Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks

Youngwook Kim, *Senior Member, IEEE*, and Taesup Moon, *Member, IEEE*

***Abstract*—We propose the use of deep convolutional neural networks (DCNNs) for human detection and activity classification based on Doppler radar. Previously, proposed schemes for these problems remained in the conventional supervised learning paradigm that relies on the design of handcrafted features. Whereas these schemes attained high accuracy, the requirement for domain knowledge of each problem limits the scalability of the proposed schemes. In this letter, we present an alternative deep learning approach. We apply the DCNN, one of the most successful deep learning algorithms, directly to a raw micro-Doppler spectrogram for both human detection and activity classification problem. The DCNN can jointly learn the necessary features and classification boundaries using the measured data without employing any explicit features on the micro-Doppler signals. We show that the DCNN can achieve accuracy results of 97.6% for human detection and 90.9% for human activity classification.**

***Index Terms*—Convolutional neural network, deep learning, human activity classification, human detection, micro-Doppler.**

I. Introduction

TARGET classification using micro-Doppler signatures has found many applications in defense, surveillance, and private sector [1], [2]. Specifically, Doppler radar has been widely used for moving-object detection and target classification because it can suppress clutter and detect only a nonstationary target. To recognize and classify a target, micro-Doppler signatures produced from various non-rigid-body motions of a target can be a key feature for exploitation [3]. Recently, the human detection and tracking problem has been extensively addressed in conjunction with the unique micro-Doppler signature. The micro-Doppler signature that is time-varying can be clearly observed in a joint time—frequency domain.

Several research efforts have been exerted to recognize the micro-Doppler signatures for target classifications. In the early study stage, spectral analysis of micro-Dopplers has been focused on distinguishing targets without time-dependent information [4]. This method could identify targets that have a distinctive Doppler shift compared with those of the other methods. Then, the time-varying signatures extracted from a spectrogram were exploited to recognize various human activities in [5]. The use of empirical mode decomposition was also successful in recognizing target types [6]. The principal component analysis and linear discriminant analysis have been also used to extract feature vectors [7]. The linear predictive code was also suggested for real-time processing because the computational cost could be significantly reduced [8].

Whereas the aforementioned schemes achieved high accuracy results in the considered problems, their approach remained in the conventional supervised learning paradigm, i.e., they require preprocessing of raw micro-Doppler signals to devise discriminative features necessary for the classification algorithms. Such dependence on the domain knowledge of micro-Doppler signals limits the scalability of the proposed algorithms to other research problems. We therefore consider an alternative deep learning approach to overcome such limitation.

In this letter, we propose the use of deep convolutional neural networks (DCNNs) to recognize micro-Doppler signatures in spectrograms for target classification problems. Deep learning algorithms, which typically use hierarchical neural networks, have recently revolutionized several applications such as image or speech recognition; they significantly outperform the previous state-of-the-art schemes that mainly relied on domain knowledge-based features. The main reason for such success is the ability of deep learning algorithms to jointly learn the features and classification boundaries directly from raw input data. The most informative features for a given classification problem can be automatically learned from the data while possessing the ability to capture these features that may otherwise be missed. To that end, we directly apply the DCNN to a micro-Doppler spectrogram with two objectives: 1) human detection and 2) human activity classification. For the first objective, targets, which include a human, a dog, a horse, and a car, are measured by Doppler radar, and the DCNN is trained for the classification according to the generated spectrogram. Second, the same data set used in [5] is tested for the human activity classification. Seven different human activities measured by Doppler radar are used as targets, and the DCNN classification performance is investigated. To the best of our knowledge, the deep learning approach has not been used in the radar community, particularly for target recognition with Doppler signatures. We present brief backgrounds on deep learning and CNN, application of

Manuscript received August 4, 2015; revised September 10, 2015; accepted October 8, 2015. Date of publication November 2, 2015; date of current version December 24, 2015.

Y. Kim is with the Department of Electrical and Computer Engineering, Fresno State Lyles College of Engineering, California State University, Fresno, CA 93740 USA (e-mail: youngkim@csufresno.edu).

T. Moon is with the Department of Information and Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu 711-873, Korea (e-mail: tsmoon@dgist.ac.kr).

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

Digital Object Identifier 10.1109/LGRS.2015.2491329

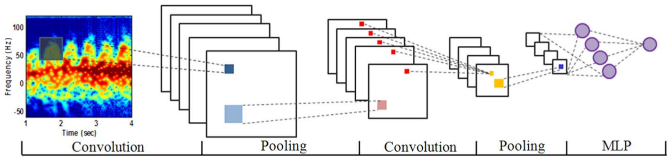

Fig. 1. Convolution in the CNN.

the CNN to the Doppler data, our experimental results, and discussions.

II. DCNNs

Deep learning is a subfield of machine learning that has recently gained large interest. In contrast to relying on the extracted features from data, the deep learning algorithms can find features and classification boundaries through optimizing a certain loss function. For this purpose, these algorithms employ a *deep* neural network structure, which stacks multiple layers of simple neural network architecture, to extract hierarchical abstractions and generalization from the data. Whereas the main concept of deep learning has been around for a few decades, it has lately regained the spotlight owing to the achievement of excellent empirical performance in several different domains of applications such as speech recognition, image recognition, and natural language processing [9]–[11]. The main reason for such renaissance is attributed to several facts: the advent of high-performance computing processors such as the Graphic Processing Unit (GPU) and the continuation of algorithmic innovation. For a more detailed overview of deep-learning algorithms, we refer the readers to [12] and the references therein.

DCNN is one of the most successful deep learning algorithms. It is based on the classical convolution neural network devised by LeCun *et al.* [13] in the late 1980s. It is a supervised learning algorithm that attempts to learn mapping between the input data point and its corresponding label provided by human annotators. The hierarchical structure of the DCNN, which is exemplified in Fig. 1, is inspired by the visual cortex of a human brain that efficiently recognizes objects. The DCNN consists of three main components. The first component is the *convolution filter* that works on small local receptive fields of input data in a sliding-window fashion. Each filter can be considered as a specific feature detector. In each layer, multiple convolution filters exist that work in parallel. The second component is the *nonlinear activation* function imposed on the convolution filter output that enables nonlinear transformation of a data space so that the discrimination among classes can become easier. The conventional choice for the activation function was the sigmoid function: $f(x) = 1/(1 + \exp(-x))$. More recently, the Restricted Linear Units (ReLU), i.e., $f(x) = \max(0, x)$, was shown to achieve better empirical results when used as an activation function [14]. The third component is the *pooling* that reduces the data size. The final prediction can become robust to the translation of input data through pooling. Fig. 2 shows the schematic of the 4 × 4 convolution filter and the 2 × 2 pooling operation. The combination of convolution filters, nonlinear activation, and pooling is regarded as one

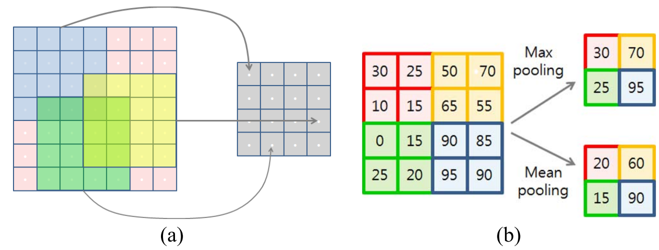

Fig. 2. (a) Process of applying a 4 × 4 convolution filter to the input data (in pink) to generate the output (in gray). (b) Examples of 2 × 2 pooling (max or mean pooling) that reduces the data dimension by half.

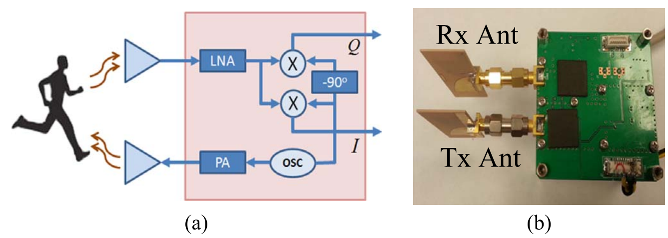

Fig. 3. (a) Configuration of Doppler radar. (b) Photograph of the radar.

layer in the DCNN, and multiple layers (sometimes, more than 20) are consecutively employed in modern CNNs; hence, the term *deep* CNN. When the DCNN is used as a classifier, the final few layers assume the form of a usual perceptron, i.e., all input nodes of the layer are connected to all output nodes. Fig. 1 shows a simple DCNN architecture with two convolution layers and one final fully connected layer.

The DCNN parameters, i.e., the coefficients of convolution filters and the final fully connected layers, are trained by backpropagation with stochastic gradient descent (SGD). To speed up the optimization process, a momentum method [15] is used as a standard choice. In addition, to prevent overfitting, dropout [16] is widely used as a regularization scheme; it randomly omits hidden nodes with a predefined probability that is independent of each training sample. Through the dropout, co-adaptation among the neural network nodes can be mitigated. The DCNN has recently revolutionized the image recognition community by significantly outperforming the previous state-of-the-art methods. However, most of the advanced DCNNs developed so far have focused on the natural RGB image recognition, and the effectiveness of the DCNN in the recognition of micro-Doppler signatures in a spectrogram has not been extensively investigated yet.

III. Radar Target Classification Through Micro-Doppler Signatures Using the CNN

A. Human Detection

We applied the DCNN to the measured Doppler data to classify target types. Several targets, including a human, a dog, a horse, and a car, were measured by Doppler radar operating at 7.25 GHz when the targets approached the radar. The radar block diagram and a picture are shown in Fig. 3. The measurements were performed in an outdoor environment under a line-of-sight condition, as shown in Fig. 4. Each target

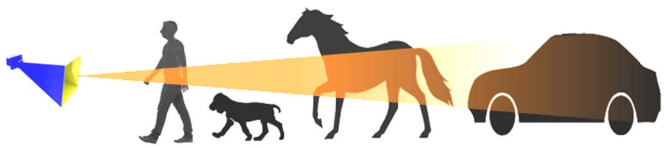

Fig. 4. Outdoor measurements. (a) Human. (b) Dog. (c) Horse. (d) Car.

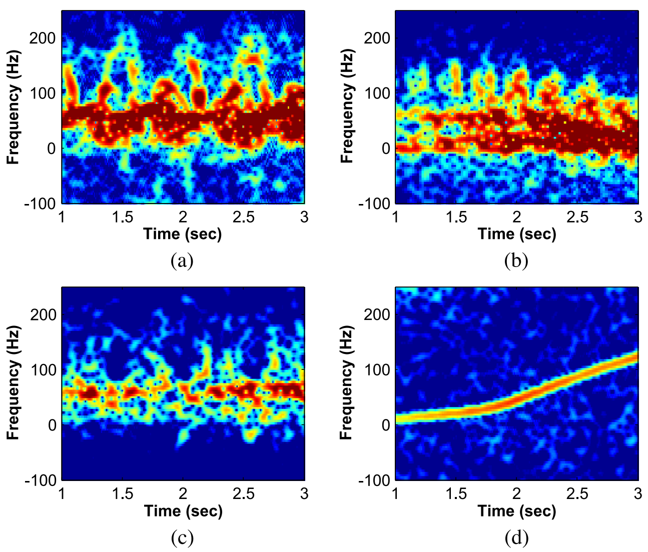

Fig. 5. Sample spectrograms. (a) Human. (b) Dog. (c) Horse. (d) Car.

was measured ten times, and four spectrograms with a 2-s time window were extracted, resulting in 40 data per target. For a joint time—frequency analysis, the measured data were processed with the short-time Fourier transform [17], and their sample spectrograms are shown in Fig. 5. We set the fast Fourier transform size to 256 ms and the overlapping time step to 10 ms in this study considering the speed of human motions.

As shown by the spectrograms, different targets present their own unique micro-Doppler signatures. In [18], human detection among other targets has been addressed by investigating the physical characteristics of the targets. The classification was mainly based on the estimation of the length of a leg and stride size. The limitation of the approach described in [18] was that the technique could not discriminate a human against a horse because they have similar leg lengths and strides, although the micro-Dopplers show different features, as shown in Fig. 5.

To utilize the micro-Doppler signatures in this study, we employed the 2-s spectrogram itself as input to the DCNN. Thus, we interpret the spectrogram classification as an image recognition problem. By using the 2-s window, the periodic micro-Doppler signatures could be captured. The size of the spectrogram was normalized to 100×100. Among the 160 data, 80% of the spectrograms of each target were used as the training data set, and the other 20% were used as the validation data set. Because the size of the training data set is not large, we used relatively small CNNs for this experiment; we used two convolution layers, where each layer had four convolution filters of size of 5×5. For pooling, we used the 2×2 max pooling for the first layer and the 4×4 max pooling for the second layer. Furthermore, we had one fully connected layer that directly connects the output of the second pooling and the target classes.

TABLE I
ACCURACIES OF THE DCNN FOR EACH FOLD AND THEIR AVERAGE

Fold 1	Fold 2	Fold 3	Fold 4	Average
93.75%	100%	100%	96.87%	97.6%

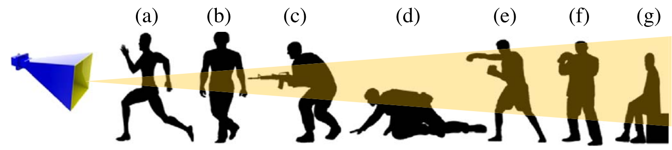

Fig. 6. Setup for human activity measurements.

The usual sigmoid function was used as activation functions. We used the MATLAB toolbox developed by Palm [19] for the experiment, and the DCNN can be successfully trained in MATLAB at a reasonable time cost. The number of epochs was set to 100. Because the number of data and the size of the CNN are not big, we used an Intel i5 2.27-GHz CPU with a 4-GB memory. The training time was 127 s. The resulting classification accuracy was 97.6%, as listed in Table I.

B. Human Activity Classification

We investigated the performance of the DCNN in the classification of human activities employing the data set used in [5]. The data were collected using a Doppler radar test bed operating at 2.4 GHz. A human being moving directly toward the radar was measured in an indoor environment under line-of-sight conditions. Twelve humans were measured for 3 s as they performed seven different activities. The activities consist of (a) running, (b) walking, (c) walking while holding a stick, (d) crawling, (e) boxing while moving forward, (f) boxing while standing in place, and (g) sitting still. Each activity was measured four times per subject, and three spectrograms were extracted from each measurement, resulting in 1008 data points. The size of the extracted spectrogram was 300×140. The measurement setups are shown in Fig. 6, and the sample spectrograms are shown in Fig. 7.

We used a fourfold cross validation to evaluate the classification performance of the DCNN, similar to that performed in [5]. The training and test sets in each fold contained 756 and 252 samples, respectively. The number of convolution filters in each layer, size of the convolution filter, and number of hidden nodes in the fully connected layer were hyperparameters chosen via the cross validation. ReLU was used for the activation function, and a 2×2 max pooling was used.

Because the data and network sizes we employed are significantly larger than those of the first experiment, we used the open-source toolkit Caffe [20], which uses the NVIDIA GPU and CUDA library (e.g., cuDNN [21]) to speed up the computation. For learning, we used the mini-batch SGD with a learning rate of 0.001 and a batch size of 84. The momentum method was also used with a weight of 0.9, and dropout was applied for the final fully connected layer with a probability of 0.5. We used the NVIDIA GeForce GTX Titan Black edition GPU (with a 6-GB memory) and 2.5 GHz Intel Xeon CPU E5-2609 v2 in our experiments. The training time for each fold with 400 epochs was about 1420 s on average.

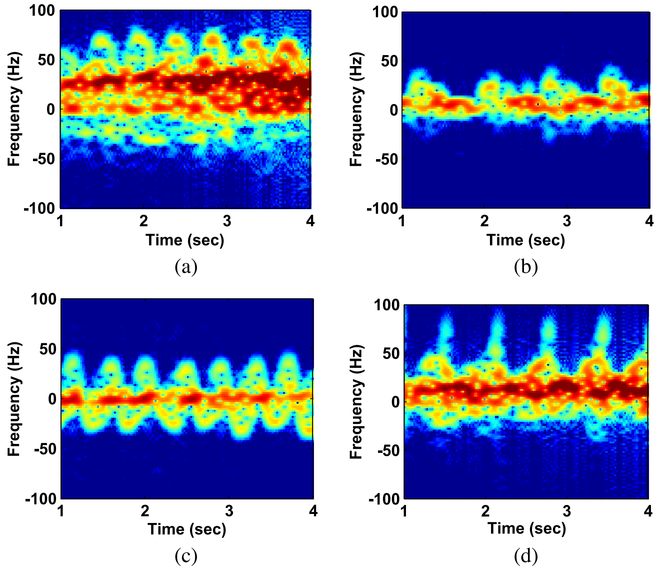

Fig. 7. Sample spectrograms of human activities. (a) Running. (b) Crawling. (c) Boxing still. (d) Boxing forward.

TABLE II
ACCURACY RESULTS OF THE DCNN FOR EACH FOLD AND THEIR AVERAGE

Fold 1	Fold 2	Fold 3	Fold 4	Average
92.9%	85.3%	93.3%	89.7%	90.3%

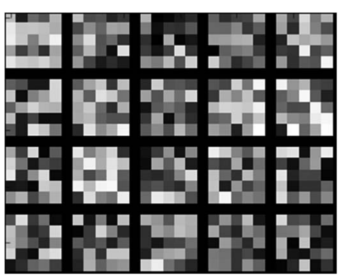

Fig. 8. Visualization of the 20 convolution filters of size 5×5 in the first layer.

After a heuristic search, we report the result of the best model: three convolution layers, where each layer has 20 filters with a size of 5×5, and two fully connected layers with 500 hidden nodes in the first fully connected layer. Table II lists the summary of the accuracy results we obtained from each fold, as well as their average of 90.3%. Whereas our result is slightly worse than that of [5], we believe that it shows the potential of the DCNN in micro-Doppler-based target classification in that we have not used any preprocessing for feature extraction and extensive hyperparameter tuning such as random search [22].

We show the learned 20 convolution filters of the first layer in Fig. 8. The visualization of the convolution filters indicates hierarchical structures, whereas obtaining the physical insights into our case is difficult. In our future research, we plan to further attempt to visualize the higher layers' convolution filters as in [23], so that we may obtain better insights into the learned representations of DCNNs.

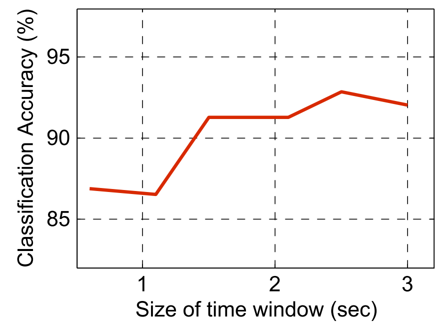

Fig. 9. Accuracy results of Fold 1 with varying time window in the spectrogram data. Note that we do not lose much even by halving the time window of the spectrogram.

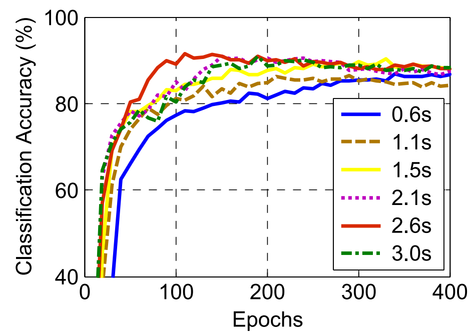

Fig. 10. Test error curves for Fold 1 with varying time windows. The x-axis index corresponds to training epochs, and the y-axis corresponds to the accuracy. Note that the test errors in Fold 1 converge at approximately 200 epochs.

Moreover, we varied the time window of the spectrogram data and tested how it affects the classification accuracy. We picked the Fold 1 training/test data and compared the test results with the data with 140 (3.0 s), 118 (2.6 s), 100 (2.1 s), 70 (1.5 s), 50 (1.1 s), and 30 (0.6 s) time stamps from the beginning of the spectrogram data. We trained the same data set for five times and show the averaged accuracy in Fig. 9. We observe that from only half of the data (i.e., 70 time stamps), 99.24% of the accuracy of that using the full data can be achieved.

Finally, Fig. 10 shows the six test error curves of Fold 1 that yielded the results shown in Fig. 9. We observe that at least approximately 200 epochs are needed for the test errors to converge while avoiding significant overfitting. One phenomenon we find is that the accuracy for 2.6 s is slightly higher than that of 3.0 s around 100 epochs. While intriguing, we think the phenomenon could be simply due to the variance in the data since the gap is not very significant and the trend does not last when the curve converges.

IV. CONCLUSION

In this letter, the DCNN has been applied for target classification problems based on the micro-Doppler characteristics in a spectrogram. The DCNN was employed to efficiently extract and recognize micro-Doppler features. By using the DCNN, human beings can be successfully classified with 97.6% accuracy among other targets, including dog, horse, and car. The seven human activities were successfully classified with 90.9% accuracy. In this letter, we did not use any explicit domain knowledge for extracting features, and the spectrogram itself served as input data to the DCNN. We believe our results show a potential of deep learning for a number of applications in radar signal processing problems. The proposed method may also have some limitations. Since the shape of the micro-Doppler signature is the key for classification, the performance can degrade if there exist variations due to the irregularity in motions. In addition, the computational complexity of DCNNs is usually higher than that of data-driven models from regular machine learning algorithms. Therefore, the computational complexity of DCNNs should be carefully considered in applications that require real-time processing.

Acknowledgment

The authors would like to thank Dr. H. Ling for permitting the use of the radar data.

References

[1] D. Tahmoush and J. Silvious, "Radar micro-Doppler for long range front-view gait recognition," in *Proc. IEEE 3rd Int. Conf. Biometrics, Theory, Appl. Syst.*, Washington, DC, USA, Sep. 28–30, 2009, pp. 1–6.

[2] P. van Dorp and F. C. A. Groen, "Human walking estimation with radar," *Proc. Inst. Elect. Eng.— Radar, Sonar Navig.*, vol. 150, no. 5, pp. 356–365, Oct. 2003.

[3] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-Doppler effect in radar: Phenomenon, model, and simulation study," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 2–21, Jan. 2006.

[4] A. G. Stove and S. R. Sykes, "Doppler-based automatic target classifier for a battlefield surveillance radar," in *Proc. IEEE Int. Radar Conf.*, Edinburgh, U.K., Oct. 15–17, 2002, pp. 419–423.

[5] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, May 2009.

[6] D. Fairchild and R. Narayanan, "Classification of human motions using empirical mode decomposition of human micro-Doppler signatures," *IET Radar, Sonar, Navig.*, vol. 8, no. 5, pp. 425–434, Jun. 2014.

[7] J. Li, S. Phung, F. Tivive, and A. Bouzerdoum, "Automatic classification of human motions using Doppler radar," in *Proc. IEEE IJCNN*, Brisbane, Qld., Australia, Jun. 10–15, 2012, pp. 1–6.

[8] J. Rios and Y. Kim, "Application of linear predictive coding for human activity classification based on micro-Doppler signatures," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1831–1834, Oct. 2014.

[9] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 29, pp. 82–97, Nov. 2012.

[10] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, vol. 25, pp. 1090–1098.

[11] T. Mikolov, A. Deoras, D. Povey, L. Burget, and J. Cernocky, "Strategies for training large scale neural network language models," in *Proc. Autom. Speech Recognit. Understand.*, Waikoloa, HI, USA, Dec. 11–15, 2011, pp. 196–201.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[13] Y. LeCun *et al.*, "Handwritten digit recognition with a back-propagation network," in *Proc. Adv. Neural Inf. Process. Syst.*, 1990, pp. 396–404.

[14] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, Fort Lauderdale, FL, USA, 2011, pp. 315–323.

[15] Y. Nesterov, "A method of solving a convex programming problem with convergence rate O(1/sqr(k))," *Soviet Math. Doklady*, vol. 27, pp. 372–376, 1983.

[16] N. Nrivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdino, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

[17] V. Chen and H. Ling, *Time–Frequency Transforms for Radar Imaging and Signal Analysis*. Norwood, MA, USA: Artech House, 2002.

[18] Y. Kim, S. Ha, and J. Kwon, "Human detection using Doppler radar based on physical characteristics of targets," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 289–293, Jul. 2014.

[19] R. B. Palm, "Prediction as a candidate for learning deep hierarchical models of data," M.S. thesis, Dept. Informat. Math. Model., Tech. Univ. Denmark, Kongens Lyngby, Denmark, 2012.

[20] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. ACM MM*, 2014, pp. 675–678.

[21] S. Chetler *et al.*, "cuDNN: Efficient primitives for deep learning," unpublished manuscript, 2014. [Online]. Available: http://arxiv.org/abs/1410.0759

[22] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, 2012.

[23] M. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland, Sep. 6–12, 2014, pp. 818–833.