European Journal of Molecular & Clinical Medicine

ISSN 2515-8260 Volume 09, Issue 02, 2022

An optical fall detector adaption at nursing homes using Flow based CNN

[1]C. Vasantha Kumari,

Assistant Professor Department of Medical Surgical Nursing Sri Venkateswara College of Nursing, Chittoor – 517127, AP

[2]Prof. V. Sujatha,

Dean, & Professor,Department of OBG Nursing, Sri Venkateswara College of Nursing, Chittoor – 517127, AP

[3]Prof.E. Sumalatha,

Professor Department of OBG, Sri Venkateswara College of Nursing, Chittoor – 517127, AP

[4]Prof. C. Rathiga,

Department of Community Health Nursing, Sri Venkateswara College of Nursing, Chittoor – 517127, AP

[5]Dr.S. Maha Lakshmi,

Professor Department of Community Health Nursing, Sri Venkateswara College of Nursing, Chittoor – 517127, AP

Abstract - It's difficult to identify falls in senior care facilities. Because the resident, who may be suffering from mental illness, is not instrumented by vision-based fall detection systems, it is a major advantage. Fall detection technology are being used in nursing homes as part of this research. A Convolutional Neural Network (CNN) trained to optimise a sensitivity-based metric is the basis of the suggested solution, which is based on Deep Learning techniques. It discusses the medical criteria and how they affect the CNN tuning in this study. The findings demonstrate the relevance of the timing component of a fall. Consequently, the medical team's aims are best served by a specific measure tailored to this use case and an implementation of a decision-making process.

1. INTRODUCTION

Fall is the primary cause of trauma-related death in senior care facilities because residents fall an average of 1.7 times each year in France [1]. After a few hours, medical experts may be able to find a person who has fallen [2]. For a fall detector to be used in this context, false alarms that impede medical personnel's work are unacceptable. Based on study with medical teams, patients, and families in three diverse nursing homes, a solution must be able detect as many falls as possible with no false alarms, be non-intrusive and easily reconfigurable and be able to adapt to varied people.

The two main types of fall detection technology are sensor-based and vision-based. Because wearable sensor-based solutions do not meet the requirements of medical professionals, this initiative focuses on vision-based solutions instead. In reality, they fall short when coping with the growing number of older people suffering from mental diseases. Medical teams, patients and their

families typically support the use of cameras for patient safety and self-reliance, according to the findings of the study.

In most cases, a fall leads in a change in velocity and position of the human body. Such features as 2D human body posture estimations [3], movement vectors or figure silhouettes using background removal techniques [4] may be retrieved from photographs in image-based methods. GMM and SVM [5–7] classifiers utilise these traits to identify the person and calculate their spatial body orientation. Locals started to consider the temporal character of falls as a result of the difficulty in explaining them. Long-Short-Term Memory (LSTM) [8] networks have been introduced, and they provide advancements in the field of vision-based fall detection. It's also possible to employ a 3D Convolutional Neural Network (CNN) for this purpose [9] and [10].

There are two types of falls among the elderly: hard falls (falls that begin from a standing position) and soft falls (falls that begin from a different starting point) (disabilities or old age). Therefore, they are a challenge to model.

A neural network-based approach that takes into consideration the variability of human body motion therefore looks to be well-suited to the problem at hand. Using a CNN that was trained to maximise sensitivity, the suggested technique use optical flow as the basis for its analysis. In this paper, we provide a unique training approach and a decision-making process modification that reduces false alarms and ensures a reasonable accurate detection rate for medical staff demands, based on realistic metrics.

2. METHODOLOGY

There are three steps to the overall solution shown in Fig. 1. Dense optical-flow TV-L1 technique [12] is used to construct two optical flow pictures from two sequential RGB images from a camera.

As in [13], a custom VGG-16 CNN created and pre-trained in [11] is the second part of the technique. It makes a fall prediction based on a stack S of L = 10 successive optical flow pictures as input. Finally, in order to take use of the fall temporal characteristic, the CNN output is subjected to a bespoke temporal filter and a prediction threshold. Therefore, if a single or consecutive S stacks are categorised as fall at the temporal filter output, medical staff will get a fall alarm.

A. Data base and training procedure

For both training and testing, three fall video datasets are used. Falling or not, they are produced from videos. A non-fall annotator is one who focuses on actions that individuals do on a daily basis. There are three distinct elements to a fall film: the events leading up to the fall, the actual fall, and the aftermath, which typically depicts a character lying on the ground. The CNN classifier may provide two potential output classes: Fall and No-Fall. This class includes pre-fall, post-fall and non-fall video sequences. In many fall-action databases, there are more mundane events than actual falls. A weighted binary cross-entropy loss function (1) is used in [11] to overcome unbalanced data and fine-tune the classifier, which is constructed from two final Fully-Connected (FC) layers of the CNN. P is the network's prediction, and t is the ground truth; w0 is the fall class weight and w1 is the No-Fall class weight in this equation.

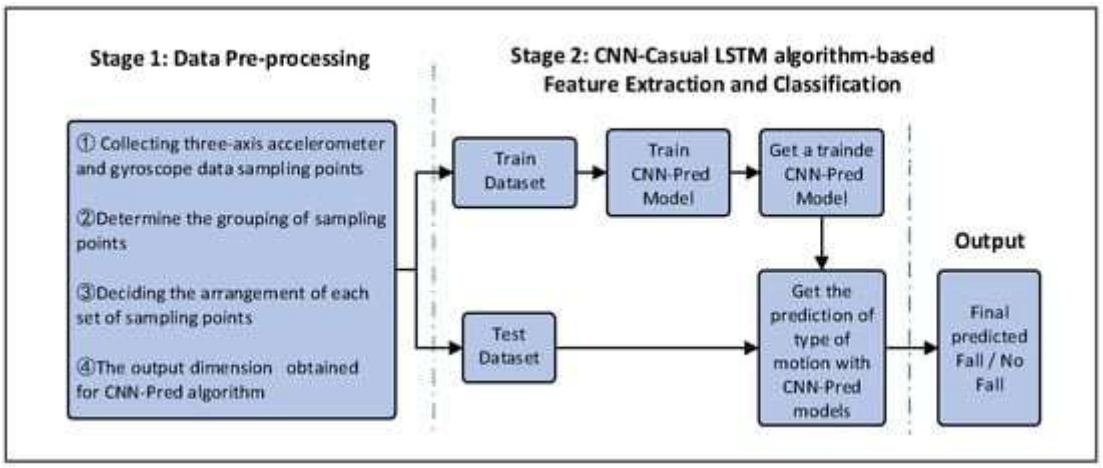

Fig. 1. Using a convolutional neural network to detect falls

$$loss(p,t) = -(w_1 \,.\, t \,.\, log(p) + w_0 \,.\, (1-t) \,.\, log(1-p)) \quad (1)$$

By using transfer-learning, fall detection may be achieved even with a little amount of data. The whole network has been frozen, with the exception of the latest two FC levels. Finally, using the 5-cross fold validation, the final two FC were trained using autumn datasets. As a result of these differences, our training method varies from that of [11] in three ways:

• During the 5-fold validation, each video sequence is stored in a separate fold. As a consequence, the stacks on the train set and the test set will be different. Both the train and validation sets are stocked with a selection of movies in order to avoid overfitting. Segments from a certain initial fall video sequence are stacked on top of one another in the same fold.

•During testing, a Transition class is used in addition to fall and No-Fall classes in order to provide students with a more realistic environment. This lesson focuses on the transitional and post-fall frames of thought. In addition, there is a set of Prefall frames provided. These precise testing frames are not mentioned in [11] or [13].

Grid search algorithm is employed with the following hyperparameter ranges to maximise training efficiency: Learner's rate of progress is: 102, 103, 10–4, 10–5, 10–6 There are four options for batch size Bs: 128, 256, or even 1024.

Classes w0: 1, 2, 5, 10, 15, 20; w1: 1

A fact about the activation function of classifiers: ELU, ReLU

When evaluating configurations, we look at their specificity sp (2), sensitivity se (3), and precision p (4), with T P denoting True Positives, T N denoting True Negatives, F P denoting False Positives, and F N denoting False Negatives. The ideal hyper-parameters configuration is determined by computing these metrics over stack forecasts.

$$sp = TN/(TN + FP) \quad (2)$$

$$se = TP/(TP + FN) \quad (3)$$

$$p = TP/(TP + FP) \quad (4)$$

To achieve greater sensitivity, authors in [11] focus on reducing detail and correctness. Specificity and accuracy are important in our situation, but sensitivity must be sacrificed in order to meet the medical staff standards outlined in Section I.

B. Fall evaluation

Classifying a fall may be difficult since it is difficult to pinpoint exact beginning and conclusion of event.

When evaluating predictions, it is also necessary to consider the duration of the decline. In Section II-A, the network states that it uses L consecutive optical flow images to make a forecast. Table I shows that the average fall time in the studied datasets is 1.11 seconds. For a 30 FPS recording, a fall prediction is created for 1/3 of the normal fall time. When a fall is filtered, false alarms and safe signals are eliminated, enhancing the time component of a fall.

TABLE I DATABASES PROPERTIES

Database	Frame rate (FPS)	Avg. fall duration (frames - seconds)	Number of falls
URFD	30	30 - 1.00	30
FDD	25	24 - 0.96	99
Multicam	30	41 - 1.36	200
Avg.	28	32 - 1.11	-

Predictions are no longer seen as stacks but as consecutive identical stack prediction types in the temporal analysis step. Detecting accurately observed falls is indicated by the letter "T," whereas false alarms are signalled by the letter "F" or "F Na," respectively. A gate function is used to describe the built convolution filter. The filter's prediction threshold Tpred and its width W (measured in frames or seconds) dictate its properties. If the prediction falls below Tpred, it is labelled "fall."

The goal is to reduce the frequency of false alerts while not missing any falls. As in [18], the capacity to measure this capability, F (7), is a function of the alarm precision spa (5) and the alarm sensitivity sea (6). False alarms are emphasised more when 0 1 and less so when the measure is greater than 1. A fall detector may be evaluated realistically in terms of medical requirements if this parameter is included.

$$p_a = TP_a/(TP_a + FP_a) \tag{5}$$

$$se_a = TP_a/(TP_a + FN_a) \tag{6}$$

$$F_\beta = (1+\beta^2) \cdot \frac{p_a \cdot se_a}{(\beta^2 \cdot p_a) + se_a} \tag{7}$$

3. Results

A. Hyper parameters choice

An acceptable balance between high specificity and sensitivity is achieved by adjusting the hyper-parameters of CNN training, as previously described.

The learning rate is the first hyper-parameter that may be changed. From the values evaluated, 102 s is too high, which causes the model to deviate from the truth. The model's convergence is too slow for a learning rate of less than 104.

Concerning Bs, [11]'s (i.e. 1024) batch size selection may not have resulted in an extremely converged model. We found that a smaller batch size (e.g., 128 or 256) led to a better degree of precision. Because of a rise in the F N, the sensitivity of the sensor decreases, although the alarm sensitivity is only slightly affected.

ELU activation function provides more sensitivity, whereas the ReLU activation function provides greater specificity. Our use case necessitates the employment of ReLUs, hence they are chosen.

Finally, a Receiver Operating Characteristic (ROC) analysis is performed on the class weight w0 (w1 is arbitrarily set to 1) to concentrate on the Fall class and chose the configuration that provides the best specificity. In reality, when w0 is more than 5, it indicates that the results are unstable, regardless of whether the data in each class is balanced or imbalanced. Overfitting on the F all class is avoided when w0 is adjusted to 2, as shown in [11].

Prediction outputs during training are analysed using metrics specified in section II-C to improve model performance. F Pa and Fall labels, shown in Fig. 2 as offset, may be used to categorise false alarms by counting the number of frames between the two labels.

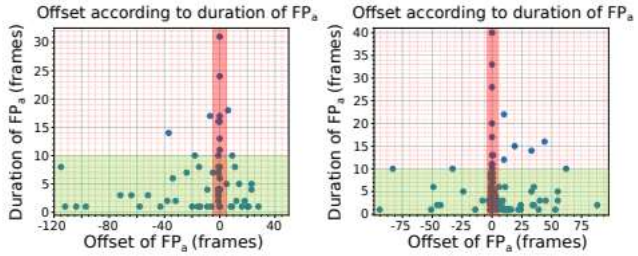

Fig. 2. Offset fall class by F Pa. The horizontal zone (green) contains F Pa under 10 frames.

F Pa offset is less than 5 frames in the red vertical zone. The URFD and FDD are on the left and right, respectively.

39 percent of the F Pa in the three datasets had an offset of less than 5 frames from the actual fall. Truth be told, these predictions are incorrect, although there is some wiggle room. Indeed, from a human perspective, they might be seen as either the beginning or the end of the related collapse. First of all, 86 percent of F Pa are shorter than 10 frames, thus they will be removed by the temporal filtering algorithm.

All CNN output predictions were evaluated using Tpred = 0.5, which was utilised to produce the theoretical accuracy metric p, the alarm sensitivity sea, and the alarm precision pa. In all databases, pa is substantially lower than p due to absence of the temporal element.

4. CONCLUSION

As part of our study, we looked at fall detection systems from a nursing home perspective. Because of CNN's vision, a new training strategy based on a realistic alert rate statistic and a decision-making process has been established that fulfils the requirements of medical experts. On average, the examined datasets showed that the proposed solution correctly detected 86.2 percent of all falls, while only 11.6 percent of all false alarms were generated by it. In the majority of instances, false alarms are the result of someone sitting down too forcefully, getting up too quickly after a fall, or bending over to pick something up off the ground.

We plan to employ a spatial filter like semantic background segmentation to enhance our results and to increase the volume and diversity of data we gather in the future. Currently, the device is being updated for clinical testing, which will take place on actors simulating fall scenes. It's also a

good idea to use the Multicam database's inner results analysis to combine data from many cameras, which shows that a fall is always detected by at least one of the cameras.

References

[1] Espinosa, R., Ponce, H., Gutiérrez, S., Martínez-Villaseñor, L., Brieva, J., & Moya-Albor, E. (2019). A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset. Computers in biology and medicine, 115, 103520.

[2] Khraief, C., Benzarti, F., & Amiri, H. (2019). Convolutional neural network based on dynamic motion and shape variations for elderly fall detection. Int. J. Mach. Learn. Comput, 9, 814-820.

[3] Vishnu, C., Datla, R., Roy, D., Babu, S., & Mohan, C. K. (2021). Human fall detection in surveillance videos using fall motion vector modeling. IEEE Sensors Journal, 21(15), 17162-17170.

[4] Zahan, S., Hassan, G. M., & Mian, A. (2021). Modeling Human Skeleton Joint Dynamics for Fall Detection. In 2021 Digital Image Computing: Techniques and Applications (DICTA) (pp. 01-07). IEEE.

[5] Gan, H., Li, S., Ou, M., Yang, X., Huang, B., Liu, K., & Xue, Y. (2021). Fast and accurate detection of lactating sow nursing behavior with CNN-based optical flow and features. Computers and Electronics in Agriculture, 189, 106384.

[6] De, A., Saha, A., & Kumar, P. (2022). Fall detection approach based on combined displacement of spatial features for intelligent indoor surveillance. Multimedia Tools and Applications, 1-24.

[7] Kong, Y., Huang, J., Huang, S., Wei, Z., & Wang, S. (2019). Learning spatiotemporal representations for human fall detection in surveillance video. Journal of Visual Communication and Image Representation, 59, 215-230.

[8] Grant, J. M. (2018). Analysis of Crowd Behavior Based on Optical Flow: Detection, Classification, and Clustering. University of Notre Dame.

[9] Oudah, M., Al-Naji, A., & Chahl, J. (2021, June). Computer Vision for Elderly Care Based on Deep Learning CNN and SVM. In IOP Conference Series: Materials Science and Engineering (Vol. 1105, No. 1, p. 012070). IOP Publishing.