

Classification of Ground Targets Based on Radar Micro-Doppler Signatures Using Deep Learning and Conventional Supervised Learning Methods

Peibei CAO, Weijie XIA, Yi LI

Key Laboratory of Radar Imaging and Microwave Photonics (Nanjing Univ. Aeronaut. Astronaut.), Ministry of Education, College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, 29 Jiangjun Street, Nanjing 211100, China

caopeibei@nuaa.edu.cn, nuaaxwj@nuaa.edu.cn, liyi1995@nuaa.edu.cn

Submitted January 31, 2018 / Accepted May 30, 2018

Abstract. Radar has great potential in military and civilian areas, including automobile anti-collision, battlefield surveillance, etc., due to its high penetration and all-weather capability. On the basis of traditional targets detection, targets classification can be realized. In this paper, a comparison of targets classification between deep learning (Deep Convolutional Neural Networks (DCNNs)) and conventional supervised learning methods (Support Vector Machine (SVM), Naive Bayes (NB) and SVM-Bayes fusion algorithm) has been made. Furthermore, several factors affecting the accuracy of classifying targets including SNR, decrease of samples, have been researched and discussed. We employ a K-band Doppler radar to acquire the raw signal due to its stationary clutter-rejection, movement detection ability and short wavelength. Then Short-time Fourier Transform (STFT) is applied to the raw signal to characterize micro-Doppler signatures which is the fundament of the classification process. We adopt the DCNNs to deal with the spectrograms directly, while features have been designed and extracted for classification with conventional supervised learning methods. It is shown that the DCNN can achieve average accuracy approximately 99.4% followed by SVM-Bayes fusion algorithm reaching around 95.8%, while the accuracy for SVM and NB is about 94.4% and 91% respectively.

Keywords

Targets classification, micro-Doppler, DCNNs, CW Doppler radar, SVM, Naive Bayes, SVM-Bayes fusion

1. Introduction

As the growing concerns of security and surveillance, targets classification [1–3] is drawing increasing attention. Traditional targets classification methods are usually based on camera surveillance system and video surveillance sys-

tem, which have high requirements of lighting in test scenarios. Besides, these methods invade people's privacy to a certain extent. These factors limit the use of traditional targets classification methods under many circumstances. On the contrary, Doppler radar is excellent in detecting movement in no light and bad weather conditions compared with the optical surveillance system, which makes it a popular tool to detect and classify targets. Meanwhile, a Doppler radar is easy to build and cost-effective, which makes it easy to be used widely [27]. The Micro-Doppler refers to the additional frequency components to the main Doppler shift, which is caused by rotating or vibrating parts of moving targets such as wheels of vehicles, the rotor of a helicopter, and the swinging limbs of human targets. Micro-Doppler has been investigated for various applications [4–6] including automobile anti-collision, battlefield surveillance and etc. In [24], a comprehensive review of micro-Doppler signatures based on different kinds of targets together with its importance and applications, was given. In [25], micro-Doppler signature generated from a target's micro-motions has been extracted using Forward Scattering Radar. But the targets classification based on Micro-Doppler signatures is still in current research.

In previous works, conventional supervised learning methods are preferred for investigating targets classification, for instance, Xiaoran Shi utilized the SVM to classify humans and vehicles based on time-frequency spectrograms [7], in [8] the SVM was also employed as the classifier for the classification of a human, a dog and a horse based on time-velocity spectrograms and in [9] Bayesian Classifier was used to classify humans and vehicles. However, the numbers of the classified species in the researches are small and the process of using conventional supervised learning methods relying on extracted features is very complicated, which limit their practical applications. In [28], the authors employed Bayes linear, k-nearest neighbor, and support vector machine to classify targets based on Gabor features. [24] utilized DCNN to recognize only one human, one dog, one horse and one car. Compared with these re-

searches, in this paper, five species have been detected, with each species having four different targets, which has increased the difficulty of detection, but a higher accuracy has been achieved by our algorithm. We employ both the deep learning method and the conventional supervised learning methods to deal with targets classification problem based on Micro-Doppler signatures. Furthermore, a fusion of the conventional supervised learning methods (SVM-Bayes) has also been implemented to study whether a better result will be achieved using fusion algorithm. In addition, the effect of SNR as well as the decrease of samples have been investigated.

The DCNN acts as the representative of the deep learning method while the SVM and NB which are excellent among conventional supervised learning methods, and frequently used by previous works, are chosen as the representative of conventional supervised learning methods. The SVM, based on Statistical Learning Theory (SLT), is one of the best conventional supervised learning methods. It performs well in solving small sample pattern recognition, nonlinear pattern recognition and high dimensional pattern recognition. The NB is a directed acyclic graph model, which can represent the causal dependence of attribute sets. Additionally, it takes full advantages of prior knowledge and possesses strong ability of learning and predicting. Although the conventional supervised learning methods have achieved good performance in a relative wide field, the process of selecting and extracting features, which requires domain knowledge of each problem, is quite complex and difficult. Whereas the DCNN observably outperforms them in several applications like pattern recognition, image recognition and speech recognition [10–12] without any feature extraction process. The reason for such success is the ability of the DCNN to jointly learn the features and classification boundaries directly from raw input data. Furthermore, its significant characteristics including nonlinear, high parallelism and robustness are accounted. Therefore we expect to yield good results for targets classification problem by exploiting the DCNN. We will present our experimental results and brief backgrounds of these four methods.

The remaining paper is organized as follows. Section 2 illustrates experiments setup and data processing. Brief backgrounds on DCNN, SVM, NB and SVM-Bayes fusion algorithm will be described in Sec. 3. Section 4 presents the experiments results analysis and discussion. Section 5 concludes the paper.

2. Experiment Setup and Data Analysis

2.1 Experiment Setup

The experiments performed in the parking lot of the campus, which is shown in Fig. 1 (a), (b) and (c). While Figure 1(d) shows the equipment and system deployment

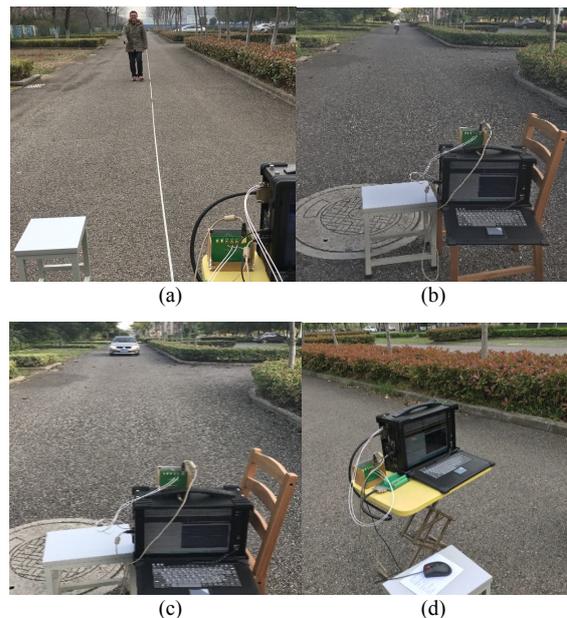


Fig. 1. Data collection setup and experiment scenario.

of our experiments. The system, which was employed to collect experimental micro-Doppler signatures of targets, included IVS-179 radar, M2i.4912 eight-channel parallel data acquisition card and ACME industrial personal portable computer. The IVS-179 Doppler radar worked at 24 GHz in the CW (continuous wave) mode without modulation while the ACME data recording industrial samples at 2 kHz.

We collected data from four humans, four dogs, four bicycles, four cars and four trees respectively. During the process, there was no disturbance created by other objects.

Each target moved along the line-of-sight path of the radar for 100 times and at each time the target moved for 5 seconds. In the experiments, the cars started from approximately 100 meters in front of the radar, the bicycles started from approximately 40 meters in front of the radar, the persons and the dogs started from approximately 20 meters in front of the radar.

2.2 Micro-Doppler Signatures and Choice of Features

According to the Doppler effect, a moving target relative to the wave source will cause a change in frequency or wavelength of a wave. Meanwhile, if the moving target has rotating or vibrating parts, the additional frequency components in addition to the main Doppler shift will be observed, which are called the micro-Doppler effect. When targets are moving, micro-Doppler will be generated in radar signatures which can be clearly observed in the joint time-frequency space [13]. Therefore, Short-time Fourier Transform (STFT) is exploited to characterize micro-Doppler signatures.

If window function is $g(t)$ which will slide along the time line, the time-domain Doppler signal is designated as

$x(t)$; its STFT [14] can be expressed as

$$F_{\text{STFTX}}(t, f) = \int_{-\infty}^{+\infty} x(u)g^*(u-t)\exp(-j2\pi fu)du. \quad (1)$$

In our work, we choose Gaussian window, then $F_{\text{STFTX}}(t, f)$ can be expressed as:

$$F_{\text{STFTX}}(t, f) = \int_{-\infty}^{+\infty} x(u)\exp\left(-\frac{t-u}{2\sigma^2}\right)\exp(-j2\pi fu)du. \quad (2)$$

Proper time-window size and sliding step is vital in capturing particular features of target in Doppler domain. After repeated practice, we choose 0.132 s as the time-window while the sliding step size is 1/2000 s, which is appropriate to recognize the micro-Doppler characteristics in the time-frequency domain. The spectrograms of five subjects are shown in Fig. 2. Since the process of DCNN requires a mass of data, when we employ it to deal with the classification problem, the data of one time was divided into five parts by average. The number of spectrograms for one target was 500, and that for each species was 2000, since the data was gathered from four different targets for one species. Furthermore, we are collecting data and expanding the dataset actively for future researches and applications.

As shown in Fig. 2, the five spectrograms of targets are different from each other. Figure 2(a) is the spectrogram of a moving bicycle, the strongest return comes from the bicycle body and torso while periodic waveforms surrounding it come from limb movements and rotation of wheels. The spectrogram of a moving car is shown in Fig. 2(b), it is observed that the vibration of the car body and the rotation of the wheels can hardly be observed, since they are too weak compared with the car body reflection. Figure 2(c) is the spectrogram of a stationary tree, the micro-Doppler of it is caused by the waggle of leaves and trunk. The spectrograms of a walking dog and a walking person, shown in Fig. 2(d) and (e) respectively, are very similar. Nonetheless, the bandwidth of a person without micro-Doppler is smaller than that of a dog owing to the fact that human has a smaller motion amplitude during the working process.

Features adopted in this paper are listed as follows.

- (1) Torso/body radial velocity.
- (2) Offset of total Doppler.
- (3) Total bandwidth of Doppler signal.
- (4) Bandwidth without micro-Doppler.

It is universally acknowledged that the radial velocities of humans, bicycles, dogs, cars and trees are usually different. In general, cars are the fastest, followed by bicycles then dogs and human, finally trees, as trees are usually stationary. The beat signal of the CW radar can be expressed as:

$$S_b(t) = A_m \exp(j\phi) = A_m \exp\left(\frac{j4\pi R}{\lambda}\right). \quad (3)$$

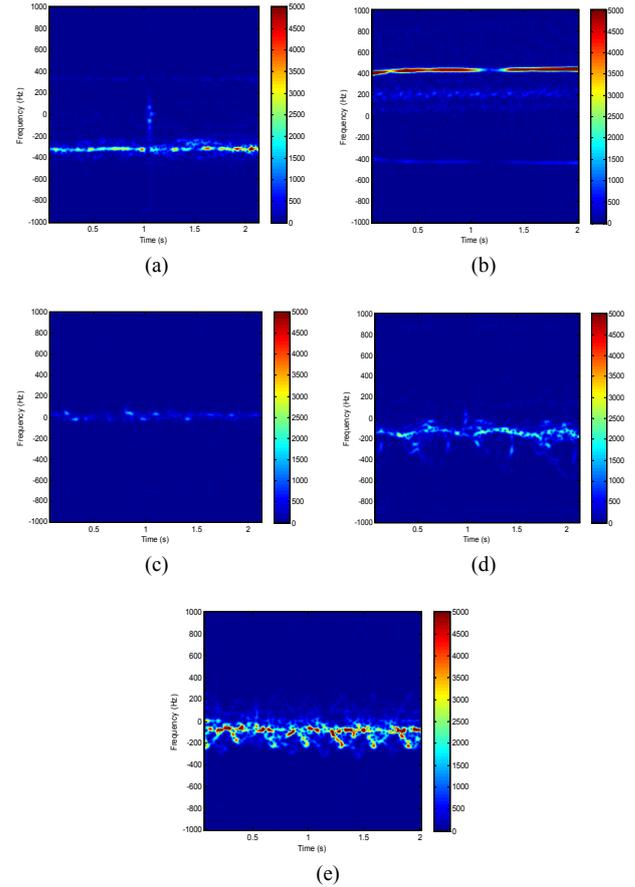


Fig. 2. Sample spectrograms of different targets. (a) a moving bicycle, (b) a moving car, (c) a stationary tree, (d) a walking dog, (e) a walking person.

Here, ϕ is phase of beat signal. λ is the wave length and R is the distance. Then we can get the distance term R is linear relationship with the phase term ϕ and the distance difference is linear with the phase difference. Comparing the unfolded phase of two data points from the same receiving channel, the distance difference between the two points can be derived as follows:

$$D = \frac{\lambda\Delta\phi}{4\pi}. \quad (4)$$

Here, D is the distance difference and the radial displacement and $\Delta\phi$ is the phase difference. Considering the interval between two adjacent points is very short, the speed of a target is almost constant during the interval. Then the radial velocity of an interval and average radial velocity can be expressed as:

$$v_i = \frac{D_i}{t} = \frac{\lambda\Delta\phi_i}{4\pi} \cdot f_s, \quad (5)$$

$$v = \frac{1}{n} \sum_{i=1}^n v_i \quad (6)$$

where f_s is the sampling frequency, n is the total number of intervals over a period of time.

It's apparent that the offset of total Doppler, total bandwidth of Doppler signal and bandwidth without micro-Doppler of different targets are different, which can be observed from the spectrograms in Fig. 2. Although the vibration of the car body and the rotation of the wheels are small, these still generate the micro-Doppler signatures, which are clear enough to be differentiated and analyzed. Moreover, from Fig. 2(d) and (e), bandwidth of the dogs without micro-Doppler is larger than that of human since the dog has a greater motion amplitude than human when moving. Even though the energy of targets is different from each other in Fig. 2, it still can't become a classification criterion for its being affected by distance to a certain extent, which means it is not universal.

3. Classifiers

3.1 Deep Convolutional Neural Networks

Deep Learning is used to identify new samples or predict the possibility intelligently by learning the underlying patterns and characteristics from existing data. It usually adopts deep neural network structure, which is constituted by interconnected neural network architecture, to extract hierarchical abstractions and generalization from the data. In recent years, deep learning has performed well in a wide range of fields such as image recognition and speech recognition with numerous researchers' study.

Deep Convolutional Neural Networks (DCNN) [15–17], which is based on the classical convolution neural network, is one of the most successful deep learning algorithms. It is a kind of multilayer supervising learning neural network. The key components of the DCNN are the convolution and pooling in the hidden layer. In the feature extraction part of the network, convolution and pooling will be implemented alternatively. Multiple convolution filters work in parallel on input data to get the feature maps in convolutional layer followed by pooling layer. Figure 3 shows the schematic of the convolution filter and the pooling operation. The layers after feature extraction part of this network are full connection layers, which includes logistic regression classifier. The input of the full connection is the output of the last pooling layer and the output of the full connection is classification results. In our work, we choose softmax regression to classify the characteristics. Softmax is developed from logistic regression in order to solve multi-class problems. The function of softmax can be expressed as

$$h_{\theta}(x) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ p(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k | x^{(i)}; \theta) \end{bmatrix} \quad (7)$$

$p(y^{(i)} = j | x^{(i)}; \theta)$ represents the probability of the input $x^{(i)}$ of the i th sample belonging to the category j . The loss

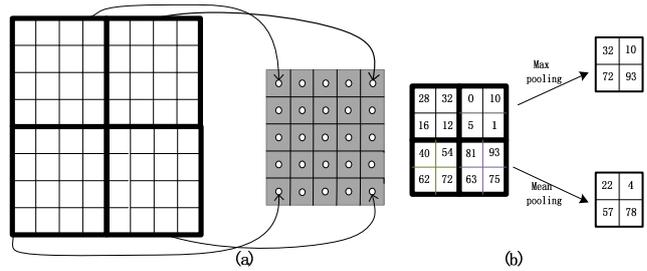


Fig. 3. (a) Process of applying a 4 × 4 convolution filter to the input data to generate the output (in gray). (b) Examples of 2 × 2 pooling (max or mean pooling).

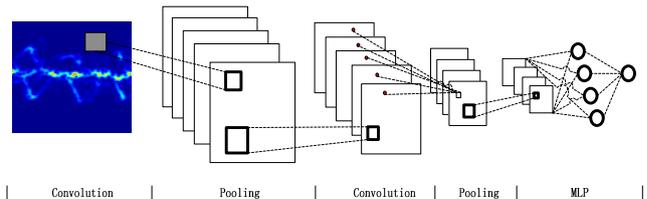


Fig. 4. A simple DCNN architecture.

function of the softmax classifier can be expressed as

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y_i = j\} \log \frac{\exp(\theta_j^T x_i)}{\sum_{l=1}^k \exp(\theta_l^T x_i)} \right]. \quad (8)$$

Then the stochastic gradient descent (SGD) is used to minimize the regulation of loss function $J(\theta)$ in back propagation until the network converges or reaches the maximum iteration number. To prevent overfitting effectively, dropout is widely used. The neuron that is in the state of dropout won't participate in the forward propagation nor in the back propagation. In this way, the neural network is like trying a new structure for every input sample, which reduces the complex interrelationship of neurons. A simple DCNN architecture with two convolution layers and one fully connected layer is shown in Fig. 4.

3.2 Support Vector Machine

The Support Vector Machine (SVM), a conventional supervised learning method based on Statistical Learning Theory (SLT) [18], has been shown working well in many areas. It's a binary classification model, which defines the maximal margin hyper-plane in the feature space, followed by the convex quadratic programming optimization algorithm [19]. When classifying targets, the hyper-plane is utilized to separate a given set of binary labeled training data. In cases where no linear separation exists, the technique of 'kernels' that automatically realizes a non-linear mapping to a feature space is introduced, which will result in a non-linear decision boundary in the input space.

The j th input point $\mathbf{x}^j = (x_1^j, \dots, x_n^j)$, labeled by the random variable $Y^j \in \{-1, 1\}$, is the realization of the random vector \mathbf{X}^j . $\phi(\mathbf{x})$ is the eigenvector of \mathbf{x} after the mapping with the technique of 'kernels'. In this paper we choose

Gaussian kernels $K(\mathbf{x}^i, \mathbf{x}^j) = \exp\left(-\frac{\|\mathbf{x}^i - \mathbf{x}^j\|^2}{\sigma^2}\right)$, σ is kernel parameter.

In the feature space the corresponding decision function is

$$f(x) = \mathbf{w}^T \phi(\mathbf{x}) + b, \quad (9)$$

\mathbf{w} is the maximal margin hyper-plane:

$$\mathbf{w} = \sum_{i=1}^m \alpha_i y^i \phi(\mathbf{x}^i). \quad (10)$$

While α_i are positive real numbers that maximize

$$\sum_{i=1}^m \alpha_i - \sum_{ij=1}^m \alpha_i \alpha_j y^i y^j (\phi(\mathbf{x}^i), \phi(\mathbf{x}^j)) \quad (11)$$

subject to

$$\sum_{i=1}^m \alpha_i y^i = 0, \quad \alpha_i > 0. \quad (12)$$

The decision function can be equivalently expressed as:

$$f(x) = \sum_{i=1}^m \alpha_i y_i \phi(\mathbf{x}^i)^T \phi(x) + b. \quad (13)$$

The complexity of the classifier is not related to the number of samples of the training set. The only influencing factor is the number of the supported vectors, which indicates that the SVM has a simple system structure. Furthermore, the learning and forecasting process of the SVM is no time-consuming, which makes it very popular in pattern recognition, image recognition and many other fields.

3.3 Naive Bayes

The Naive Bayes (NB), a classification method based on Bayes [20], [21] theorem, has the smallest misclassification rate when the conditional independence assumption is established. Although the assumption limits its application to some degree, in practical application, not only the complexity of the Naive Bayesian model reduce exponentially, but also the considerable robustness and efficiency have been shown in many fields opposition to the assumption. Considering its efficient evaluation, high accuracy and solid theory foundation, it has been successfully applied to data mining tasks including classification, clustering and selection of models.

Abstractly, the probability model for Naive Bayes classifier can be expressed as follows: For a given dataset $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, firstly, the algorithm learns the joint probability $P(\mathbf{X}, \mathbf{Y})$ of input and output based on attribute conditional independence assumption. On this basis, for a given input \mathbf{X} , the method figures out the output \mathbf{Y} with the maximum posterior probability using Bayes theorem. Naive Bayes can be equivalently expressed as:

$$y = f(x) = \arg \max_{c_k} \frac{P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)}{\sum_k P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)}. \quad (14)$$

When the input is x , choosing the category with the highest conditional probability as the class to be classified after calculating the conditional probability of all categories. Since the denominator of formula (14) is the same for all c_k , formula (14) can be simplified as:

$$y = f(x) = \arg \max_{c_k} P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k). \quad (15)$$

3.4 SVM-Bayes Fusion Algorithm

To explore whether better results will be achieved by the fusion of different conventional supervised learning methods, A SVM-Bayes fusion algorithm has been employed to process the data. Figure 5 shows the framework of this algorithm. In the process of Bayesian inference, when there is no empirical data available, the subjective probability can be used to substitute the prior probability as well as the likelihood function of the hypothetical event.

Assumed that there are m kinds of features extraction methods, n species categories, and let Θ represents the collections of species categories $\Theta = \{O_1, O_2, \dots, O_n\}$. For an unknown sample ($O \in \Theta$) the fusion algorithm based on Maximum a posteriori (MAP) can be expressed as:

$$P(O_{MAP} | O^1, O^2, \dots, O^m) = \max_{O \in \Theta} P(O | O^1, O^2, \dots, O^m). \quad (16)$$

O^i is the decision for an unknown sample of SVM using the i th feature extraction method. O_{MAP} is the decision of the sample based on multi-feature extraction methods. $P(O_{MAP} | O^1, O^2, \dots, O^m)$ is joint probability density function of SVM based m kinds of feature extraction methods.

Then the formula (16) can be expressed as:

$$O_{MAP} \triangleq \arg \max_{O \in \Theta} P(O | O^1, O^2, \dots, O^m). \quad (17)$$

According to the Bayes formula in Sec. 3.3, the fusion algorithm can be expressed as:

$$O_{MAP} \triangleq \arg \max_{O \in \Theta} \left\{ \frac{P(O^1, O^2, \dots, O^m | O) P(O)}{\sum_{O \in \Theta} P(O^1, O^2, \dots, O^m | O) P(O)} \right\}. \quad (18)$$

$P(O)$ is prior probability of $O \in \Theta$.

The denominator of the equation is full probability formula which is not affected by the value of O , then it can be simplified as:

$$O_{MAP} \triangleq \arg \max_{O \in \Theta} \{P(O^1, O^2, \dots, O^m | O) P(O)\}. \quad (19)$$

Since the feature extraction methods are independent, the equation can be expressed as:

$$O_{MAP} \triangleq \arg \max_{O \in \Theta} \left\{ P(O) \prod_{i=1}^m P(O^i | O) \right\}. \quad (20)$$

$P(O^i | O)$ is the likelihood function of SVM based on the i th feature extraction method. As logarithm of this function is monotone increasing, the logarithm on the right side of the

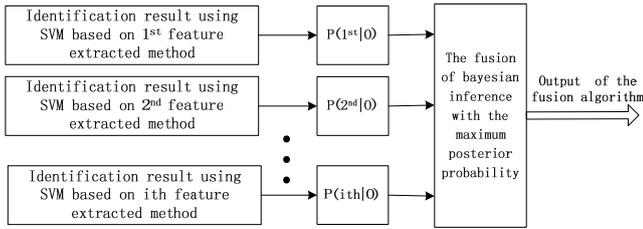


Fig. 5. The framework of the SVM-Bayes fusion algorithm.

formula (20) can simplify the operation without affecting the fusion recognition decision.

Then the equation can be expressed as:

$$O_{MAP} \triangleq \arg \max_{O \in \Theta} \left\{ \ln P(O) + \prod_{i=1}^m \ln P(O^i | O) \right\}. \quad (21)$$

$P(O)$ and $P(O^i|O)$ need to be confirmed. $P(O)$ is prior probability, and in this case, it is assumed that each species has the same probability. In this case, likelihood function $P(O^i|O)$ means that when the species is certain, the probability that it can be identified by SVM correctly. Its value can be approximated by the recognition experiment of training samples.

4. Classification and Discussion

4.1 The Classification Result of DCNN

We employed the DCNN to the measured, experimental spectrograms directly to classify targets. Then the targets classification problem was transformed into image recognition problem. 80% of the spectrograms of each target were used as training data and the rest were used to test. We employed Caffe [22], i.e., Convolutional architecture for fast feature embedding, which is open-source and speeded up by the NVIDIA GPU and CUDA library, as the platform to analyze the spectrograms. The GPU we used is the NVIDIA Quadro M4000.

We adopted AlexNet [23] in which there were five convolution layers, two fully connected layers with 4096 hidden nodes in the first fully connected layer and an output layer as shown in Fig. 6(a). The size of the input spectrograms of the whole AlexNet must be 256×256 and we used the tools of Caffe to normalize the spectrograms directly. The partial internal structure of the network is shown in Fig. 6(b) and (c). Figure 6(b) shows the first two convolutional layers and pooling layers while Figure 6(c) shows the first full connection layer. Rectified Linear Units (ReLU) was used as activation function and followed by max pooling in each layer. We fine-tuned the parameters of the network according to our experiments, which was recorded in the configuration file of Caffe. The data of learning rate was adjusted to 0.001 because when the learning rate was larger the loss wouldn't converge and when the learning rate was smaller the process would be time-consuming. In our experiments, the data of maximum iteration

number was changed to 4000 for stable results. The learning rate was reduced to $0.001 \times 0.9^{\lfloor 4000/2000 \rfloor}$ after every 2000 iteration, since we employed gradient descent method to solve the optimization problem. The weight decay was changed to 0.0005. The definition of the network for training and validation as well as the parameters for every layer are recorded in a particular file. In this file the spectrograms were resized to 227×227 , since the size of the input spectrograms of the first convolution layer must be 227×227 . The batch sizes in "Train" part and "Test" part were adjusted to 32 and 16, respectively. Batch size represents the number of samples taken in each iteration. The average of the gradient of these samples is used to update the parameters of the network. The batch size determines the direction of gradient descent and the effect and rate of convergence, as well as memory utilization. In the full connected layer, we adapted the learning rate of bias to 10 and the learning rate of weight to 20 to speed up

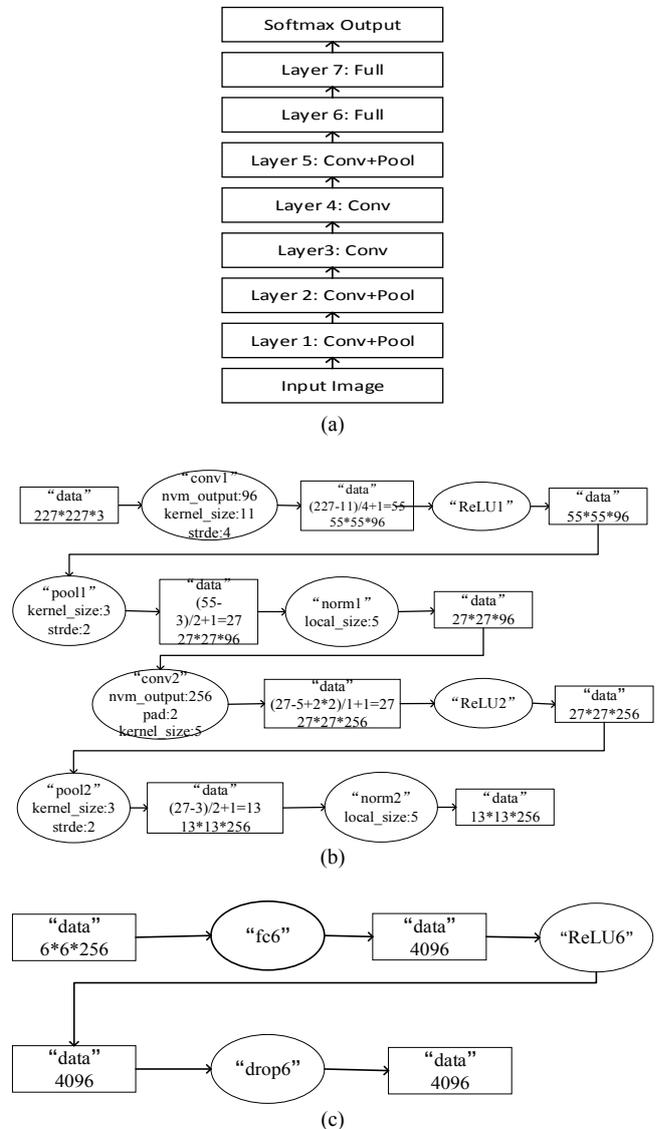


Fig. 6. (a) The structure diagram of AlexNet, (b) the first two convolution layers and pooling layers, and (c) the first full connection layer.

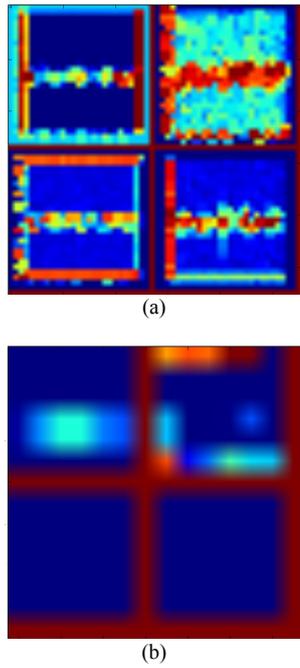


Fig. 7. The feature maps: (a) the first pooling layer, (b) the fifth pooling layer.

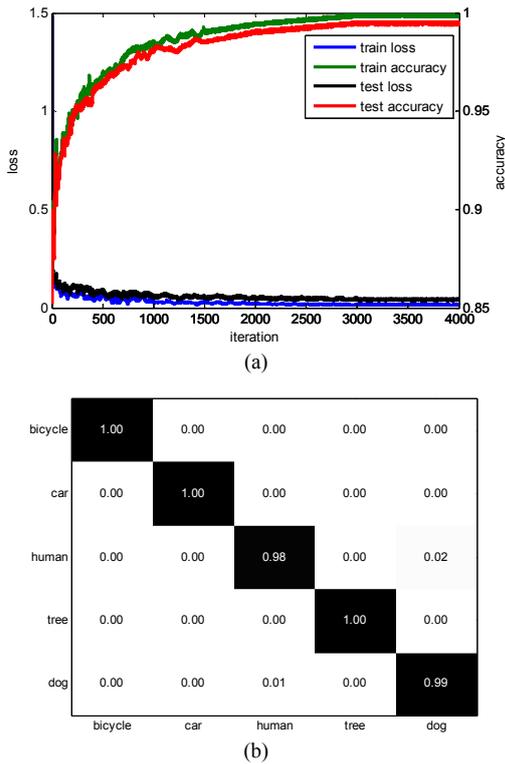


Fig. 8. (a) The accuracy and loss of classification process for 5 species, (b) confusion matrix for DCNN. The x and y axis indicate the species of the subjects.

the learning rate in this layer. After 4000 iterations, our own network was generated storing some model parameters like weight according to our spectrograms.

We assessed the performance of our own network under the scenario where we recognized one species from five different species with every species having four differ-

ent targets. We randomly selected 100 spectrograms for each species from the test data to verify the stability of our own network, and calculated the time it took to classify one species. The whole process was repeated for 100 times. The variance of the 100 times was 0.004, which means that the repeated 100 times varied slightly with each other, and our network was very stable.

The feature maps of the first pooling layer and the fifth pooling layer are shown in Fig. 7. Compared with (a) in Fig. 7, (b) is already a little abstract for us to obtain the physical insights into our case. In the future, we plan to study the learned features of the DCNN, which will help us get better insights. Figure 8(a) shows the accuracy as well as the loss of training and testing process for the species classification process. The loss showed is the value of the loss function in formula (8). As it converges to a number tending to zero, the network doesn't overfit. Figure 8(b) shows the confusion matrices of a random time for classification results of the DCNN. Since the results of the repeated 100 times varied slightly, result of one time could represent the accuracy to some extent. The average accuracy is 99.4%, which shows that one species can be classified from five species with high possibility.

4.2 The Classification Result of SVM, NB and SVM-Bayes

The SVM and NB, two typical conventional supervised learning methods, were applied to the extracted features respectively for the targets classification problem. The amount of data required in this part is relatively small compared with that utilized in the DCNN, which is one of the advantages of conventional supervised learning methods. Figure 9 is a 3D graph showing the estimated values of the features. It seems like that the features of bicycles, cars and trees are far from each other while the features of the dog and humans mixed together. This indicates that it's promising for the classification of bicycles, cars and trees but difficult to classify human and dog. Figure 10(a), (b) show the confusion matrices of SVM and NB, respectively. The average accuracy of SVM is higher than that of NB, which is 94.4% and 91%, respectively. Though both of them are satisfying, they both make misclassification of the

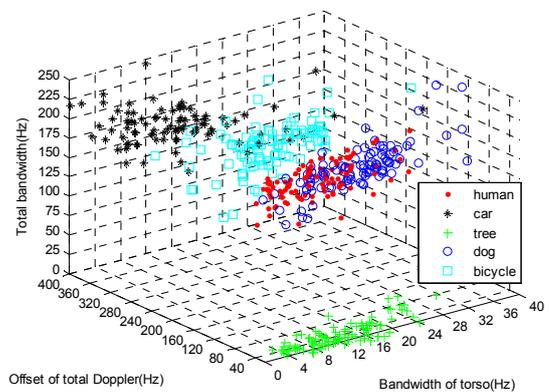


Fig. 9. Scatter plot of three features.

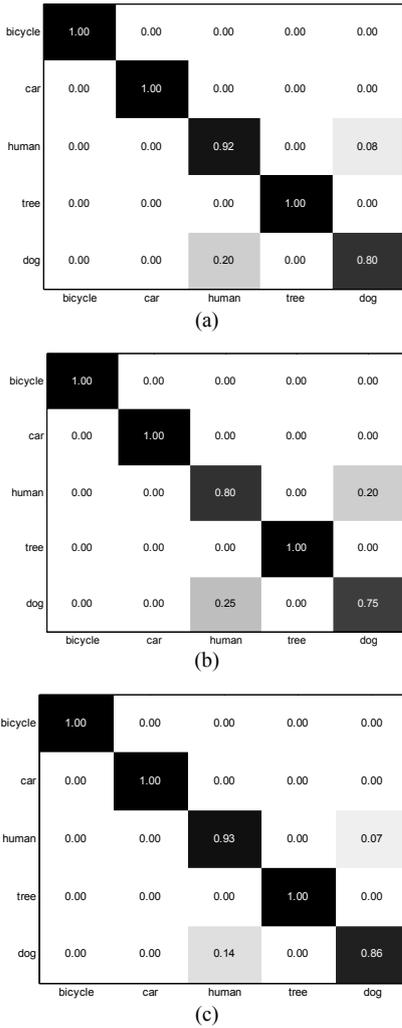


Fig. 10. (a) Confusion matrix for SVM. (b) Confusion matrix for NB. (c) Confusion matrix for SVM-Bayes fusion. The x and y axis indicate the species of the subjects.

persons and the dogs with a relatively higher probability compared with the DCNN.

Regarding to the results of SVM-Bayes fusion algorithm, two different set of features have been utilized, the features that we employed in Sec. 2.2 worked as the first set of features. The second features set we chose were Principal Component Analysis (PCA) based features.

We chose three features extracted from PCA on spectrograms. The first feature referred to as the latent feature was the principal component variance of the covariance matrix. While the second feature was Hotelling’s T-squared statics, which is “a statistical measure of the multivariate distance of each observation from the center of data set.” The third feature, “explained”, represents the percentage of how much the variance explains. The top five values of each vector were taken. Figure 10(c) shows the confusion matrices of SVM-Bayes fusion. Table 1 compares the results of the four algorithms. The accuracy of SVM-Bayes fusion is higher than that of SVM and NB but lower than that of DCNN. It means fusion of different conventional supervised learning methods may lead to better results with

Methods	Accuracy
DCNN	99.4%
SVM	94.4%
NB	91%
SVM-Bayes	95.8%

Tab. 1. Comparison between the results of DCNN, SVM, NB and SVM-Bayes.

constant optimization like using different fusion operation, different feature extracted methods or different classification methods. However, these methods rely severely on the extracted features which means different features will lead to different results, and require domain knowledge of each problem. Furthermore, the process of choosing and extracting features is very complex. These factors limit their use by people who are not familiar with related fields. However none of them is a problem for DCNN, which means DCNN is easier to be used widely in real life.

4.3 Effect of Noise

In this sub-part, the anti-noise performances of the four algorithms have been studied respectively. In detecting the anti-noise performance of the DCNN, our own network was employed. We added different grade ($SNR = 29\text{ dB}\sim 30\text{ dB}$, $20\text{ dB}\sim 21\text{ dB}$, $15\text{ dB}\sim 16\text{ dB}$, $10\text{ dB}\sim 11\text{ dB}$, $0\text{ dB}\sim 1\text{ dB}$) of random noise to the echo respectively. For example, we added random noise ($SNR = 0\text{ dB}\sim 1\text{ dB}$) to the raw signals that we collected in Sec. 2. And after that, we randomly chose 100 spectrograms of each target at each noise grade to test the noise immunity of the algorithm. For each noise grade, the whole process was repeated for 100 times, and the average accuracy was taken as the final result. As for detecting the anti-noise performance of the SVM, NB and SVM-Bayes fusion algorithm, we also added different grades $SNR = 29\text{ dB}\sim 30\text{ dB}$, $20\text{ dB}\sim 21\text{ dB}$, $15\text{ dB}\sim 16\text{ dB}$, $10\text{ dB}\sim 11\text{ dB}$, $0\text{ dB}\sim 1\text{ dB}$ of random noise to the signals that we collected in Sec. 2 respectively, followed by features extraction.

The classification results are shown in Fig. 11. It’s obviously that the classification performance of the four methods can be compromised when the noise increases. The accuracy of the DCNN, however, decreases very slowly

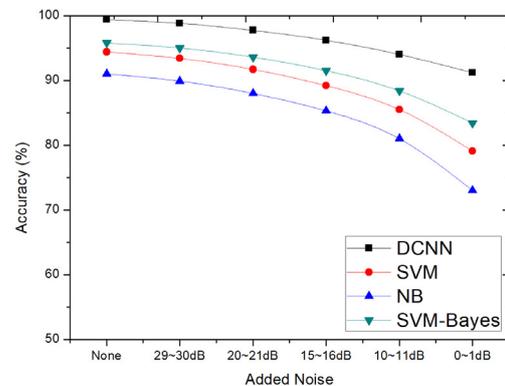


Fig. 11. The impact of noise of the four classification algorithm.

considering the degree of noise increasing, which demonstrates that the algorithm has good anti-noise performance. The noise immunity of the SVM-Bayes fusion algorithm is similar to but better than that of SVM, and both are better than that of Naive Bayes. Whereas the performance of the three algorithms cannot compare with that of the DCNN.

4.4 Discussion of Computational Time and Decrease of Samples

It took about 15 minutes to train the network, and after that our own network was generated, however, it only took about 0.391 s to classify one spectrogram, meaning that the trained network could be used in real-time monitoring as the time it took to classify one spectrogram was so short. Regarding to the time for the other three classifiers, it took about 5.567 s for SVM-Bayes fusion algorithm, 1.895 s for SVM and 1.592 s for NB to complete the whole classification progress based on extracted features.

Although the time it took to train the network was far more than that used by the other three algorithms to classify species, the time it took to classify one spectrogram by the trained network was very short, and the accuracy was particularly high. Meanwhile, the progress of selecting as well as extracting features were very complicated and time-consuming, which definitely extended the total time for conventional supervised learning methods. As for which

method is better, it depends on the amount of the data, the accuracy and the computational speed that are required. If the volume of the data is large, and is constantly updated and expanded, the DCNN is more suitable. Under this circumstance, a database needs to be set up, and the parameters of the network need to be updated along with update of the data. However, if the volume of data is very limited, and the accuracy doesn't need to be very high, conventional supervised learning methods will be better.

It is well known that the sample size DCNN needing is very large. If the sample size is insufficient, the basic image transformation methods including translation, rotation, zoom, mirroring and cropping will be applied to the spectrograms to augment the dataset, as the basic nature of the images won't be altered by these transformations, which means the classification results won't be influenced. To study the performance of the method when the sample sizes are reduced, two circumstances have been taken in to consideration, where sample size for every species is 1000 and 500, respectively. Then the basic image transformation methods are used to augment the dataset to 2000 for both situations. Figure 12 shows the results of the two circumstances. The accuracy decreases with the samples size reduced, though we augmented the dataset. It is inevitable, since some features have not been learned. So more researches need to be done to optimize the frame of the network reducing its severe dependence on big data.

5. Conclusion

In this paper, we classify ground targets based on micro-Doppler signatures using the DCNN, SVM, NB and SVM-Bayes fusion algorithm as well as studying several factors that will affect results including SNR, decrease of samples. By using the four methods, targets can be classified successfully. To be specific, the average accuracy is about 99.4% for the DCNN, 94.4% for the SVM, 91% for the NB and 95.8% for the SVM-Bayes algorithm, indicating that the DCNN performs the best in targets classification. In addition, when it comes to noise immunity of the algorithms, the DCNN also performs better than the other three methods.

Though the accuracy of the DCNN is high, the amount of data that it needs is relatively large, while conventional supervised learning methods can get a satisfying accuracy with a small dataset. However, the conventional supervised learning methods relying on the extracted features require domain knowledge of each problem. Furthermore, their development tends to be saturated, and they have low flexibility. Oppositely, none of them is a problem for the DCNN, which shows the potential of the DCNN for targets classification problems based on radar micro-Doppler signatures.

In the future, further researches will be done for the purpose of promoting the classification performance. First, the condition we considered in the experiments, where the

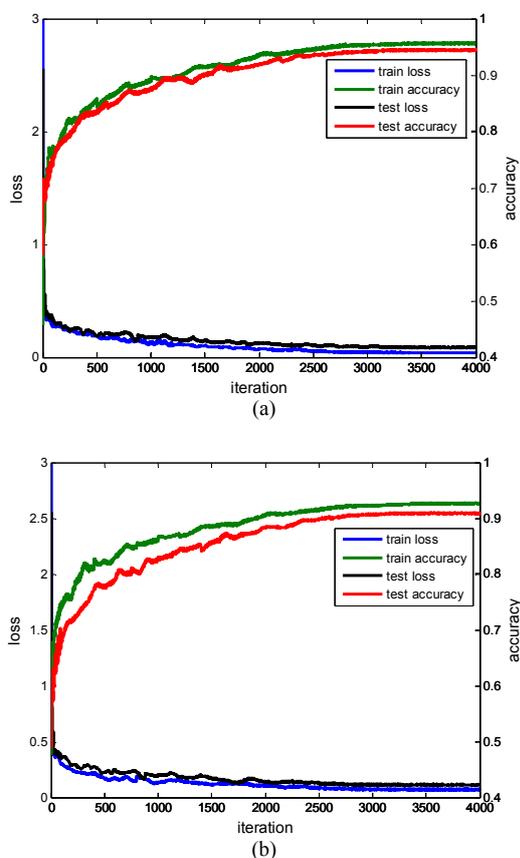


Fig. 12. Impact of decrease of samples. (a) Sample size is 1000, (b) sample size is 500.

targets moved along the line-of-sight path of the radar, is simple compared with realistic moving pattern. In the future, we will do further researches about non-LOS scenarios. Second, in our experiments one species is classified from a group of five species. To apply our system to the practical conditions we also need to expand the maximum number of species. In addition, the optimization for the frame of the network also lies in our future research plan.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (Grant no. 6150010825), the Fundamental Research Funds for the Central Universities (Grant no. NS2016040), the Fundamental Research Funds for the Central Universities (Grant no. NJ20150020) and the China Scholarship Council (Grant no. 201606835062).

References

- [1] MOLCHANOV, P., ASTOLA, J., EGIAZARIAN, K. Classification of ground moving radar targets by using joint time-frequency analysis. In *IEEE Radar Conference*. Atlanta (GA, USA), 2012, p. 0366–0371. DOI: 10.1109/RADAR.2012.6212166
- [2] VILLEVAL, S., BILIK, I., GURBUZ, S. Z., et al. Application of a 24 GHz FMCW automotive radar for urban target classification. In *IEEE Radar Conference*. Cincinnati (OH, USA), 2014, p. 1237–1240. DOI: 10.1109/RADAR.2014.6875787
- [3] TAHMOUSH, D., SILVIOUS, J., BURKE, E. A radar unattended ground sensor with micro-Doppler capabilities for false alarm reduction. In *Conference on Unmanned/Unattended Sensors and Sensor Networks VII*. Toulouse (France), 2010, p. 20–22. DOI: 10.1117/12.864959
- [4] CHEN, V. C. *The Micro-Doppler Effect in Radar*. Norwood, Mass: Artech House 2011. ISBN: 9781608070572
- [5] CHEN, V. C., TAHMOUSH, D., MICELI, W. J. *Radar Micro-Doppler Signatures: Processing and Applications*. Institution of Engineering and Technology, Stevenage, U.K., Jan, 2014. ISBN: 9781849197168.
- [6] STANKOVIC, L., STANKOVIC, S., OROVIC, I. Time frequency analysis of micro-Doppler signals based on compressive sensing. In *Compressive Sensing for Urban Radar*, 2014. p. 283–326. DOI: 10.1201/b17252–10
- [7] SHI, X. R., ZHOU, F., LIU, L., et al. Textural feature extraction based on time-frequency spectrograms of humans and vehicles. *IET Radar, Sonar & Navigation*, 2015, vol. 9, no. 9, p. 1251–1259. DOI: 10.1049/iet-rsn.2014.0432
- [8] BJÖRKLUND, S., JOHANSSON, T., PETERSSON, H. Target classification in perimeter protection with a micro-Doppler radar. In *The 17th International Radar Symposium (IRS)*. Krakow (Poland), 2016, 5 pages. DOI: 10.1109/IRS.2016.7497363
- [9] NANZER, J. A., ROGERS, R. L. Bayesian classification of humans and vehicles using micro-Doppler signals from a scanning-beam radar. *IEEE Microwave and Wireless Components Letters*, 2009, vol. 19, no. 5, p. 338–340. DOI: 10.1109/LMWC.2009.2017620
- [10] HINTON, G., DENG, L., DONG, Y. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Processing Magazine*, 2012, vol. 29, no. 6, p. 82–97. DOI: 10.1109/MSP.2012.2205597
- [11] KRIZHEVSKY, A., SUTSKEVER, I., HINTON, G. ImageNet classification with deep convolutional neural networks. In *International Conference on Neural Information Processing Systems*. Lake Tahoe (NE, USA), 2012, vol. 25, no. 2, p. 1097–1105. DOI: 10.1145/3065386
- [12] HUANG, R., XIE, X., FENG, Z., et al. Face recognition by landmark pooling-based CNN with concentrate loss. In *IEEE International Conference on Image Processing (ICIP)*. Beijing (China), 2017, p. 1582–1586. DOI: 10.1109/ICIP.2017.8296548
- [13] JAVIER, R. J., KIM, Y. Application of linear predictive coding for human activity classification based on micro-Doppler signatures. *IEEE Geoscience and Remote Sensing Letters*, 2013, vol. 11, no. 10, p. 781–785. DOI: 10.1109/LGRS.2014.2311819
- [14] CHEN, V., LING, H. *Time-Frequency Transforms for Radar Imaging and Signal Analysis*. Norwood (MA, USA): Artech House, 2002. ISBN: 9781580532884
- [15] WANG, T., WU, D. J., COATES, A., et al. End-to-end text recognition with convolutional neural networks. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR)*. Tsukuba (Japan), 2012, p. 3304–3308. Print ISBN: 9781467322164
- [16] OUYANG, W., WANG, X., ZENG, X. DeepID-net: Deformable deep convolutional neural networks for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2016, vol. 39, no. 7, p. 1320–1334. DOI: 10.1109/TPAMI.2016.2587642
- [17] JI, S. W., XU, W., YANG, M. 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, vol. 35, no. 1, p. 221–231. DOI: 10.1109/TPAMI.2012.59
- [18] CHERKASSKY, V., MULIER, F. Statistical learning theory. Chapter 4 in *Learning from Data: Concepts, Theory, and Methods*, 2006, p. 99–150. DOI: 10.1002/9780470140529.ch4
- [19] FUREY, T. S., CRISTIANINI, N., DUFFY, N., ET AL. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 2000, vol. 16, no. 10, p. 906–914. DOI: 10.1093/bioinformatics/16.10.906
- [20] KIM, C. J., HWANG, K. B. Naive Bayes classifier learning with feature selection for spam detection in social bookmarking. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*. Antwerp (Belgium), 2008.
- [21] POP, I. An approach of the Naive Bayes classifier for the document classification. *General Mathematics*, 2006, vol. 14, no. 4, p. 135–138.
- [22] JIA, Y., SHELHAMER, E., DONAHUE, J. Caffe: Convolutional architecture for fast feature embedding. In *The 22nd ACM International Conference on Multimedia*. Orlando (FL, USA), 2014, p. 675–678. DOI: 10.1145/2647868.2654889
- [23] KRIZHEVSKY, A., SUTSKEVER, I., HINTON, G. E. Imagenet classification with deep convolutional neural network. In *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS'12)*. Lake Tahoe (NE, USA), 2012, vol. 1, p. 1097–1105. DOI: 10.1145/3065386
- [24] TAHMOUSH, D. Review of micro-Doppler signatures. *IET Radar, Sonar & Navigation*, 2015, vol. 9, no. 9, p. 1140–1146. DOI: 10.1049/iet-rsn.2015.0118
- [25] ABDULLAH, R. S. A. R., ALNAEAB, A., SALAH, A. A., et al. Micro-Doppler estimation and analysis of slow moving objects in forward scattering radar system. *Remote Sensing*, 2017, vol. 9, no. 7, 23 pages. DOI: 10.3390/rs9070699

- [26] KIM, Y., MOON, T. Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks. *IEEE Geoscience Remote Sensing Letters*, 2016, vol. 13, no. 1, p. 2–8. DOI: 10.1109/LGRS.2015.2491329
- [27] REZNICEK, M., BEZOUSEK, P. Commercial CW Doppler radar design and application. In *Proceedings of the 27th International Conference Radioelektronika*. Brno (Czech Republic), April 2017. DOI: 10.1109/RADIOELEK.2017.7937577
- [28] LEI, J., LU C. Target classification based on micro-Doppler signatures. In *IEEE International Radar Conference*. Arlington (VA, USA), 2005, p. 1–5. DOI: 10.1109/RADAR.2005.1435815

About the Authors ...

Peibei CAO was born in China in 1993. She received the B.S. degree in Information Engineering from the Nanjing University of Aeronautics and Astronautics, College of Electronic and Information Engineering, Nanjing, China in 2016. She is currently in M.Sc. degree in Nanjing University of Aeronautics and Astronautics, College of Electronic

and Information Engineering, Nanjing, China. Her research interests are in radar signal processing and machine learning. She is doing the researches about combining radar signal processing with machine learning.

Weijie XIA was born in China in 1979. He received his B.S. degree, M.Sc. degree, and Ph.D. in Information Engineering from the Nanjing University of Aeronautics and Astronautics, College of Electronic and Information Engineering, Nanjing, China. In 2004, Prof. Xia joined the Nanjing University of Aeronautics and Astronautics (NUAA), China, as a Professor of Information Engineering. Prof. Xia's current research interests are in radar signal processing, and particularly in Doppler feature extraction and electromagnetic modeling. Prof. Xia has published more than 20 papers in refereed journals and conference proceedings.

Yi LI was born in China in 1995. He is currently in B.S. degree in the Nanjing University of Aeronautics and Astronautics, College of Electronic and Information Engineering, Nanjing, China.