DCNN Based Human Activity Recognition Using Micro-Doppler Signatures

Ajay Waghumbare
School of Computer Engg. &
Mathematical Sciences
DIAT, Pune
Student Member IEEE
Waghumbareajay@gmail.com

Upasna Singh
School of Computer Engg. &
Mathematical Sciences
DIAT, Pune
Senior Member IEEE
upasna.diat@gmail.com

Nihit Singhal
School of Computer Engg. &
Mathematical Sciences
DIAT, Pune
nihitsinghal2@gmail.com

***Abstract*—In recent years, Deep Convolutional Neural Networks (DCNNs) have demonstrated some promising results in classification of micro-Doppler (m-D) radar data in human activity recognition. Compared with camera-based, radar-based human activity recognition is robust to low light conditions, adverse weather conditions, long-range operations, through wall imaging etc. An indigenously developed "DIAT-μRADHAR" human activity recognition dataset comprising micro-Doppler signature images of six different activites like (i) person fight punching (boxing) during the one-to-one attack, (ii) person intruding for pre-attack surveillance (army marching), (iii) person training (army jogging), (iv) person shooting (or escaping) with a rifle (jumping with holding a gun), (v) stone/hand-grenade throwing for damage/blasting (stone-pelting/grenades-throwing), and (vi) person hidden translation for attack execution or escape (army crawling and compared performance of this data on various DCNN models. To reduce variations in data, we have cleaned data and make it suitable for DCNN model by using preprocessing methods such as re-scaling, rotation, width shift range, height shift range, sheer range, zoom range and horizontal flip etc. We used different DCNN pre-trained models such as VGG-16 , VGG-19 , and Inception V3. These models are fine-tuned and the resultant models are performing efficiently for human activity recognition in DIAT-μRadHAR human activity dataset.**

Keywords*—*micro-Doppler signatures, human activity recognition, Deep Convolutional neural network based classification

I. Introduction

Human activity recognition(HAR) one of the most important and emerging study areas in the field of human-computer interface, intelligent driving, human-robot interaction, health monitoring, and public violent protests' /attacks' early detection, care for the elderly, child monitoring, education, entertainment, environmental support, sports, etc., [1], [2]. People performs various activities in different ways and some activities are composed of subactivities i.e. complex activities due to these reasons, HAR is challenging task. The motion of human body generates micro-Doppler effects such as boxing, walking, running,etc., and it can be used to characterize human activity [3]. Most of HAR systems are available based on surveillance video, infrared, thermal, and acoustic sensors but the limitations of vision based sensors are that they require light, operate at short target range, narrow field of view, and there performance decreases in adverse weathers etc. Collecting data for vision based sensors require more work and take too much time to complete. Radar based human activity recognition has several advantages: it gives good accuracy even without illumination, robust against adverse weather conditions, capable of atmospheric diffusion, and has the privacy of a person, and so on. Micro Doppler radars can detect motion and suppress stagnant clusters in the background with ease. Most common human activities are (i) boxing, (ii) army marching, (iii) army jogging, (iv) jumping with holding gun, (v) stone throwing, (vi) army crawling [4], [5]. Recently, to recognize human activities like walking with rifle, crawling, jumping, running, carrying laptop, patient walking, etc., using radar-based systems are becoming popular. In [4], authors reported the classification of human activities like running, walking, and crawling classes using SVM classifier. Deep learning based methods automatically learns necessary features for classification tasks and achieve higher accuracy [6]-[10].

Many researchers have used machine learning techniques; like principal component analysis (PCA), multilayer perceptron, linear predictive coding (LPC), and support vector machine (SVM), etc., which are retrieve/classification algorithms [11]. However, features extracted by these traditional techniques for classification are not efficient and limited by the intricacy and prior information of categorization problems [12]. Also, performance of models using these techniques degrades if there is similarity between classes of m-D signatures. SVM classifier mixed with decision tree to perform multi class classification [13]. In [14] authors summarized the improvements of HAR models happened according to the CV domain and also discuss the new trends and achievements in the improvement of HAR models. An thorough analysis of all vision-based HAR related information is provided in this survey and also shown us the recent development made in this area. At the same time, they also discussed the main problems and challenge we can face and also the scope for future developments. Due to aforementioned problems, deep learning (DL) gains researchers attention since it does automatic feature extraction, self regulation, and self control. In [15], researchers employed transfer learning in image classification for small aerial targets using deep convolutional neural network (DCNN) i.e., AlexNet, VGG16, and VGG19 on Defence Institute Of Advanced Technology (DIAT) created a dataset of small aerial targets using 10Ghz (CW) Radar consisting of 4849 spectrogram images of 6 classes. Analyzed VGG16 as Deep Convolutional Neural Network to classify 25,000 images on dogs and cats with and without standard data augmentation and achieved a 16.20% jump in accuracy by using data augmentation [16]. In this paper, we have compared the performance of DCNN models by fine tuning them. The experimental results show that performance of fine tuning models are better than pre-trained models.

978-1-6654-9291-1/22/$31.00 ©2022 IEEE

The rest of this paper is organized as follows: data description and preprocessing described in Section II. Section III gives details of model implementation. Performance of models are discussed in Section IV, and finally, this paper concluded in Section V.

II. Data description and preprocessing

A. Data Description

To investigate performance of models in radar based human activity recognition DIAT-RadHAR dataset is used [17]. To generate diversified dataset, males and females of varying weights performed suspicious activities in open field at different orientations: 0 °, ±15°, ±30°, ±45° in front of radar, at a distance of 10m to 0.5 km, and all experiments are conducted for 3 s to 4 s. Micro-Doppler signatures of different human activities are generated such as: (i) boxing, (ii) walking, (iii) jogging, (iv) jumping, (v) throwing, and (vi) crawling. Thus, dataset (DIAT-RadHAR) containing micro Doppler signatures of suspicious activities, consist of total 3780 images for 6 classes.

B. Data Preprocessing

The preprocessing aids to remove the noise from images to improve the training process of CNN. The size of images in the dataset is different for images. The first step is to reduce the size of images in dataset. The preprocessing module support DCNN to prepare input spectrogram pictures. The input images are first re-sized to the dimension 224 x 224 x 3, then normalized with a scale factor of 1/255 by the preprocessing module. Pixel normalization is done so that all parameters of input remain same everywhere. By applying "ImageDataGenerator" , data is augmented using re-scaling, rotation, width shift range, height shift range, sheer range, zoom range, and horizontal flip etc. Operations. These minor variations does not change data so much and also increase the size of data to train model more precisely.

III. Methodology

DIAT-μRadHAR dataset contains micro-Doppler signatures of 6 classes in the form of spectrograms, which are given to the model as an input image which processed according to the model. Processed data is fed to fine-tuned DCNN models like VGG-16, VGG-19, Inception-V3 etc. for training which is then analyzed and evaluated model performance.

A. VGG-16 Implementation

Input image size for VGG-16 is 224 x 224 in RGB which is passed through VGG model. Pre-trained VGG-16 [18] model consist of total 16 layers (13 convolutional , 3 fully connected layers). It is better to partially tune the network especially with small training dataset due to over-fitting concerns. VGG-16 consist of two fully connected layers. We fine tune the model by freezing 13 layers, replacing first (4096 feature vector), second (4096 feature vector) fully connected layer replaced by 128, 64 dense layer respectively and Global Average Pooling (GAP) added before first dense layer to reduce number of parameters so that model becomes light weight as well as to prevent overfitting. At the end, 1000 (number of classes) replaced by 6 (activity classes). To avoid neuron dying problem, ReLU is replaced with LeakyRelu activation function. LeakyReLU does not make negative values to zero. Softmax activation function is used to predict the output in terms of probability. After freezing first 13 layers, remaining layers are retrained. Introduced dropout layer before softmax activation to introduce stochasticity because of that, model achieves good generalization ability. To optimize performance of model, RMSProp is used with learning rate of 0.0001 to limit over-fitting. Fine tuned model consist of 14.7 million parameters which are very less compare to 138 million parameters of base model. Due to this, model performs faster compare to base model.

B. VGG-19 Implementation

Size of 224 x 224 RGB image was given as input to this network which means that the matrix was of shape 224 x 224 x3. VGG19 is just like VGG16 the only difference is that it have three extra layers than VGG-16 model. It consist of two fully connected layers each 4096 feature vector, replaced by 128, 64 dense layer. Global Average Pooling introduced before dense layer to reduce number of parameters of network. Six classes are predicted using softmax activation function, which gives output in probability. After every layer, ReLU replaced by LeakyReLU activation function to avoid dying ReLU problem. All the hyper-parameters are same as fine-tuned VGG-16 model. Performance of model optimized using RMSProp optimizer with learning rate 0.0001.

Just like VGG-16, dropout is used before softmaz activation function to introduce randomness and prevent overfitting. First 16 layers are freezed and remaining layers are retrained. Base VGG-19 model consist of about total 22 million parameters but our fine tuned model consist of 20 million parameters which makes model lightweight and model converges faster.

C. Inception-V3

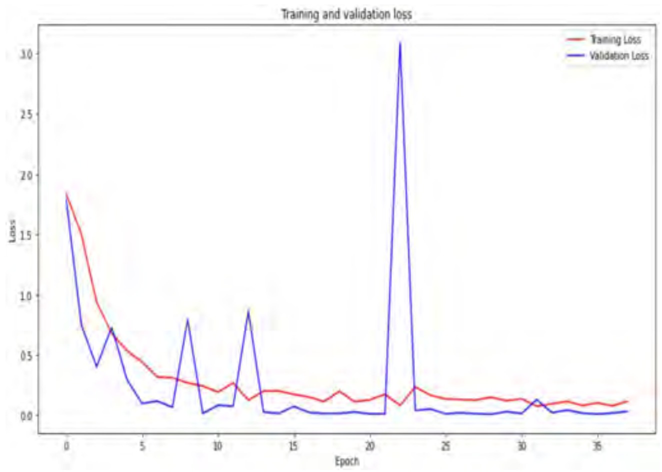
Base model of Inception-V3 is already very deep because of its numerous blocks so to reduce number of parameters of model we applied Global Average Pooling (GAP) [19]. To avoid dying ReLU problem, LeakyReLu is used. Base model consist of one fully connected layer (1024 feature vector) replaced by two dense layers of 128, 64 neurons which reduces parameters of model. GAP is used before first dense layer. GAP used not only to reduce number of parameters but also act as a regularizer which is used to prevent over-fitting. In this model, only two dense layers are retrained and remaining layers are freezed. Fine-tuned model is optimized using RMSProp optimizer with learning rate 0.0001. We added dropout layer before six states classifier i.e. softmax activation function, which gives good generalization and prevent over-fitting. After fine tuning model, consist of 22 million parameters which are 2 million less compare to base model.

IV. Evaluation and results

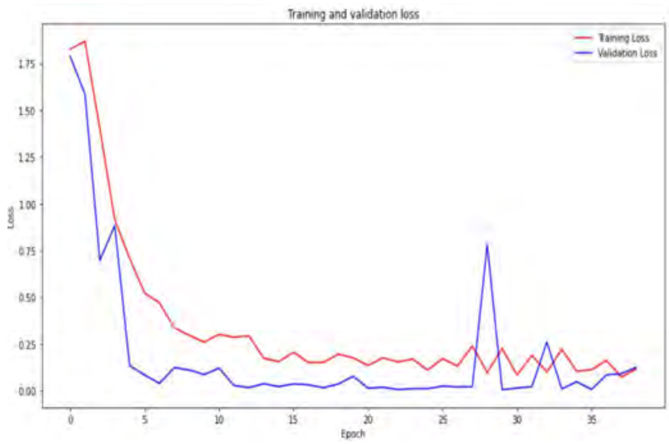
In this section, we explore the performance of the different DCNN models like VGG-16, VGG-19, Inception-V3 etc. and show the validation accuracy as well as loss curves. Lastly, we show the confusion matrix of all fine-tuned models.

A. Training And Validation Results: VGG-16

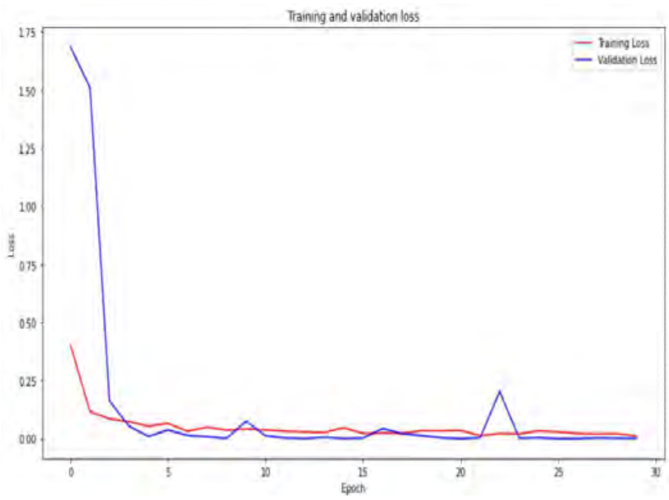
Fine tune VGG-16 model has achieved 98.95% accuracy on validation data and 99.00% accuracy on training data. Fig. 1 (a) shows loss curve of training and validation data on VGG-16 model where X-axis shows number of epochs and Y-axis shows loss values for every epoch. Training and validation loss of model is 9%, 7% respectively. As it can be observed, at epoch number 23 loss function value increases suddenly after that it decreases and give steady performance. Model trained for 40 epochs on 3780 images, where data divided in 80% for training and 20% for validation. Fig. 2 (a) shows training and validation accuracy curves. From fig. 2 (a), we can say that at epoch 23 accuracy of model decreases after that curve becomes stable.



(a)



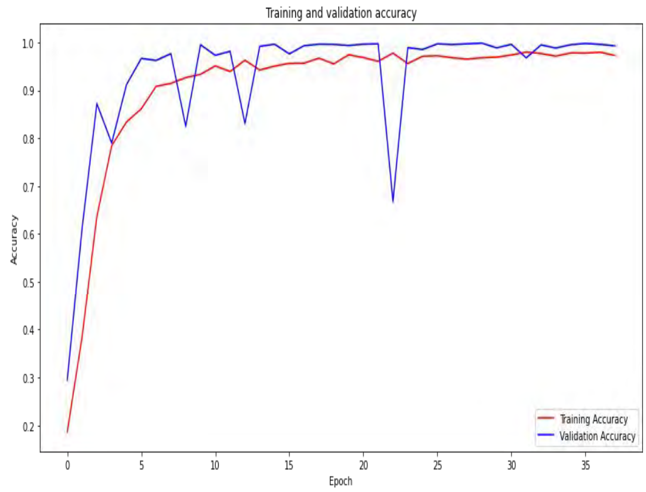
(b)



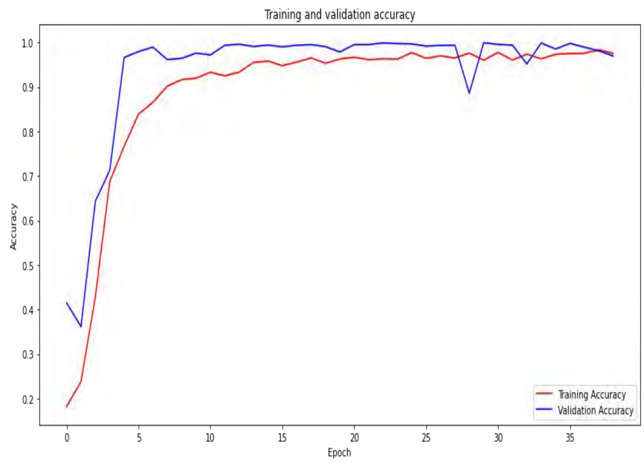
(c)

Fig. 1. Training and validation loss curve : (a) VGG-16 (b) VGG-19 (c) Inception-V3

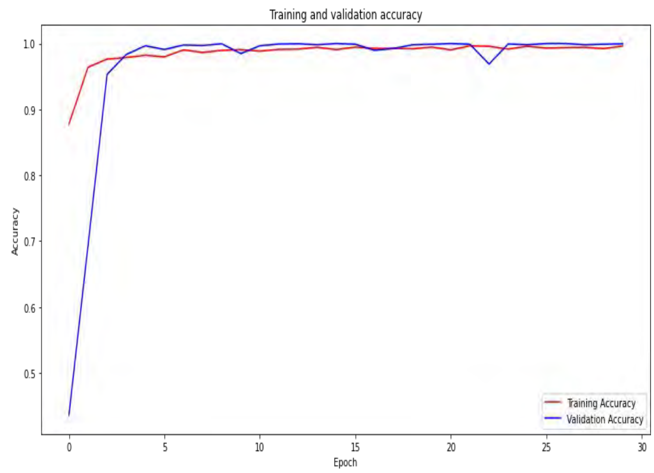
Table I shows list of hyperparameters used to train the model. After that, model tested on 572 unseen images. Fig. 3 (a) shows confusion matrix of VGG-16 model on unseen data. We can say that model performed well on jumping, jogging, walking classes with 100% accuracy.



(a)



(b)



(c)

Fig. 2. Training and validation accuracy curve : (a) VGG-16 (b) VGG-19 (c) Inception-V3

Model achieved 99.30% testing accuracy on unseen data. When pre-trained model train on same data then it gives validation and training accuracy 97.50% , 98.75% respectively. On unseen data it gives 98.85% accuracy which is less compare to fine tune VGG-16 model. We can say that, performance of fine tune model is better than base model of VGG-16. Table II shows that precision, recall, f1-score of fine tune VGG-16 model is better than pre-trained model.

Table I shows list of hyperparameters used to train the model. After that, model tested on 572 unseen images. Fig. 3 (a) shows confusion matrix of VGG-16 model on unseen data. We can say that model performed well on jumping, jogging, walking classes with 100% accuracy. Model achieved 99.30% testing accuracy on unseen data. When pre-trained model train on same data then it gives validation and training accuracy 97.50% , 98.75% respectively. On unseen data it gives 98.85% accuracy which is less compare to fine tune VGG-16 model. We can say that, performance of fine tune model is better than base model of VGG-16. Table II shows that precision, recall, f1-score of fine tune VGG-16 model is better than pre-trained model.

B. Training And Validation Results: VGG-19

Dataset is divided into 80:20 ratio, where 80% for training and 20% for validation. Accuracy curves of training and validation data of fine-tuned VGG-19 model is shown in fig. 2 (b), which achieves accuracy 95.45%, 94.90% respectively. Same as above VGG-16 model, this model trained for 40 epochs. Fig. 1 (a) shows loss curves of training and validation data. Loss starts at very high, then reduces drastically. Just like previous model same hyperpar-

TABLE I. LIST OF HYPERPARAMETERS

Hyperparameters	***Values***
Learning Rate	0.0001
RMSProp	Rho=0.9,momentum=0.0
Batch Size	32
Epochs	40
LeakyReLU	α = 0.3

TABLE II. PERFORMANCE OF FINE-TUNED MODELS

Models	**Fine-tuned results (%)**			
	Precision	***Recall***	***F1-Score***	***# parameters***
VGG-16	99.11	99.13	99.30	14.78 M
VGG-19	96.26	95.80	95.77	20 M
Inception-V3	97.53	97.38	97.38	22 M

TABLE III. PERFORMANCE OF PRE-TRAINED MODELS

Models	**Pre-trained results (%)**			
	Precision	***Recall***	***F1-Score***	***# parameters***
VGG-16	98.45	98.60	98.50	138 M
VGG-19	94.10	93.85	93.90	22 M
Inception -V3	93.83	90.17	90.38	24 M

-ameters are used from table I. Training and validation loss of the model is 23%, 22%, which is less compare to pre-trained model, 26%, 24% respectively. Confusion matrix of model on test data shown in fig. 3 (b), we can say that jumping, jogging, throwing classes are classified with 100% accuracy. Performance of model for crawling class is not good. As mentioned in table II, Pre-trained VGG-19 model achieved accuracy on training, validation data, test data of 94%, 93.30%, 93.85 % respectively. We can say that, performance of fine-tuned model (table II) on unseen data better than pre-trained model (table III). Fine-tuned model test on unseen data of total 572 images, which gives precision, recall, f1-score of 0.9626, 0.9580, 0.9577, more than pre-trained VGG-19 as shown in table II and table III.

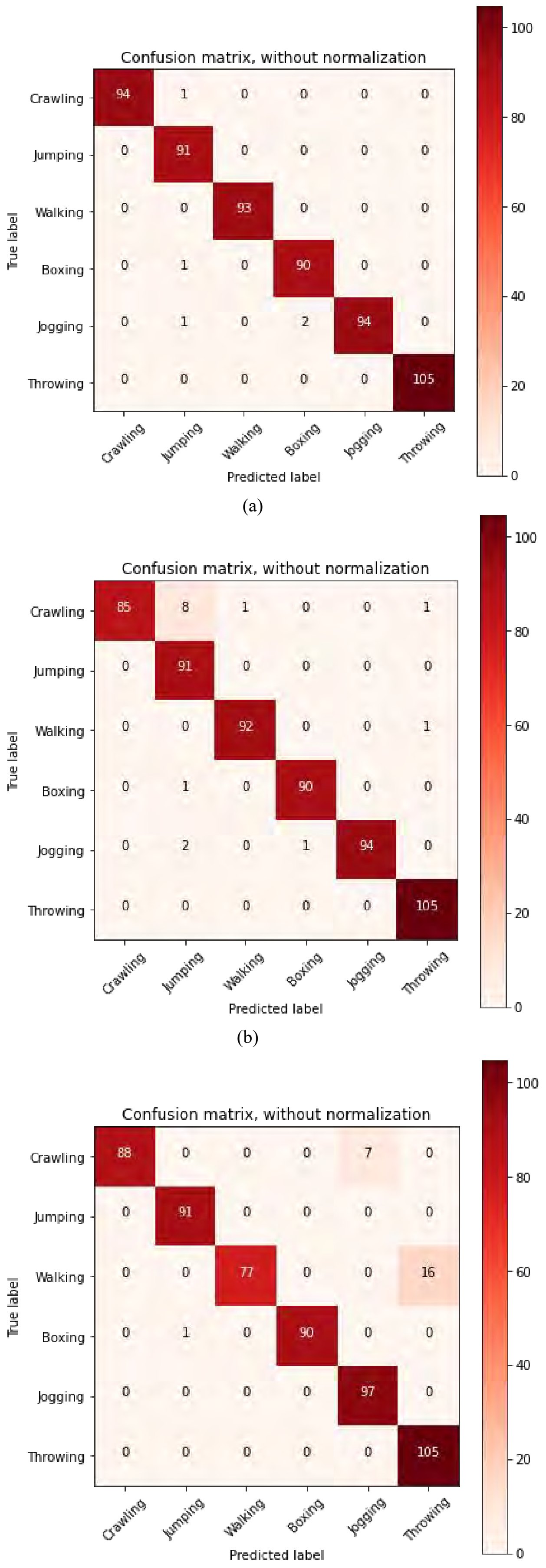
C. Training And Validation Results: Inception-V3

Performance of fine tune model is shown in fig. 1(c), 2 (c). Model trained for 30 epochs, which achieves accuracy of 98.25%, 98% on training, validation data respectively. From loss curve, we can say that at epoch 23 loss increases then converges for remaining epochs. Apart from epochs, remaining hyperparameters are same from table I. Pre-trained model does not converge well, gives accuracy of 91%, 90% on training, validation data respectively. Fig. 3 (c) shows confusion matrix of model on test data of 572 images, achieves accuracy 97.40%. Pre-trained Inception-V3 model achieved only 90 % on test data, which is very less compare to performance of fine tune model. From table II, we can say that precision, recall, f1-score of fine-tuned model is more than pre-trained inception model.

All DCNN models implemented using NVDIA Quadro RTX 3000 graphics card, Keras 2.9, python 3.10, matplotlib 3.5.3, and TensorFlow-GPU. We used keras ImageDataGenerator to augment images to increase number of images in dataset.

V. CONCLUSION

We successfully performed human activity recognition using m-D signatures on DIAT-μRadHAR dataset and tested different DCNN fine tune models such as VGG-16, VGG-19, Inception-V3 etc.All DCNN models were modified then, fine tune and trained without transfer learning.



(a) (b) (c)

Fig. 3. Confusion matrix: (a) VGG-16 (b) VGG-19 (c) Inception-V3

VGG16, VGG19, and Inception-V3 performed better after applying fine tuning, gives accuracy of 99.3% , 95.8%, and 97.38% respectively, more than pre-trained models. Dataset does not have enough data so that DCNN models can face overfitting problem. To solve this, large dataset should be created with more types of classes using GANs, which is one of our ongoing research work.

REFERENCES

[1] 1.O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," IEEE communications surveys & tutorials, vol. 15, no. 3, pp. 1192–1209, 2012.

[2] F. Fioranelli, S. A. Shah, H. Li, A. Shrestha, S. Yang, and J. Le Kernec,"Radar sensing for healthcare," Electronics Letters, vol. 55, no. 19, pp.1022–1024, 2019.

[3] .J. L. Geisheimer, E. F. Greneker, and W. S. Marshall, "High-resolution Doppler model of the human gait," Proc. SPICE, vol. 4744, pp. 8–19, Jul. 2002

[4] M. Zenaldin and R. M. Narayanan, "Radar micro-doppler based human activity classification for indoor and outdoor environments," in Radar Sensor Technology XX, vol. 9829. International Society for Optics and Photonics, 2016, p. 98291B.

[5] H. Chen and W. Ye, "Classification of human activity based on radar signal using 1-d convolutional neural network," IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 7, pp. 1178–1182, 2019.

[6] 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, 2015.

[7] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, "Personnel recognition and gait classification based on multistatic micro-doppler signatures using deep convolutional neural networks," IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 5, pp. 669–673, 2018.

[8] S. Abdulatif, Q. Wei, F. Aziz, B. Kleiner, and U. Schneider, "Micro-doppler based human-robot classification using ensemble and deep learning approaches," in 2018 IEEE Radar Conference (RadarConf18). IEEE, 2018, pp. 1043–1048.

[9] .H. Chen and W. Ye, "Classification of human activity based on radar signal using 1-d convolutional neural network," IEEE Geoscience and Remote Sensing Letters, 2019.

[10] J. Zhu, H. Chen, and W. Ye, "A hybrid cnn–lstm network for the classification of human activities based on micro-doppler radar," IEEE Access, vol. 8, pp. 24 713–24 720, 2020

[11] W. Gao, L. Zhang, W. Huang, F. Min, J. He, and A. Song, "Deep neural networks for sensor-based human activity recognition using selective kernel convolution," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–13, 2021.

[12] I. Alnujaim, D. Oh, and Y. Kim, "Generative adversarial networks for classification of micro-doppler signatures of human activity," IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 3, pp. 396–400, 2019.

[13] Youngwook Kim,and Hao Ling."Human Activity Classification Based on Micro Doppler Signatures Using a Support Vector Machine."IEEETransactions on GeoscienceandRemote Sensing,vol.47,no.5,May2009,pp.13281337,10.1109/tgrs.2009.2012849.Accessed 23Oct.2020.

[14] Kumawat,HarishC.,etal."DIAT-SAT:Small Aerial Targets'Micro-Doppler Signatures and Their Classification Using CNN."IEEE Geoscience and Remote Sensing Letters, 2021, pp.1–5,10.1109/lgrs.2021.3102039.Accessed12Dec.2021.

[15] Beddiar,Djamila Romaissa,et al."Vision-Based Human Activity Recognition:A Survey."Multimedia Tools and Applications,vol.79,no.41-42,15 Aug.2020,pp.30509–30555,10.1007/s11042-020-09004-3. Accessed 11 Mar.2021.

[16] Tammina,S.(2019).Transfer learning using VGG-16 with 'Deep Convolutional Neural Network for Classifying Images' International Journal of Scientific and Research Publications (IJSRP),9(10),p9420.https://doi.org/10.29322/ijsrp.9.10.2019.p9420

[17] M. Chakraborty, H. C. Kumawat, S. V. Dhavale and A. A. B. Raj, "DIAT-μ RadHAR (Micro-Doppler Signature Dataset) & μ RadNet (A Lightweight DCNN)—For Human Suspicious Activity Recognition," in IEEE Sensors Journal, vol. 22, no. 7, pp. 6851-6858, 1 April1, 2022, doi: 10.1109/JSEN.2022.3151943.

[18] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. The 3rd International Conference on Learning Representations (ICLR2015). https://arxiv.org/abs/1409.1556

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.