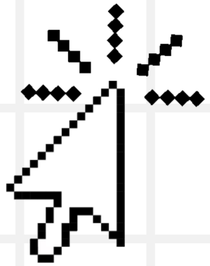
FPT UNIVERSITY

SKELETON-BASED FALL DETECTION USING COMPUTER VISION



Duong Thi Phuong Dung
Pham Le Anh Duc
Can Thi Hoa Mai
Supervisor: Assoc. Prof. Phan Duy Hung

AIP490_G6 | 2023

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

1 INTRODUCTION

MOTIVATION

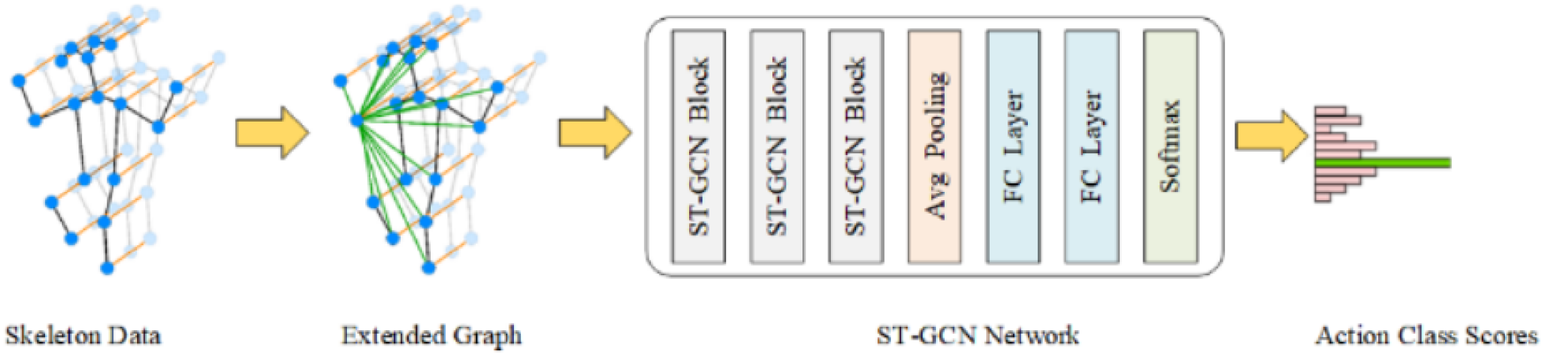
- An estimated 684,000 people die from falls
- Cause many aftereffects for the patient and their family

DETECT FALLS timely is extremely necessary

RELATED WORKS

Fall detection & Human action recognition

- **Spatio-temporal graph convolution networks (ST-GCN):**
  - is presented by Yan et al. in 2018
  - human skeleton's spatial characteristics are extracted using a GCN, and the same joint's continuous time edge is subjected to a time convolution network.
  - original ST-GCN is created by ST-GCN blocks that alternately apply temporal and spatial graph convolutions to a skeleton graph



RELATED WORKS

Fall detection & Human action recognition

➡ Based on ST-GCN, a lot of research in the recently have great improvements:

- **Adaptive graph convolution networks (AGCN)** by Wang et al.:
  - have a learning metric for graph update

- **Multi-scale skeleton adaptive weighted graph convolution networks(MS-AWGCN)** by Xu et al. :
  - propose a new human body representation more comprehensive human motion relationship model
  - apply a learning-weighted strategy that enhances the features while combining

RELATED WORKS

Fall detection & Human action recognition

With respect to the fall detection problem, Oussema Keskes et al. introduced a general **Vision-based fall detection using ST-GCN**:

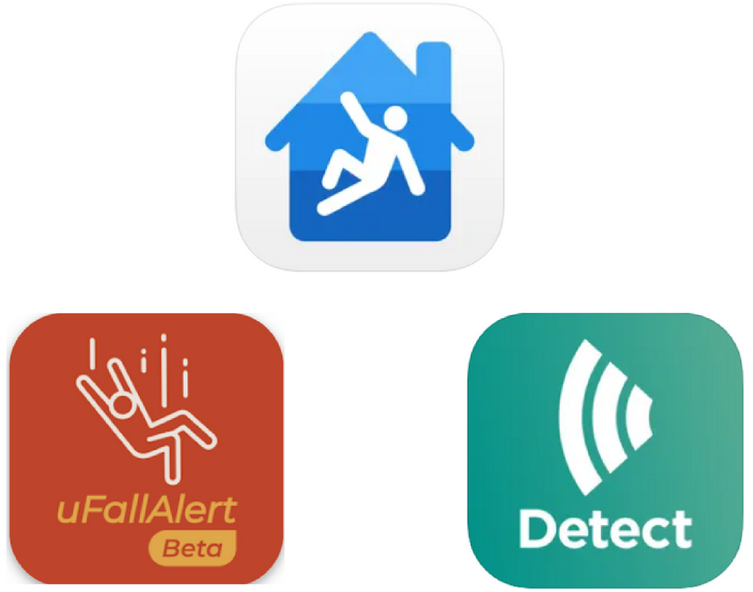
- employed the **transfer learning technique** to aid the model in utilizing features extracted from the task of human action recognition, which is related to fall detection
- some experiments achieved 100% accuracy on fall datasets

RELATED WORKS

Fall detection & Human action recognition

In the last few years, there have been several applications developed in the fall detection field. Here are some remarkable example

Fall detection on the Apple Watch

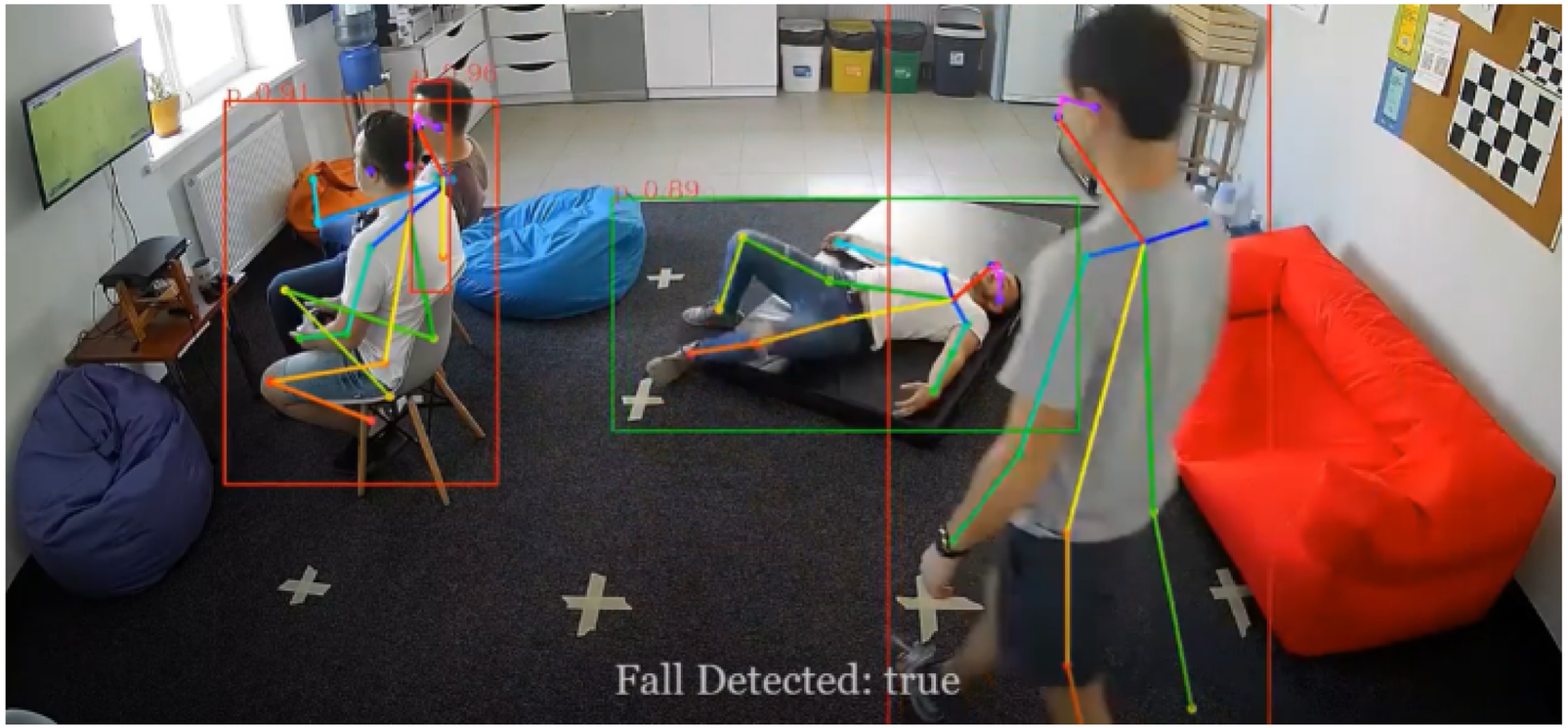


Some applications on smartphone

RELATED WORKS

Fall detection & Human action recognition

- Another example is the video analytics module developed by Abto Software.

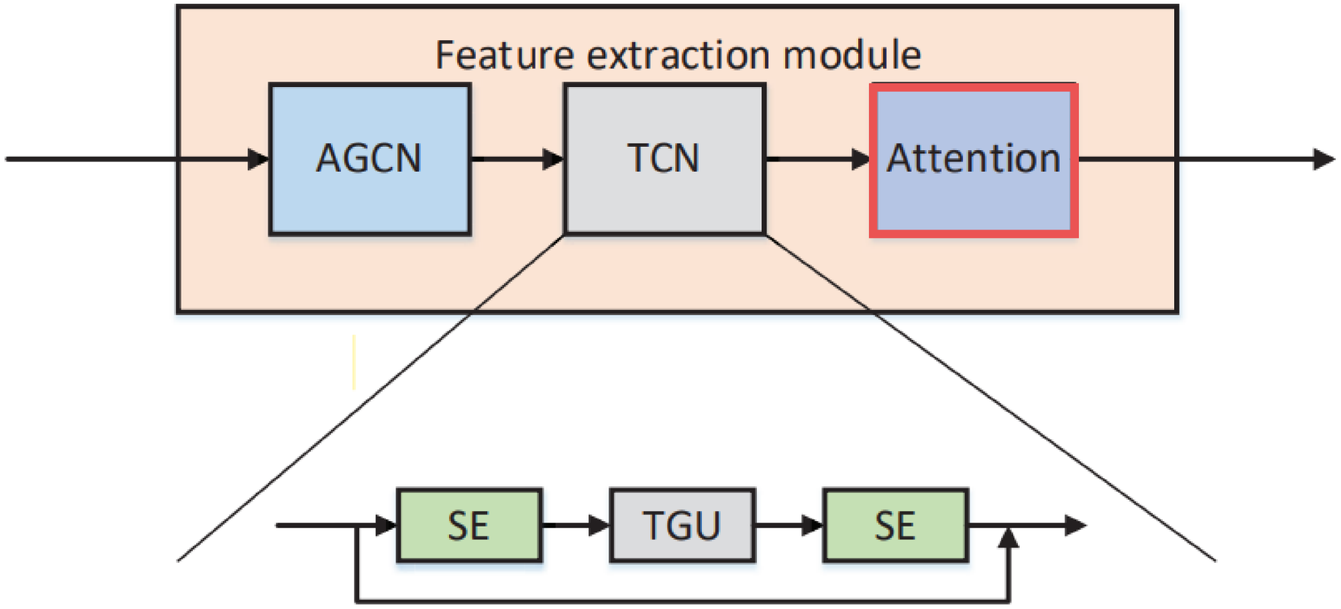


A video analytics solution in real-time was produced by Abto Software

RELATED WORKS

Fall detection & Human action recognition

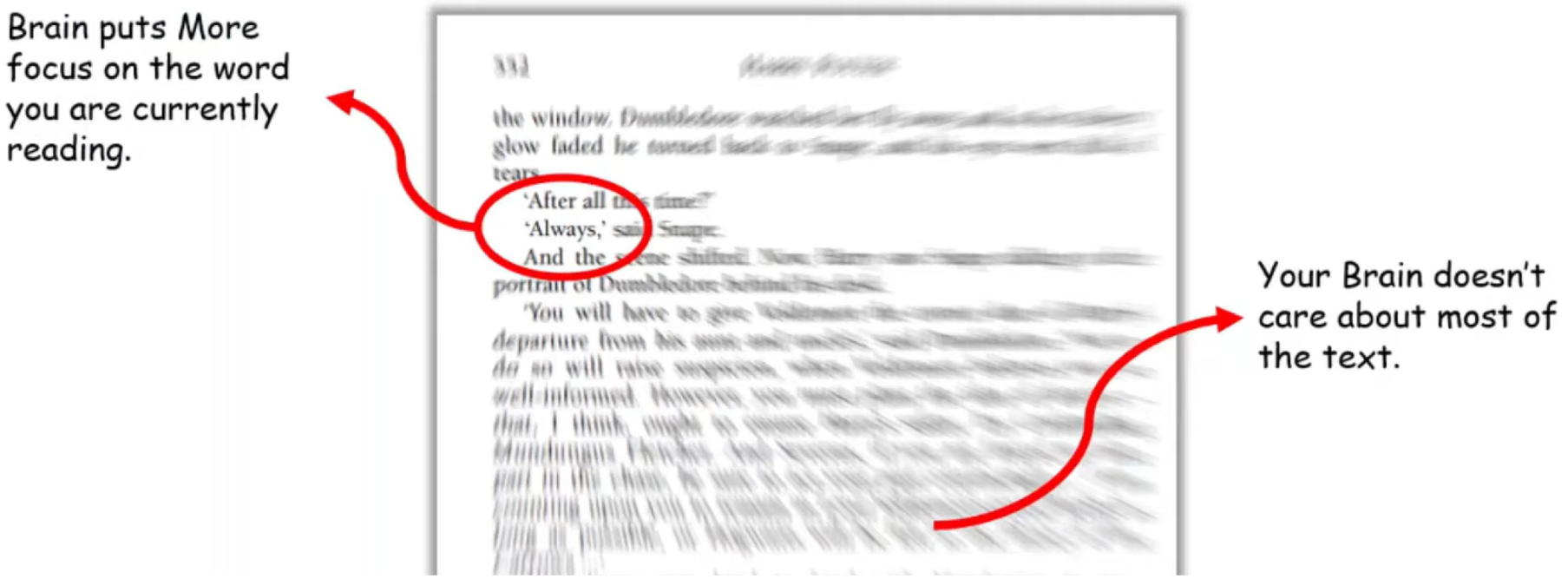
- Skeleton Action Recognition Based on Temporal Gated Unit and Adaptive Graph Convolution by Zhu et. al (2022)



In the action of “falling”, the action is similar in most frames and changes significantly in only a few frames

RELATED WORKS

Attention mechanism



Attention in human visual processing system

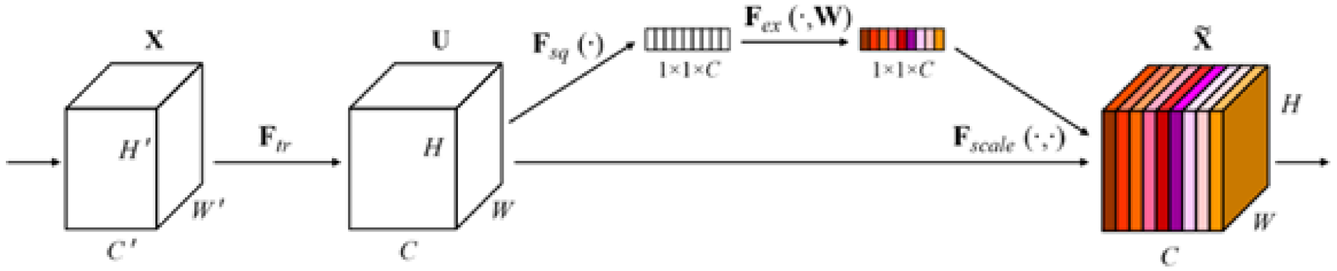
RELATED WORKS

Attention mechanism

→ The basic idea of attention mechanism is to assign different weights for each part of the sequence

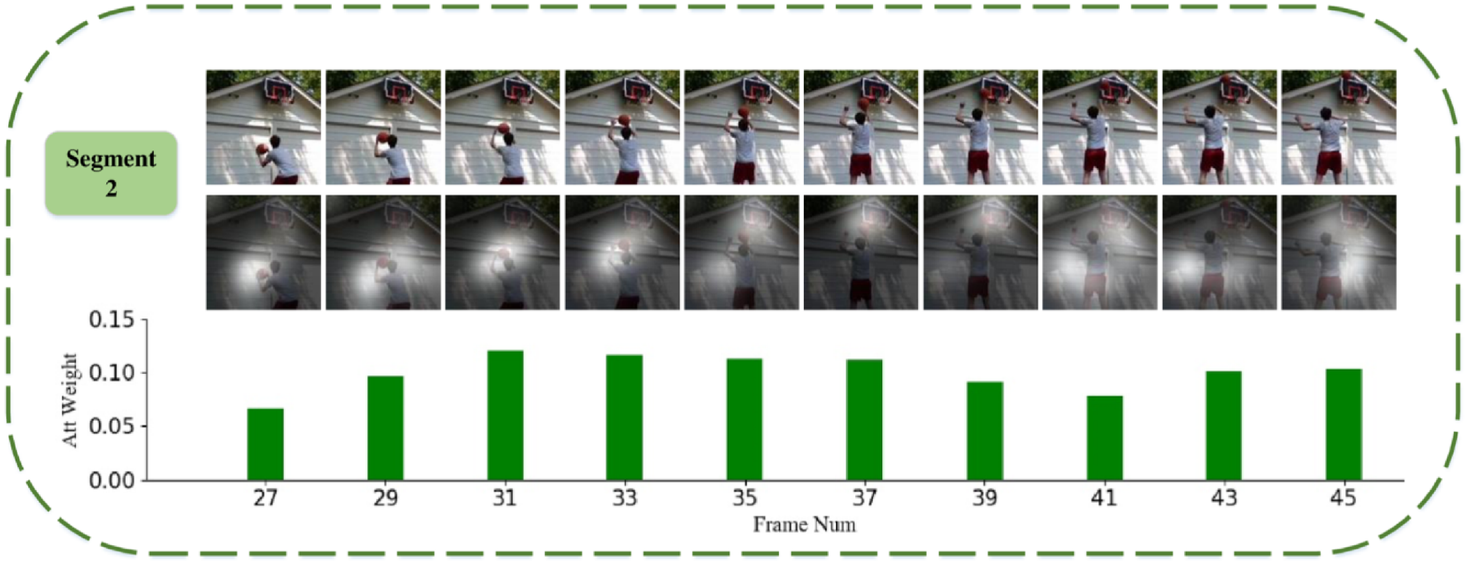
RELATED WORKS

Attention mechanism



Squeeze and Excitation Networks
by Hu et al. (2018)

Channel attention



Spatial & Temporal attention

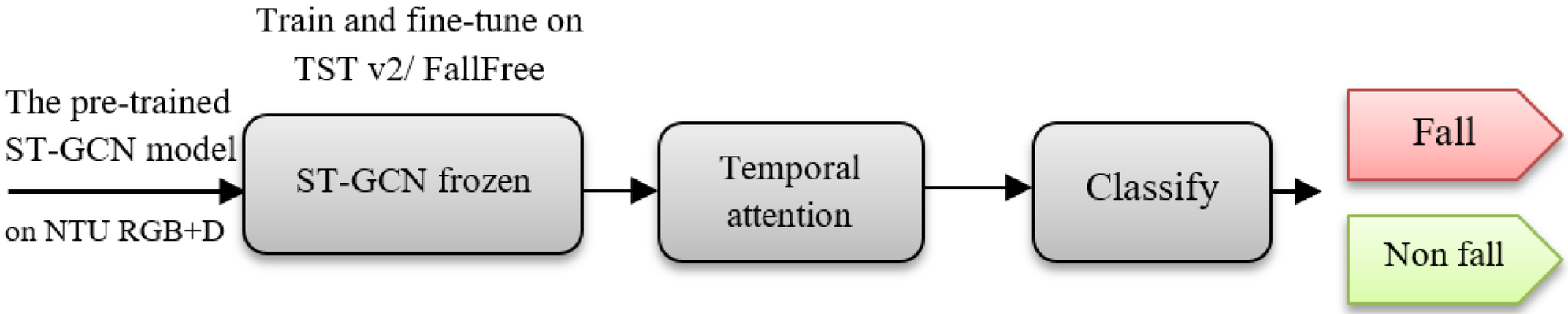
Spatial-Temporal Attention Temporal Segment Network for action recognition in video
by Yang et al. (2022)

CONTRIBUTION

- Combine transfer learning techniques with attention mechanism in problem of fall detection
- Propose a temporal attention module to highlight important frames in the input sequence

2 METHODOLOGY

METHODOLOGY



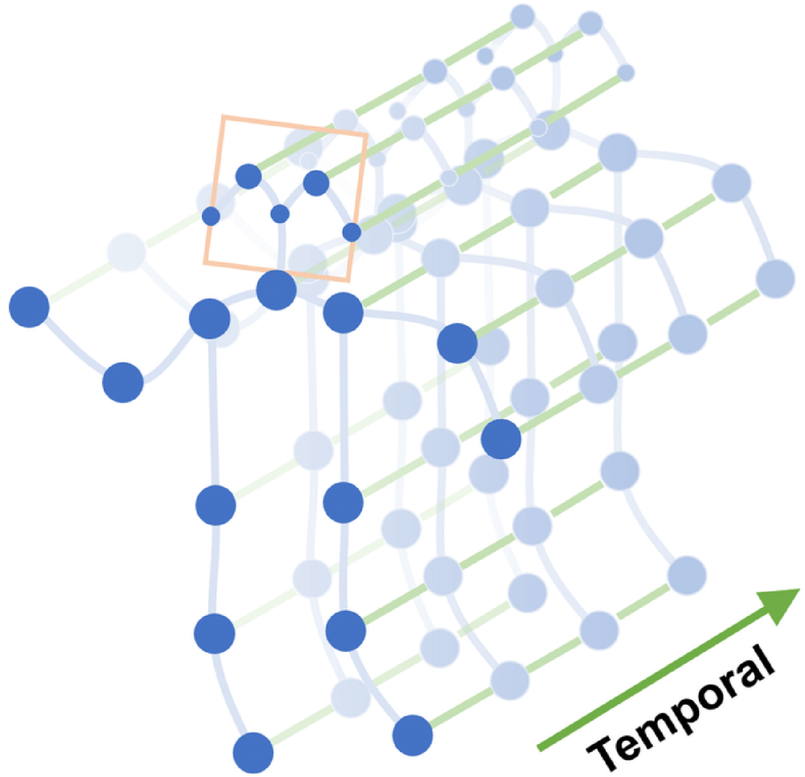
The pipeline of our proposed model

METHODOLOGY

ST-GCN

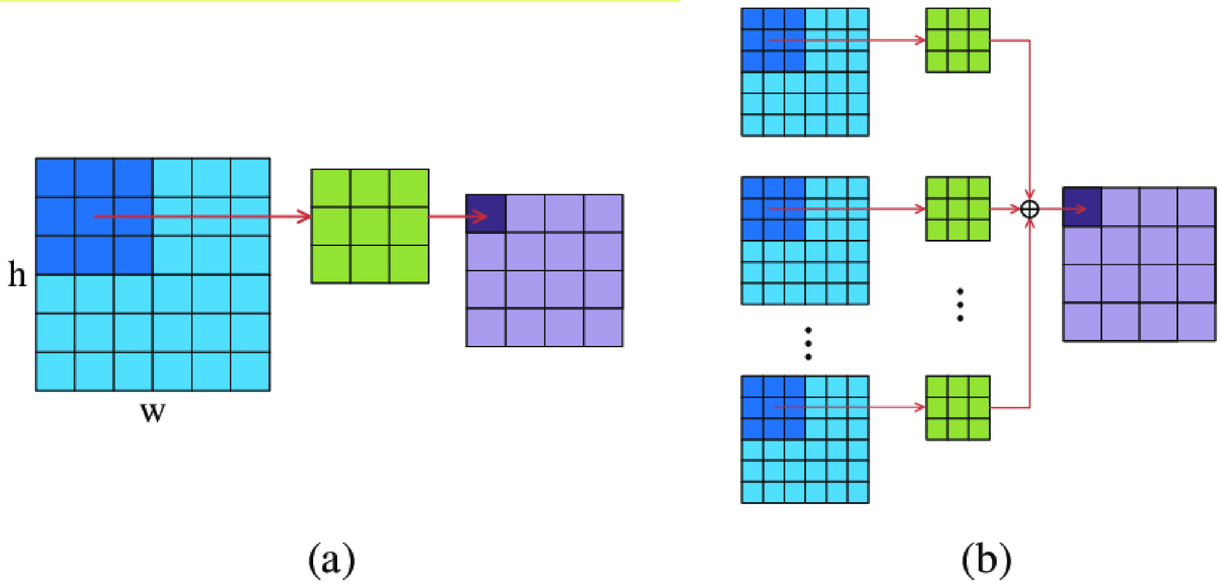
- A deep learning architecture that is specifically designed for processing spatiotemporal data, such as human actions in videos or sensor data from wearable devices.

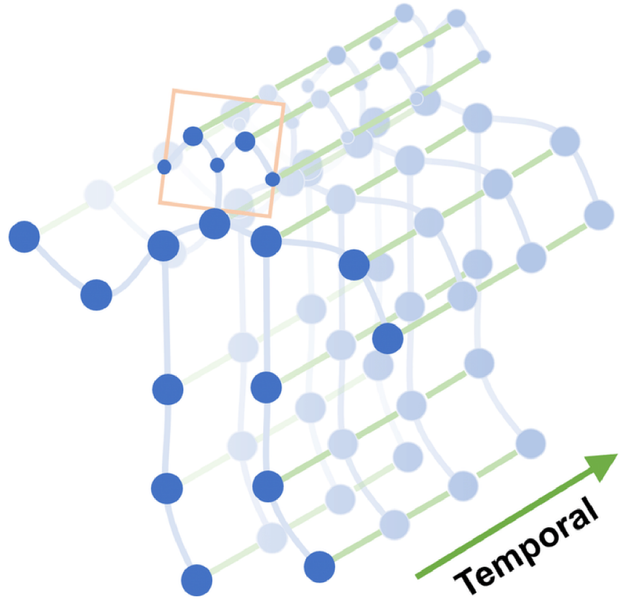
- The input data is represented as a spatiotemporal graph. Each node in the graph corresponds to a specific body joint, and the edges represent the relationships between them.



METHODOLOGY

ST-GCN





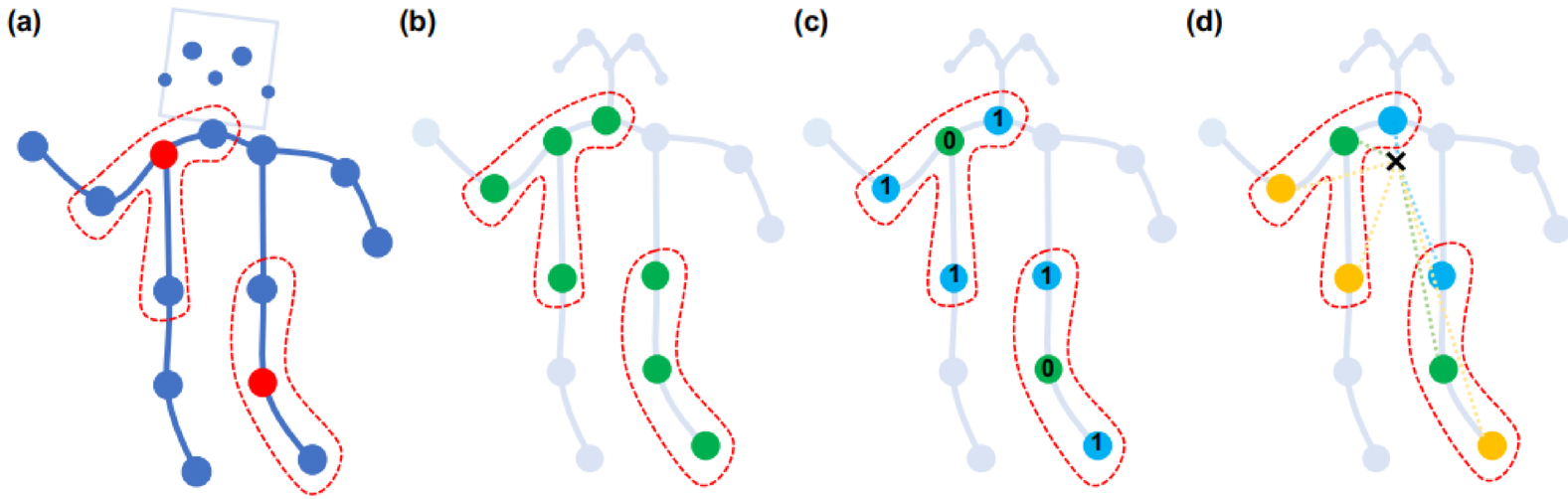
- In 2D convolution, a rigid grid naturally exists around the center location, so we can use the weights of a kernel to map with the fixed neighbors of the current pixel.

→ **Need to define a graph labeling process in the neighbor graph around the root nodes**

METHODOLOGY

ST-GCN

- For each node, we have 3 methods to define neighbor nodes and label the weight coefficient for them: uni-labeling partition, distance partition, and spatial configuration partition.



(a) an illustration of an input skeleton frame; (b) Uni-labeling partition; (c) Distance partitioning; (d) Spatial configuration partitioning

METHODOLOGY

Pre-trained ST-GCN model

- The model have a ST-GCN networks which is a list of 10 layers of ST-GCN units and a fully connected layer for a classifier
- Then froze the first layers of ST-GCN networks and changed the number of classes in the output layers to 2 (for fall and no fall)

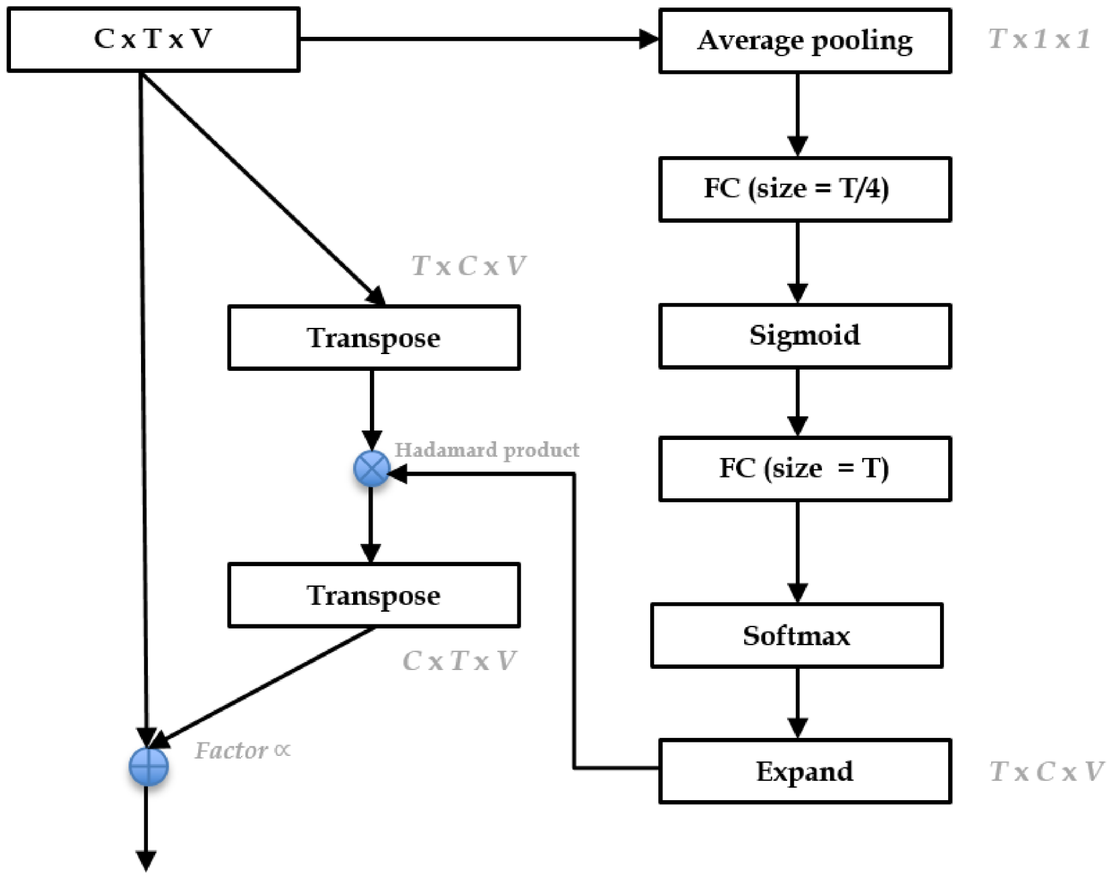
METHODOLOGY

Temporal attention module

(1) Average pooling to extract global features along the temporal dimension

(2) FC layers are to learn different levels of importance for each frame

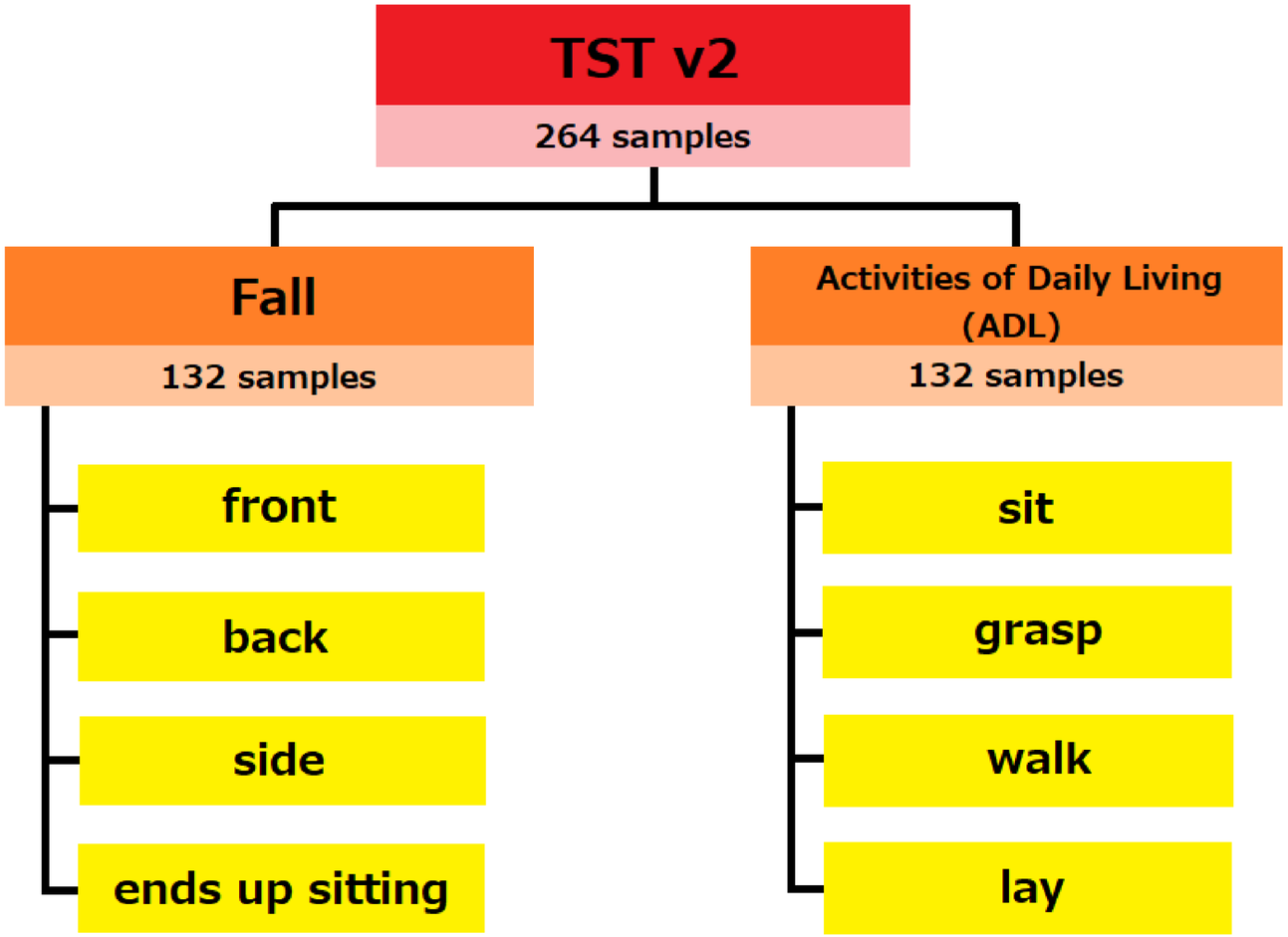
(3) Softmax layer is to generate the distribution of weights



3

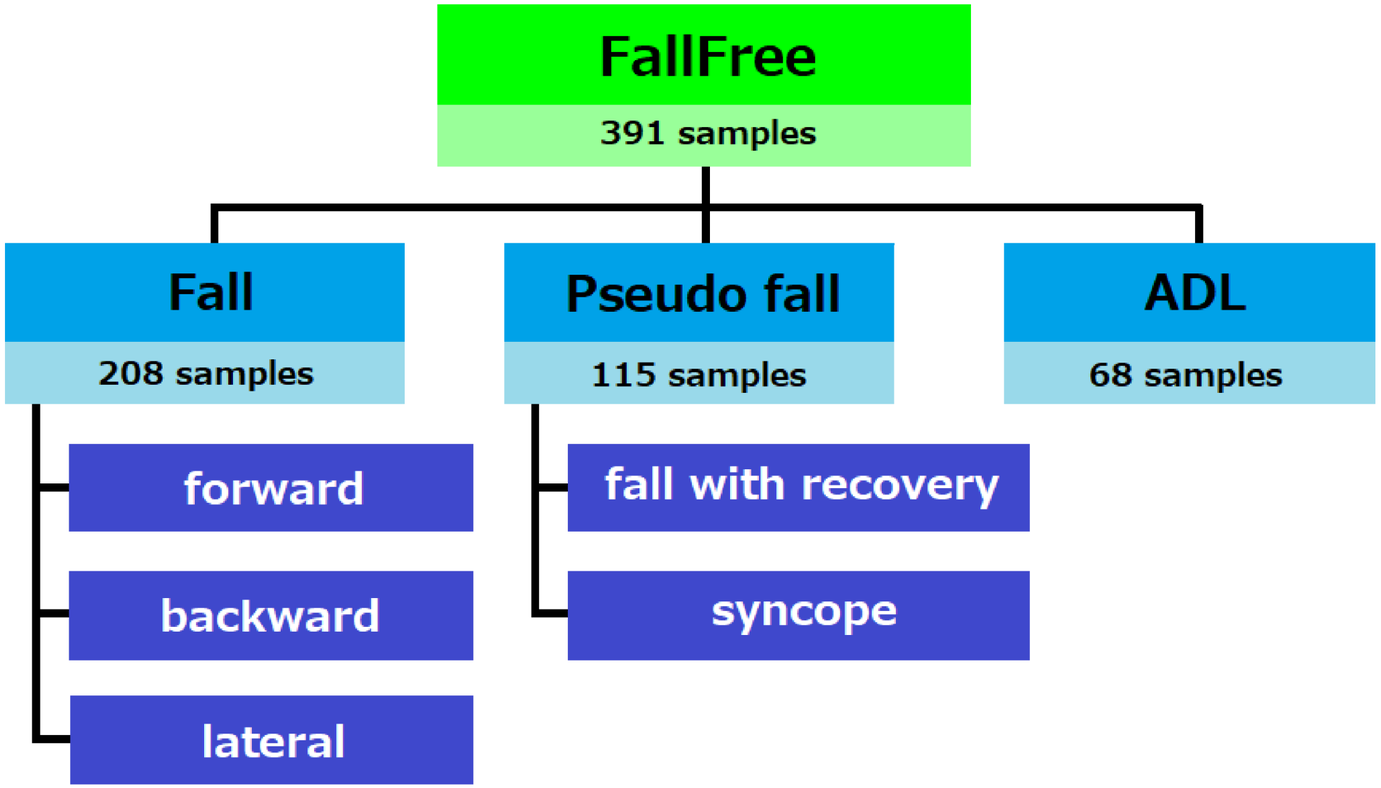
EXPERIMENT

DATASET 1



- The dataset was published in 2015
- It was gathered by using the Microsoft Kinect v2 camera with the inertial measurement unit (IMU)
- 11 actors

DATASET 2



- The dataset has been public since 2017
- It was collected using the Microsoft Kinect v2 camera

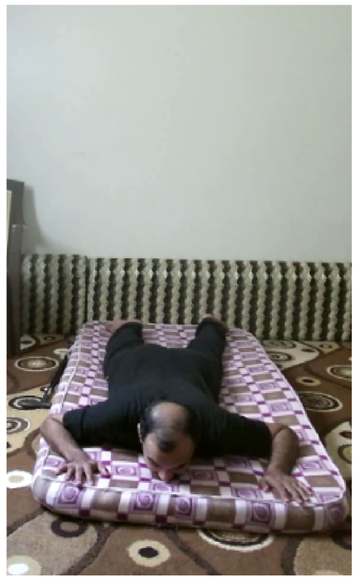
DATASET 2

Fall

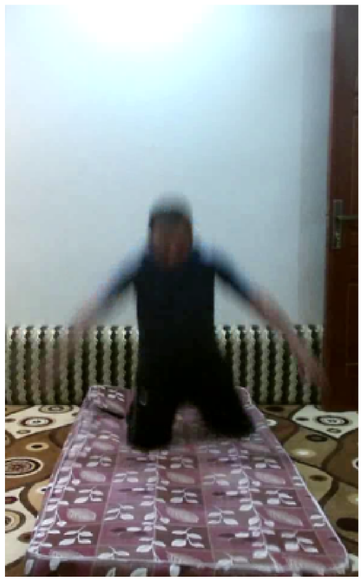
208 samples

- forward
- backward
- lateral

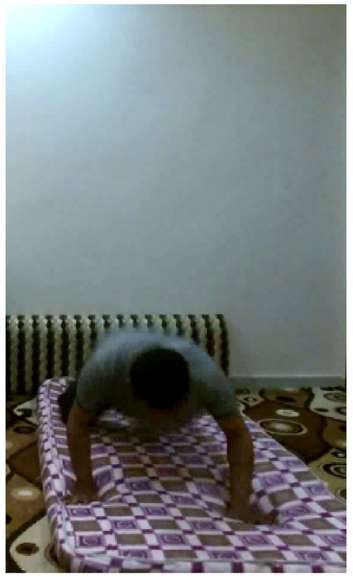
Fall then end up lying



Fall then end up lying



Forward fall
on the knees



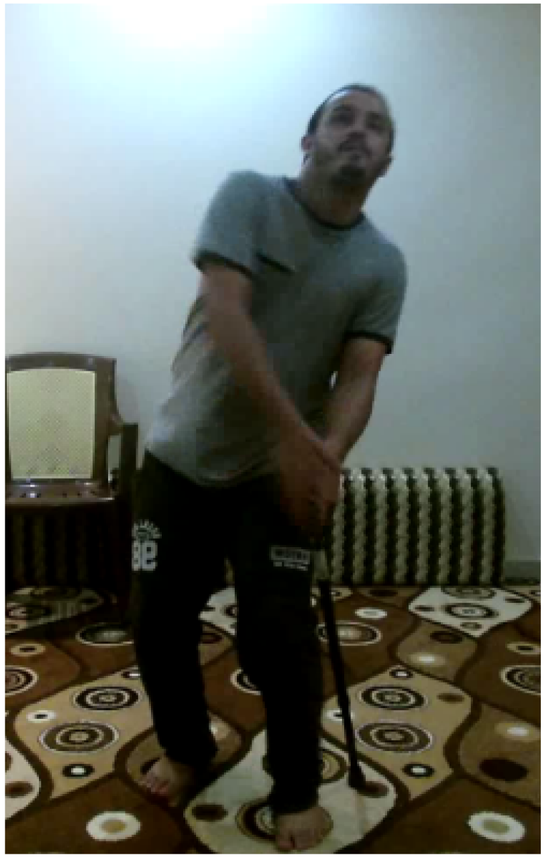
Fall with forward
arms protection

DATASET 2

Pseudo fall

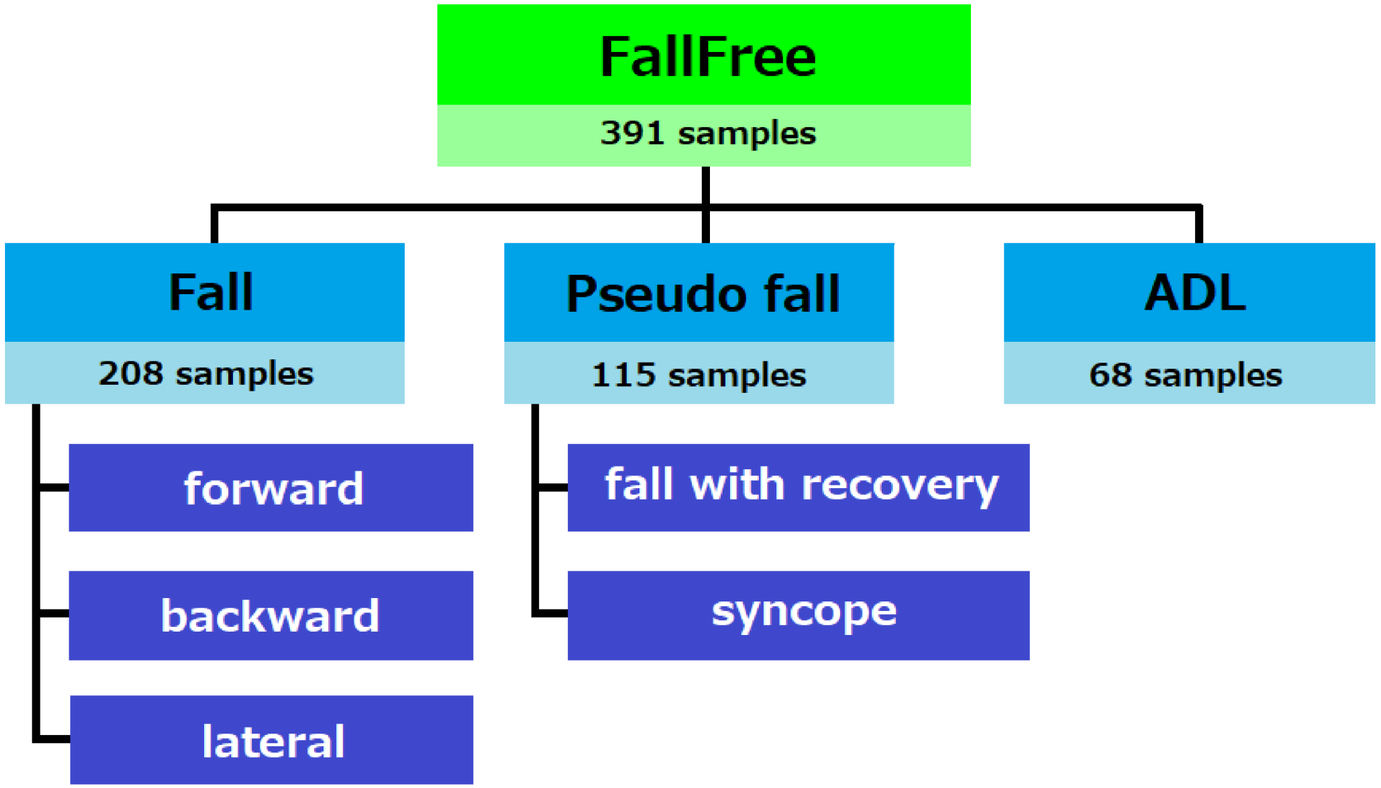
115 samples

- fall with recovery
- syncope



Fall with recovery

DATASET 2



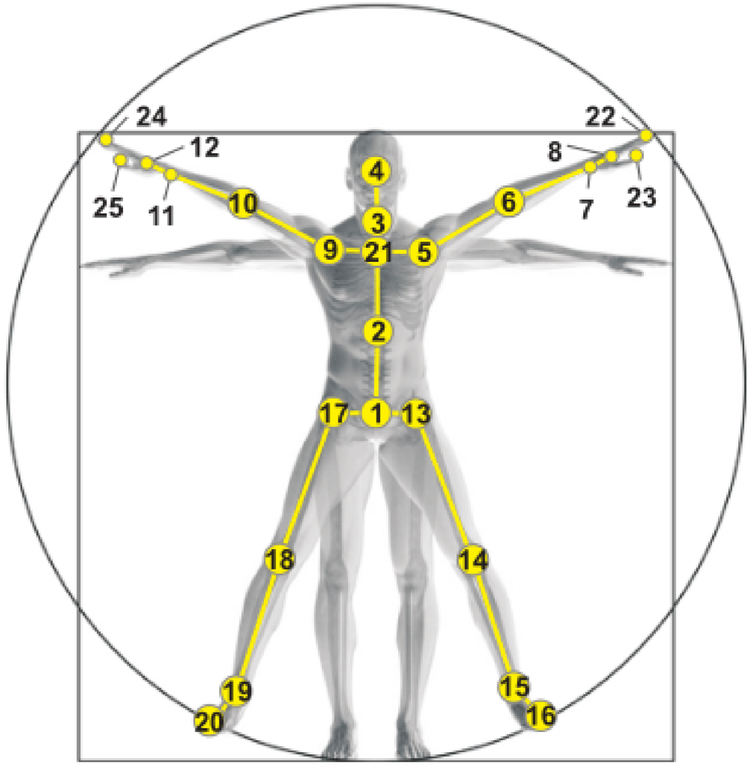
- The dataset has been public since 2017
- It was collected using the Microsoft Kinect v2 camera
- 2 actors

DATASETS

RGB

Depth

Skeleton



25 body joints in the 3D skeleton data

DATASETS

Table 1. Descriptions of the datasets

Dataset	# Actions	# Subjects	# Fall samples	# Samples	Publish year
NTU RGB+D	60	40	276	56,880	2016
TST v2	5	11	132	264	2015
FallFree	10	2	208	391	2017

EXPRIMENT SETTINGS

Environment

- **Pytorch & Cuda 12**
- **Python 3.9**
- **Anaconda 3**
- **GPU NVIDIA GTX 1650**

Parameter	Value
base lr	0.1
batch size	8
num epoch	50
optimizer	SGD
weight decay	0.0001
factor α	0.5

RESULT

Table 2. Results of our proposed method on two fall datasets

Dataset	Accuracy	Sensitivity	Specificity	FPR	F1-score	ROC AUC
TST v2	89.58%	97.22%	85%	12.5%	87.5%	91.11%
FallFree	100%	100%	100%	0%	100%	100%

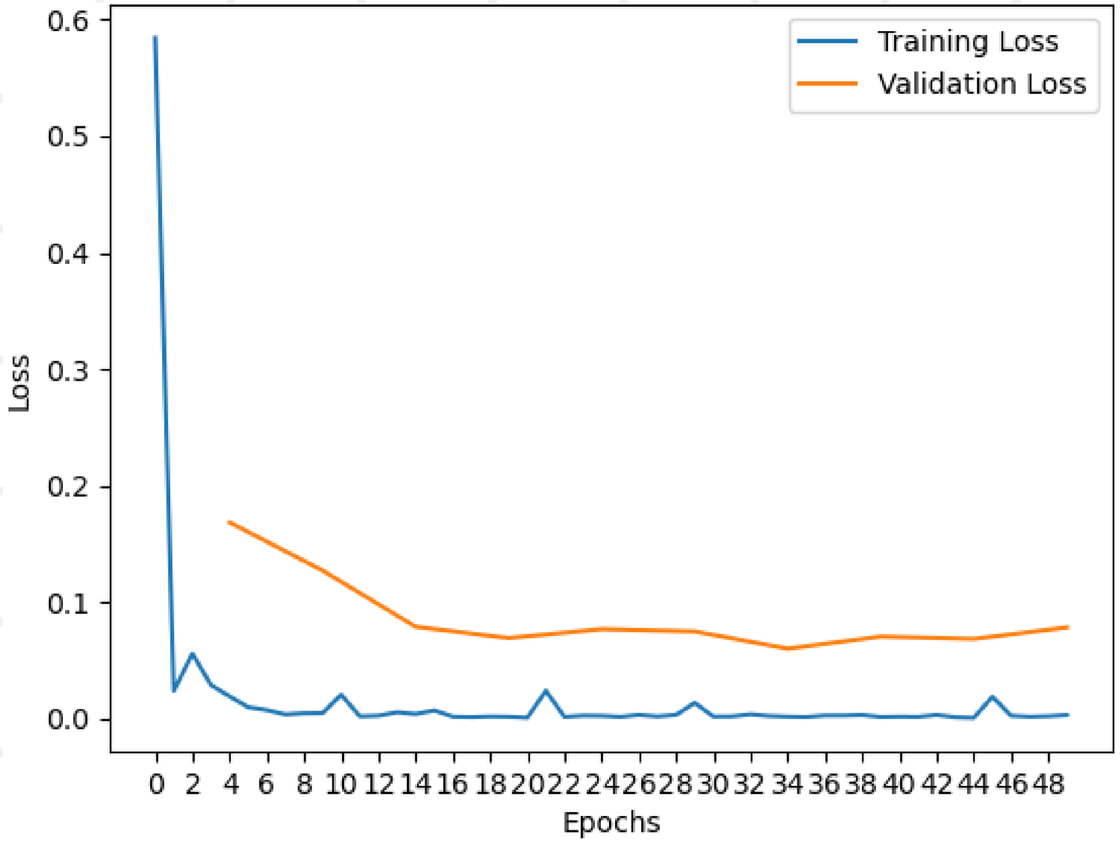
Table 3. Comparison of our method and others on the TST v2 dataset

Method	Algorithms	Evaluation method	Accuracy
Min et al., 2018	SVM	2/3 of data for training, 1/3 of data for testing	92.05%
Yao et al., 2019	SVM	70% of data for training, 30% of data for testing	93.56%
Seredin et al., 2019	SVM and CUSUM	Leave-one-person-out	91.7%
Keskes et al., 2021	Pre-trained ST-GCN	Cross subject	100%
Keskes et al., 2021*	Pre-trained ST-GCN	Cross subject	86.46%
Our method	**Pre-trained ST-GCN and attention mechanism**	**Cross subject**	**89.58%**

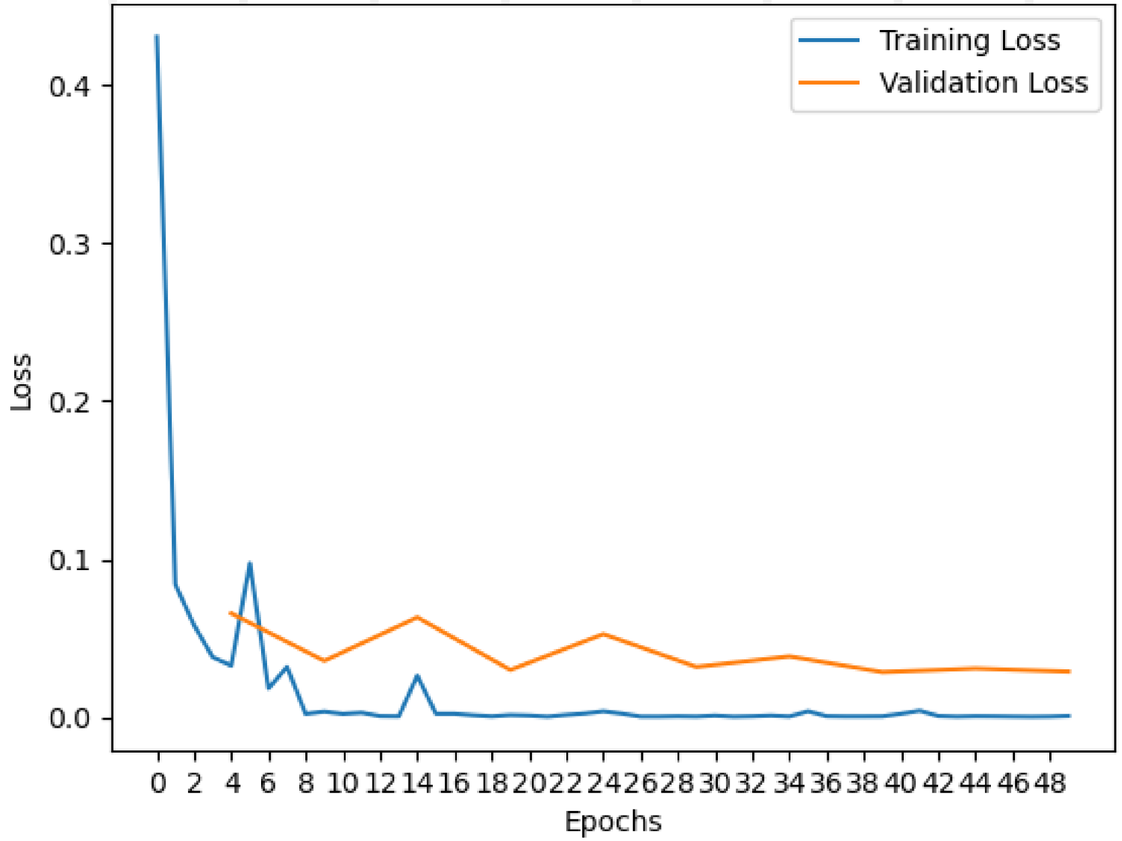
* The result we obtained after implementing the method of Keskes et al.

Table 4. Comparison of our method and others on the FallFree dataset

Method	Algorithms	Evaluation method	Accuracy
Alzahrani et al., 2019	NCA for feature selection and RF classifier	70% of data for training, 30% of data for testing	99.5%
Keskes et al., 2021	Pre-trained ST-GCN	Cross subject	97.33%
Our method	**Pre-trained ST-GCN and attention mechanism**	**Cross subject**	**100%**



Without attention



With attention

4

CONCLUSION & FUTURE WORKS

CONCLUSION

- This thesis proposes a model for fall detection by combining the transfer learning technique and the temporal attention mechanism.

- Our method resulted in improved accuracy of 3.12% in the TST and 2.67% in the FallFree dataset.

FUTURE WORKS

- Improve the results on TST v2 dataset
- Combine temporal attention with other attention mechanisms
- Build a real-time fall detection system to detect falls timely and send urgent warnings to the patient's family.

REFERENCE

1. Yan, S., Xiong, Y., & Lin, D.: Spatial Temporal Graph Convolutional Networks for skeleton-based action recognition, arXiv:1801.07455 (2018)
2. Xu, W., Wu, M., Zhu, J., & Zhao, M.: Multi-scale skeleton adaptive weighted GCN for skeleton-based human action recognition in IoT. Applied Soft Computing, 104(107236), 107236 (2021)
3. Wang, Q., Zhang, K., & Asghar, M. A.: Skeleton-based ST-GCN for human action recognition with extended skeleton graph and partitioning strategy. IEEE Access: Practical Innovations, Open Solutions, 10, 41403–41410 (2022)
4. Verger, R.: The Apple Watch learned to detect falls using data from real human mishaps, Popular Science, https://www.popsci.com/apple-watch-fall-detection/ (Accessed 01/03/2023)
5. Keskes, O., & Noumeir, R.: Vision-based fall detection using ST-GCN. IEEE Access: Practical Innovations, Open Solutions, 9, 28224–28236 (2021)
6. Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E.: Squeeze-and-Excitation Networks, arXiv:1709.01507 (2017)
7. Yang, G., Yang, Y., Lu, Z., Yang, J., Liu, D.: STA-TSN: Spatial-Temporal Attention Temporal Segment Network for action recognition in video. PLOS ONE 17(3): e0265115 (2022)
8. Zhu, Q., Deng, H., & Wang, K.: Skeleton action recognition based on temporal gated unit and adaptive graph convolution. Electronics, 11(18), 2973 (2022)
9. Min, W., Yao, L., Lin, Z., & Liu, L.: Support vector machine approach to fall recognition based on simplified expression of human skeleton action and fast detection of start key frame using torso angle. IET Computer Vision, 12(8), 1133–1140 (2018)
10. Yao, L. Y., Yang, W., & Huang, W.: ‘An improved feature-based method for fall detection. Gazette, 26(5), 1363–1368 (2019)
11. Seredin, O. S., Kopylov, A. V., Huang, S.-C., & Rodionov, D. S.: A skeleton features-based fall detection using Microsoft Kinect v2 with one class-classifier outlier removal. ISPRS - International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences, XLII-2/W12, 189–195 (2019)
12. Alzahrani, M. S., Jarraya, S. K., Ben-Abdallah, H., & Ali, M. S.: Comprehensive evaluation of skeleton features-based fall detection from Microsoft Kinect v2. Signal, Image and Video Processing, 13(7), 1431–1439 (2019)

5 DEMO

THANK YOU