Human Activity Recognition Based on Deep Learning Method

Xiaoran Shi, Yaxin Li, Feng Zhou*, Lei Liu
The Ministry Key Laboratory of Electronic Information Countermeasure and Simulation
Xidian University
Xi'an, Shaanxi, China
shixr_xidian@163.com; erlangkaochi@qq.com; fzhou@mail.xidian.edu.cn; liulei_xidian@163.com

***Abstract*—With the increasing demand of security defense, anti-terrorism investigation and disaster rescue, human activity classification and recognition have become a hot research topic. When a human is illuminated by electromagnetic waves, a Doppler signal is generated from his or her moving parts. Indeed, bodily movements are what make humans' micro-Doppler signatures unique, offering a chance to classify human activities. Classification needs a lot of samples for training, however, in the real application, there is a certain gap between the simulated data and the real data, and the measured data is often difficult to obtain. Due to the non-stationary characteristic for human radar echoes, the spectrograms for the human activities show different micro-Doppler signatures. Therefore, we proposed a method of human activity classification based on spectrograms using deep learning techniques, including deep convolutional generative adversarial network for expanding and enriching training set and a transfer-learned deep convolutional network (DCNN) for feature extraction and classification, which is based on a DCNN pre-trained by a large-scale RGB image data set-that is, ImageNet. Finally, the simulation results verified the effectiveness of the proposed method.**

Keywords—human activity classification; generative adversarial networks; deep convolutional neural networks; transfer learning.

I. INTRODUCTION

With the increasing demand of security defense, anti-terrorism investigation and disaster rescue, human detection and human activity classification have become a hot topic [1-4]. When the human is moving, the torso introduces Doppler effect, and the swing arms will induce plentiful micro-Doppler (m-D) information, which is distributed around the main Doppler and broaden the Doppler spectrum [5]. The m-D effect can reflect the unique structural and kinematic characteristics of the target. Therefore, it can be used for target classification and recognition.

Recently, the research of human activity classification has become more and more popular. Kim Y *et al.* utilized the support vector machine (SVM) classifier [6] to classify seven human activity, running, walking, walking while holding a stick, crawling, boxing while moving forward, boxing while standing in place, and sitting still. [7] classified the walking, skiing and riding humans based on the difference of their m-D signatures. [2] adopted different classifiers to recognize human motions using empirical mode decomposition method.

Nowadays, with the rapid development of machine learning technique, lots of deep learning methods have been widely used in the application of human activity classification. It solved the problem of poor generalizability and time-consuming for handcrafted features. [8] utilized the deep convolutional neural networks (CNN) to classify seven human activities in [6]. [9] discussed the m-D signatures of human aquatic activity and classified them through transfer learning. However, all of these need adequate training samples to obtain good classification result. [10] utilized generative adversarial network (GAN) to synthesize new synthetic aperture radar image. Inspired by this, a method of human activity classification using deep convolutional GAN (DCGAN) and deep CNN (DCNN) is proposed to get better result under the condition of small training sample number.

The remainder of this paper is as follows. Section II presents the human moving rules and the main methodology of DCGAN. Section III introduces VGG16, one of the classical pre-trained DCNN architectures, to classify different human activities. Section IV shows some experiment results. Section V gives the conclusion.

II. SAMPLE EXPANSION BASED ON DCGAN

A. The Boulic kinematic model of human

Different from other non-ridge body targets, the movements of human segments are rich and coordinated. In this paper, a human kinematic model based on biomechanical experimental data presented by Boulic *et al.*[5] is established, aiming to provide the three-dimension space position and movement trend of the human. The human body is divided into 17 segments. The moving trajectory of each joint can be calculated according to the degree of free (DOF) in Table 4.1 of Ref. [5]. These DOFs describe not only the moving characteristics but also the m-D signatures. According to the DOFs, the trajectory of each joint could be obtained using the experienced equations and parameters to calculate human echo.

According to Ref. [5], trajectories are described in three methods. Six trajectories are given by sinusoidal expressions (one of them by a piecewise function), and six trajectories are represented by cubic spline functions passing through control points located at the extremities of these trajectories. Each segment has its unique curve trajectory. When human is moving, his arms and legs are swing periodically with torsos. The kinematic mechanism and Doppler signatures can be summarized as follows.

i) The bulk motion of the human torso, which brings Doppler frequency shifts;

ii) The micro-motion of swinging arms and legs, which generates m-D frequency side bands.

According to the above analysis, it is obvious seen that the human echo belongs to the classical non-stationary signal. Traditional Fourier transform (FT) cannot precisely describe the local characteristic or the frequency variation with time. Joint time-frequency (TF) technique must be taken into consideration here to analyze the human echo signal. Due to the total radar cross section (RCS) of human is the linear summation of the RCS of each segment, the short-time Fourier transform (STFT) [11] is chosen here for its superiority of linear property and no cross terms.

B. Samples expansion based on DCGAN

GAN is a powerful class of generative models introduced in 2014 by Goodfellow *et al.* [12]. This paper proposes a TF spectrogram simulation method based on DCGAN. This method applies a generative model that does not require any prior assumptions on the TF spectrogram but directly based on real samples. Because GAN itself is a rapidly developing method, we focus on its potential in human activity classification applications, rather than the bound of performances.

GAN adapts a confrontation of generator and discriminator to realize the approximation of a mixed Gaussian model $P_G(x;\theta)$, where θ is the parameters of the mixed Gaussian model including mean and variance, to the real model $P_{\text{data}}(x)$. The generator tries to capture the potential distribution of real samples, and generates new data samples. The discriminator is often a binary classifier, discriminating real samples from the generated samples as accurately as possible. The input of the generator G is a noise vector z, the output $G(z)$ has the similar structure of x. The input of the discriminator D is x or $G(z)$, the output $D(x)$ or $D(G(z))$ is a scalar between 0 and 1, which indicates the probability that the discriminator will judge the input x or $G(z)$ as the real data. The optimization process of GAN is a minimax game process, and the optimization goal is to reach Nash equilibrium [4], as shown in (1).

$$\min_G \max_D V(G,D) = E_{x\sim P_{data}}[\log D(x)] + E_{x\sim P_G}[\log(1-D(x))] \quad (1)$$

When the network training is complete, we can get a good generator to generate highly realistic data.

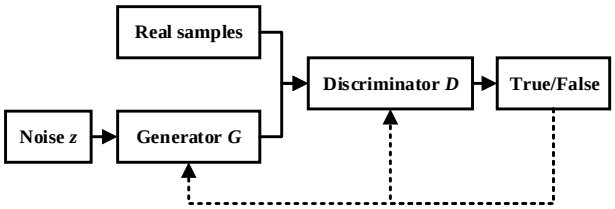


Fig. 1 The architecture of GAN.

Fig. 1 shows the architecture of the basic GAN. Both the generator and the discriminator can adopt the structure of currently popular DCNN, so GAN here is called DCGAN [13]. To improve the image quality and the rate of convergence, we have made some changes for the traditional CNN. Details are as follows.

i) The pooling layer is cancelled. The generator uses the transpose convolution to carry on the sampling, the discriminator adds the convolution of the stride instead of pooling.

ii) The batch normalization is utilized in both the generator and the discriminator.

iii) The full connected layer is removed, and the network becomes a full convolution network.

iv) The ReLu is chosen as the activation function, the last layer adopts the tanh function in the generator. The activation function of the discriminator is leaky ReLu.

III. Human Activity Classification Based on DCNN

AlexNet, VGGNet, Coogle InceptionNet and ResNet are the typical CNNs. VGGNet is a DCNN developed by researchers at the University of Oxford Computer Vision Group and Google DeepMind. It explores the relationship between the depth of a CNN and its performance. By repeating stacking small convolution kernels of 3×3 and the maximum pooing layers of 2×2, VGGNet successfully constructed a DCNN with the number of layers of 16 to 19. It has very good scalability, the generalization of the migration to other image data is very good.

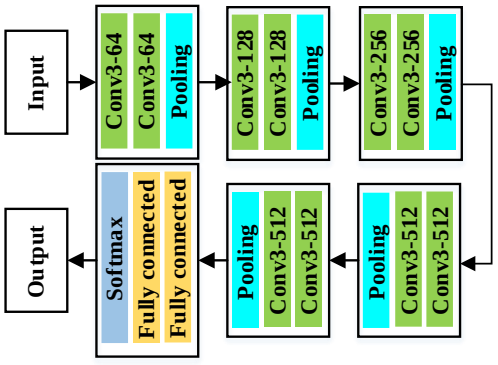


Fig. 2 The architecture of VGG16.

In this paper, the effectiveness of transfer learning using a DCNN model pre-trained on ImageNet—namely, VGG16, is discussed. The network includes 13 convolutional layers and 3 fully connected layers, which is used for classifying

different human activities based on the TF spectrograms. The architecture of VGG16 is shown in Fig. 2. The network has 138 million parameters and achieves a much lower error rate on the ImageNet test set than other nets did. The size of the input image is 224×224. We set the learning rates at 0.00001 for VGG16, the batch size is set at 50.

IV. Experiments and Simulations Analysis

To verify the effectiveness of the proposed method, the following experiments are done. In this paper, we use the Boulic kinematic model to construct a database of three types of human activities. Parameters are listed in TABLE I. Boulic model can only describe the human walking model. Here, when $0 < rv \le 0.5$, regard the activity as slow walking; when $0.5 < rv \le 1.3$, define it as fast walking; when $1.3 < rv \le 3.0$, think it really fast walking. The human height ranges from 1.6 to 1.8. The sample number of the three types of human activities are 341, 671, and 1881 respectively.

TABLE I. Parameters for Real Sample Set Generation

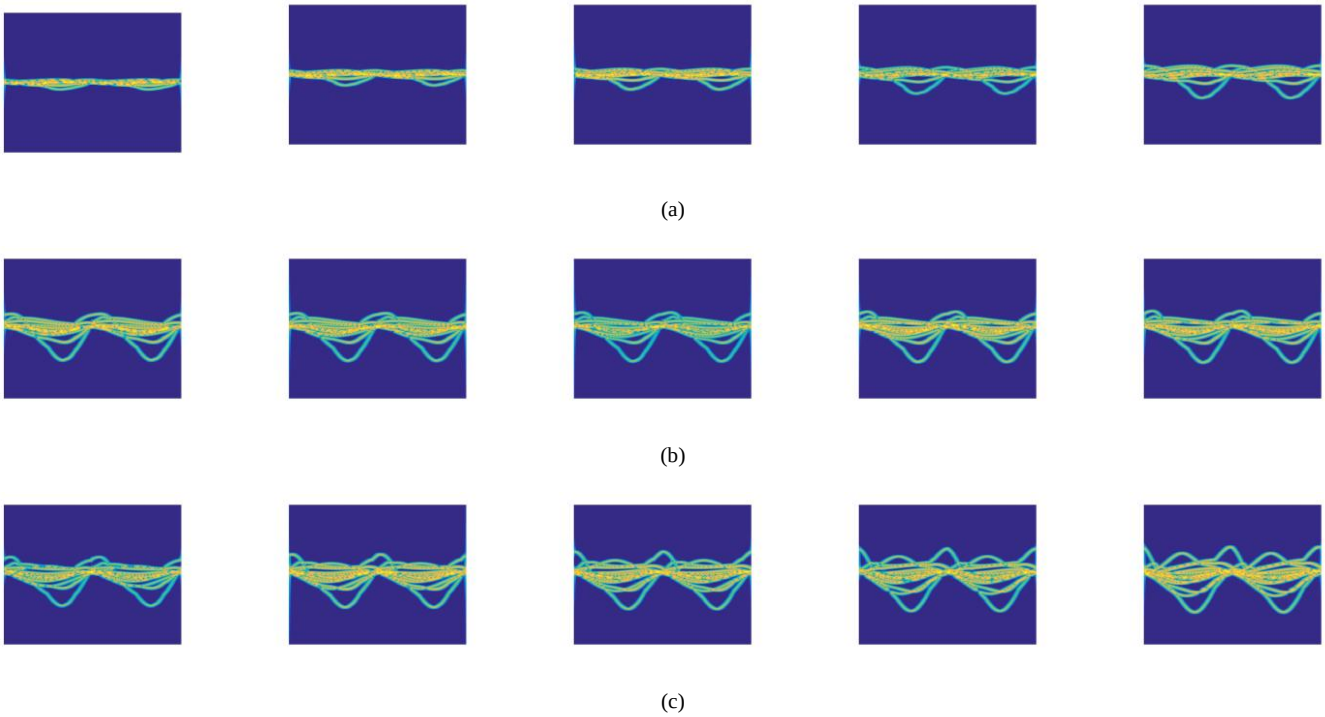
	Height(m)	Velocity(m/s)	Numbers
Slow walking	1.6:0.02:1.8	0.1:0.01:0.4	341
Fast walking		0.5:0.01:1.1	671
Really fast walking		1.3:0.01:3	1881

Fig. 3 shows some TF spectrogram samples of three different types of human activities. Fig. 3 (a) presents the TF spectrograms of slow walking human, Fig. 3 (b) gives the spectrograms of fast walking human, and Fig. 3 (c) illustrates the spectrograms for really fast walking human. From Fig. 3, the different m-D signatures of the three types of human activities are seen clearly, including the m-D frequency bandwidth, the curve, and so on.

TABLE II. Parameters for DCGAN

	Slow walking	Fast walking	Really fast walking
Number of training samples	341	671	1881
Number of epoch	260	200	115
Batch size	64	**Learning rate**	0.0005
Size of input image	600×600	**Size of output image**	512×512

TABLE II gives the parameters of DCGAN, the batch size of the input image is 64, the learning rate is set at 0.0005, the number of the epoch for the three types of human activities are 260, 200 and 115, respectively. The sizes of the input and output image are 600×600 and 512×512 respectively.



(a)

(b)

(c)

Fig. 3 Real samples of different human activities. (a) Slow walking; (b) Fast walking; (c) Really fast walking.

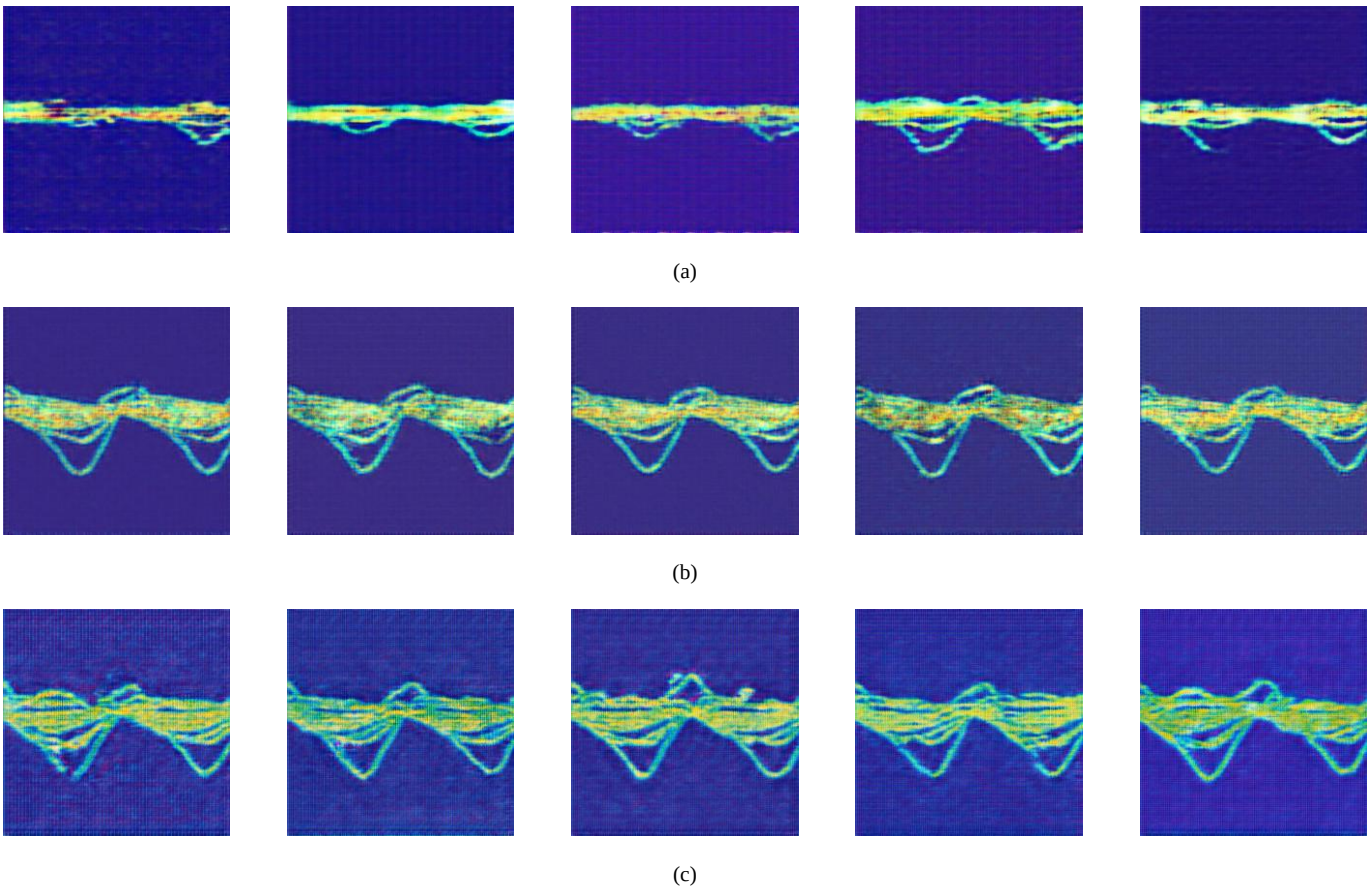


Fig. 4 Some generated samples of different human activities based on DCGAN. (a) Slow walking; (b) Fast walking; (c) Really fast walking.

TABLE III. EXPERIMENT RESULTS.

Experiment I: DCNN				
		Slow walking	**Fast walking**	**Really fast walking**
Training set	**Number**	205	403	1129
	Source	Real samples		
Testing set	**Number**	136	268	752
	Source	Real samples		
Accuracy		99.82%		

Experiment II: DCGAN+DCNN							
		Slow walking		**Fast walking**		**Really fast walking**	
Training set		Real	Fake	Real	Fake	Real	Fake
	Number	205	1295	403	1097	1129	371
	Source	Real and fake samples					
Testing set	**Number**	136		268		752	
	Source	Real samples					
Accuracy		99.91%					

--"Fake" donates the generated samples using DCGAN.

Due to the characteristics of the training samples are few, there is a big similarity between the classes. After training for many times, the values can be adjusted to minimize their own generator error state. Therefore, the network has good reconstruction ability. However, it reduces the random of generation. Therefore, for each kind of training, five models of convergence are selected for the generation of samples, which can increase the diversity of the generated samples, as shown in Fig. 4. Fig. 4 (a) is the spectrograms of slow walking humans, Fig. 4 (b) is the spectrograms of fast walking humans, and Fig. 4 (c) is the spectrograms of really fast walking human. Compared with Fig. 3, the signal-to-noise rate of the samples generated by DCGAN is lower than the real samples, but the main signatures of the real samples are obtained. This is because the new samples are achieved from a Gaussian white noise. If we use these newly generated images as training samples, this will improve the network's noise resistance.

The classification results are listed in TABLE III. In experiment I: The training sample is 60% of the total sample, and the remaining 40% is the test samples. Both the training samples and the testing samples are real samples. The accuracy is 99.82%, which verified the effectiveness of VGG16, the pre-trained DCNN on ImageNet. In experiment II, the training set consists of two parts, the real samples are the same with experiment I, the fake samples are generated by DCGAN. The number of each class is 1500. To increase

the speed of training, set the number of samples in each class to the same. The testing set is the same with experiment I. And the accuracy is 99.91%. It can be concluded that DCGAN enriches the training set and improves the accuracy of the network.

V. CONCLUSION

In this paper, a method of human activity classification using deep learning techniques is proposed. The DCGAN is utilized for expanding and enriching training set to avoid overfitting and get better training results. Some changes have been made for the CNN in the generator and discriminator to improve the quality of the image and the rate of convergence. And a transfer-learned DCNN, called VGG16, is used for human activity classification. The accuracy can be 99.91% based on the combination of DCGAN and DCNN. The simulation results verify the effectiveness of the DCGAN and transfer-learned DCNN.

ACKNOWLEDGMENT

We would like to thank the editors for taking the time to consider our paper. This paper is funded in part by China Postdoctoral Science Foundation (2017M613076, 2016M602775); in part by the National Natural Science Foundation of China (61201283, 61471284, 61571349, 61631019); in part by the NSAF under Grant U1430123, and by the Fundamental Research Funds for the Central Universities (XJS17070, K5051202001, K5051302047, NSIY031403, 3102017jg02014).

REFERENCES

[1] M. K. McDonald, "Discrimination of human targets for radar surveillance via micro-Doppler characteristics," IET Radar, Sonar & Navigation, vol. 9, no. 9, pp. 1171-1180, 2015.

[2] D. P. Fairchild and R. M. Narayanan, "Classification of human motions using empirical mode decomposition of human micro-Doppler signatures," IET Radar, Sonar & Navigation, vol. 8, no. 5, pp. 425-434, 2014.

[3] M. G. Amin, F. Ahmad, Y. D. Zhang, and B. Boashash, "Human gait recognition with cane assistive device using quadratic time-frequency distributions," IET Radar, Sonar & Navigation, vol. 9, no. 9, pp. 1224-1230, 2015.

[4] S. S. Ram, C. Christianson, Y. Kim, and H. Ling, "Simulation and analysis of human micro-Dopplers in through-wall environments," IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 4, pp. 2015-2023, 2010.

[5] V. C. Chen, The micro-Doppler effect in radar. 2011.

[6] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler dignatures using a support vector machine," IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 5, pp. 1328-1337, 2009.

[7] G. Garreau, N. Nicolaou, C. Andreou, C. D. Urbal, G. Stuarts, and J. Georgiou, "Computationally efficient classification of human transport mode using micro-doppler signatures," in 2011 45th Annual Conference on Information Sciences and Systems, 2011, pp. 1-4.

[8] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 1, pp. 8-12, 2016.

[9] Y. Kim, J. Park, and T. Moon, "Classification of micro-Doppler signatures of human aquatic activity through simulation and measurement using transferred learning," in SPIE Defense Security, vol. 10188, p. 6: SPIE, 2017.

[10] J. Guo, B. Lei, C. Ding, and Y. Zhang, "Sythetic aperture radar image synthesis by using generative adversarial nets," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 7, pp. 1111-1115, 2017.

[11] L. Stankvic, M. Dakovic, and T. Thayaparan, Time-frequency signal analysis with application. 2013.

[12] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F. Y. Wang, "Generative adversarial networks: introduction and outlook," IEEE/CAA Journal of Automatica Sinica, vol. 4, no. 4, pp. 588-598, 2017.

[13] P. L. Suárez, A. D. Sappa, and B. X. Vintimilla, "Infrared Image Colorization Based on a Triplet DCGAN Architecture," in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 212-217.