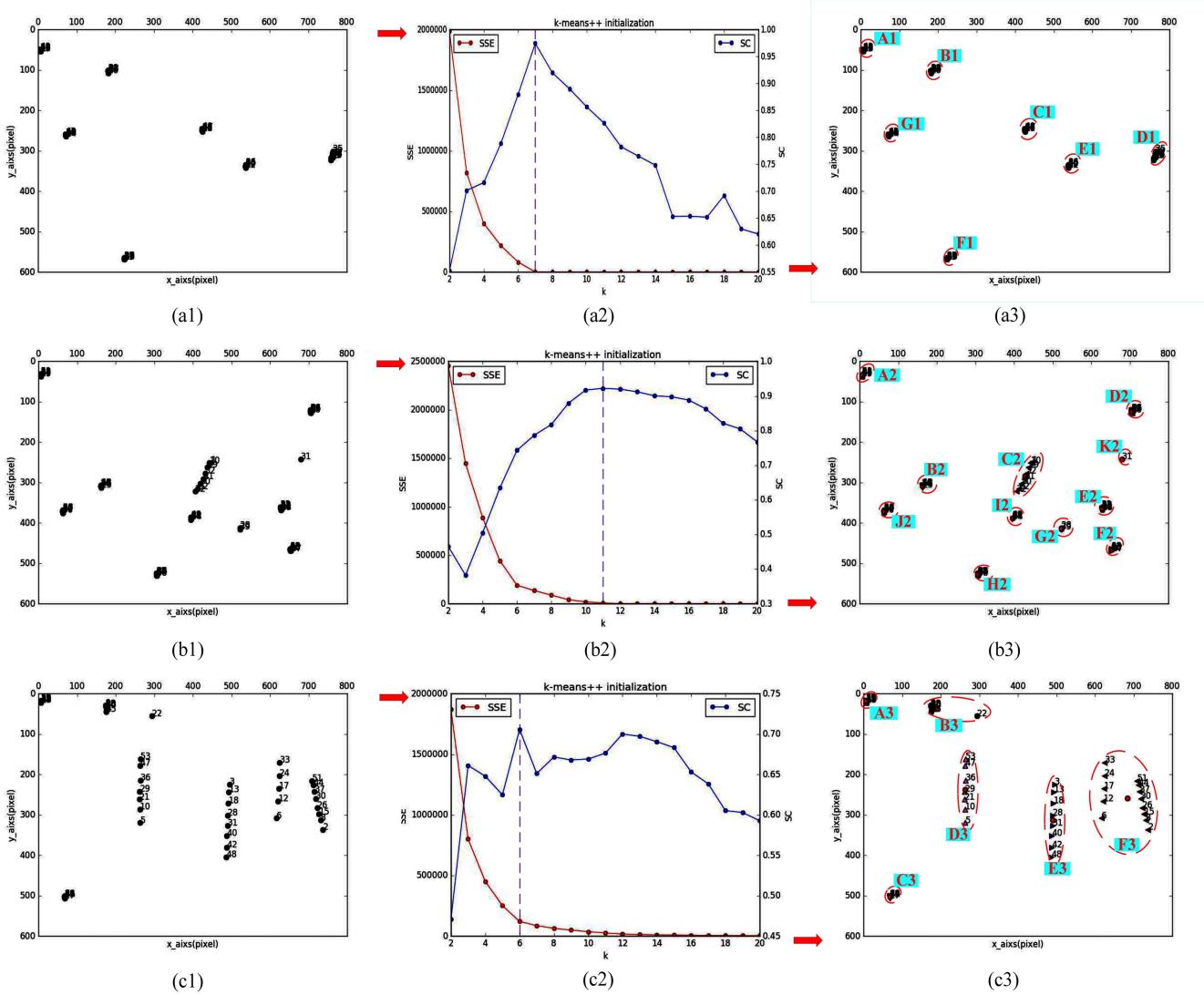


Fig. 7. Cluster analysis. (a1), (b1), and (c1) Initial data obtained by the Faster R-CNN. (a2), (b2), and (c2) Broken line statistic chart of SSE and SC under different k values. (a3), (b3), and (c3) Corresponding clustering result based on the selected k [(a3): $k = 7$; (b3): $k = 11$; (c3): $k = 6$].

shown as in Fig. 7(b3). Finally, in Fig. 7(c1), the SC obtains a maximum value of 0.706 when $k = 6$ in Fig. 7(c2), and the

SSE is also basically in a small range. The clustering results are shown in Fig. 7(c3).

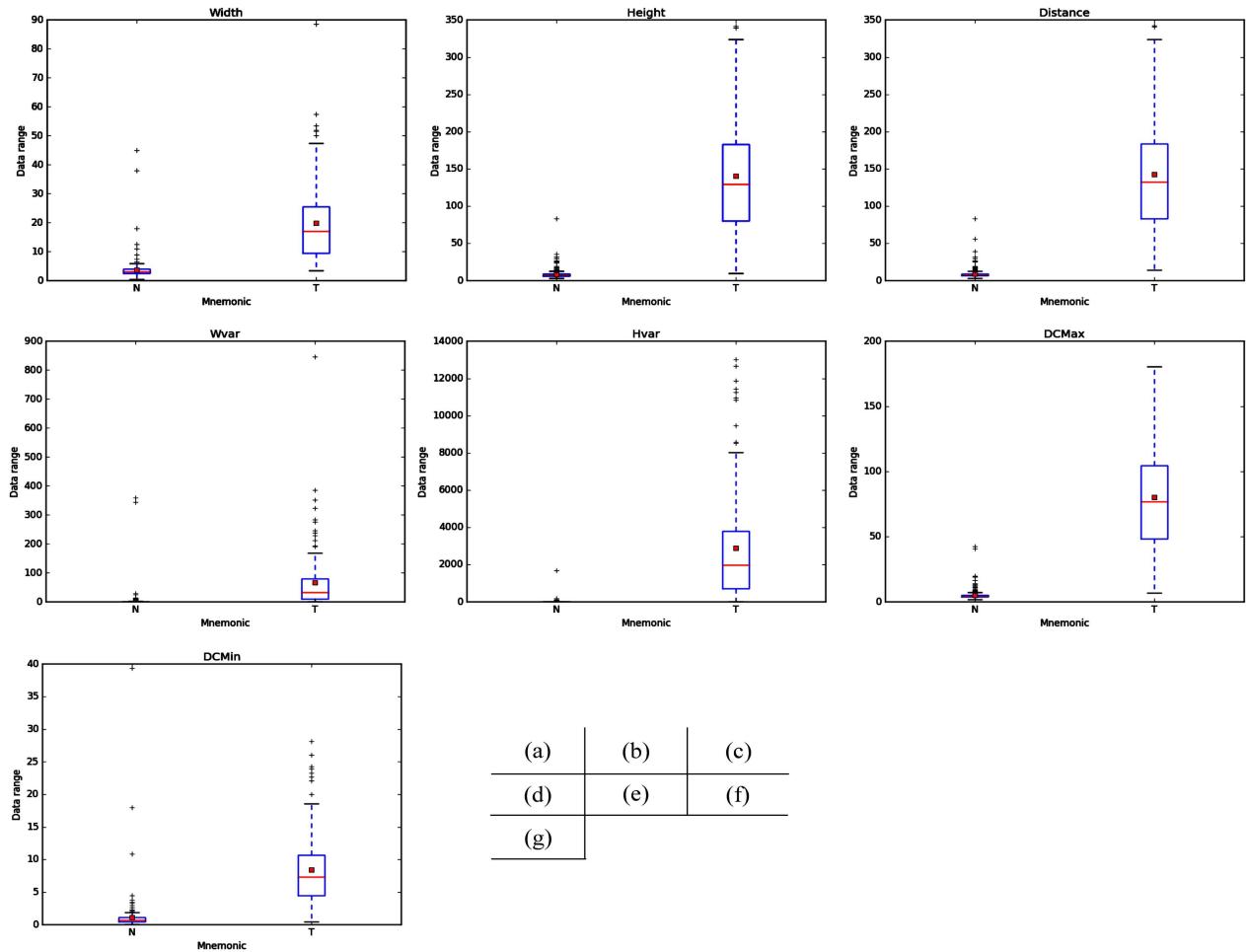


Fig. 8. Analysis of feature attribute distribution of noise and foreign matter, where N and T represent random noise and foreign bodies, respectively. (a) “Width.” (b) “Height.” (c) “Distance.” (d) “Wvar.” (e) “Hvar.” (f) “DCMax.” (g) “DCMin.”

C. Trajectory Feature Extraction and Training Classifier

According to the clustering, the trajectories are obtained, and their respective eigenvectors are calculated according to formula (5). However, here, we first verify the feature vectors that we constructed on the training data set. In this experiment, the training data set of the classifier has been manually classified and annotated. We obtained 426 trajectories based on a total of 520 training set images of 65 groups.

1) *Features Analysis:* The feature box plot of the training data is shown in Fig. 8. For “Width,” the maximum value of the distance between the abscissas of the points in the random noise trajectories is mostly small, and only a few outliers reach a large value. However, the distance between the abscissas of the points in the trajectories of the foreign particles is generally larger. “Height” represents the maximum distance of the trajectory point ordinate. The “height” of random noises is mostly small and the “height” of foreign matters is still generally larger. In terms of “distance,” the value of the noise trajectories is smaller and the value of the foreign matters trajectories is larger, which shows that the random noise has a small range of motion, while the range of motion of the foreign matters is larger. For “Wvar” and “Hvar,” random noise has a small value due to small-range motion, while

the foreign matter has a large range of motion, resulting in a large value. Moreover, the “Hvar” value of the foreign matters is larger than that of the “Wvar” value, which indicates that the range of motion of the foreign matter in the longitudinal direction is large. For “DCMax” and “DCMin,” it is still that the value of the noise trajectory is small and the value of the foreign matter trajectory is large. It is worth mentioning that the “DCMax” of the noise is even smaller than most of the “DCMin” of the foreign matter. Finally, in the random noise, each element of the eigenvector is in a small range of values, and the close range of values between the elements also indicates that the motion of the random noise does not have obvious directionality. The elements of the eigenvector in the foreign matter trajectory have a large range of values. Compared with the vertical and horizontal axes, most of the values of “Height” are greater than the “Width” and “Hvar” is larger than “Wvar,” which indicates that the motion of the foreign matter has directionality and its direction of motion is substantially parallel to the longitudinal axis. In general, random noise has a large difference in the foreign matter on the trajectory eigenvector that we construct.

2) *Training Classifier:* The 426 trajectories obtained are used as the training data set of RF classifier. We manually label the foreign matter in the training data as a positive

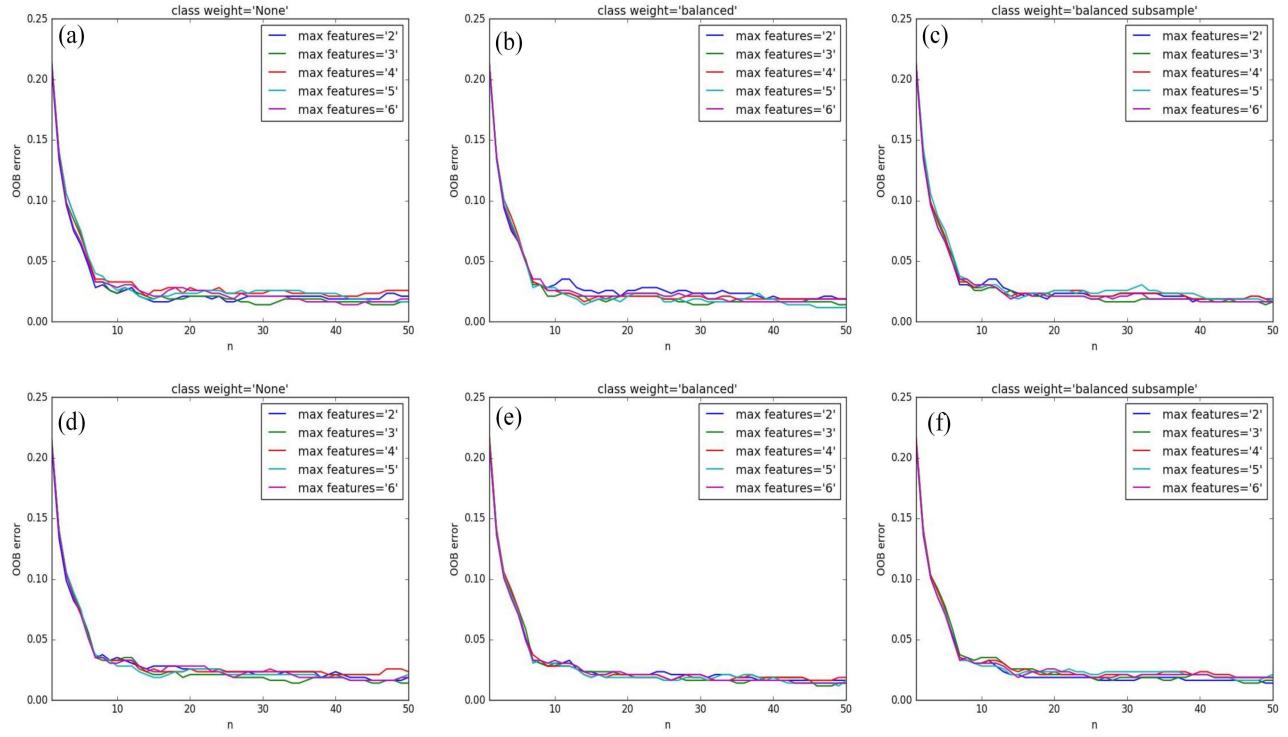


Fig. 9. RF parameter analysis. n represents the number of base learns, with values from 0 to 50, and OOB error represents the error score of the training data set obtained using an out-of-bag estimate. (a)–(c) Gini index. (d)–(f) Information gain.

sample with label “1” and the noise as a negative sample with label “0.” Each trajectory classification attribute is calculated according to (5). In our training data set, the negative samples are 288 and the positive samples are 138. The positive and negative sample ratio is approximately 0.5. The total amount of data is small and the imbalance between positive and negative samples. Since the RF training uses a bootstrapping method, there is no need to divide the data set during training. In fact, each base learner trains only about 63.2% of the original data, and the remaining 36.8% of the samples are used as a validation set to estimate the generalization performance (“out-of-bag estimate,” oob). The RF generalization oob error ratio is evaluated as

$$\epsilon^{\text{oob}} = \frac{1}{|D|} \sum_{(x,y) \in D} \mathbb{I}(H^{\text{oob}}(x) \neq y) \quad (18)$$

$$\text{where } H^{\text{oob}}(x) = \arg \max_{y \in Y} \sum_{t=1}^T \mathbb{I}(h_t(x) = y) \cdot \mathbb{I}(x \notin D_t). \quad (19)$$

D is the original data set, where (x, y) is the feature attribute and the corresponding label. D_t represents the training sample set actually used by the h_t base learner, and $H^{\text{oob}}(x)$ represents the “out-of-bag” prediction of sample x . Y is a collection of y .

RF Parameter Analysis: Some key parameters have a significant impact on the performance of RF. The number of base learners in the RF, the splitting function to measure the quality of a split, the influence of the max features to consider when looking for the best split, and the class weight on the classification model are analyzed here. In the experiment,

TABLE III
COMPARISON OF CLASSIFICATION METHODS

method	accuracy	F1	Precision	Recall
N=1(Ordinary decision tree)	92.86%	86.96%	83.33%	90.91%
N=1(bootstrap sampling)	92.86%	88%	78.57%	100%
N=30	97.62%	95.65%	91.67%	100%

TABLE IV
PERFORMANCE COMPARISON OF THE PROPOSED METHOD AND OTHER DETECTION METHODS

	SRC	HJSRC	ID-ELM	FRCRF	ALW-CSM
Miss rate	3.77%	2.25%	5.59%	1.86%	—
Error rate	4.63%	2.90%	5.33%	2.38%	2.75%
Consumed time(s)	10.17	3.02	3.18	3.87	>1.45

because our data set is small, we designed the number of decision tree base learner from 1 to 50 to experiment. The difference in the max features as an attribute disturbance in the classification model is an important factor in reducing the similarity between the base learners. Considering that there are fewer attributes in our experiment, we test the different max features of $k = \{2, 3, 4, 5, 6\}$ directly. Fig. 9 shows the trend of oob (“out-of-bag estimate”) error ratio changes for different split criterion, max feature, and class weight as the base learner changes continuously from 1 to 50. The class weight of “none” means that two classes are supposed to have weight one. The “balanced” mode uses the values of the label to adjust weights inversely proportional to class frequencies in the input data. The “balanced subsample” mode is the same as “balanced” except that weights are computed based on the bootstrap sample for every decision tree grown [42]. The splitting function of the decision tree of Fig. 9(a)–(c) is the

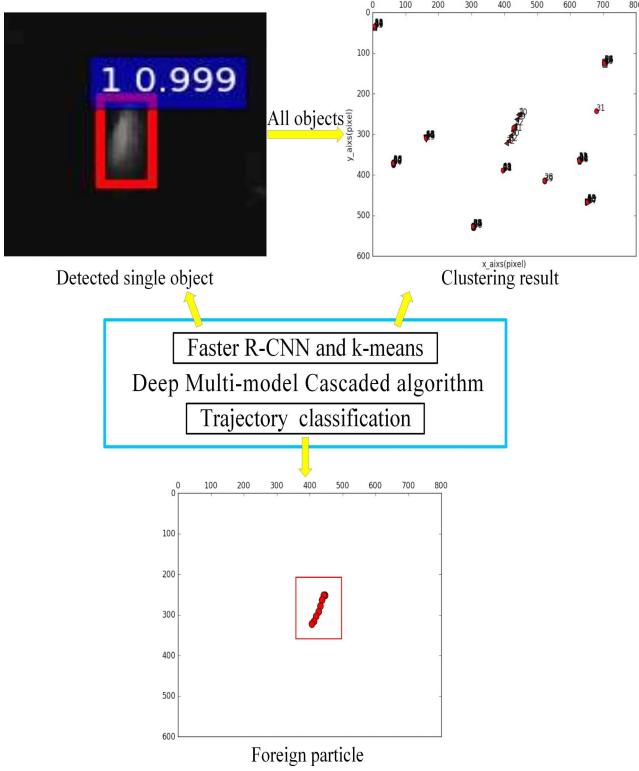


Fig. 10. Detection and classification of our method. The detection results only demonstrate a single object, while the clustering and classification results are actually the results of eight sequential images.

Gini index, and the splitting function of the decision tree of Fig. 1(d)–(f) is the information gain. As the number of base learners increases, the oob error ratio decreases significantly and eventually converges to less than 5%. When the splitting function is information gain, the curve converges more gradually and has less fluctuation than the Gini index. Also, considering the sufficient convergence and time consuming, the split function of the decision tree selected in this article is information gain, the number of decision trees is 30, and the maximum feature number is determined to be 3. To increase the difference between base learners (decision tree), the class weight is balanced subsample mode. Finally, based on the earlier parameters, the average score of the training data set using an out-of-bag estimation was about 98%.

3) RF Trajectory Classification Analysis: In this section, we validate the performance of the classifier on the test data set. The test data set is obtained by clustering after the Faster R-CNN detection. We tested the three classifications shown in Table III, where N is the number of decision trees or base learners, and the recorded data are averages calculated after counting 100 times. The specific calculation is as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (20)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (21)$$

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (22)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (23)$$

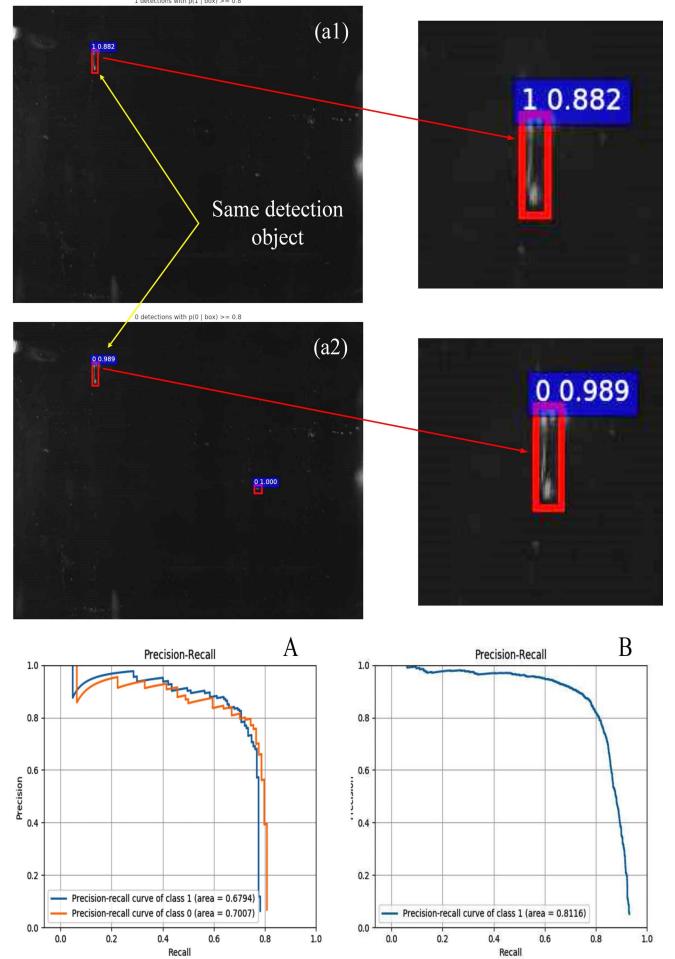


Fig. 11. Comparison of detection results. A is a classified P-R curve, and B is not classified. “1” represents foreign matter label, and “0” represents noise label.

where true positive (TP), true negative (TN), FP, and false negative (FN) correspond to the count of defective unqualified, defect-free qualified, defective qualified, and defect-free unqualified products, respectively. It can be observed from Table III that the data of RF ($n = 30$) is optimal. In fact, RF ($n = 30$) is more robust than ordinary decision trees and a single base learner.

V. DISCUSSION

In this section, we first describe the overall detection classification method and then discuss why a single classification of suspected foreign matter is used in the detection, rather than two classifications of noise and foreign matter. It is mainly analyzed from the results of the experiment, which is easier to understand intuitively. Finally, the experimental results of the whole method are analyzed.

A. Detection and Classification of Our Method

As shown in Fig. 10, the Faster R-CNN is first used to detect these objects. Faster R-CNN detects both foreign particles and random noise as suspected objects with the

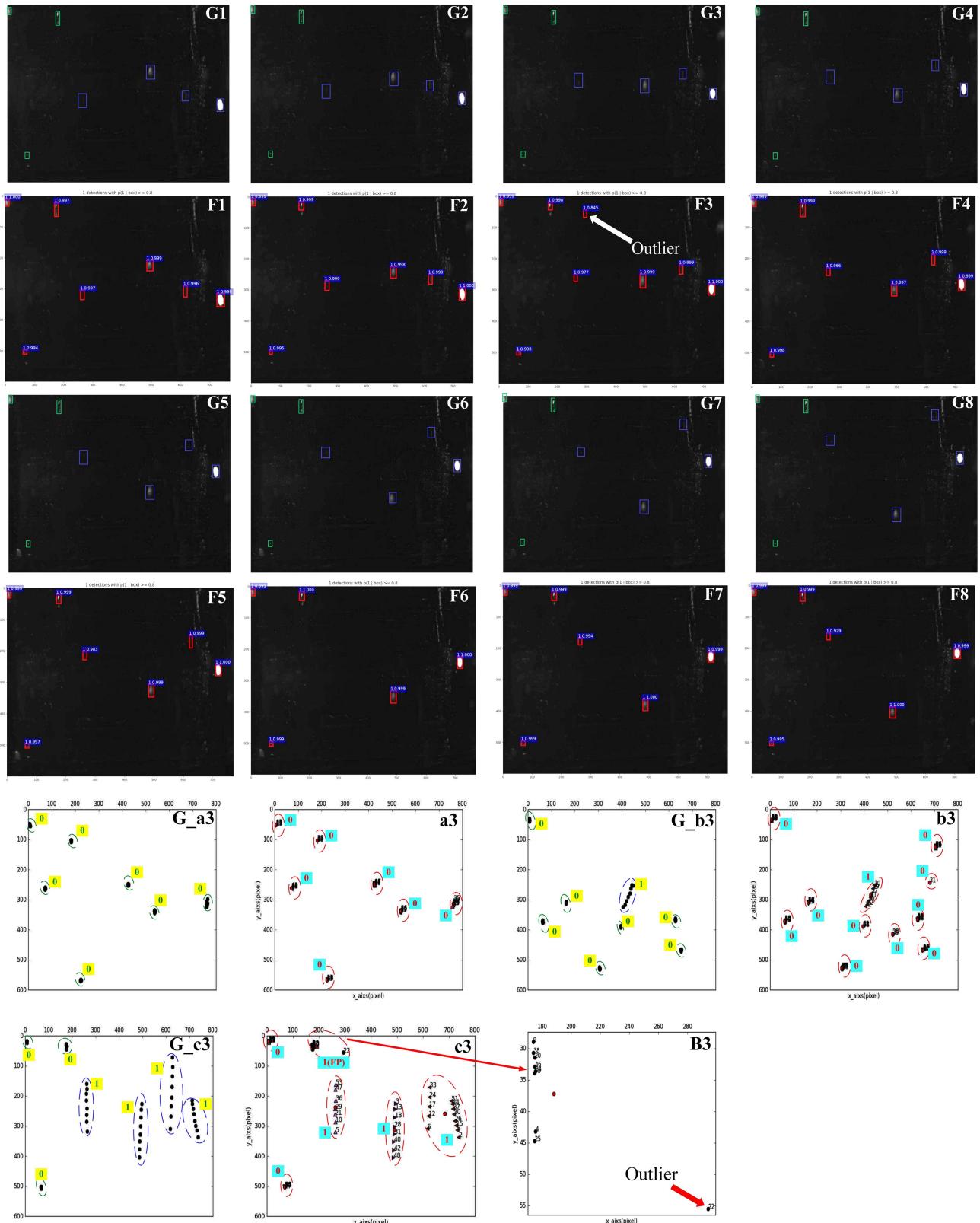


Fig. 12. G1–G8: original images and its ground truth. Green box represents noise, and blue box represents the real foreign matter. F1–F8: detection results corresponding to G1–G8. (G_a3), (G_b3), and (G_c3) original trajectories and its ground truth. (a3), (b3), and (c3) Classification results. “1” represents the real foreign matter label, and “0” represents the noise label.

label of “1” and then clusters all the detected objects to get the trajectory. Finally, the eigenvectors of each trajectory are

calculated and then accurately classified according to the RF classifier.

B. Comparison of Experimental Results

Fig. 11(a1) and (a2) shows the classification detection results of two consecutive frames of images. The detection results of the same detection object in Fig. 11(a1) and (a2) are foreign matter and noise, respectively, that is to say, due to the similarity between foreign matter and noise, the detection results will be completely different.

In addition, by comparing the results of the two classification methods, the detection result of the classification detection is actually more missed, and it is also reflected from the P-R curve A. The detection results of unclassified detection are relatively better and the P-R curve B is also superior to A, and the mAP is also different. The mAP of the foreign matter is about 0.679 and the mAP of the noise is about 0.701 when classified, but the mAP is about 0.812 when not classified. In general, due to inconsistent detection results and more missed detection objects, classification detection is difficult to accurately and effectively complete foreign matter detection.

C. Overall Analysis of Detection and Classification Results

Fig. 12 shows the detection results of a set of sequence images. Fig. 12(a3), (b3), and (c3) shows the classification results corresponding to the clustering results in Fig. 7, and the detection result of F1–F8 is also corresponding to the results in Fig. 7(c3). The classification results shown in Fig. 12(a3) and (b3) are accurate, but it can be observed in the detection result of F1–F8, which not only has extra false detection but also has missed detection in F6–F8. In the classification result shown in Fig. 12(c3), although two trajectories are divided into one cluster, the correct classification is still not affected. However, in Fig. 12(c3), there is an FP classification result compared to the ground truth of (G_c3). The FP point trace is specifically shown in Fig. 12(B3), and its classification result is deviated due to the occurrence of outlier 22. It is worth mentioning that although there is a missed detection, due to the trajectory classification, this can reduce the impact of missed detection.

In addition to the proposed method, we have also compared with some existing pharmaceutical liquid particles detection algorithms on the same data set. The results are compared according to the average miss rate, the average error rate, and the consumed time of each group of eight sequence images

$$\text{Miss rate} = 1 - \frac{N_{\text{det}}}{N_{\text{GT}}} \quad (24)$$

$$\text{Error rate} = 1 - \text{accuracy} \quad (25)$$

where N_{GT} is the number of ground truth, N_{det} is the number of objects detected in the N_{GT} range ($N_{\text{det}} \leq N_{\text{GT}}$), and the accuracy is calculated by (20).

- 1) *Sparse Representation Classifier (SRC)*: This method uses the block principal component tracking algorithm to detect the moving object and then classifies it according to the SRC [43].
- 2) *Hierarchical Joint Sparse Representation (HJSRC)*: Improved hierarchical joint SRC algorithm based on SRC.

- 3) *ID-ELM*: Pruning based on the influence degree for extreme learning machine [44], [45]. This method first performs the original image filtering and foreign matter segmentation, and then, the classification is performed by the constructed ID-ELM classifier according to the selected features.
- 4) *FRCRF*: The proposed method described in Section III.
- 5) *ALW-CSM*: The automatic inspection system with multicarousel grips, adaptive local weighted-collaborative sparse model tracking algorithm [22].

It can be observed in Table IV that the proposed method is the most accurate. Compared with HJSRC, SRC, and ID-ELM algorithms, the missed detection rate is decreased by 0.39%, 1.91%, and 3.73%, respectively, and the error detection rate is decreased by 0.52%, 2.25%, and 2.95%, respectively. Compared with our proposed FRCRF algorithm, the error rate of ALW-CSM is 0.37% higher and the consumed time(s) of ALW-CSM is >1.45 s higher. Although our method is slightly longer than some algorithms in terms of consumed time, it is basically on the same level.

VI. CONCLUSION

This article presented a method to detect pharmaceutical foreign matter. The proposed method consists of a three-stage architecture, and first, the suspected foreign matter in the eight sequential images is detected and located by the Faster R-CNN algorithm. In order to utilize the spatiotemporal information of eight sequential images, the k -means is used to cluster the sequence of suspected object points to obtain the corresponding trajectories. Since it is difficult to distinguish noise and foreign matter effectively from the morphological characteristics, this article uses the trajectory and extracts the features from it to classify. Therefore, the image object classification problem is transferred to the data domain for processing. Finally, the RF is used to classify according to the feature vector. We thoroughly analyzed the parameters of the RF and obtained the parameters that fit our data set by comparison. The experimental results show that the proposed method has a better performance than other detection methods and its detection efficiency has reached the general requirements. Thus, the proposed approach can effectively and efficiently perform the accurate detection of pharmaceutical foreign matter. Further research focuses on improving the design of classifiers, considering the introduction of feature extraction networks for the classification of sequence-structure data.

REFERENCES

- [1] S. E. Langille, "Particulate matter in injectable drug products," *J. Pharmaceutical Sci. Technol.*, vol. 67, no. 3, pp. 186–200, May 2013.
- [2] H. Grindinger, H. Neusser, N. Seidenader, and V. Wedershoven, "Product testing apparatus." U.S. Patent Appl. 10/928433, Jun. 2, 2005. [Online]. Available: <https://www.google.com/patents/US20050117149>
- [3] Y. Wang, J. Ge, H. Zhang, and B. Zhou, "Intelligent injection liquid particle inspection machine based on two-dimensional Tsallis Entropy with modified pulse-coupled neural networks," *Eng. Appl. Artif. Intell.*, vol. 24, no. 4, pp. 625–637, Jun. 2011.
- [4] T. Taoa, H. Mei-Zhen, L. Ming-Liang, J. Jun-Bob, L. Tian-Yuan, and L. Pei-Huab, "Signal processing methods for detecting impurities in ampoules," *J. Shanghai Jiaotong Univ.*, vol. 47, no. 6, pp. 998–1002, 2013.

- [5] H. Dong, K. Song, Y. He, J. Xu, Y. Yan, and Q. Meng, "PGA-Net: Pyramid feature fusion and global context attention network for automated surface defect detection," *IEEE Trans. Ind. Informat.*, to be published.
- [6] Z. Zhou, Y. Wang, Q. M. J. Wu, C.-N. Yang, and X. Sun, "Effective and efficient global context verification for image copy detection," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 1, pp. 48–63, Jan. 2017.
- [7] Q. Li and S. Ren, "A real-time visual inspection system for discrete surface defects of rail heads," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2189–2199, Aug. 2012.
- [8] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Trans. Instrum. Meas.*, to be published.
- [9] D. Zhang, Q. Li, Y. Chen, M. Cao, L. He, and B. Zhang, "An efficient and reliable coarse-to-fine approach for asphalt pavement crack detection," *Image Vis. Comput.*, vol. 57, pp. 130–146, Jan. 2017.
- [10] G. Kang, S. Gao, L. Yu, and D. Zhang, "Deep architecture for high-speed railway insulator surface defect detection: Denoising autoencoder with multitask learning," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2679–2690, Aug. 2019.
- [11] C.-Y. Chang and W.-C. Wang, "Integration of CNN and faster R-CNN for tire bubble defects detection," in *Proc. 13th Int. Conf. Broadband Wireless Comput., Commun. Appl.* Taichung, Taiwan: Springer, 2018, pp. 285–294.
- [12] X. Gibert, V. M. Patel, and R. Chellappa, "Deep multitask learning for railway track inspection," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 153–164, Jan. 2017.
- [13] Z. Liu, L. Wang, C. Li, and Z. Han, "A high-precision loose strands diagnosis approach for isoelectric line in high-speed railway," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1067–1077, Mar. 2018.
- [14] L. Shang, Q. Yang, J. Wang, S. Li, and W. Lei, "Detection of rail surface defects based on CNN image recognition and classification," in *Proc. 20th Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2018, pp. 45–51.
- [15] F.-C. Chen and M. R. Jahanshahi, "NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4392–4400, Oct. 2017.
- [16] K. Huang and B. Zhang, "Fine-grained vehicle recognition by deep convolutional neural network," in *Proc. 9th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2016, pp. 465–470.
- [17] L. Zhang, F. Yang, Y. Daniel Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3708–3712.
- [18] Y.-J. Cha, W. Choi, and O. Büyükoztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 5, pp. 361–378, May 2017.
- [19] J. Huang et al., "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7310–7311.
- [20] 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.
- [21] G. Yuan, P. Sun, J. Zhao, D. Li, and C. Wang, "A review of moving object trajectory clustering algorithms," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 123–144, Jan. 2017.
- [22] H. Zhang et al., "Automated machine vision system for liquid particle inspection of pharmaceutical injection," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 6, pp. 1278–1297, Jun. 2018.
- [23] M. Zhao, H. Zhang, L. Liu, Z. Liang, and G. Deng, "Joint deep learning and clustering algorithm for liquid particle detection of pharmaceutical injection," in *Proc. 8th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Nov. 2018, pp. 1–6.
- [24] T. W. Liao, "Clustering of time series data—A survey," *Pattern Recognit.*, vol. 38, no. 11, pp. 1857–1874, 2005.
- [25] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier, 2011.
- [26] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-means clustering algorithm," *Appl. Statist.*, vol. 28, no. 1, p. 100, 1979.
- [27] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.
- [28] Z. Feng and Y. Zhu, "A survey on trajectory data mining: Techniques and applications," *IEEE Access*, vol. 4, pp. 2056–2067, 2016.
- [29] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "T-Drive: Enhancing driving directions with taxi drivers' intelligence," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 220–232, Jan. 2013.
- [30] Y. Zheng, X. Xie, and W.-Y. Ma, "GeoLife: A collaborative social networking service among user, location and trajectory," *IEEE Data Eng. Bull.*, vol. 33, no. 2, pp. 32–39, Jun. 2010.
- [31] J. Krumm and E. Horvitz, "LOCADIO: Inferring motion and location from Wi-Fi signal strengths," in *Proc. MobiQuitous*, 2004, pp. 4–13.
- [32] T. Sohn et al., "Mobility detection using everyday GSM traces," in *Proc. 8th Int. Conf. Ubiquitous Comput.* Berlin, Germany: Springer, 2006, pp. 212–224.
- [33] Y. Zhu, Y. Zheng, L. Zhang, D. Santani, X. Xie, and Q. Yang, "Inferring taxi status using GPS trajectories," 2012, *arXiv:1205.4378*. [Online]. Available: <http://arxiv.org/abs/1205.4378>
- [34] L. Liao, D. J. Patterson, D. Fox, and H. Kautz, "Learning and inferring transportation routines," *Artif. Intell.*, vol. 171, nos. 5–6, pp. 311–331, Apr. 2007.
- [35] L. Breiman, *Classification and Regression Trees*. Evanston, IL, USA: Routledge, 2017.
- [36] R. E. Schapire, "The strength of weak learnability," *Mach. Learn.*, vol. 5, no. 2, pp. 197–227, 1990.
- [37] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 1995, pp. 231–238.
- [38] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [39] X. Chen and A. Gupta, "An implementation of faster RCNN with study for region sampling," 2017, *arXiv:1702.02138*. [Online]. Available: <http://arxiv.org/abs/1702.02138>
- [40] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [41] D. Frossard. *VGG in TensorFlow*. Accessed: 2017. [Online]. Available: <https://www.cs.toronto.edu/frossard/post/vgg16>
- [42] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
- [43] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Apr. 2010.
- [44] Y. Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, and A. Lendasse, "OP-ELM: Optimally pruned extreme learning machine," *IEEE Trans. Neural Netw.*, vol. 21, no. 1, pp. 158–162, Jan. 2010.
- [45] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.



Hui Zhang (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in pattern recognition and intelligent system from Hunan University, Changsha, China, in 2004, 2007, and 2012, respectively.

He is currently a Professor with the School of Robotics, Hunan University, Changsha, China. He is also a Professor with the College of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha, China. His research interests include machine vision, deep learning, defect detection.



Miao Zhao was born in Yueyang, China, in 1994. He received the B.S. degree in automation from the Central South University of Forestry and Technology, Changsha, China, in 2017. He is currently pursuing the master's degree with the College of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha.

His current research interests include machine learning and deep learning.



Li Liu was born in 1984. He received the Ph.D. degree from Hunan University, Changsha, China, in 2019.

He holds a post-doctoral position at the College of Electrical and Information Engineering, Hunan University. His current research interests include robot vision measurement, robot path planning, and intelligent control.



Hang Zhong received the B.S. and M.S. degrees in automation from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree in control theory and application with Hunan University.

His current research interests include robotics modeling and control, visual servo control and path planning of the aerial robots.



Xianen Zhou received the B.S. degree in electronic information engineering from the Hunan University of Science and Engineering, Yongzhou, China, in 2010, the M.S. degree in circuits and systems from the East China University of Technology, Nanchang, China, in 2013, and the Ph.D. degree in pattern recognition and intelligent system from Hunan University, Changsha, China, in 2019.

He is currently a Post-Doctoral Fellow with Hunan University. His current research interests include machine vision, machine learning, and automatic visual detection systems.



Zhicong Liang was born in Zhongshan, China, in 1994. He received the B.S. degree in automation from the Changsha University of Science and Technology, Changsha, China, in 2017, where he is currently pursuing the master's degree with the College of Electrical and Information Engineering.

His current research interests include image processing, deep learning, and industry inspection.

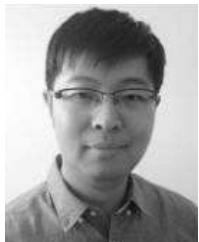


Q. M. Jonathan Wu (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Wales, Swansea, U.K., in 1990.

He has been with the National Research Council of Canada, Ottawa, ON, Canada, for ten years since 1995, where he became a Senior Research Officer and a Group Leader. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. He has authored over 300 peer-reviewed

articles in computer vision, image processing, intelligent systems, robotics, and integrated microsystems. His current research interests include 3-D computer vision, active video object tracking and extraction, interactive multimedia, sensor analysis and fusion, and visual sensor networks.

Dr. Wu is a fellow of the Canadian Academy of Engineering. He is the Tier 1 Canada Research Chair in automotive sensors and information systems. He has served on technical program committee and international advisory committee for many prestigious conferences. He is also an Associate Editor of the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON CIRCUIT AND SYSTEM FOR VIDEO TECHNOLOGY, *Cognitive Computation*, and *Neurocomputing*.



Yimin Yang (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Hunan University, Changsha, China, in 2013.

From 2014 to 2018, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. He is currently an Assistant Professor with the Computer Science Department, Lakehead University, Thunder Bay, ON, Canada. He has authored or coauthored more than 50 refereed articles. His research interests are in artificial neural networks, machine learning, and robotics.

Dr. Yang was a recipient of the Outstanding Ph.D. Thesis Award of Hunan Province and the Outstanding Ph.D. Thesis Award Nominations of Chinese Association of Automation, China, in 2014 and 2015, respectively. He has been serving as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, a reviewer for international journals of his research field, and a program committee member of some international conferences.



Yaonan Wang received the Ph.D. degree in electrical engineering from Hunan University, Changsha, China, in 1994.

Since 1995, he has been a Professor with the College of Electrical and Information Engineering, Hunan University. From 1994 to 1995, he was a Post-Doctoral Research Fellow with the Normal University of Defense Technology, Changsha. From 1998 to 2000, he was supported as a Senior Humboldt Fellow by the Federal Republic of Germany at the University of Bremen, Bremen, Germany. From 2001 to 2004, he was a Visiting Professor with the University of Bremen. His research interests include robotics and image processing.

Dr. Wang is a fellow of the Chinese Academy of Engineering.