WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Human Fall Detection System Based on LSTM

Ms.V.Narmadha[1], Bilal.A[2], Karthik.K[2], Avaneeth.A[2]

[1]Professor, Department of IT, Sri Sairam Engineering College,
narmadha.it@sairam.edu.in
[2]UG Student, Department of IT, Sri Sairam Engineering College

ABSTRACT

Fall is one of the most difficult challenges for elderly, pregnant, and young children who live alone at home.This type of fall might result in serious injuries or even death. For senior persons, detecting a fall is quite crucial.According to the study, the majority of fall detection based on video monitoring is complicated and redundant, which has an influence on the real-time and accuracy of detection.In light of the aforementioned issues, this work provides a video-based fall detection approach in a complex environment, with the goal of detecting fall behavior more correctly and promptly.This paper's key contribution is as follows: First, a detection technique based on the LSTM network model is developed. Second, using the Pascal VOC data set format, the human fall detection data set is created.The method model is then adjusted and trained in a deep learning server using GPUs (graphic processing units).Finally, testing results show that our detection approach has a substantial recognition impact when compared to our LSTM network model and other detection techniques.

I. INTRODUCTION

The health of the elderly will not be regular all of the time.Because they may be suffering from one of the frequent ailments, an unintentional fall might result in serious harm or even death.There are already sensor-based gadgets, such as an accelerometer and a watch, that can alert them when they require assistance.However, occasionally elderly persons forget to wear these sensor devices or are unable to hit the button due to unconsciousness following a rapid fall.This quick fall shock might cause a heart attack in the aged people.According to the data, over billion elderly persons are treated as emergencies as a result of fall injuries.Heart disease, diabetes, depression, stroke, cancer, and a variety of other chronic disorders are among the most frequent chronic diseases among the elderly.Some people may lose their balance, faint or lose consciousness, or experience dizziness as a result of the chronic disorders indicated above, which can lead to a fall.Although ambience devices are used to detect old people's movements, the main issue is that it also detects non-object noise signals, resulting in false alarms.The accelerometer in wearable sensors will detect a fall; however, the main problem is that elderly people may forget to wear the sensors, and if they fall while not wearing the sensors, no one will notice.One or more cameras will be positioned to detect the fall using computer vision.The cameras are here to keep an eye on the elderly's actions at all times. The fundamental issue is that real-time films necessitate a lot of processing power, and everyday acts like napping and leaning down might be misinterpreted as falls. These sorts of actions should not be considered falls by this technique.Deep learning technology has advanced fast in recent years, because of ongoing improvements in processing performance. The deep convolutional neural network was proposed by Krizhevsky et al. [1].From R-CNN (region CNN) to Fast R-CNN [3], Faster R-CNN [4], Girsick et al. [2] has done a series of effective work in conjunction with depth neural network and image target recognition.Redmon et al. [5] introduced a novel LSTM (you only look once) target identification approach, while Liu et al. [6] proposed the SSD algorithm, which combines the LSTM regression notion with the Faster R-CNN anchor box mechanism.Then, Redmon et al. [7] improved the LSTM algorithm and presented the LSTM method.The LSTM method is proposed in this work, and it has the benefits of excellent accuracy and cheap computing complexity.The premise of the LSTM algorithm, the method of detecting falls, model training, and testing using a self-built data set will all be covered in the next chapters.The detection findings are more accurate, and the detection time is faster than other traditional approaches, according to this study, which improves the accuracy of the current popular LSTM method.

II.ARCHITECTURE

1.Structure of Darknet-53

At the moment, deep learning-based target detection tasks largely consist of two types of detection methods:Fast-RCNN is a two-stage detection model based on region suggestion to extract target candidate areas, whereas LSTM, SSD, and others are one-stage detection models based on regression thinking.Although the accuracy of a one-stage detection model is slightly lower, it is faster at detecting real-time events.A human fall is an unexpected event. At the quickest possible speed, the fall target must be precisely detected.As a result, this work employs the LSTM algorithm, which considers both speed and detection accuracy.This network is made up of a sequence of 1*1 and 3*3 convolution layers, and it's termed Darknet-53 since it contains 53 convolutions.

2.Darknet Residual Component

The residuals are linked to the residuals structure of the resnet network and provide quick links between some layers, minimizing deep network degradation and allowing for a more complex network structure.The input to Darknet-53's network is 256*256*3, and the numbers 1, 2, 8 in the left column denote the number of duplicate residual components.Two convolution layers and a fast connection are present in each residual component.

3.Network Structure of LSTM

LSTM takes the feature fusion pyramid notion of FPN (feature pyramid networks) as a starting point, and then uses up sampling and feature fusion to produce three types of feature graphs (13*13, 26*26, 52*52) with varied sizes.Multi scale feature detection can improve the detection impact of big and small objects by increasing the richness of features .LSTM keeps the K-means clustering algorithm from LSTMv2, but raises the number of anchors to nine and assigns three separate anchors to each scale feature graph.The network's capacity to recognise tiny things is considerably increased as a result of this.In addition, the prediction layer's softmax function has been replaced with a logistic function, which may handle the output of multi-label objects.The anchor parameter is incorporated into the LSTM algorithm in this work. The LSTM algorithm is a series of predefined boxes with predetermined width and height values.The size of the preceding boxes has a direct impact on the speed and accuracy of target recognition, therefore while training the fall data, it's critical to establish the network parameters according to the intrinsic features of the fall labels.The K-means clustering technique is used to examine the dimension of human fall labels in order to obtain the optimal training impact.

4.Loss function

The parameters in the network are continually modified during the model's training process, the loss function is optimized to the smallest value, and the model's training is finished.The prediction error of the center point (x, y), the prediction error of the width and height (w, h), the confidence error, and the classification prediction error make up the loss function of LSTM.The target location offset loss Lloc (l, g), target confidence loss Lconf(o,c), and target classification loss Lcla (O,C) are the three sections of the LSTM loss function, with 1,2,3 as the balancing coefficient.The target confidence may be thought of as the chance of the target being found within the target's rectangle box.The binary cross entropy loss is used for the target confidence loss and the target category loss.The adjustment total of the difference between the true deviation value and the forecasted deviation value is used to calculate the target positioning loss.

5.Feature Extraction

The picture is first scaled to an uniform form of 416 pixels in length and width, which serves as the network's input.Second, features are extracted using the Darknet-53 network, and convolution operations are performed using 3*3 and 1*1 convolution kernels, respectively.The 77th, 84th, and 94th layers' outputs of 13*13*512 dimension, 26*26*768 dimension, and 52*52*384 dimensions are taken as three features, which are delivered into the system after dimension reduction.The final weight model is created in the LSTM layer by training three scales. Finally, a marked picture of the human fall test is generated.

6.Network Prediction Process

The LSTM algorithm's network structure is used to identify human falls, and the specific detection technique is as follows:

1) Image preprocessing is performed to process the fall data in the training set, and the unified image is then utilized as the input to the whole training network.
2) For fall feature extraction, the processed picture is submitted to the Darknet-53 network.
3) The first feature is taken from the 77th layer output, which is twisted and sampled once.
4) The 83rd and 61st layer outputs are spliced together to create the second feature, which is twisted and sampled once.
5) The third feature is created by combining the outputs of the 93rd and 36th layers.
6) Three features are passed to the LSTM layer for training, and the iteration is halted when the training times are completed to obtain the final weight model.
7) Input the test set's picture into the same network, use the training weight model to identify the test set's image, and output the results.

III. IMPLEMENTATION

1.Data Set

Set of data The quality of the data collection has a direct impact on the final detection effect in deep learning.At the same time, the network requires a sufficient number of samples to properly understand the features of the object to be identified.The experiment's data set, which totaled 2600 images, was compiled using the open data set and additional photographs found on the Internet.To begin, the photos in the data set were sorted according to the VOC2007 data set format, and then separated into two groups at random: training set and test set.Second, the labeling tool was used to mark each picture in the training set one by one, and an XML format target box position information file was created.Finally, a python application was created to normalize the object frame's location information in XML format and convert it to TXT format as the data set label for human fall detection.

2.Evaluating Indicator

The mainstream mAP (mean average precision) metric was chosen as the assessment indicator in this work.In this study, the mainstream mAP (mean average precision) measure was used as the assessment indicator.When the maximum accuracy is more than or equal to these recall values, move on to the next step. Take the highest precision under this recall as the independent variable, and the recall value as the dependent variable. As the AP of this target, draw the curve and the area under the curve.The missed detection rate of recognising a target is referred to as the recall rate, while the accuracy rate is referred to as the precision rate.The technique of computation is as follows: (2) (3). The number of correctly identified target classes is true positive, the number of wrongly identified target classes is false positive, and the number of incorrectly recognised target classes is false negative.

3.Model Training

The model training setting in this article is Ubuntu 18.04, python programming, tensorflow open source framework 1.14, keras open source framework 2.2.4, cuda10.0 cudnn7.4, video card 2070s.The process of parameter fitting in a model is known as model training. The error between the network output score and the predicted score is calculated using supervised learning in this research.To lower the error, the acquired error is utilized to adjust the network's internal settings.The loss function is optimized during the training process using the batch random gradient descent technique, with a total of 200000 batches learned.The initial learning rate is set to 0.001, the weight attenuation value is set to 0.0005, the batch size is set to 16, and the average loss tends to zero, signifying convergence. The loss curve in training

4.Test result

The performance of existing common target detection methods was previously reviewed.On the COCO dataset, the test results reveal that the LSTM algorithm has a higher mAP and a faster inference time.On the ImageNet dataset, a comparison of the effects of Darknet-53 and other networks is demonstrated.It can be observed that Darknet-53 has a similar impact as ResNet-152, although it is twice as fast.

The matching test data set's findings are displayed.The test findings are divided into two categories: We can demonstrate that our approach outperforms other algorithms on the VOC2007 and VOC2012 datasets by comparing the detection results of this network model with those of other techniques.person and down. The down class has an AP of 0.971, the person class has an AP of 0.689, and the resulting mAP is 0.83.

IV. CONCLUSION

This suggested system Methodology of Fall Detection is primarily intended for elderly persons who are unable to always be in the care of others.This category may also include children whose parents are outside, as well as pregnant women who can't always do their work, especially during that time, and we can't predict what will happen to these people when they are at home, and unfortunately, if something bad happens, these people won't be able to contact them, and they may be unable to get up.Imagine someone always being by your side and assisting you when anything unexpected happens at this important time.The end-to-end detection technique LSTM is used in this study, and the anchor parameters of LSTM are improved using the K-means clustering algorithm, and the network model is trained and evaluated on a GPU server with self-created data.The mAP of the method is 0.83, and the AP of down is 0.97, indicating that it is superior to other standard algorithms in terms of detection effect and resilience.
Human fall detection behavior and human health protection are extremely important.

REFERENCES

1. Alex Krizhevsky, Ilya Sutskever,Geoffrey E. Hinton, “Imagenet Classification with Deep Convolutional Neural Networks,” Advances in Neural Information Processing Systems, vol.25, Jan. 2012, pp. 1097- 1105, doi:10.1145/3065386.
2. United Nations Population Fund (UNFPA) and Help Age International. “Aging in the Twenty–first Century: A Celebration and a Challenge,” New York, 2012.
3. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhad, “You Only Look Once: Unified Real- Time Object Detection,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, June.2016, pp. 779- 788, doi: 10.1109/CVPR. 2016.91.
4. Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan., “Feature Pyramid Networks for Object Detection,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, Apr.2017, pp. 936- 944, doi:10.1109/CVPR. 2017.106.
5. Joseph Redmon, Ali Farhadi, “LSTM: An Incremental Improvement,” arXiv: 1804.02767[cs.CV], Apr.2018.
6. J. Willems, G. Debard, B. Vanrumste, and T. Goedeme. “A video-based algorithm for elderly fall detection,” World Congress on Medical Physics and Biomedical Engineering, Springer Berlin Heidelberg, 2009, pp. 312-315.
7. Panagiotis Tsinganos, and Athanassios Skodras. “A Smartphone-based Fall Detection System for the Elderly,” 10th International Symposium on Image and Signal Processing and Analysis, 2012
8. S.J. Preece, J.Y. Goulermas, L.P.J. Kenney, D. Howard, K. Meijer, and R. Crompton. “Activity identification using body- mounted sensors - a review of classification techniques,” Physiological Measurement, 2009.
9. akumi Kaneko, and Meifen CAO. “Human Fall Detection Using CHLAC Features with Skeletal Image Sequences,” IEEE International Symposium on Robotics and Intelligent Sensors (IRIS2016), 2016.
10. Bogdan Kwolek, and Michal Kepski. “Human fall detection on embedded platform using depth maps and wireless accelerometer,” Computer Methods and Programs in Biomedicine, 117(3), 2014, pp. 489-501.