Human Fall Detection using Convolutional Neural Network

Shahna Sherin P
M. Tech Student
Dept.of.CSE
LBS Institute of Technology for Women
Trivandrum, India

Anju.J
Assistant Professor
Dept. of. CSE
LBS Institute of Technology for Women
Trivandrum,India

***Abstract*—Human falls could be highly dangerous especially when the fallen people lay unattended or unobserved for a long time. Depending on the impact of the fall, the casualties may also increase. Prolonged blood loss can also lead to early death or major health problems. Thus it is necessary to take measures to help the victims, who unfortunately are mostly old or disabled.Even after showing good accuracy rates, wearable sensor based fall detection systems are not used by many potential users due toa variety of personal and physical reasons. Thus detection of falls using cameras is a very reliable solution to address this problem This paper proposes an efficient fall detection system based on camera vision using convolutional neural networks (CNN).Convolutional Neural Networks have proven to be an efficient means of classification in the field of image processing. Here we propose an efficient fall detection design using Inception model CNN by applying transfer learning and aim to achieve state-of-the-art accuracy in multiple datasets such as URFD(UR Fall Detection Dataset) and FDD(Fall Detection Dataset).**

Keywords*—*Computer vision, Convolutional Neural Network, Fall detection, Image Processing

I. INTRODUCTION

Accidental or unintentional falls are a major problem for people especially those who are weak due to old age, disabilities or other injuries. The reports from World Health Organization shows that serious falls account for the death of about 6,46000 people every year and most of these reports are from countries that are poor, that too among the 65 and above age group. In the case of older people biological changes in their body make them more weak and thus more prone to fainting and falling. The effects of such falls could be highly critical, even leading to long term hospitalization or much worse, early death. Criticality of the fall depends on the impact of the fall. A person lying on the floor after a fall for too long indicates that something is definitely wrong with his or her health condition. It could be a minor injury or a major injury but its the responsibility of onlookers to help the victim. However what happens when no one helps, or the victims go unnoticed by others? Here is where a fall detection system could possibly help. The fall detectors work towards giving fast detection of the fall event and giving quick alerts to nearby help centres or concerned individuals so that the people could be saved at the earliest.Fear of fall is always associated with the fall.Old people who have experienced falls earlier tend to refrain from daily physical activities as negative feelings of helplessness develop in their minds.There are several factors that cause falls like changes in the body due to aging,illness etc.Apart from these reasons,wet floors,open drainages,sewage channels,potholes on the roads etc are also major threats and almost people are prone to such severe accidents irrespective of their age,health and other matters.

Brownsel et al.[1] analysed the effect of fall detection systems on the fear of falling by experimenting on people who have gone through falls in the last six months.The study showed that people who used fall detectors were less tensed than the others.We know that for almost all accident cases,timely assistance is the thing that matters the most.If the fall is severe,lying down for a long time can even lead to death and it is highly unfortunate that thousands of people die this way ,unattended.It is thus very important to develop efficient fall detection systems that would help save lives.The automated fall detection systems detect falls and send alerts so that medical assistance could be provided at the earliest.Timely detection and timely alerts can definitely reduce the grave situation of people dying unattended.Fall detection systems can be classified based on different criterias.Some classifications are based on the sensors that are used to record the fall data such as ambience sensors,pressure sensors,camera etc.Some are based on the different phases after the fall and some based on the impact of the fall.The basic structure of all fall detection systems are same i.e firstly the sensors have to record the fall data and secondly efficient differentiation has to be made between the daily life activities such as sitting,walking,standing etc and falling. The ambience sensor based fall detection could be implemented only in places equipped with the ambience sensors.This creates the problem of noise often getting mixed with the sensed data.Wearable device based systems are found to be used more in case of outdoor fall detection systems.However when it comes to older people they might not be that interested in being watched or monitored always.Apart from that,people may forget to wear such sensing devices and it also has the added disadvantage of long term exposure to skin.Vision based fall detection systems overcome this problem of ambience based and wearable device based sensors to an extent.It is very complex and has also various factors that limit its performance such as effects of lighting on the photos or video ,quality of the camera used for vision,multiple objects that appear in the background etc.However with the introduction of latest technologies such as convolutional neural networks,vision based fall detection techniques have improved a lot.

Vision based fall detection or sensing systems can be seen as a category of context aware systems.The context aware systems as the name suggests involve sensors that record or sense data from the surroundings by continous monitoring.Cameras and sensors embedded on floors or on other surfaces etc thus come under this category of context aware systems.Comparison of different fall detection systems is very difficult as they cannot be easily combined under one single section.Eventhough the fall detection based on computer vision is difficult,it has proven to be one of the best ways to detect falls as cameras could be placed anywhere indoors or outdoors.Here we propose an automated fall detection system based on camera vision,where the falls are detected with the help of inception model of CNN(Convolutional Neural Network).

II. RELATED WORK

Vision based fall detection systems have to efficiently classify the actions or events in the specified input frame or image as "fall" or "not fall".Numerous studies have occurred in this area of fall detection and is an area of constant research as its important to develop a high quality fall detection system that would help people especially the elderly.The fall detection techniques could vary depending upon the features analysed such as spatiotemporal features,head position,body posture,body shape change ,body state change etc.Lesya Anishchenko in [11] applied deep learning and transfer learning concepts in about 30 sets of videos for the classification of fall events.The method showed good results in terms of sensitivity and specificity values. Nadi et al in [12] introduced a method in which LDA (Linear Discriminant Analysis) was used.Shadows in the frames were deleted using noise removal. Zhou et al.[9] developed a fall detection system combined the concepts of convolutional neural networks and multi sensor fusion.Radar and Optical camera were both used in this system.Using Short Time Fourier Transform,the characteristics of radar such as micromotion, time frequency(TF) were obtained.For classification and recognition of the human action two kinds of CNNs namely,Alex-Net and Single Shot Multi-box Detector(SSD)Net were used.

In Lu et al.[17]'s design of an intelligent human fall detection system that could detect falls from video surveillance,Vibe algorithm was used to detect humans and SVM(Support Vector Machine) classifier.Lu et al.[13] designed a fall detector system in which kinetic data alone was used for training an automatic feature extractor.From the two dimensional images spatial features were obtained and inorder to get motion informations,three dimensional convolution operations were performed. Ko et al.[15] used a combination of three dimensional depth tracking and EKF(Extended Kalman Filter).EKF was used to get the depth informations from an image.Markov Random Field (MRF) was used to generate or compute depth maps from the input images.Bian et al.[3] proposed a depth camera based fall detection method in which changes in the light's illumination did not matter.Three dimensional joints were extracted using Improved Randomized Decision Tree algorithm.Depth cameras helped in reducing the ambiguity of silhouettes.SVM classifier was then applied on these extracted 3D joints to confirm falls.Suneung Kim et al.[5] developed a method using Extended Kalman Filter in which both two dimensional and three dimensional informations of a dynamic scene were utilised.Depth maps were generated from RGB input frames using learning techniques.PSO(Particle Swarm Optimization) was used to track human bodies and EKF was used for three dimensional human tracking.CNN(Convolutional Neural Network) is one of the best classifiers available in the area of image processing,natural language processing and in many other areas.And inorder to develop a highly efficient and accurate fall detection system here we used CNN as our classifier.

III. CONVOLUTIONAL NEURAL NETWORK

A neural network is designed similar to the working of neurons in the human brain. In case of traditional neural networks,the images would have to be fed to the network as reduced-resolution pieces and they could not be utilised effectively for image classification. In CNNs the arrangement of neurons is similar to the visual stimuli of humans and other animals.This helps in overcoming the problems faced by traditional neural networks.CNNs are deep artificial neural networks that help in accurate image classifications.A ConvNet succeeds in highly efficient classifications of the input image by the application of numerous filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better.They can identify humans,animals,other objects,street signs etc.They can be applied in various fields such as text recognition,sound recognition(when sound is represented as spectrogram),text analytics etc.The convNet helps in such a way that the features that are very critical for prediction are not lost.This ability of convNet is what it makes scalable to be applicable to many large datasets apart from learning features from the input images well.

The efficiency of convolutional nets (ConvNets or CNNs) in image recognition and classification has motivated people around the world to learn more about deep learning.CNNs are nowadays extensively used in computer vision (CV), which has a lot of applications for robotics,security, medical diagnosis, and treatments for the visually impaired,self-driving cars etc. The convolutional Layer is the most important layer of the Convolution Neural Network and the main purpose of a convolution operation is to extract high level features from the input image such as edges.ConvNets can have more than one convolution layer.Lower level features like gradient orientation,color etc are learned by the first convLayer.These convolutional layers have sets of independent filters and the image is convolved with each of these filters independently.So,if there were 7 filters in the convolution layers,then the output would be 7 feature maps.The output that we obtain by convolving an image with a filter is called a feature map.As we add more and more layers,the convNet learns more higher layer features from the input images just like humans would.If a number of convolution layers are arranged in sequence,all the filters in there are randomly intitialized and these become parameters,which the network learns subsequently.The filters in a convolution layer do dot products to the input of the previous convolution layer or in other words they are taking small portions of the features and making larger features out of them.

Parameter sharing and local connectivity are two important concepts of CNN.Parameter sharing refers to the sharing of weights by all neurons in a single feature map and local connectivity means that each neuron is connected to only a portion or only a subset of the input image,unlike the fully connected neural networks.These concepts make the CNN computationally efficient.Pooling layer is another important building block of CNN.It also helps to reduce the computational overhead and the amount of parameters by reducing the spatial size of the representation in a progressive way.Max Pooling is the most commonly used approach.Activation function or Transfer function is a function used to output function used to get output at nodes.The activation functions could be linear or non linear.Non linear activation functions are the most commonly used activation functions and there are different types of them such as ReLu(Rectified Linear Unit),Sigmoid,tanh activation functions etc.Another most commonly used activation function used for classification of features is softmax activation function which is also non linear.Here we selected one of the best models of CNN ,that is Inception model which suited our task of detecting falls the best than the other options available.

IV. MATERIALS AND METHOD

The main objective behind the design of the system is to develop a human fall detector that could reduce the large number of steps of normal image processing systems using Inception V3 CNN model and achieve the state-of-the-art accuracy using different standard publicly available fall datasets such as URFD(UR Fall Detection Dataset) and FDD(Fall Detection Dataset).As discussed earlier,the fall detection systems based on ambience sensor and wearable sensors have a lot of drawbacks that makes it use not so applicable for daily life usage.The fall detection design based on the Inception model CNN helps to improve the accuracy of the fall detection.In this way the designed system could get rid of the false alarms or reduce the number of false positives and could be used to make a more reliable fall detection system in the real world environment.

A. *Architecture of the Neural Network*

CNN has proven to be one of the best image classifiers and is being widely used in image and video processing.The decision of choosing the concept of deep learning and CNN has been very crucial for the system design.A number of architectures have been introduced in the area of computer vision for classification of images.LeNet-5,a seven level CNN was used by several banking institutions in 1998 for recognizing hand-written numbers on cheques.LeNet-5 was applied to recognize and classify the digits,however for efficient classification of higher resolution images,more number of convolution layers are required and thus this method is constricted by the availability of computational resources.AlexNet was introduced in the year 2012 ,with an architecture deeper than LeNet ,with more number of filters in the layers and stacked convolution layers.AlexNet included 11*11,5*5,3*3 convolutions and ReLu activation function for feature learning.Later ZFNet was introduced with the same architecture as that of AlexNet but included more deep learning elements.Then came VGGNet with 16 convolution layers and a higher accuracy hoewever the number of parameters were about 138 million.Inception or GoogleNet performed better than all the other models with parameters reduced to about 4 million parameters.Residual Neural Network,ResNet outlived the early versions of Inception by featuring heavy batch normalization.However the modified versions of Inception model were able to perform better than ResNet in the recent years.

Here we use Inception v3 for detecting falls from the input video frames.Inception v3 is an image recognition model that consists of two major parts.The first one is feature extraction which is done by a convolutional neural network and the second part is classification of features which is done by fully-connected and softmax layers.Prior to the introduction of Inception network the CNNs used to stack convolution layers deeper and deeper inorder to get high accuracy.But the major task was selecting the important or salient portions of the image.In the case of detecting a fall from an image or a frame,the human in the frame could occupy any part of the frame and the CNN network has to determine where the object of interest is,inorder to perform the convolution operations.The location information of the object in an image varies from frame to frame and thus it is important to choose the right kernel.For informations that are distributed locally small sized kernels are used and large sized kernel would suit images with more globally distributed information.When the networks are very deep,the problem of overfitting may arise and cost of computation might also be very high.The Inception network is a more complex network,which is heavily engineered and something that could achieve better results in terms of speed and accuracy.In the Inception network instead of deeper layers, filters of multiple sizes are made to operate on the same level.The idea used here is to create a more wider network than a deeper network.

The number of input channels in the network are limited in the basic inception module by adding extra 1*1 convolutions before adding the 3*3,5*5 convolutions.The architecture of Inception V1 is built based on the inception module whose dimensionality is reduced.About 9 inception modules are included in the Inception network and to avoid the “dying out” of the network’s middle portion 2 auxiliary classifiers are added.In the Inception version 2 representational bottlenceks are reduced and in Inception version 3 batch normalization is added to the auxiliary classifiers,7*7 convolutions are factorized,RMSProp optimizer is used and label smoothing is also done.Inception V3 is a CNN model that is pretrained with ImageNet dataset which has over thousands of classes.So here we are doing supervised learning technique that is teaching the machine using well labelled data.Inception V3 helps us to reduce the time of computation as its pretrained on ImageNet and hence on comparison with training from scratch,we could save a lot of time.That is here we are applying the concept of transfer learning,applying previously learnt knowledge to learn new things.We already know that neural networks are organised just like the neurons of our brain and the input signal has to pass through multiple layers until it reaches the output layer.By the time it reaches the last hidden layer there would be enough information to be passed to the next set of layers which would classify the input in to specific labels.For the design of our fall detector we used the FD dataset.So here we remove the old layer of the network and add a new layer on the frames or images of our dataset

B . Image Preprocessing

The preprocessing stage is done to process all the images equally and store them in appropriate sub folders under the 'train' category.Here we selected four labels which are "stand","sit","bend" and "fall".So by the end of the preprocessing stage,we have all the images in the dataset stored under their specific categories.Data augmentation is done to tackle the problem of overfitting, if it occurs.We perform the operations in both of the specified datasets,i.e FDD and URFD.

C . Object Detection

The next step after preprocessing is to determine whether the object in question is human or not.Here we use Mobilenet for detection of humans from the input frames as it best suites the object detection and only if the object of interest is a human,the fall would be detected.

D. Retraining And Fine Tuning

Retraining is the most important part of the algorithm in any of the image processing jobs.This step trains a bottleneck layer which has information to classify the images.The 'bottleneck' is just a term that we call the layer just before the final output layer that performs the actual task of classification.It is also called as the image feature vector and bottleneck caches are stored inorder to avoid repeated recalculations as each image is used repeatedly,a number of times during the process of training.The bottleneck layer receives a 2048-dimensional vector for each of the input images.A softmax layer is also added on top of this which has N labels and thus it ends up learning about N + 2048 * N parameters.During retraining, JPEG decoding is also performed and we take the training images in JPEG format.A series of outputs are obtained during actual training of the top layer that showed training accuracy,validation accuracy,cross entropy values etc.We configured the following things correctly:

a. Image directories - The path to our labeled images.
b. Output directories - The places where we save output files such as graphs,intermediate graphs,output labels etc
c. Distortion features- Configured the system so as to not lose the prediction accuracy while dealing with blurred or dimly lit input frames.
d. Number of training steps are set accordingly as 500 and also the learning rate were selected differently for URFD and FDD.
e. After successful training ,we tested the system using a fresh set of images and also displayed the confidence values for each class on the output.We repeated the same process by training with URFD and testing the system with a new set of images.Learning rate,number of epochs etc are a few of the deterministic features that we can adjust in the way we please.

V. RESULTS

Learning rate,number of epochs etc are a few of the deterministic features that we can adjust in the way we please For the better evaluation of the fall detection system we chose the best performance metrics such as sensitivity,specificity etc.Sensitivity is also called as recall value or true positive rate and Specificity is the true negative rate.

Specificity = TN/(TN + FP)
Sensitivity = TP/(TP + FN)
Accuracy = TP/(TP + TN + FP + FN)

TP means the true positive rate which means that the prediction of the action in the input frame as fall was completely right.TN refers to the true negative rate which means that the action in the input frame was not fall and the prediction also came out as not fall.FP or false positive is when an input that is not fall is concluded as fall and False Negative is when the input is actually fall but prediction comes as not fall.

Dataset	FDD	URFD
Learning Rate	10^{-2}	10^{-5}
Train Batch Size	100	64
Activation Function	ReLu	ReLu
Sensitivity	93.47 %	100%
Specificity	97.2%	94%
Accuracy	98.5%	91.5%

Table I : Comparison of the efficiency of the system on datasets URFD and FDD.

The accuracy rates obtained for the fall detection in Fall Detection Dataset is very high however for the UR Fall Detection Dataset,the value is less due to the highly imbalanced nature of the dataset.So that's why here we took specificity and sensitivity as the key metrices for performance measurement. It could be clearly seen that the system showed excellent results in predicting true positives for the URFD dataset whereas the system predicted true negative rate well in FDD more than that in URFD.And also we were able to achieve state-of-the-art accuracies for both of these public fall datasets with our system design.After testing the calculated confidence values.

VI. CONCLUSION

In this paper we studied how to successfully apply transfer learning for creating a fall detection system which is based on computer vision and obtained state-of-the-art accuracies in the public fall detection datasets such as FDD and URFD.We trained the CNN in different datasets and this kind of retraining helped in learning a variety of generic features,reducing the large number of steps normally involved in complex image processing systems.Transfer learning could successfully tackle the problem of low number of samples in the datasets.However for successful deployment of such a design of a fall detection system in the real world,many aspects have to be considered in detail.Inorder to learn more generic features the layers in CNN could be modified more deeply,paying attention to the fact that it doesn't end up being more specific.All the publicly available fall datasets have only one single actor and one of the most promising research direction in the area of fall detection would be focusing on systems that could detect multiple persons and analyze their fall behavior.

REFERENCES

[1]. Brownsell S ,Hawley MS "Automatic fall detectors and the fear of falling", 2004 , Journal of Telemedicine and Telecare. Amuzuvi, C. K., & Addo, E. (2015).

[2]. Rui-dong Wang,Yong-liang Zhang ,Ling-ping Dong, Jia-wei Lu , Zhi-qin Zhang, Xia He , 2015, "Fall detection algorithm for the elderly based on human characteristic matrix and SVM",15th International Conference on Control, Automation and Systems (ICCAS).

[3]. Zhen-Peng Bian, Junhui Hou, Lap-Pui Chau, Nadia Magnenat-Thalmann ," Fall Detection Based on Body Part Tracking Using a Depth Camera", 2015 , IEEE Journal of Biomedical and Health Informatics (Volume: 19,Issue: 2,March 2015).

[4]. Zhen-Peng Bian, Junhui Hou, Lap-Pui Chau, Nadia Magnenat-Thalmann ," Fall Detection Based on Body Part Tracking Using a Depth Camera", 2015 , IEEE Journal of Biomedical and Health Informatics (Volume: 19,Issue: 2,March 2015).

[5]. Suneung Kim, Myeongseob Ko, Kyungchai Lee, Mingi Kim, Kwangtaek Kim , "3D fall detection for single camera surveillance systems on the street" , 2018, IEEE Sensors Applications Symposium (SAS).

[6]. Homa Foroughi , Alireza Rezvanian ,Amirhossien Paziraee,"Robust Fall Detection Using Human Shape and Multi-class Support Vector Machine", 2008, Sixth Indian Conference on Computer Vision, Graphics and Image Processing.

[7]. Chih-Yang Lin , Shang-Ming Wang , Jia-Wei Hong , Li-Wei Kang , Chung-Lin Huang,"Vision-Based Fall Detection Through Shape Features",2016 , IEEE Second International Conference on Multimedia Big Data.

[8]. Caroline Rougier, Jean Meunier, Alain St-Arnaud, and Jacqueline Rousseau ,"Robust Video Surveillance for Fall Detection Based on Human Shape Deformation ", 2011, IEEE Transactions on circuits and systems for video technology,vol.21,No.5,May.

[9]. Xu Zhou , Li-Chang Qian , Peng-Jie You, Ze-Gang Ding , Yu-Qi Han, "Fall Detection Using Convolutional Neural Network With Multi-Sensor Fusion" , 2018 ,IEEE International Conference on Multimedia and Expo Workshops (ICMEW).

[10]. S.-G. Miaou, Pei-Hsu Sung, Chia-Yuan Huang, "A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information",2006 ,1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare. D2H2.

[11]. Lesya Anishchenko ," Machine Learning In VideoSurveillance For Fall Detection", 2018, Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT).

[12]. C Mai Nadi ,Nashwa El-Bendary,Aboul Ella Hassanien,Tai-hoon Kim ,"Falling Detection System Based on Machine Learning" , 2015 ,4th International Conference on Advanced Information Technology and Sensor Application (AITS).

[13]. Na Lu , Xiaodong Ren,Jinbo Song,Yidan Wu ,"Visual guided deep learning scheme for fall detection" , 2017 ,13th IEEE Conference on Automation Science and Engineering (CASE).

[14]. Kamal Sehairi, Fatima Chouireb, Jean Meunier, "Elderly fall detection system based on multiple shape features and motion analysis", 2018, International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, pp. 1-8.

[15]. Myeongseob Ko ,Suneung Kim, Kyungchai Lee,Mingi Kim, Kwangtaek Kim , 2017 , "Single camera based 3D tracking for outdoor fall detection toward smart healthcare" , 2nd International Conference on Bio-engineering for Smart Technologies (BioSMART).

[16]. Lykele Hazelhoff,Jungong Han and Peter H.N. de ,"Video-Based Fall Detection in the Home Using Principal Component Analysis" ,2008, 10th International Conference, ACIVS , Juan-les-Pins, France, October 20-24. Proceedings.

[17]. Hong Lu,Bohong Yang,Rui Zhao,Pengliang Qu,Wenqiang Zhang "Intelligent Human Fall Detection for Home Surveillance" , 2014, IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops.