Deep Learning Methods for Personnel Recognition based on Micro-Doppler Features

Yuming Shao*, Yulong Dai, Longzhi Yuan, Weidong Chen*
Key Laboratory of Electromagnetic Space Information, Chinese Academy of Science
University of Science and Technology of China
Hefei, Anhui, P.R. China
Email: {ymshao, dyl25984, longzhi}@mail.ustc.edu.cn, {wdchen}@ustc.edu.cn

ABSTRACT
In this paper, we investigate the use of human gait micro-Doppler features for personnel recognition with a deep learning approach. Compared with conventional methods for radar-based human recognition, most existing schemes remain in discussing the distinction of different human motions. The proposed method employs a deep convolutional neural network (DCNN), which combines the inception parallel structure, to learn the necessary features of the time-frequency complex tensor that can be used to address the problem of recognizing different human subjects. Real data is collected relating to eight human subjects using a continuous wave radar operating at K-band. The experimental results illustrate that by using the human gait micro-Doppler features, the DCNN can achieve an accuracy rate of 96.9% on personnel recognition.

CCS Concepts
• **Information systems→ Mobile information processing systems**

Keywords
Personnel recognition; micro-Doppler; convolutional neural network; deep learning.

1. INTRODUCTION
Due to the increasing requirement of human identification and protection monitoring, the research on personnel recognition attracts much attention in recent years. Compared with other sensors, feature extraction and recognition based on radar, which can be operated in adverse weather and penetrate opaque obstacles have gradually become an indispensable technical method [1]-[3].

The human micro-Doppler features refer to the frequency modulations on top of the main Doppler shift caused by a moving person [4]. It provides a unique signature of the target and can be potentially used for target classification and identification. At present, most existing schemes using the unique micro-Doppler signatures remained in discussing the discrimination of different human motions, such as walking, running, crawling, and carrying objects [5]-[10]. The data processing usually consists of three steps. First of all, the micro-Doppler features are extracted by time- frequency transform methods including Short Time Fourier Transform (STFT), Wigner-Ville distribution and Cohen's class distribution [11]-[12]. After that, some traditional machine learning algorithms, such as SVM [5], naive Bayesian [9] and KNN [13], are used to learn the handcrafted features from the spectrogram. At last, a recognition model is trained time and again so that it has a better generalization ability for the test data.

All above methods are concerned with the classification of different motions. We all have the experience that the people whom we are familiar with can be distinguished by their walking gait features. Similar to radar-based human motions classification, the micro-Doppler signatures which are related to the swinging micro-movements of legs, arms, torso and head of the person can be considered as a unique feature to identify a person. Therefore, the human walking gait micro-Doppler features can also provide a viable means for personnel recognition. In [14], Svante presented an experiment which concludes that the radar micro-Doppler signatures of walking humans likely contain the information to distinguish among different human individuals. Ricci in [15] analyzed a set of experimental data of four different human targets walking and running on a treadmill using a continuous wave radar operating at 10 GHz. Fioranelli [16] investigated the classification performance of the singular value decomposition feature and of novel features based on the centroid of the micro-Doppler signatures to address the identification problem of specific individuals.

During the last decade, deep learning approaches [17] have obtained state-of-the-art results in the area of computer vision and speech recognition. In the field of radar signal processing, using the concept of deep learning, several research results have been proposed in recent years. Mason [18] proposed a recurrent neural network framework to address the autofocus problem in SAR imaging and present numerical simulations to demonstrate it. Liu introduced a novel neighborhood preserved deep neural network for polarimetric synthetic aperture radar feature extraction and classification in [19]. Feng [20] utilized a deep network called Stacked Corrective Autoencoder (S-CAE) for HRRP-based radar automatic target recognition. Kim investigated the classification of human hand gestures and motions with a DCNN network [21]-[22]. The benefit of deep neural network is that it can learn the necessary features jointly and automatically through a series of non-linear mapping, which has a great advantage over other machine learning methods which rely on handcrafted features, so that it can get excellent results in many tasks.

In this paper, we carry out a research on the use of human gait micro-Doppler features for personnel recognition with a deep convolutional neural network. The previous work in [21]-[22] utilized micro-Doppler spectrogram as the training and testing data of the DCNN, which may miss some information during the compression and transform of the spectrogram. Meanwhile, they only make use of the magnitude but lose the phase information of

ICSPS 2017, November 27–30, 2017, Auckland, New Zealand

ACM ISBN 978-1-4503-5384-7/17/11…$15.00
https://doi.org/10.1145/3163080.3163095

the features. The human gait micro-Doppler features which we use in this paper are composed of a complex tensor in which the real and imaginary parts are viewed as a channel, respectively. The deep convolutional neural network which we exploit is inspired by the GoogLeNet Inception architecture [23]-[24]. The inception architecture employs the multiple size of the convolutional kernels parallel structure in the same layer to increase the representational power of neural network. Compared with the previous research, we apply the inception architecture to personnel recognition based on human gait micro-Doppler features for the first time.

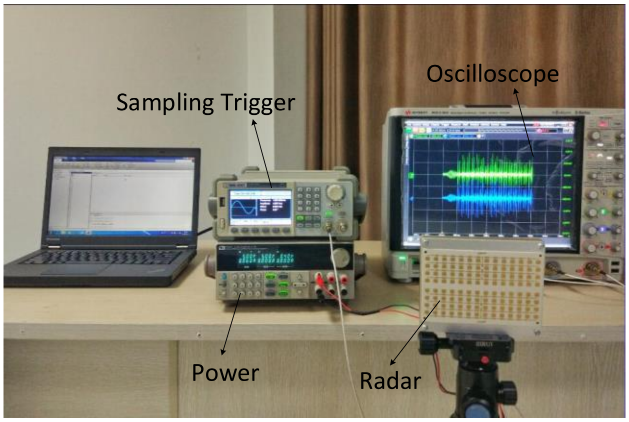

Figure 1. Schematic of the experimental configuration.

The rest of this paper is organized as follows. The walking gait measurement data collection and the experimental setup are drawn in Section II. In section III, the architecture and components of our deep convolutional neural network are proposed in detail. Section IV discusses and analyzes the training and testing results of the personnel recognition. And conclusions are presented in Section V.

2. Measurement of Human Walking and micro-Doppler Features

We employ the radar which is a K-Band VCO transceiver centered at 24 GHz. The radar includes two analogue channels, which provide the in-phase (I) and quadrature (Q) components of the received signals. The output power of the radar is 20 dBm. As shown in Fig.1, the radar is placed at an elevation of 80 cm above the ground. People walk in a normal situation towards the radar with an average distance of about 10 meters. The data of eight human subjects are collected by an Agilent Oscilloscope at a sampling rate of 4 kHz. Each target is recorded for a duration of 6 seconds and the data collection is repeated for several times. The physical characteristics of the human subjects are given in Table 1.

As the radar back-scattered signals from walking people are non-stationary, the most common choice to analyze the signals is the Short Time Fourier Transform (STFT), which is one of the most powerful time-frequency analysis methods. The equation of STFT is defined as follows

$$STFT(n,k) = \sum_{m=0}^{M-1} x(n+m)h(m)e^{-j2\pi\frac{mk}{M}} \tag{1}$$

where $h(\cdot)$ is a window function of length M, $x(\cdot)$ is the time-domain radar echo of length N, n and k represent the discrete time and frequency.

Because of the small value of the human Radar-Cross Section (RCS), the radar back-scattered signal is usually relatively weak. It is difficult to observe the micro-Doppler features directly in the

Table 1. Physical characteristics of the analyzed human targets.

Target	Gender	Height/(m)	Weight/(kg)	Leg Length/(m)	Arm Length/(m)
(a)	Male	1.76	54	1.01	0.76
(b)	Male	1.73	62	0.94	0.78
(c)	Male	1.78	68	0.95	0.81
(d)	Male	1.83	70	1.07	0.75
(e)	Male	1.80	68	1.02	0.85
(f)	Female	1.68	55	0.90	0.70
(g)	Female	1.56	53	0.87	0.58
(h)	Female	1.59	40	0.93	0.62

process of human walking after time-frequency analysis of the echo. So some contrast enhancement methods should be exploited after STFT in order to enhance the weak micro-Doppler features. Here, we use the Naka-Rushton equation [25] because it not only enhances the weak micro-Doppler amplitudes but also suppresses the small amplitudes, which usually represent the noise. The contrast enhanced equation is given as follows

$$\widetilde{W}(i,j) = \frac{W(i,j)^{r}}{W(i,j)^{r} + \mu^{r}} \tag{2}$$

where $W(i, j)$ represents the absolute value of the $STFT(n,k)$ and μ is the mean value of the patch. In this paper, the parameter r is set to 1.

The micro-Doppler features of eight human walking gaits are presented in Fig. 2. A 32 *ms* Hamming window is used in the STFT and the time step of non-overlapping samples to 5 *ms*. It can be seen that the strongest echo of each person returns from the torso and the Doppler frequency is about 200 Hz, which can be calculated that the human walking speed is about 1.25 m per second. The periodic micro-Doppler modulations surrounding the torso come from the movements of arms and legs. In contrast to the male subjects, female subjects usually swing arms more slightly when they are walking, which can be observed in the spectragram. In addition, the target (h) is the lightest in the 8 subjects, which is reflected in the spectragram that the micro-Doppler modulations surrounding the torso is weak due to her small RCS. All these differentiated features provide some references for the personnel recogniton. However, it is noteworthy that compared with the distinction of different human motions, recognizing different people by the handcrafted walking gait micro-Doppler features is still a challenging problem due to the subtile disparity.

Owing to the similarity of each person's micro-Doppler features of walking gaits, traditional supervised learning paradigms which rely on the feature extraction of the raw micro-Doppler signatures for human intervention are hard to achieve good results in the personnel recognition problem. To differentiate and recognize these features, a forceful pattern recognition technique without handcrafted features is necessary.

3. Architecture of Deep Convolutional Neural Network

In this section, we first introduce the characteristic and general layout of our Deep Convolution Neural Networks (DCNNs), and then discuss the training process of parameters in details. Finally, we will show the specific configuration in our implementation.

Generally, starting with LeNet-5 [29], CNNs have typically had a standard structure: stacked convolutional layers, pooling layers and followed by one or more fully-connected layers. The series connection into a network such as AlexNet [26] have achieved excellent results in many tasks. Recently, in the field of radar target recognition and human motions classification, most the architecture of networks which exploit deep learning methods also use the kind of series structure. In order to obtain a better representation, the usual way is to deepen the network. However, with the deepening of the kind of stacked series network, the computational bottlenecks are apparent in the case of limited training data.

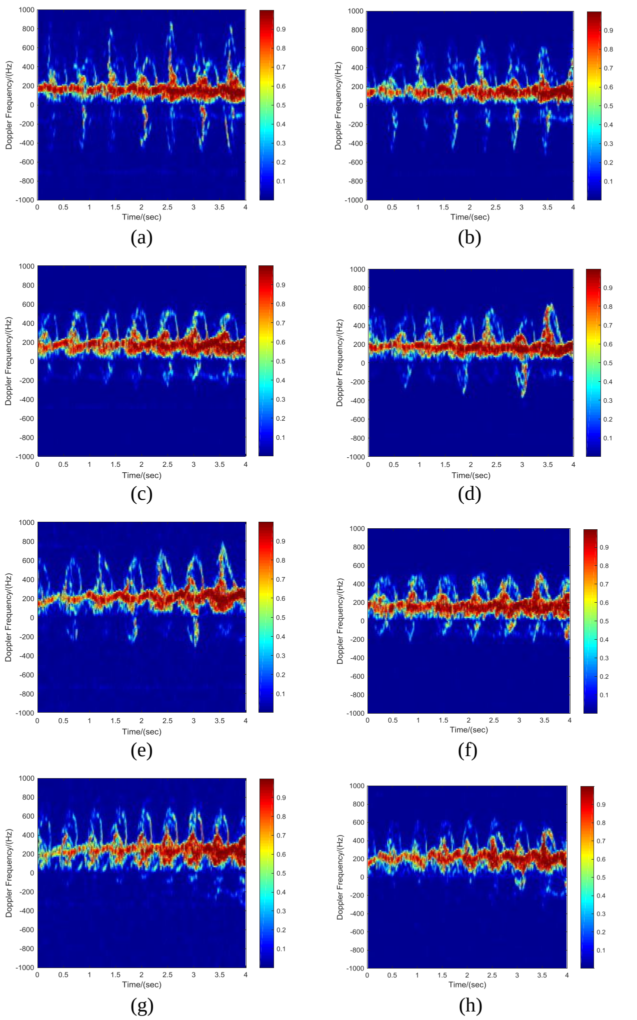

Figure 2. Micro-Doppler features of eight human walking gait.

Inspired by the GoogLeNet Inception module [23]-[24], we want to utilize the parallel connection to widen the network in order to extract more features. The inception architecture in Fig. 3 uses multi-scale convolutional kernels to convolve the same feature map which can extract multi-size features. The convolution kernel of 1×1 can significantly reduce the amount of computation by fusing the information of each channel and can increase the nonlinearity of the network.

The architecture of our proposed deep neural networks are shown in Fig. 4. The Conv2d_BN layer in our network contains a 2-D convolution layer, which is followed by a restricted linear unit (ReLU) activation function to increase the nonlinearity in the network and then using the Batch Normalization method [27] to normalize the middle outputs. Each of the Conv2d_BN layer is followed by a pooling operation which is a down-sampling process so that the extracted features are more concentrated. After the initial feature extraction, we get a series of feature maps which have an appropriate size. These feature maps are passed through the Inception module parallel structure in Fig. 3 to achieve more abstract feature extraction. At the end of our network, we add a dropout [28] layer, which is an extremely effective technique to prevent overfitting. The softmax function is applied in the output layer to solve the case of multi-class recognition after a series of feature extraction.

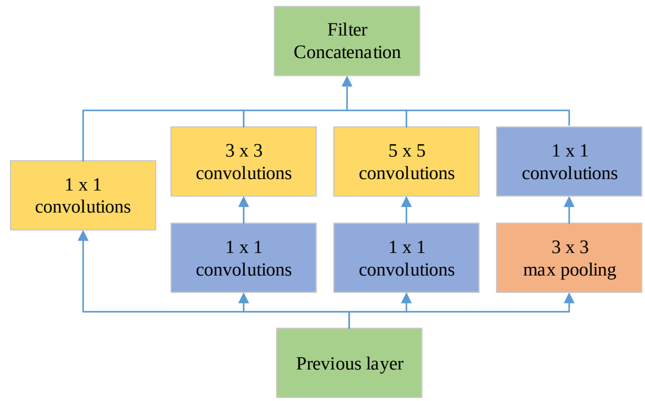

Figure 3. The inception module of GoogLeNet.

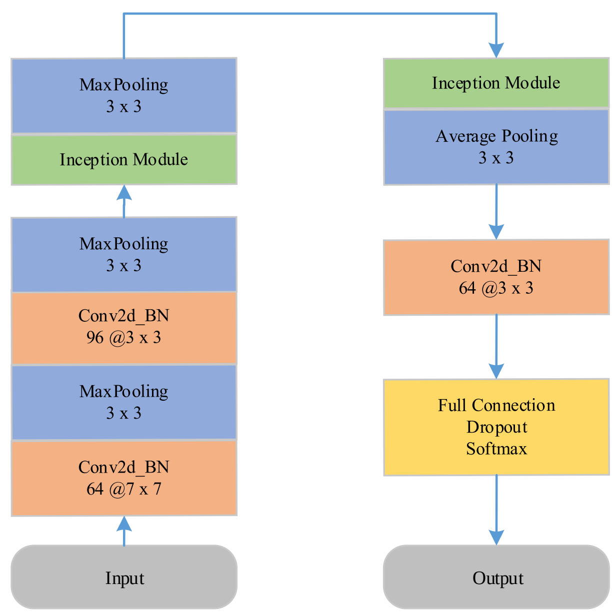

Figure 4. Architecture of our proposed neural network.

The input of our network is a 3-D tensor whose size is $225 \times 225 \times 2$, in which the first dimension represents the micro-Doppler signatures, the second dimension shows the time information, and the third dimension includes the real and imaginary parts after the time-frequency analysis. Most existing applications of neural network to solve human motions recognition problems use the real-valued spectrogram as the input of the network, whose resolution is determined by the image compression algorithm and some original information will be lost during the image compression process. Compared with the real-valued spectrogram, we employ the complex 3-D tensor which contains not only the magnitude information but also the phase information.

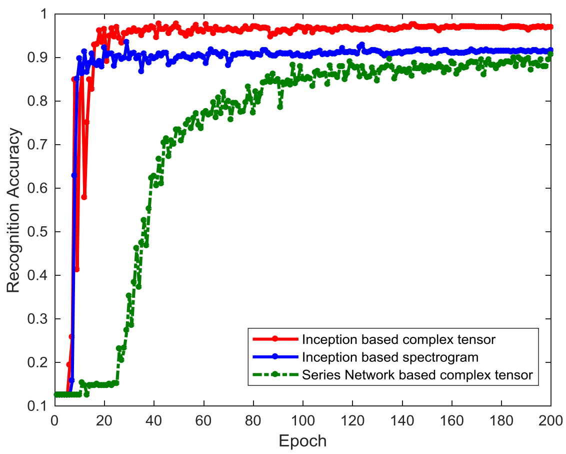

Figure 5. Recognition accuracy for test data.

4. Experimental Results and Related Analysis

In the experiment, we collected a total of eight human subjects' experimental data, including five males and three females, each of which was measured about 250 times by walking towards the radar.

We acquired a total of 2000 pieces of data approximately for a duration of 6 seconds. The data of the middle four seconds were cut into two segments and then the micro-Doppler features of the Time-Frequency (TF) distributions are obtained by STFT. We collected the data for 5 days continuously, of which the first four days were regarded as train sets and the last day were the test sets. For training, we used the open-source toolkit Keras with the Tensorflow backend. The whole network was trained using the stochastic gradient descent (SGD) method with a mini-batch size of 35 and a learning rate of 0.001. The momentum method was employed with a weight of 0.9 and a weight decay parameter of 0.004. All the weights were initialized from Gaussian distributions with zero mean and a standard deviation of 0.01. The training time with 200 epochs was about 10 minutes on average.

The recognition accuracy rate for test data is presented in Fig. 5, and the abscissa indicates the number of epochs. The solid line represents the results by using the inception parallel structure network where the red and blue curves express the results by using the complex tensor and the real-valued spectrogram as the inputs respectively. The green dotted curve denotes the result of using the traditional stacked series convolution neural network. It shows that compared with the series convolutional neural network, our approaches, i.e. the red curve, which combined the inception parallel structure have improved the recognition accuracy by nearly 7 percentage points and the convergence is faster. Furthermore, the complex tensor inputs have also improved the recognition results compared with the real-valued spectrogram. Because the former can obtain the magnitude and phase of the micro-Doppler features by combining the real and imaginary channels, which is more abundant than the real-valued spectrogram.

For analysis of misclassification, the confusion matrix of test sets using our approach is presented in Table 2. Each row in the confusion matrix represents the actual human subjects, and each column denotes the subject predicted by our network. Label (a) - (h) stands for the people listed in Table 1, respectively. From the table, it can be seen that (a) and (h) have an excellent recognition results indicating that the gait features of these two people are quite significant and the daily walking posture is relatively similar. The male subject (b) has the lowest recognition accuracy in this experiment and we note that all the false identities are predicted to the subject (a). The reason may be due to the fact that his walking gait features are similar to the subject (a). The more likely reason we supposed is that there is a difference in the gait of daily walking for some people and it may lead to the accuracy lower than other subjects.

Table 2. Confusion matrix for test set

Subjects	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	Acc
(a)	80	0	0	0	0	0	0	0	1.000
(b)	6	74	0	0	0	0	0	0	0.925
(c)	0	0	79	0	0	1	0	0	0.986
(d)	1	2	0	77	0	0	0	0	0.963
(e)	2	0	1	0	77	0	0	0	0.963
(f)	0	0	1	3	0	75	0	1	0.938
(g)	0	0	0	0	0	1	78	1	0.975
(h)	0	0	0	0	0	0	0	80	1.000
Average									0.969

The performance of our network which is compared with other convolution neural networks and traditional supervised learning paradigms is listed in Table 3. It can be seen that the deep learning methods, using the neural networks, have better effects in the recognition task of feature extraction than the traditional handcrafted feature classification method such as SVM. Furthermore, adding the inception parallel module to the network can further improve the accuracy of the recognition, and the use of complex tensor instead of real-valued spectrogram as the input of the network can also increase the micro-Doppler feature information to enhance the accuracy.

Table 3. Our networks versus conventional methods

Methods	Average Recognition Accuracy
SVM: spectrogram-based	0.798
Series CNN[22]: spectrogram-based	0.864
Series CNN[22]: complex tensor-based	0.906
Our network: spectrogram-based	**0.916**
Our network: complex tensor-based	**0.969**

5. Conclusion

In this paper, the micro-Doppler features have been applied to radar-based personnel recognition problems using deep convolutional neural networks. Real data measurements were collected corresponding to eight human subjects walking towards the radar. Compared with most existing stacked series convolutional neural network applying in the field of radar signal processing, we add the inception parallel module to the network and choose the complex micro-Doppler tensors as the input for training instead of real-valued spectrogram, which can obtain more detailed features in order to achieve the purpose of personnel recognition. The network's architecture, training process and common rules for setting parameters are described in this paper. By employing the inception parallel module, human subjects can be successfully recognized with the accuracy of 96.9%.

Inspired by this research, we plan to measure more human subjects including both male and female as well as at all ages. Meanwhile, we hope to optimize the network architecture so that we can use less training costs, i.e., computing resources and training samples, to get more robust personnel recognition performance.

6. REFERENCES

[1] Z. Sun, J. Wang, J. Sun, and P. Lei, “Parameter estimation method of walking human based on radar micro-Doppler,” *IEEE Radar Conference*, pp. 0567-0570, 2017.

[2] F. Fioranelli, M. Ritchie, S. Z. Gürbüz, and H. Griffiths, “Feature Diversity for Optimized Human Micro-Doppler Classification Using Multistatic Radar,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, pp. 640-654, 2017.

[3] E. Cippitelli, F. Fioranelli, E. Gambi, and S. Spinsante, “Radar and RGB-Depth Sensors for Fall Detection: A Review,” *IEEE Sensors Journal*, vol. 17, pp. 3585-3604, 2017.

[4] V. C. Chen, F. Li, S. S. Ho, and H. Wechsler, “Micro-Doppler effect in radar: phenomenon, model, and simulation study,” *IEEE Transactions on Aerospace and electronic systems*, vol. 42, pp. 2-21, 2006.

[5] Y. Kim and H. Ling, “Human activity classification based on micro-Doppler signatures using a support vector machine,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, pp. 1328-1337, 2009.

[6] R. M. Narayanan and M. Zenaldin, “Radar micro-Doppler signatures of various human activities,” *IET Radar, Sonar & Navigation*, vol. 9, pp. 1205-1215, 2015.

[7] D. P. Fairchild and R. M. Narayanan, “Multistatic micro-doppler radar for determining target orientation and activity classification,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, pp. 512-521, 2016.

[8] S. Björklund, H. Petersson, and G. Hendeby, “Features for micro-Doppler based activity classification,” *IET radar, sonar & navigation*, vol. 9, pp. 1181-1187, 2015.

[9] F. Fioranelli, M. Ritchie, and H. Griffiths, “Centroid features for classification of armed/unarmed multiple personnel using multistatic human micro-Doppler,” *IET Radar, Sonar & Navigation*, vol. 10, pp. 1702-1710, 2016.

[10] B. Tekeli, S. Z. Gurbuz, and M. Yuksel, “Information-theoretic feature selection for human micro-doppler signature classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, pp. 2749-2762, 2016.

[11] 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, pp. 1224-1230, 2015.

[12] B. Jokanovic, M. G. Amin, and F. Ahmad, “Effect of data representations on deep learning in fall detection,” *IEEE Sensor Array and Multichannel Signal Processing Workshop*, pp. 1-5, 2016.

[13] Z. Zhang, R. Zhang, W. Sheng, Y. Han, and X. Ma, “Feature extraction and classification of human motions with LFMCW radar,” *IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition*, pp. 1-3, 2016.

[14] S. Björklund, H. Petersson, and G. Hendeby, “On distinguishing between human individuals in micro-Doppler signatures,” *14th International Radar Symposium*, pp. 865-870, 2013.

[15] R. Ricci and A. Balleri, “Recognition of humans based on radar micro-Doppler shape spectrum features,” *IET Radar, Sonar & Navigation*, vol. 9, pp. 1216-1223, 2015.

[16] F. Fioranelli, M. Ritchie, and H. Griffiths, “Performance analysis of centroid and SVD features for personnel recognition using multistatic micro-Doppler,” *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 725-729, 2016.

[17] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, pp. 504-507, 2006.

[18] E. Mason, B. Yonel, and B. Yazici, “Deep learning for radar,” *IEEE Radar Conference*, pp. 1703-1708, 2017.

[19] H. Liu, S. Yang, S. Gou, D. Zhu, R. Wang, and L. Jiao, “Polarimetric SAR Feature Extraction With Neighborhood Preservation-Based Deep Learning,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, pp. 1456-1466, 2017.

[20] B. Feng, B. Chen, and H. Liu, “Radar HRRP target recognition with deep networks,” *Pattern Recognition*, vol. 61, pp. 379-393, 2017.

[21] Y. Kim and B. Toomajian, “Hand Gesture Recognition Using micro-Doppler Signatures with Convolutional Neural Network,” *IEEE Access*, vol. 4, pp. 7125-7130, 2016.

[22] 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, pp. 8-12, 2016.

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, and D. Anguelov, “Going deeper with convolutions,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9, 2015.

[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818-2826, 2016.

[25] K. Naka and W. Rushton, “S-potentials from colour units in the retina of fish (Cyprinidae),” *The Journal of physiology*, vol. 185, pp. 536, 1966.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, pp. 1097-1105, 2012.

[27] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International Conference on Machine Learning*, pp. 448-456, 2015.

[28] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929-1958, 2014.

[29] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, “Backpropagation applied to handwritten zip code recognition,” *Neural computation*, vol. 1, pp. 541-551, 1989.