TL-FALL: CONTACTLESS INDOOR FALL DETECTION USING TRANSFER LEARNING FROM A PRETRAINED MODEL

Hamidreza Sadreazami †, *Member, IEEE*, *Miodrag Bolic* †, *Senior Member, IEEE*,
and Sreeraman Rajan ‡, *Senior Member, IEEE*

†School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada
‡Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada
E-mail: {hsadreaz@uottawa.ca, mbolic@eecs.uottawa.ca, sreeramanr@sce.carleton.ca}

ABSTRACT

A new radar-based fall detection method is proposed using the recent advances in deep neural networks. An ultra-wideband radar is used to monitor human daily activities and identify the occurrence of falls from radar data. A transfer learning approach is employed based on a pre-trained model on ImageNet dataset to realize a robust feature extraction from radar data. The architecture and depth of the model are fine-tuned to radar time-frequency representations. From the results, it is observed that the proposed transfer learning based method can achieve a detection accuracy for fall incidents higher that those of the other methods even using a small-size dataset.

Index Terms— Ultra-wideband radar, transfer learning, time-frequency analysis, fall detection.

I. INTRODUCTION

Falling down is major public health concern for seniors that might lead to serious injury or even death. Developing efficient and robust technologies for fall detection is of great importance. Several fall detection systems have been developed in past years based on different detection approaches based on wearable sensors and and video cameras [1]. Contactless radar-based indoor monitoring is an emerging field in recent years [2]–[5], since it is non-invasive and privacy-friendly as compared to wearable devices and vision-based techniques [1]. For instance, in [6] and [7], by incorporating micro-Doppler signatures, human activity recognition and gait abnormalities have been investigated. In [8], a Doppler radar has been used to devise a fall detection method using features extracted from wavelet domain. A fall detection method has been developed in [9] by extracting a set of features from time-frequency representation of the radar Doppler signatures. An auto-encoder has been employed in [10] to extract features from radar time-frequency representations. A radar signal recognition has been proposed in [12] using extracted features by a restricted Boltzmann machine.

In this work, a radar-based fall detection method is proposed when feature learning is equipped with learnt features from a pre-trained model. The motivation behind using a pre-trained model is that there is not adequate labeled training data to train a network built from scratch. The proposed transfer learning-based approach uses time-frequency representations, derived from an ultra-wideband radar signals, to fine-tune the pre-trained VGG16 network. This network was constructed by stacking convolutional blocks and trained on ImageNet. To effectively extract features from the radar data, all the weights in the VGG networks are frozen except for the last convolutional block. The network is then retrained and a global average pooling is used to flatten signal at the end of the convolutional blocks. This pooling strategy prevents the network from overfitting. A softmax classifier is used in the output layer to identify the probability of different classes. The performance of the proposed transfer-learning-based approach is investigated against that provided by the other existing approaches.

The paper is organized as follows. The proposed fall detection method is presented in Section II. Results and conclusion are given in Sections III and IV, respectively.

II. FALL DETECTION APPROACH

The proposed fall detection approach is comprised of preprocessing and feature learning stages, as presented in the following.

II-A. Preprocessing

A supervised learning fall detection approach from radar return signals is studied. The development kit employed in the experiments is Xethru X4M03, operating in the range of $5.9 - 10.3$ GHz. Fall and non-fall activities are included in the dataset collected in a room environment, simulating a realistic elderly house. Each 15-second experiment is digitized at a rate of 200 samples/second. The resulting scattering matrix consists of columns of spatial samples from different ranges and rows of observations recorded at different time intervals.

After removing clutter, by mean removal, from the radar returns, radar time series signal is derived from the scattering matrix by averaging over all the columns. A time series $x[n]$ is then obtained. A time-frequency representation is then obtained by applying the short-time Fourier transform [13] to the radar time series as given by

$$STFT[n,k] = \sum_{r=-\infty}^{\infty} x[r]W[r-n]\exp(-j2\pi rk/N), \quad (1)$$

where $W[.]$ is a Hamming window of size 256 samples, $k = 0, 1, ..., N-1$ is the frequency index and N is the number of frequency points. 80% overlap between adjacent windows is considered. The spectrogram is obtained by taking the squared magnitude of $STFT[n,k]$ as

$$SP = |STFT[n,k]|^2 \quad (2)$$

It is noted that the spectrogram SP represents the energy distribution of motion signature at specific time and frequency. Fig. 1 shows time-frequency spectrograms of the falling down and standing up activities, where the horizontal axis is time and vertical axis is frequency. It is seen from this figure that the energy content of these activities are distinguishable in their time-frequency signatures.

The obtained spectrogram set is denoted by $\{(SP_i, L_i)\}$, for $(1 \leq i \leq N_{tr})$, where N_{tr} is the number of data samples and L_i is the label for the ith experiment; "fall" or "non-fall".

II-B. Feature Learning and Classification

Deep learning has shown a great potential for applications requiring automatic feature extraction and classification. It is known that transfer learning provides robust model with respect to data variability and produces a generalizable classifier [14]. In this work, the use of a pre-trained network is proposed for fall detection through fine-tuning from a large scale dataset to small scale and domain-specific dataset. More specifically, the VGG16 [15] network, which has been trained on ImageNet [16], is employed. In the following, we explain this network.

VGG16 constructs a deeper network structure than the well-known AlexNet [17]. VGG16 network is composed of convolution modules followed by fully-connected layers. This network is constructed by repeatedly stacking a 3×3 convolutional filter and a pooling layer of size 2×2. In order for a fine-tuning process to be realized, different layers of the network are initialized by properly-trained weights. In other words, adding randomly initialized fully-connected layers on top of the bottleneck features of the pre-trained VGG network may result in a large update for gradient value in each training iteration, and thus, the previously-learned weights may be destructed. In view of this, the last convolutional block of VGG16 network is also fine-tuned, i.e., weights will be updated. It is noted that fine-tuning the entire network is not recommended as it may cause

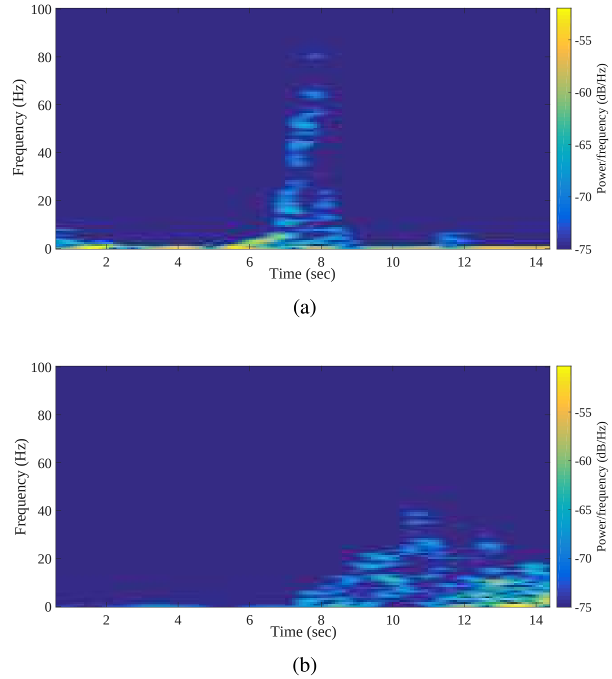

Fig. 1. Time-frequency analysis: spectrograms resulted from different activities; (a) Falling down, (b) Standing up.

overfitting. This may be due to the fact that the number of parameters in the model is high, and thus, the network stores large amount of information, i.e., large entropic capacity. In general, the low-level convolutional blocks learn less abstract features than those in the higher levels. In view of this, the first four convolutional blocks are kept fixed and only the last convolutional block is fine-tuned. Fig. 2 shows the block diagram of the proposed fall detection method using transfer learning.

III. RESULTS

In this Section, the results of the proposed fall detection method is presented. 206 radar return samples were collected in a cluttered room from five different subjects including 121 fall and 85 non-fall activities. The activities are as follows: (1) the subject walks in front of the radar line of sight and falls down at different distances to the radar, (2) the subject stands still and falls down at different distances to the radar, (3) the subject stands and falls down with with 90 degree angle to the radar, (4) the subject lies down with or without side rolling or other movements, and (5) the subject stands up from lying down position in front of the radar and with 90 degree angle to the radar. A time series is derived by averaging over all columns of the radar back-scattered matrix. Radar time series is used to derive the radar time-frequency image, i.e., spectrogram. The spectrograms are

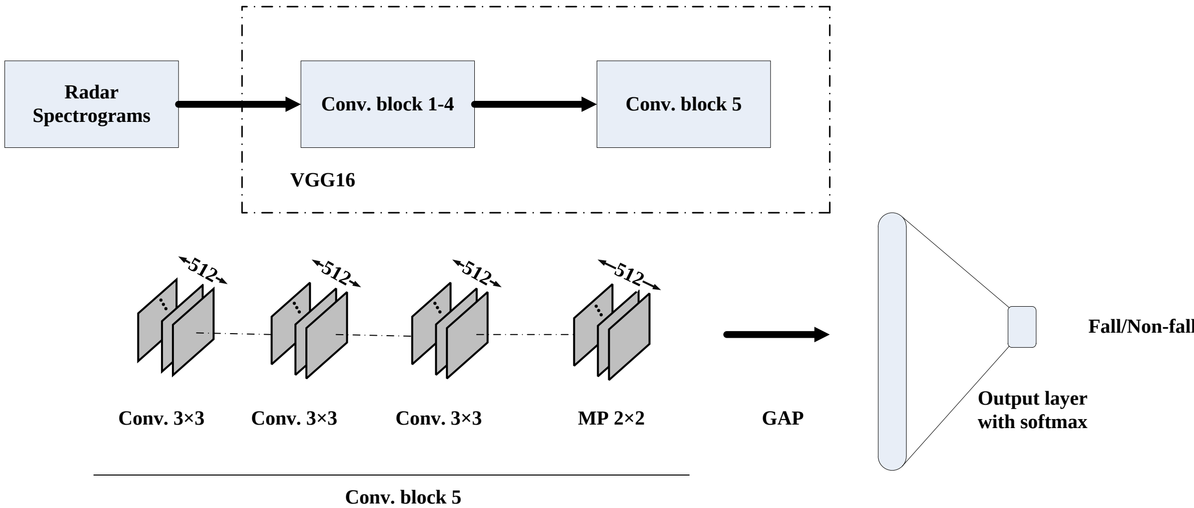

Fig. 2. Block diagram of the proposed radar-based fall detection method using transfer learning.

Table I. Accuracy, precision and sensitivity values obtained using the proposed transfer learning-based method, when fine-tuning the VGG16 model with or without convolutional layers in a 3-fold cross-validation sense.

	Metrics		
Method	Accuracy	Precision	Sensitivity
2 Conv+MP+GAP+Output	95.64	96.12	96.73
1 Conv+MP+GAP+Output	95.64	96.12	96.73
MP+GAP+Output	89.80	90.72	92.37
GAP+Output	89.32	90.89	91.66

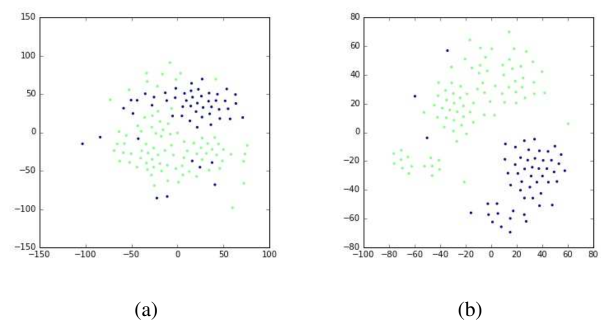

Fig. 3. Visualization of bottleneck feature clusters extracted from the pre-trained VGG16 network for training set, (a) before and (b) after fine-tuning.

then employed to retrain the last convolutional block and top layer classifier of the VGG16 pre-trained network.

Table I gives classification accuracies obtained using the VGG16, when the model is fine-tuned for the top-level layers with or without a convolutional layer, i.e., the last two convolutional layers in convolutional block 5, shown in Fig. 2. It is seen from this table that fine-tuning the model with one or two convolutional layers is better than only fine-tuning the top-level layers like max-pooling (MP) and global average-pooling (GAP) layers. This can be attributed to the fact that the radar data, i.e., spectrograms, is different from ImageNet dataset used to train VGG16, and thus, transferring the weights of lower layers of the network to extract low-level features results in a better classification accuracy than directly classifying the data based on the entire trained networks. It should also be noted that fine-tuning with more number of convolutional layers may result in over-fitting.

It should be noted that a global spatial average pooling layer is added to the last convolutional block to flatten the feature map and to avoid overfitting. The use of a GAP was proposed in [18] as a replacement for the fully-connected layers. The GAP precludes the need for a separate flatten layer and produces better results. The learning rate for fine-tuning is set to 0.001, since the magnitude of the updates should be kept small to preserve the previously-learned features. The batch size is 16. In the output layer, the softmax activation is used. The loss function employed is the categorical cross-entropy. The network is retrained by minimizing the error using the stochastic gradient decent optimizer via backpropagation. Fig. 3 shows the bottleneck features of the pre-trained VGG16 network extracted from training set before and after fine-tuning. T-distributed stochastic neighbor embedding [19] is used for visualization purpose. It is seen

Table II. Accuracy, precision and sensitivity values obtained using the proposed transfer learning based method and those provided by LSVM, GSVM and KNN in a 3-fold cross-validation.

	Metrics		
Method	Accuracy	Precision	Sensitivity
LSVM	80.01	82.64	83.34
GSVM	79.13	85.12	80.46
KNN	78.64	82.64	81.30
Proposed	**95.64**	**96.12**	**96.73**

from this figure that fine-tuning has considerable impact on feature learning as it adjusts the upper layers of a pre-trained network very precisely to fit with new dataset under study, i.e., spectrograms.

In order to investigate the performance of the proposed transfer learning-based method, other methods based on K-nearest neighbors (KNN), Gaussian support vector machine (GSVM) and linear support vector machine (LSVM) were also implemented. To this end, the radar time series obtained from the radar back-scattered matrix is used as input for these classifiers. Different K values were tested to optimize the performance of a KNN-based classifier. It is found experimentally that $K = 5$ results in higher classification metrics. Table II gives classification accuracy, precision and sensitivity of various methods in a 3-fold cross validation. It can be seen from this table that the transfer learning-based method is superior to other methods as evidenced by its higher classification metrics.

IV. CONCLUSION

A new fall detection method has been proposed. The proposed method has been developed by using the signals received from an ultra-wideband radar sensor. In particular, time-frequency representations from the radar time series have been derived and used to fine-tune the last convolutional block of the pre-trained VGG16 network. To this end, the network weights up to the last convolutional block have been kept unchanged and the rest are used for retraining with the spectrograms of the radar returns of human activities. Experimental results have shown that the proposed method is superior to other methods in discriminating falls from non-fall activities. In addition, it has been observed that a transfer learning approach can be promising for feature extraction from the radar returns and classifying them. Future direction of our research is toward the use of a pre-trained network directly applied to the radar time series.

V. REFERENCES

[1] A. T. Ozdemir and B. Barshan, “Detecting falls with wearable sensors using machine learning techniques,” *Sensors,* vol. 14, pp. 10691-10708, 2014.

[2] M. G. Amin, Y. D. Zhang, F. Ahmad and K. C. D. Ho, “Radar signal processing for elderly fall detection: The future for in-home monitoring,” *IEEE Signal Processing Magazine,* vol. 33, no. 2, pp. 71-80, 2016.

[3] I. Nejadgholi, H. Sadreazami, S. Rajan and M. Bolic, “Classification of Doppler radar reflections as preprocessing for breathing rate monitoring,” *IET Signal Processing,* vol. 13, no. 1, pp. 21-28, 2019.

[4] H. Sadreazami, M. Bolic and S. Rajan, “On the use of ultra wideband radar and stacked LSTM-RNN for at home fall detection,” in *Proc. IEEE Life Sciences Conference,* pp. 255-258, 2018.

[5] H. Sadreazami, M. Bolic and S. Rajan, “Residual network-based supervised learning of remotely sensed fall incidents using ultra-wideband radar,” in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS),* 2019.

[6] S. Bjorklund, H. Petersson and G. Hendeby, “Features for micro-Doppler based activity classification,” *IET Radar, Sonar & Navigation,* vol. 9, no. 9, pp. 1181-1187, 2015.

[7] A.-K. Seifert. M. G. Amin and A. M. Zoubir, “New analysis of radar micro-Doppler gait signatures for rehabilitation and assisted living,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* pp. 4004-4008, 2017.

[8] B. Y. Su, K. Ho, M. J. Rantz and M. Skubic, “Doppler radar fall activity detection using the wavelet transform,” *IEEE Transactions on Biomedical Engineering,* vol. 62, no. 3, pp. 865-875, 2015.

[9] Q. Wu, Y. D. Zhang, W. Tao and M. G. Amin, “Radar-based fall detection based on Doppler time-frequency signatures for assisted living,” *IET Radar, Sonar & Navigation,* vol. 9, no. 2, pp. 164-172, 2015.

[10] B. Jokanovic, M. G. Amin and F. Ahmad, “Radar fall motion detection using deep learning,” in *Proc. IEEE Radar Conference (RadarConf),* pp. 1-6, 2016.

[11] Y. Wang, Y. Chen, Md. Z. A. Bhuiyan, Y. Han, Sh. Zhao and J. Li, “Gait-based human identification using acoustic sensor and deep neural network,” *Future Generation Computer Systems,* vol. 86, pp. 1228-1237, 2018.

[12] D. Zhou, X. Wang, Y. Tian and R. Wang, “A novel radar signal recognition method based on a deep restricted Boltzmann machine,” *Engineering Review,* vol. 37, no. 2, pp. 165-171, 2017.

[13] L. Stankovic, M. Dakovic and T. Thayaparan, *Time-frequency signal analysis with applications,* Artech House, 2013.

[14] S.-J. Kim, et al., “Deep transfer learning-based hologram classification for molecular diagnostics,” *Scientific Reports*, vol. 8, no. 1, pp. 17003, 2018.

[15] K. Simonyan, and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv:1409.1556*, 2014.

[16] O. Russakovsky, et al., “Imagenet large scale visual recognition challenge,” *arXiv:1409.0575,* 2014.

[17] A. Krizhevsky, I. Sutskever and G. E. Hinton, “Imagenet, classification with deep convolutional neural networks,” in *Proc. Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.

[18] L. Min, Q. Chen and Sh. Yan, “Network in network,” *arXiv:1312.4400*, 2013.

[19] L. V. D. Matten and G. Hinton, “Visualizing data using t-SNE”, *Journal of machine learning research*, vol. 9, pp. 2579-2605, 2008.