Activity recognition of FMCW radar human signatures using tower convolutional neural networks

A. Helen Victoria¹ · G. Maragatham²

Accepted: 31 May 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Human activity recognition has become an obligatory necessity in day to day life and possible solutions can be provided with the technological advancement of sensing field. Radar based sensing with its unbeatable unique features has been a promising solution for identifying and distinguishing human activities in recent years. The ascent of loss of life among elderly people in care homes during COVID-19 is mainly due to poor monitoring services, that was not able to track their daily life activities. This has even more emphasized the need for savvy activity monitoring and tracking system. In this work, we have used a dataset that has captured six daily life activities of people from different locations during different times under realistic environments, unlike an regular controlled data collection environment. We have proposed a novel tower based convolutional neural network architecture that has employed parallel input layers with individual color channel images sent as inputs to the model. We have concatenated all the unique signature features from each channel to have better and robust feature representation to the model. We have analyzed the proposed model with different color spaces like RGB, LAB, HSV as inputs and found that our chosen input type performs better with the proposed model with significant test accuracy results. We have also compared our proposed model with other existing state of art architectures for radar based human activity recognition.

Keywords Human activity recognition · Radar · COVID-19 · Convolutional neural networks · Color spaces

1 Introduction

Human activity recognition has been an active research area in recent times with broad range of applications like surveillance [1], video analytics [2], daily life activities monitoring [3] and in human computer interactions [4]. The activities are captured using smart sensing devices like smart cameras, optical sensors and other wearable sensors yielding good results in the past [5]. In recent years, radar based sensors has been deployed as a sensing technology for human activity recognition surpassing other sensing

technologies with its unique features. The radar sensing is not influenced by the light, fog and other changing environmental conditions, unlike other optical sensors [6]. Another major advantage of radar based sensing is that it captures the range and velocity information of the target activities under study without sending their actual photographs or videos, whereas the existing vision based sensors has the visual images leading to data privacy issues. Moreover, radar sensors are not wearable sensors and can even sense through the walls making it as an excellent choice for real life based scenarios [7]. The need for such advanced sensing can be witnessed with the increasing death rate of covid-19 in care homes, elderly living and in hospitals. According to Wall Street Journal report, nursing homes alone has witnessed nearly 26,000 deaths [8]. The covid-19 death toll statistics across the world has shown elderly persons in care homes has been so vulnerable to this pandemic disease [9]. Most of the elderly deaths were due to poor surveillance and tracking in care homes leaving them unattended when they were in need for

✉ A. Helen Victoria
helenvia@srmist.edu.in
G. Maragatham
maragatg@srmist.edu.in

¹ Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

² Department of Information Technology, SRM Institute of Science and Technology, Chennai, India

help, during this pandemic. Besides this, the elderly population in care homes are always vulnerable to other infectious disease such as scabies, influenza which demands more realistic and robust activity tracking smart mechanism [10]. Effective monitoring can provide proactive measures that can eliminate after effects even like fatal injuries in terms of health care.

In this work we have used a dataset publicly provided by University of Glasgow, acquired daily life activities like walking, sitting, standing, picking an object, drinking water and falling using frequency modulated continuous wave (FMCW) radar in indoor and outdoor environments. Previous radar data acquisitions that has been done only in indoor lab based environments. This dataset has also captured the activities of elderly people in the age group ranging from 60 to 98. Apart from the University of Glasgow, the data for the elderly has been obtained in Elderly Homes or Senior Living homes located at Age UK West Cumbria [11]. Radar signals can be viewed as micro Doppler signatures images. Hence the activity classification can be categorized as a 2D image classification problem. Deep learning techniques works fairly well for such image classification tasks which is evident from the results of ImageNet dataset challenge [12]. In terms of human activity recognition previous work with deep learning has obtained better results on classifying 12 activities compared to machine learning algorithms like support vector machine (SVM) [13]. The major attraction towards deep learning is its automatic feature extraction without requiring high level human expertise in the problem domain. [14] classified 12 activities using an convolutional auto encoder with the recognition performance of 94.2%. [15] proposed a convolutional neural network architecture for classifying human activities based on micro Doppler signatures along with Bayesian optimization. Bayesian optimization is used to optimize the hyperparameters of the deep neural network [16]. However, most of the previous research works have employed deep layered architecture or hybrid architectures stating that more deeper the architecture higher the accuracy is, in terms of complex computer vision tasks [17, 18]. Deeper networks had the issue of convergence of the network due to vanishing gradient problem [19]. In our work, we have adopted the idea of widening the network instead of deepening the network layers which has shown significant results.

In existing works, spectrogram images were used as inputs for classifying the human activities [20]. Other input image inputs were usually time-range map [21], time-Doppler map, range-Doppler map [22]. However time Doppler maps are the inputs which many previous works have used till date [23, 24]. To the best of our knowledge,

color space exploration has not been done in the radar input images.

The main contributions of our work is

1. We have built a novel convolutional neural network with unique widened parallel input layers followed by dense layers to classify the activities of the data captured under controlled and uncontrolled environments that is more suitable to real life deployment.
2. We have explored various color spaces that showed promising results in general image classification tasks as inputs to our input layer. We have chosen individual color channel images as inputs to the proposed model as those features were more robust and had significant classification results on the dataset.
3. We have validated our model results and found that our model performance was better compared to the existing radar based human activity recognition architectures.

The rest of the paper is organized in the following manner. Section 2 unveils the background study of the radar signal representation, color spaces, deep learning architectures and about the dataset used in this work. Section 3 explains the proposed methodology right from color space exploration to novel architecture design and learning. Section 4 showcases the results obtained and its comparative analysis with other state of art approaches. Section 5 explains the discussions from the proposed model design and its observed results. Conclusion is explained in the Sect. 6.

2 Background study

2.1 Radar signal representation

Frequency modulated continuous wave (FMCW) radar has been a smart choice for human activity monitoring and recognition in recent years. This is mainly due to its deployment for short range real life scenarios [25]. This radar is capable of sensing both the range and Doppler information of targets under study. The beat signals generated by FMCW radar has the combination of transmitted signal and the back scattered echo signal. The frequency delay in the obtained signal gives the range information of the target. The beat frequency (f_b) is given by the Eq. (1) in which t_d represents time delay, T_s and B_s indicates the sweep time and sweep bandwidth as shown in the Fig. 1. Apart from range given in the Eq. (2), relative velocity of the target can also be calculated.

$$fb = [(t_d * B_s) / T_s] \quad (1)$$

$$R = (c * T_s * fb) / 2 * B_s \quad (2)$$

Fig. 1 Frequency and time representation of FMCW radar

The radar signal returned viewed in joint time–frequency distribution, short time Fourier transform (STFT) is used to analyze various time instances of returned non stationary radar signals to generate range-Doppler and micro Doppler radar signatures. The STFT of a signal(x) is given by the Eq. (3).

$$X(t, f) = \sum_{w=-\infty}^{\infty} h(w) * x(t - w) e^{-j2\pi fw} \quad (3)$$

f represents the frequency, t indicates the time index and $h()$ function indicates the choice of discrete time window function for clear time and frequency resolutions. Choosing an ideal window size is important as it reflects in the resolutions of time and frequency variations. In this work, the window length is 128 samples per second for radar signal preprocessing. The signal variation with time t and frequency f can be visualized by generating spectrogram based on the Eq. (4).

$$S(t, f) = |X(t, f)|^2 \quad (4)$$

These spectrograms are usually used as a 2d image input in the classification. In this work the 2D RGB images are further processed into individual color channel images.

2.2 Color space representation

In image classification tasks, images are mostly colored images represented in Red, Green, Blue (RGB) format. There are various color spaces available which deals with different organization of colors. Most of the research works doesn't explore much with these color spaces. Basically in all computer vision tasks, the inputs are numbers even if it is an image input or a video input. Hence different color spaces are also generally numbers that represents different colors or combination of colors. Other popular color spaces are hue, saturation, value (HSV), luminance of A, B chromatic components (LAB). There are many more color spaces available. In this work, we have explored only the above mentioned two color spaces as these color spaces has

shown promising results in image classification based tasks compared to other color spaces [26]. RGB color space has the combination of shades of Red, Blue, Green color channels each of which is represented as 8 bits ranging from 0 to 255.

LAB color space has all the colors that are visible to the human eye and its major advantage is it is a device independent model. 'L' represents lightness, 'A' indicates the combination of green–red and 'B' represents the combination of blue–yellow colors. The conversion of RGB to LAB is shown in the Eqs. (5)–(8) by first converting into an XYZ color space.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.489989 & 0.310008 & 0.2 \\ 0.176962 & 0.81240 & 0.010 \\ 0 & 0.01 & 0.99 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (5)$$

$$L = 100 \sqrt{Y/Y_n} \quad (6)$$

$$A = K_a \left(\frac{X/X_n - Y/Y_n}{\sqrt{Y/Y_n}} \right) \quad (7)$$

$$B = K_b \left(\frac{Y/Y_n - Z/Z_n}{\sqrt{Y/Y_n}} \right) \quad (8)$$

Hue, saturation and value (HSV) was considered based on how human eye can perceive different colors. Hue ranges from 0 to 360 representing colors and saturation and value ranges from 0 to 255 describing shades and brightness [27]. In order to obtain the values for HSV the values of RGB has to be normalized in the range of 0 to 1 by dividing it by 255, the maximum range value of R, G, B channels individually. The conversion of RGB to HSV is given by the following Eqs. (9)–(13). The sample spectrograms of these preprocessed color spaces are shown in the Fig. 2.

$$R_i = R/255, G_i = G/255, B_i = B/255 \quad (9)$$

$$\begin{aligned} \text{Max} &= \max(R_i, G_i, B_i), \quad \text{Min} = \min(R_i, G_i, B_i), \quad D \\ &= \text{Max} - \text{Min} \end{aligned} \quad (10)$$

$$H = \begin{cases} 0^\circ & D = 0 \\ 60^\circ \left[\frac{(G_i - B_i) \bmod 6}{D} \right], & \text{Max} = R_i \\ 60^\circ \left[\frac{(B_i - R_i)}{D} + 2 \right], & \text{Max} = G_i \\ 60^\circ \left[\frac{R_i - G_i}{D} + 4 \right], & \text{Max} = B_i \end{cases} \quad (11)$$

$$S = \begin{cases} 0, & \text{Max} = 0 \\ \frac{D}{\text{Max}}, & \text{Max} \neq 0 \end{cases} \quad (12)$$

$$V = \text{Max} \quad (13)$$

Fig. 2 Color space representation of **a** RGB, **b** HSV, **c** LAB, **d** R-channel, **e** G-channel, **f** B-channel

Each image is basically represented as numbers. Each color channel in turn, the variations of image digital data gives different perception to the model which is configured to train the images. The images are down scaled from its original size to give as the input to the deep learning model. When the images are down scaled, the pixel size will be more thereby reducing the data of each image in the power of 2. Each color channel indicates different wavelengths in the light spectrum. Hence different channels gives varied information contributing to more number of features from the images under study. In radar micro Doppler signature images, the reflected signals exhibits the frequency variation based on the activity. Each macro and micro movements has frequency drift indicating the individual body parts of human. Each color channel provides unique information. The need for exploring these different color spaces is that, they are mathematical representation of colors as numbers that are sent as input to the deep learning model to identify which color space has best feature representation for radar human activity recognition. Best feature extraction paves way for better learning and generalization of the model. In our work we have analyzed RGB, LAB, HSV, individual R, G, B channel color space with the proposed deep neural network model and have found that individual color channel contribution was more compare to other color spaces. Individual channel representations helps us to retain the uniqueness of each and

every activity. Blue channel indicates intricate micro movements of hands and legs, whereas Red channel contributes for macro movements of the body. I.e. torso part of the body which contributes more while performing each activity. This will be in general for most of the activities under study. Green channel provides more pixel information yielding more unique features for the activity under study. Hence Red channel generalizes the major features, Green and Blue channel contributing to micro level movement features pertaining to that specific activity. Hence combining all these channel contributions as a concatenated feature maps enhances the learning ability of the deep learning model.

2.3 Deep convolutional neural networks

Deep convolutional neural networks in image classification has given many breakthrough results in image classification tasks [28]. Convolutional neural networks are basically hierarchical architectures which helps in high level feature extraction without having the need of manual intervention for extracting the features like traditional machine learning algorithms [29]. Larger number of convolutional blocks learns more detailed features even for complex tasks. It has multiple layers with shared parameters and learns through back propagation for building an effective model. Generally the convolutional layer is followed by the pooling

layer and the dense layer. Pooling layer is mainly used to reduce the size of the network by reducing the number of parameters and thereby the computation of the network. The fully connected dense layer is used to have intense learning among the neurons that are activated by the previous layers. Dense layer brings in non linearity to the data that is capable of generating any complex mathematical models. The detailed understanding of each and every layer is given [30, 31].

The main reason for adopting convolutional neural networks for radar based human activity recognition (HAR) is that the spatial property of CNN helps us to extract localized features considering nearby signal positions, rather than extracting from one single position alone. Moreover, it retains the extracted information of frequencies as features throughout the layers of the network [32]. This is mainly due to the scale invariant property of CNN for different frequencies of the signal [33].

2.4 Details about the dataset

We have used a publicly available dataset provided by the University of Glasgow. It has collected 6 different activities of over 56 subjects in 9 different places including open space environments and in Controlled Lab Environment. Each and every subject repeated activities like walking, standing up, sitting on a chair, bending to pick up object, drinking water from glass or bottle. Simulated frontal fall was captured in only some subjects concerning their safety reasons. Data collection was a major challenge and it was done in the past only with very few subjects for study. This is a first kind of benchmark dataset, available for researchers for radar based human activity recognition applications with more number of subjects captured in realistic environment. This data was collected using off the shelf frequency modulated continuous wave radar [34]. It is capable of recording micro Doppler signatures of the subjects under study with the given system parameters in the Table 1.

Table 1 FMCW radar parameters

System parameter type	Value
Operating carrier frequency	5.8 GHz
Operating band	C-Band
Gain	+ 17 dB
Chirp bandwidth	400 MHz
Chirp duration	1 ms
Output power	+ 18 dBm

3 Proposed methodology

3.1 Preprocessing of radar signals

The preprocessing of radar signals to reproduce as images into the neural network is explained using the following steps

1. The radar data is in the form of 1-dimensional complex array is converted to range bins using pulse compression technique.
2. Range time plot (range vs. time matrix) is generated and fast Fourier transform is applied across the time dimension.
3. Moving target indicator (MTI) Filter is used to alleviate the stationary objects echo from the back scattered radar signal [35].
4. Extract only those range bins within the specified MTI filter data for lucid connection of data.
5. Range Doppler map is generated and short time Fourier transform is applied on the selected sample values.
6. The obtained 2D matrix is viewed as micro Doppler signature images in the form of spectrogram as shown in Fig. 3.

The spectrogram images helps us to view the heavier and lighter movements of different parts of the body. The oscillations in the image indicates the movements and the signature varies according to the movement type, thereby indicating a feasibility to uniquely identify human activities through these micro Doppler signatures. But the major challenge was the similarity among the patterns for different activities. The brighter yellow and red color in the spectrograms shows the intense movements and light blue indicates less intense or lighter movements. Spectrogram of each activity like walking, sitting, standing, bending to pick up an object, drinking water from a glass and falling is given in the Fig. 4.

The line plot of spectrogram is plotted in the Fig. 5 which shows that each color in the figure indicates a

Fig. 3 Spectrogram of walking pattern

◀**Fig. 4** Spectrogram of various activities **a** walking, **b** standing, **c** sitting, **d** bending to pick an object, **e** drinking water, **f** falling

particular body part. The collective plots of all spectrogram activity is shown in the Fig. 6.

Each plot in the Fig. 6 indicates the intricate variations and similarities among patterns. Figure 6a, b indicates uniform movements of all parts of the body matching the pattern of walking and standing. Figure 6c shows more variations in the upper half of the body movement. Figure 6d, e shows more movement in the torso, foot and leg part in the first half portion of the image and the later half portion indicates the predominant movement of arms, knees and legs. Figure 6f indicates the activity of falling, shown by knees and legs in the last portion of its spectrogram image. It also indicates a drag in the first half portion contributing its uniqueness for its class recognition. Hence the image information of each and every pixel is important to classify these activity patterns effectively. This challenge requires building a model that can handle noisy images and its unique gait pattern features. So each and every pixel information is required. This brought in the

requirement of exploring individual channel features in depth. It also contributes to the model by increasing the training data size by having samples from each channels, whilst training sample size with RGB images are less comparatively. Moreover, efficient computation can be obtained in 1-dimensional array rather than using a 3-dimensional radar data array.

3.2 Preprocessing RGB images into individual color channels

The RGB image (3-dimension) is split into individual channel images and appended in an separate array of 3376 training samples for each channel (1-dimension). The preprocessed image samples are given in the Fig. 7 in the order of Red (R), Green (G), Blue (B). It is clearly seen that each channel's contribution varies by minute difference. Each channel has values ranging from 0 to 255. The Green channel has more pixel information compared to other channel's images. Blue channel is observed with some intricate hand and leg movements compared to torso movement, but with noise and red channel acts as contrast

Fig. 5 Line plot of walking spectrogram

Fig. 6 Frequency representation in line plots for all activities

map enabling basic difference among different pattern signatures.

The Fig. 7(a–c) shows the Red channel, Green channel and Blue channel of walking pattern, Fig. 7(d–f) depicts the Red, Green and Blue channel of bending and Fig. 7(g–i) represents individual channel spectrogram images of sitting pattern. Each channel of the pattern has some unique image information. Each channel's variation can give more insights for the model to have better learning curve for radar based activity recognition applications.

3.3 Proposed model architecture

The micro Doppler radar signature images are processed into 1-dimensional individual color channels and are sent into deep neural network as parallel inputs to the first convolutional layer of the model. The output layer of the model is used to classify the different human activities like walking, standing, sitting, bending to pick an object, drinking water and falling. The workflow diagram is given in the Fig. 8. The + symbol indicates the concatenation of features that is extracted from the individual image channel (Red + Green + Blue) for efficient and robust classification. Each activity has more similarities with minimal visual difference among them.

The detailed architecture of our proposed tower CNN model is summarized in the Fig. 9. The model has 3 input layers, 3 pairs of convolutional block consisting of convolutional layer followed by max pooling layer, concatenation layer and 7 dense layers. The novel or unique features contributing to this model are explained in the subsections.

3.3.1 Input layers

There are 3 three input layers, each of which is the input from each channel of an RGB preprocessed micro Doppler signature images. The first, second and third input layers corresponds to Red (R) channel, Green (G) channel, Blue channel (B) respectively. The channel images are sent in parallel to retrieve its features individually from each. It is evident from the previous section that each channel shows significant difference among them. This yields larger number of parameters that has fine grained image information without missing any important features, as some of the activities exhibits more similar Doppler signatures.

Fig. 7 Spectrograms of Red, Green, Blue channel for different activities

3.3.2 Widening the network model

Deeper the network, better the performance would be as per many previous researches. In this work we have proposed the idea of widening the convolution block that extracts features which is used in parallel for each color channel. We have used a kernel size of 3 in the first two convolutional layers across all the input layers. Having kernel size as 3 in subsequent layers imitates the output of receptive field of size 5. The third convolutional layer had a

kernel size of 4. Smaller size of the kernel in the initial layers helps to retain more information than having a larger kernel size which might skip pixel information. The pooling layer size is 2 across the convolutional blocks of all input layers.

3.3.3 Concatenating the individual channel features

The output features of each channel are concatenated with each other and is collectively passed to the dense layer.

Fig. 8 Overall proposed workflow diagram

Since the radar training and testing samples are lesser compared to other type of images, we need more features to make the model learn better. Hence individual layer outputs i.e. pooling layer outputs are added on to the next layer as concatenated features. This is given by the Eq. (14). RCo, GCo, BCo are the pooling layer outputs of Red, Green and Blue channels respectively.

$$\text{Output_Features} = \text{Concatenate}[RCo, GCo, BCo] \quad (14)$$

where output_features represents the output of channel wise concatenation. This is followed by seven dense layers. The output shape of concatenation has $5 * 5 * 48$ that leads to 1200 parameters that is fed into the dense layer. The features obtained using three different color channels and concatenated, hence those features requires proper learning which could be achieved through the dense layers. Dense layers are fully connected layers in which each neuron is connected to every other neuron. This helps in training the extracted features obtained from the convolutional layers. The last dense layer has Softmax Classifier which is used to squeeze into probability range (0–1) to identify the class of the image which is under training and validation. Soft Max acts as an output classifier to classify the six labeled activities. In total our model has 265,424 trainable parameters.

3.3.4 Learning in the model

We trained our model with Adam optimizer in contrast to the stochastic gradient optimizer [36] with the batch size of 32 with the learning rate of 0.001. Adam optimizer is chosen due to its minimal weight updations that can have better learning in the training phase. It eventually converges well with global minimum with certain number of epochs. Smaller learning rates are preferred for smoother weight updations. The rectified linear unit (RELU) is used as an activation function for activating neurons in each

layer and to introduce non linearity into the data. The firing effect of RELU activation is given by the Eq. (15).

$$\text{ReLU}(y) = \begin{cases} 1 & \text{if } y > 0 \\ 0 & \text{if } y < 0 \end{cases} \quad (15)$$

We have used the Glorot uniform initializer for the weight matrix, we have set the bias value as 1 in the convolutional layers and in the first two dense layers. Other layers are set with zero for the bias initializer. The early initialization value of bias will make the RELU filled with more positive inputs. The major pixel information is maintained in the early stages leaving way for the remaining layers with zero initialization to reduce overfitting during training. In addition, dropout layer with 20% is used after the dense layers to combat overfitting. This in turn helps in extracting robust features as well.

In contrast to the multinomial cross entropy loss function used for multilabel classification, we have used binary cross entropy loss function inspired from binary classifiers [37]. Multilabel is assigned for each image and the output of the lowest classification loss is chosen, thereby yielding better performance during back propagation. The loss function is given by the Eq. (16).

$$l(\theta) = -1/n \left[\sum_{p=1}^n \sum_{q=1}^c y_{pq} \log(P_{pq}) \right] \quad (16)$$

p, q represents the samples, classes out of the total n samples and c classes respectively. y is the label for the sample p and q class, P_{pq} represents the probability value of the label which has the minimum classification loss. The model updates the weights, bias and filter values during back propagation learning and is explained by the following Eqs. (17)–(19).

$$W_p^* = W_p - L * (\partial E / \partial W_p) \quad (17)$$

◀Fig. 9 Detailed architecture of proposed tower based convolutional neural network

$$F^* = F_p - L * (\partial E / \partial F) \quad (18)$$

$$B^* = B_p - L * (\partial E / \partial B) \quad (19)$$

W_p , F_p , B_p indicates the previous weight, filter values and bias, L indicates the learning rate and ∂E represents the error with respect to weight, filter and bias respectively. This is how our models learns through forward and backward training with the argumentation of features from individual channels yielding robust classification rate. The network was trained for 20 epochs with the training set of 3376 training samples in GPU run time environment of Google Colab with Python Keras framework.

4 Experimental results

We have trained our network with 3376 training samples and 676 testing samples in the fall data dataset. To the best of our knowledge, this is the first work on this dataset as it is a recently published dataset. We have made a detailed study of different designs for our model before choosing the best performing one. We have analysed our model with different optimizer and different activation function for better learning. We have also made a comparative study with traditional RGB based convolutional neural networks and other color spaces like LAB, HSI which promised to have good results for image classification based tasks [38]. We have also compared our model architecture with other architectures for similar kind of radar based activity recognition works considering various aspects.

4.1 Proposed model based results

The proposed tower CNN model have achieved an accuracy rate of 98.67% as training accuracy and 97.58% as the testing accuracy as shown in the Fig. 10a. The test accuracy indicates that the model learns better with new test images as well. It is noted that the test accuracy is reasonably high starting with 84% in the initial epoch itself. This is mainly due to the rich feature extraction by the three parallel input layers consisting of three channels and unique way of concatenating the features at the training phase. The loss function plot is given in the Fig. 10b. The less initial loss and substantial reduction of loss in subsequent epochs is mainly due to the binary cross entropy of individual mapping and comparison of each image with all the labels of the dataset.

Another way to validate our proposed model is by probing into the visual knowledge of feature maps obtained at the convolutional level as shown in the Fig. 11. It shows that the individual channel based pixel information is more, eliminating the gradient vanishing problem and thereby inducing good learning by avoiding overfitting of the training data. As discussed in the earlier sections, the green and blue channel image information is more compared to the red channel.

Different designs with number of convolutional layers, dropout layers, dense layers were analysed. The same model was tried with 2 convolutional layers at each input layer in which the accuracy was 95.4%. The dropout layer was placed before the classifier block and after each dense layer which has the accuracy of 95%, 96% respectively. Moreover, keeping dropout layer after convolutional layer is not a good practice as it might end up losing unique image feature information. Dropout layers are not used to increase the performance of the training data, it is mainly for achieving good recognition rates with respect to test

Fig. 10 Proposed model accuracy and loss for training and testing data

Table 2 Comparison of proposed CNN with different color space CNN architectures

Color space based model type	No of parameters	Training accuracy	Test accuracy
RGB-CNN	386,918	95.17	85.15
HSV-CNN	299,200	83.33	83.33
LAB-CNN	299,200	94.65	92.61
Proposed tower CNN	265,424	98.67	97.58

Fig. 11 Feature map outputs of convolutional layers

data or any new data. We have identified that the proposed model design shows better performance compared to different design trails.

4.2 Analyzing and comparing different color spaces with the proposed model

In order to analyse our model with other color spaces like LAB, HSI, RGB similarly designed convolutional neural network model was used. This is done with deeper network architecture by unwinding the parallel convolutional blocks into sequential ones. Besides having many color spaces available we have chosen only LAB, HSI based on the promising results of these color spaces in image classification tasks. The results are summarized in the following Table 2. HSV based CNN has 83.33% as accuracy in both the training and testing dataset which is nearly 15% and 14% less than our proposed model rate. In RGB CNN even though the training accuracy is good, the test data accuracy is 12% less than the proposed one. LAB based CNN had reasonable classification but still, the training and testing accuracy was less by 4% and 5% respectively to the proposed one. The confusion matrices of RGB, LAB and the proposed architectures are given in the Fig. 12.

The main novelty of this model design is that it has good test data learning right from the beginning. It is clearly

observed that the proposed model also has less number of parameters in turn reducing the computation complexity of the network during forward and backward learning of the model.

4.3 Comparison with other deep neural network architectures

Table 3 summarizes the results of other state of art architectures for human activity recognition across different specifications. We have compared our results with these works [39], continuous wave radar, Ancortek Software [40], FMCW radar [14, 25]. Even though the data sets are different we want to insist that our proposed model surpasses all these previous architectural approach results with high test accuracy besides having a realistic environment considering both indoor and outdoor based data acquisition obtained with 56 subjects under study, which is a first of its kind dataset. We have also compared our work with a recent work using the same dataset. The authors has made a fusion of handcrafted features and extracted features from CNN model and has obtained an accuracy of 96.65% [41]. We have obtained features from individual channels and proposed a model which has obtained 97.58% which is nearly 1% increase from the recent work. Also this increase in accuracy is obtained without having much overhead

Fig. 12 Confusion matrices of RGB CNN, LAB CNN, proposed tower CNN

compared to extracting hand crafted features which requires knowledge from a domain expertise as well.

5 Discussion

The proposed work exhibits concatenation of features at the pooling layer from individual channels. I.e. unique image features are retained at the initial convolutional layer itself with $11 * 11 * 16$ as activation size for each channel for attaining better model generalization. Those unique features were analysed in detail though the dense layers.

The test accuracy results of our model compared to other similar CNN based color spaces like LAB, HSI and other architectures like stacked auto encoder, RNN with LSTM and other hybrid architectures requires more computational resources by having large number of parameters running with more epochs taking longest time for training. In contrast to this, our model design uses optimized computational resources with less number of parameters and training time by imposing concatenation technique at the pooling layer. The model shows significant classification results with walking, sitting, standing and falling. Yet there are some mishaps in similar kind of activities like bending

Table 3 Comparison of proposed work with previous architectures

Radar type	Type of activity	Input type	No of subjects	Deep neural network type	Test accuracy (%)
Continuous wave radar [39]	4-Indoor	Denoised spectrogram grayscale image	–	Stacked auto encoder	87
Ancortek Software defined radio [40]	6-Indoor	Spectrogram	2	RNN with LSTM	92.65
FMCW radar [14]	12-Indoor	Spectrogram + physical features	11	Convolutional auto encoder	94.2
FMCW radar [25]	2-Indoor	Range maps	2	Sparse auto encoder	95.1
FMCW radar [41]	6-(Indoor + outdoor)	Handcrafted features + CNN	56	CNN	96.65
FMCW radar	6-(Indoor + outdoor)	R channel + G channel + B channel of spectrogram	56	Proposed tower CNN	97.58

to take an object and picking up a glass or bottle to drink water. They have very less difference in its micro Doppler signatures as well. Picking a bottle or glass and drinking water from it can be categorized under bending to take an object category since both has predominant activities of hands, knees and elbows. Such specific activities doesn't have consistent pattern paving way for confusion while learning unique features pertaining to specific class in the model. As discussed in earlier sections, the main contribution of the proposed tower CNN architecture is its performance at best under realistic environments similar to a real life scenario with more number of subjects, unlike other radar data collected in an controlled Lab Environment alone with less number of subjects for study.

6 Conclusion and future work

We observed that the individual color channels represents the radar signature class features better compared to direct RGB images. There exists a challenge to have huge dataset size for radar based human signatures applications. This approach has also increased the training data by splitting the images into individual channel images. Hence with this idea, we have proposed a tower based CNN model that has three parallel input layers with convolutional blocks. The obtained features from individual channels are concatenated at the pooling layer of the model for better feature representation and better learning. This model has achieved significant results with 97.58% accuracy in the test data, when compared to other similar problem based architectures. When working with huge data sets, preprocessing might add some overhead. In future we have planned to optimize the training time of the proposed model. This kind of work can be used in old age homes, rehabilitative centers and in senior living and even in defence based

applications where monitoring the activities using radar signals is more common. The limitation of this work is even though the training data is increased with the help of wide and deep proposed architecture it might end up having extra overhead in the model learning which needs optimization. We have also planned to collect more age specific micro Doppler signatures for elderly people alone, as this dataset had people eventually balanced across all ages. We have also planned to build a meta based learning algorithm with the smaller sized radar dataset without doing augmentation in the training data. This could be helpful in similar kinds of applications like face recognition, drug discovery, and other health care applications, where the data size is less, an existing challenge for deep learning algorithms.

Acknowledgements We would like to extend our sincere thanks to University of Glasgow and their researchers for their radar dataset contribution for researchers, using which we have performed our experiments. Radar signatures of human activities (2019), data cite <https://doi.org/10.5525/gla.researchdata.848>.

Declaration

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Su, B. Y., Ho, K. C., Rantz, M. J., & Skubic, M. (2015). Doppler radar fall activity detection using the wavelet transform. *IEEE Transactions on Biomedical Engineering*, 62(3), 865–875.
2. Wang, J. L., & Singh, S. (2003). (2003) Video analysis of human dynamics—A survey. *Real-Time Imaging*, 9, 321–346.
3. Seifert, A., Amin, M. G., & Zoubir, A. M. (2019). Toward unobtrusive in-home gait analysis based on radar micro-Doppler signatures. *IEEE Transactions on Biomedical Engineering*, 66(9), 2629–2640. <https://doi.org/10.1109/TBME.2019.2893528>

4. Chen, L., Zhou, M., Wu, M., She, J., Liu, Z., Dong, F., & Hirota, K. (2018). Three-layer weighted fuzzy support vector regression for emotional intention understanding in human–robot interaction. *IEEE Transactions on Fuzzy Systems*, 26, 2524–2538.
5. Shoaib, M., Bosch, S., Incel, O. D., Scholten, H., & Havinga, P. J. (2016). Complex human activity recognition using smartphone and wrist-worn motion sensors. *Sensors*, 16, 426.
6. Chen, Z., Li, G., Fioranelli, F., & Griffiths, H. (2018). Personnel recognition and gait classification based on multistatic micro-Doppler signatures using deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 15, 669–673.
7. Nkwari, P., Sinha, S., & Ferreira, H. (2017). Through-the-wall radar imaging: A review. *IETE Technical Review*, 35, 1–9. <https://doi.org/10.1080/02564602.2017.1364146>
8. Fresh Data Shows Heavy Coronavirus Death Toll in Nursing Homes. (2020). *The Wall Street Journal*. Retrieved June 5th, 2020, from <https://www.wsj.com/articles/nearly-26-000-nursing-home-residents-died-of-covid-19-federal-tally-shows-11591053033>
9. Coronavirus deaths: How big is the epidemic in care homes. (2020). Retrieved June 1st, 2020, from <https://www.bbc.com/news/health-52284281>
10. Diraco, G., Leone, A., & Siciliano, P. (2017). A radar-based smart sensor for unobtrusive elderly monitoring in ambient assisted living applications. *Biosensors (Basel)*, 7(4), 55. <https://doi.org/10.3390/bios7040055>
11. Fioranelli, D. F., Shah, D. S. A., Li, H., Shrestha, A., Yang, D. S., & Kerne, D. J. L. (2019). Radar sensing for healthcare. *Electronics Letters*, 55(19), 1022–1024. <https://doi.org/10.1049/el.2019.2378>
12. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 25, pp. 1097–1105). Curran Associates, Inc.
13. Kim, Y., & Ling, H. (2009). Human activity classification based on micro-Doppler signatures using a support vector machine. *IEEE Transactions on Geoscience and Remote Sensing*, 47(5), 1328–1337. <https://doi.org/10.1109/TGRS.2009.2012849>
14. Seyfioğlu, M. S., Özbayoğlu, A. M., & Gürbüz, S. Z. (2018). Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities. *IEEE Transactions on Aerospace and Electronic Systems*, 54(4), 1709–1723. <https://doi.org/10.1109/TAES.2018.2799758>
15. Le, H. T., Phung, S. L., Bouzerdoum, A., & Tivive, F. H. C. (2018). Human motion classification with micro-Doppler radar and Bayesian-optimized convolutional neural networks. In *Proceedings of the IEEE international conference on acoustics, speech and signal processing (ICASSP), Calgary, AB, Canada* (pp. 2961–2965).
16. Victoria, A. H., & Maragatham, G. (2020). Automatic tuning of hyperparameters using Bayesian optimization. *Evolving Systems*. <https://doi.org/10.1007/s12530-020-09345-2>
17. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
18. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *2015 IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA* (pp. 1–9). <https://doi.org/10.1109/CVPR.2015.7298594>
19. Glorot, X., & Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256).
20. Tahmouss, D. (2015). Review of micro-Doppler signatures. *IET Radar, Sonar and Navigation*, 9, 1140–1146.
21. Erol, B., Amin, M., Zhou, Z., & Zhang, J. (2016). Range information for reducing fall false alarms in assisted living. In *Proceedings of the IEEE radar conference, Philadelphia, PA, USA* (pp. 1–6).
22. Molchanov, P., Gupta, S., Kim, K., & Pulli, K. (2015). Short-range FMCW monopulse radar for hand-gesture sensing. In *Proceedings of the IEEE radar conference, Arlington, VA, USA* (pp. 1491–1496).
23. Kim, Y., & Moon, T. (2016). Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 13, 8–12.
24. Le, H. T., Phung, S. L., & Bouzerdoum, A. (2018). Human gait recognition with micro-Doppler radar and deep autoencoder. In *Proceedings of the IEEE 24th international conference on pattern recognition (ICPR), Beijing, China* (pp. 3347–3352).
25. Jokanovic, B., Amin, M., & Erol, B. (2017). Multiple joint-variable domains recognition of human motion. In *Proceedings of the IEEE radar conference, Seattle, WA* (pp. 0948–0952).
26. Shin, M. C., Chang, K. I., & Tsap, L. V. (2002). Does colorspace transformation make any difference on skin detection? In *Applications of computer vision. Proceedings. Sixth IEEE workshop* (pp. 275–279).
27. Krizhevsky, A., Nair, V., & Hinton, G. (2014). *The CIFAR-10 dataset*. <http://www.cs.toronto.edu/kriz/cifar.html>
28. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016* (pp. 770–778).
29. Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2019). A Survey of Deep learning and its applications: A new paradigm to machine learning. *Archives of Computational Methods in Engineering*. <https://doi.org/10.1007/s11831-019-09344-w>
30. Kang, M., Ji, K., Leng, X., & Lin, Z. (2017). Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection. *Remote Sensing*, 9, 860.
31. Lin, Z., Ji, K., Leng, X., & Kuang, G. (2019). Squeeze and excitation rank faster R-CNN for ship detection in SAR images. *IEEE Geoscience and Remote Sensing Letters*, 16, 751–755.
32. Zhou, Z., Cao, Z., & Pi, Y. (2017). Dynamic gesture recognition with a terahertz radar based on range profile sequences and Doppler signatures. *Sensors*, 18, 10. <https://doi.org/10.3390/s18010010>
33. Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2018). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119, 3–11.
34. Fioranelli, F., Shah, S. A., Li, H., Shrestha, A., Yang, S., & Le Kerne, J. (2019). Radar signatures of human activities. *University of Glasgow*. <https://doi.org/10.5525/gla.researchdata.848>
35. Will, C., Vaishnav, P., Chakraborty, A., & Santra, A. (2019). Human target detection, tracking, and classification using 24 GHz FMCW radar. *IEEE Sensors Journal*, 2019, 1–1.
36. Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. In *3rd international conference for learning representations, San Diego*. [arXiv:1412.6980v9](https://arxiv.org/abs/1412.6980v9)
37. Ho, Y., & Wooley, S. (2019). The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE Access, Machine Learning, Artificial Intelligence*, 8, 4806–4813. <https://doi.org/10.1109/ACCESS.2019.2962617>
38. Gowda, S., & Yuan, C. (2019). ColorNet: Investigating the importance of color spaces for image classification. In C. Jawahar, H. Li, G. Mori, & K. Schindler (Eds.), *Computer vision—ACCV 2018. ACCV 2018. Lecture notes in computer science* (Vol. 11364). Cham.

39. Jokanović, B., & Amin, M. (2018). Fall detection using deep learning in range-Doppler radars. *IEEE Transactions on Aerospace and Electronic Systems*, 54, 180–189.
40. Zhang, Z., Tian, Z., & Zhou, M. L. (2018). Dynamic continuous hand gesture recognition using FMCW radar sensor. *IEEE Sensors Journal*, 18, 3278–3289.
41. Jia, M., Li, S., Kernec, J. L., Yang, S., Fioranelli, F., & Romain, O. (2020). Human activity classification with radar signal processing and machine learning. In *2020 International conference on UK-China emerging technologies (UCET)*, Glasgow (pp. 1–5). <https://doi.org/10.1109/UCET51115.2020.9205461>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dr. G. Maragatham is an Associate Professor and Research Supervisor in SRM Institute of Science and Technology. She has good expertise in data mining, artificial intelligence, big data analytics, renewable energy. She has published many papers in reputed international journals and conferences. She has more than 20 years of experience in the field of teaching and research.

A. Helen Victoria is a Research Scholar and an Assistant Professor in SRM Institute of Science and Technology. Her research interests are deep learning, machine learning, radar signal applications, optimization techniques.