Automated Fall Detection on Smart Factory based on Deep Learning Approach

Syifa Maliah Rachmawati, Danielle Jaye S. Agron, Dong-Seong Kim, and Jae-Min Lee,
Networked Systems Lab., Department of IT Convergence Engineering,
Kumoh National Institute of Technology, South Korea.
(syifamr, danielleagron, dskim, ljmpaul)@kumoh.ac.kr

***Abstract*—The emergence of the smart environment and the Internet of Things paradigms with the increasing number of cameras in daily life, forms an optimal context for vision-based systems. This paper proposes a model to detect human falling by using deep learning algorithm for vision-based system. To improve fall-detection accuracy, a deep learning technique such as Convolutional Neural Network (CNN) combined with data augmentation and dropout layer to avoid over-fitting is proposed. It compared with existing convolutional-based architecture such as AlexNet, SqueezeNet, GoogleNet, and ResNet-18. The performance of the proposed algorithm is verified by using UR Fall Detection data set. The simulation result showed that the proposed algorithm achieves an accuracy 96.88% with validation loss 0.0638.**

***Index Terms*—Convolutional Neural Network (CNN), Fall-detection**

I. Introduction

Fall incidents in the manufacturing industry are frequent. Some of fall cases may cause serious injury such as fracture and slipped disk. To avoid such accidents, most factories install video surveillance systems to increase the safety on factory and it has become most important to design an intelligent surveillance systems, especially vision-based system which can automatically detect the accident [1]. The designed system must have high accuracy and be able to detect rapidly because fall cases not only because of a fall in general but can caused by a heart attack and that is an emergency condition.

Human fall-detection can be done using sensor-based [5] and vision-based. Several studies have been conducted to apply an action recognition for vision-based system that recognizes human activity based on Convolutional Neural Network (CNN) method [2], [3]. Vision-based systems that use video cameras do not limit normal activities as they do not use wearable devices and can avoid electromagnetic interference.

Deep networks can automatically extract features for detection after learning and analyzing a mass of data. CNN is the most commonly applied deep supervised models for image classification. CNN architectures are great for finding patterns and shapes within given images. This study propose a vision-based fall detection method based on two dimensional CNN with image augmentation.

The rest of the paper is arranged as follows. In section II, we describe our fall detection method using 3 layers of Conv2D with data augmentation. Section III shows the performance evaluation of the proposed algorithm compared with existing algorithm. The conclusion is drawn in Section IV.

II. Proposed System

In this paper, we propose three stacks of convolutional layers with data pre-processing such as image augmentation. The CNN takes an image as input, then processes it and classifies the processed image. Image data augmentation is used to artificially expand the size of a training dataset in order to improve the performance and ability of the model to generalize by creating modified versions of images in the dataset.

In general, CNN architecture consists of layers of convolution and pooling, then followed by at least one fully connected layer and softmax function for training and testing purposes. The convolution is operated by multiple filters and the pooling operation is aimed to reduce the dimensions of features mapped from the convolution operation.

The system design of this work can be described as follows:

1) The input images set as a batch of 32 images of shape 180x180x3.
2) Normalize the image pixel values by dividing by 255 and added data augmentation. In this step, the image is treated in the form of being flipped horizontally, rotates dan zoom randomly. This process is included in the data pre-processing layer.
3) Apply three stacks of Conv2D layers with a max pool layer in each of them that is activated by Rectified Linear Unit (ReLU) activation function.
4) Added dropout layer for regularization. Dropping out 20% of the output units randomly from the convolution layer.
5) Flatten or unroll the 3D output from dropout layer to 1D, then add a fully connected layer with 128 units that is activated by a ReLU activation function.
6) To complete the model, feed the last output with one more Dense layers to perform classification.
7) Compile, train, and evaluate the model.
8) The output of this system is a prediction of an image belong to 'fall' or 'not fall'.

The architecture of the proposed algorithm is shown in Fig. 1.

III. Performance Evaluation

A. Dataset

UR Fall Detection data set [4] was used in this experiment. The data set includes RGB and depth videos which are

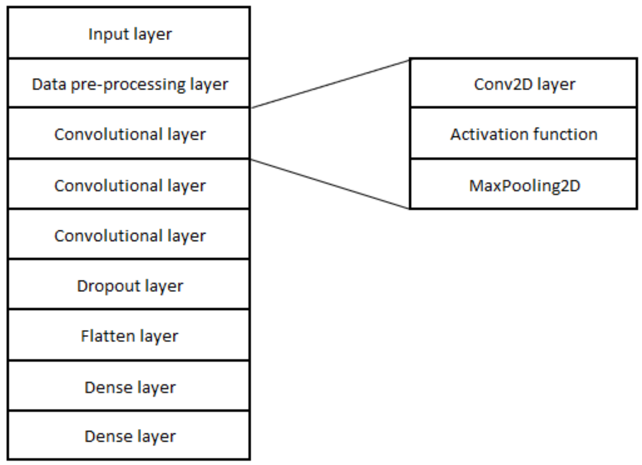

Fig. 1. Architecture of proposed algorithm

recorded by two Microsoft Kinect cameras with a frame rate of 32 fps and a resolution of 640 x 480. This dataset contains 30 falls and 40 activities of daily living sequences. We only used fall events with RGB data of camera 0. There are in total 887 fall frames and 2038 non-fall frames.

B. Experimental Result

The simulation implemented in Python based on the tensorflow framework with Keras Library. The number of stack of Conv2D is varied from 1, 2, and 3 stacks with varying number of filter, then compared and checked which number of filter will give the accurate result. All of the method were obtained from the training of 15 epochs, adam optimizer, and Sparse Categorical Cross-entropy to computes the cross-entropy loss between the labels and predictions. The validation split is used in developing the model. 80% of the images for training and 20% for validating. We trained our neural network with additional synthetically modified data. It help to increase the amount of relevant data in the dataset.

First, we try to simulate one stack of Conv2D with 128 filters followed by max pool and 2 fully connected layer. Result shows it achieves 94.5% of accuracy and 0.425 of loss. Then, we train again using two and three stacks of the Conv2D respectively. Both of the models achieve same accuracy 95.45% but, by increasing the number of Conv2D layer the loss was decrease. Combination of three stacks of Conv2D and dropout after the last stack of Conv2D with image data augmentation before it feeds into the model was achieve better accuracy, namely 96.88% with loss 0.063 as shown in Table .1. The trained model using the proposed algorithm can classify whether the image include fall activity or not. The learned features from a model contribute differently to the performance improvement of classification.

The last method which is shown in Table .1. is our proposed method. Next, we compare it to existing convolutional-based machine learning algorithm. The results are shown in Table .2. From the table, our proposed method achieve better accuracy and can minimize the loss.

TABLE I
CLASSIFICATION RESULT

Method	Accuracy	Loss
1 blocks Conv2D	94.5%	0.425
2 blocks Conv2D	95.45%	0.14
3 blocks Conv2D	95.45%	0.102
3 blocks Conv2D + Dropout	95.45%	0.248
3 blocks Conv2D + Data Augmentation + Dropout	96.88%	0.063

TABLE II
COMPARISON OF RESULT WITH EXISTING METHOD

Method	Accuracy	Loss
AlexNet	95.06%	0.34
SqueezeNet	92.59%	0.21
GoogleNet	95.30%	0.10
ResNet-18	93.83%	0.16
Proposed method	96.88%	0.063

IV. CONCLUSION

This paper presented a 2D RGB images fall-detection based on an image classifier algorithm. The three blocks of convolutional layers combined with data augmentation and dropout was adopted to decide if a sequence of frames contains a person falling in the factory. We experimented with several combinations of Conv2D filters and layers, then compared with existing algorithms. Based on the result, implementing data augmentation and dropout can reduce the over-fitting to get better result. Experimental results shown that the proposed method can achieve accuracy 96.88% and validation loss 0.063. In spite of outperforming other considered models, the algorithm should be optimized prior to implementation and validated in more scenarios.

V. ACKNOWLEDGEMENT

This research work was supported by Priority Research Centers Program through NRF funded by MEST(2018R1A6A1A03024003) and the Grand Information Technology Research Center support program (IITP-2021-2020-0-01612) supervised by the IITP by MSIT, Korea.

REFERENCES

[1] M. Mubashir, L. Shao, and L. Seed, “A survey on fall detection: Principles and approaches,” Neurocomputing, vol. 100, pp. 144-152, Jan. 2013.

[2] X. Li, T. Pang, W. Liu, and T. Wang, “Fall detection for elderly person care using convolutional neural networks,” in Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI), Shanghai, China, Oct. 2017, pp. 1-6.

[3] A. Marcos, G. Azkune, and I. Arganda-Carreras, “Vision-based fall detection with convolutional neural networks,” Wireless Commun. Mobile Comput., vol. 2017, pp. 1-16, Dec. 2017.

[4] Bogdan Kwolek, Michal Kepski,“Human fall detection on embedded platform using depth maps and wireless accelerometer,” Computer Methods and Programs in Biomedicine, Volume 117, Issue 3, 2014, pp. 489-501, ISSN 0169-2607.

[5] Cosmas Ifeanyi Nwakanma, Fabliha Bushra Islam, Mareska Pratiwi Maharani, Jae-Min Lee and Dong-Seong Kim, “Detection and Classification of Human Activity for Emergency Response in Smart Factory Shop Floo,” Applied Sciences, Vol. 11, no. 8, 3662, April 2021