The Impact of Synthetic Data on Fall Detection Application

Dr. Anne H. H. Ngu, Minakshi Debnath

Department of Computer Science, Texas State University

Introduction

- The accurate recognition of the dynamic of fall using deep learning requires a lot of data.
- Three different methods for creating realistic synthetic fall data utilizing generative AI with diffusion, fall data extraction from 2D video recordings, and traditional data augmentation techniques are explored.

Methodology

Datasets

- SmartFallMM's smartwatch data collected in our lab and the UniMib dataset[2]
- Real-life video recordings of older adults falling [1]

Fall detection learning model

- Basic LSTM deep learning model with two layers, 128 neurons and one normalization layer

Data Preprocessing, Training and Evaluation

- Input data is preprocessed via window segmentation (size 128, step size 10)
- Leave-One-Out and 5-fold cross validation is utilized.
- Standard metrics such as Precision, Recall and F1-score used for evaluation.

Synthetic Data Generation

Basic Data Augmentation

- Jittering introduces random Gaussian noise to time series data, enhancing the presence of noisy samples without significant alterations.
- Magnitude warping modifies the magnitude of each sample in a time series dataset using a cubic spline curve, allowing controlled adjustments of signal amplitudes.
- Rotation augmentation simulates variations in sensor placement in wearable sensor data, diversifying data without changing the underlying labels.

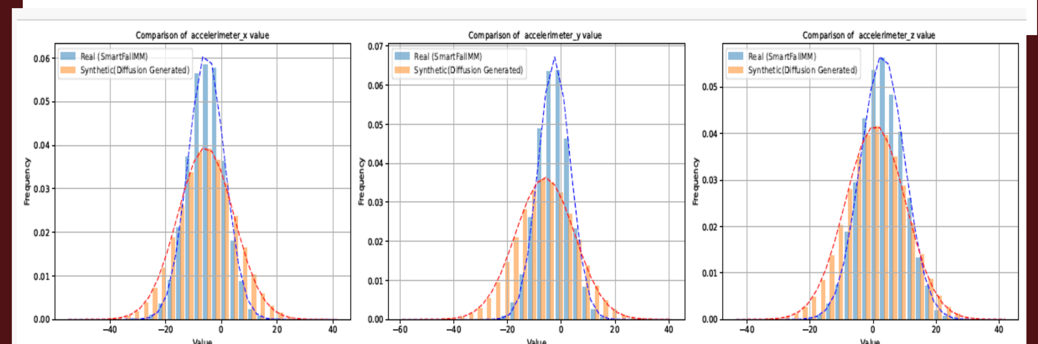

Figure 3: Comparison of SmartfallMM real fall data and synthetic fall data using diffusion method

Methodology(Contd.)

Generative AI Diffusion Model

- Leveraging Denoising Diffusion Probabilistic Models (DDPMs) [3],with a U-Net generative model to synthesize high-quality time-series data, focusing on accelerometer data for fall detection.
- The model utilizes 1D convolutional layers to map the time-series input to U-Net image.
- RMSNorm is used to stabilize training dynamics and ResNet blocks and Linear Attention units refine features and preserve temporal information, optimizing synthetic time series data generation.

Video Extraction via Pose Estimation

- Utilizing 3D pose estimation from [4], we extract fall data from 34 public videos of two long-term care facilities.
- Through cropping and adjustments, we isolate falling individuals in the videos to ensure relevance and reduce extraction time.
- Each video yields 17 joint positions, with accelerometer data obtained from specific joints based on the target dataset.
- The final acceleration data is computed from 3D keypoints

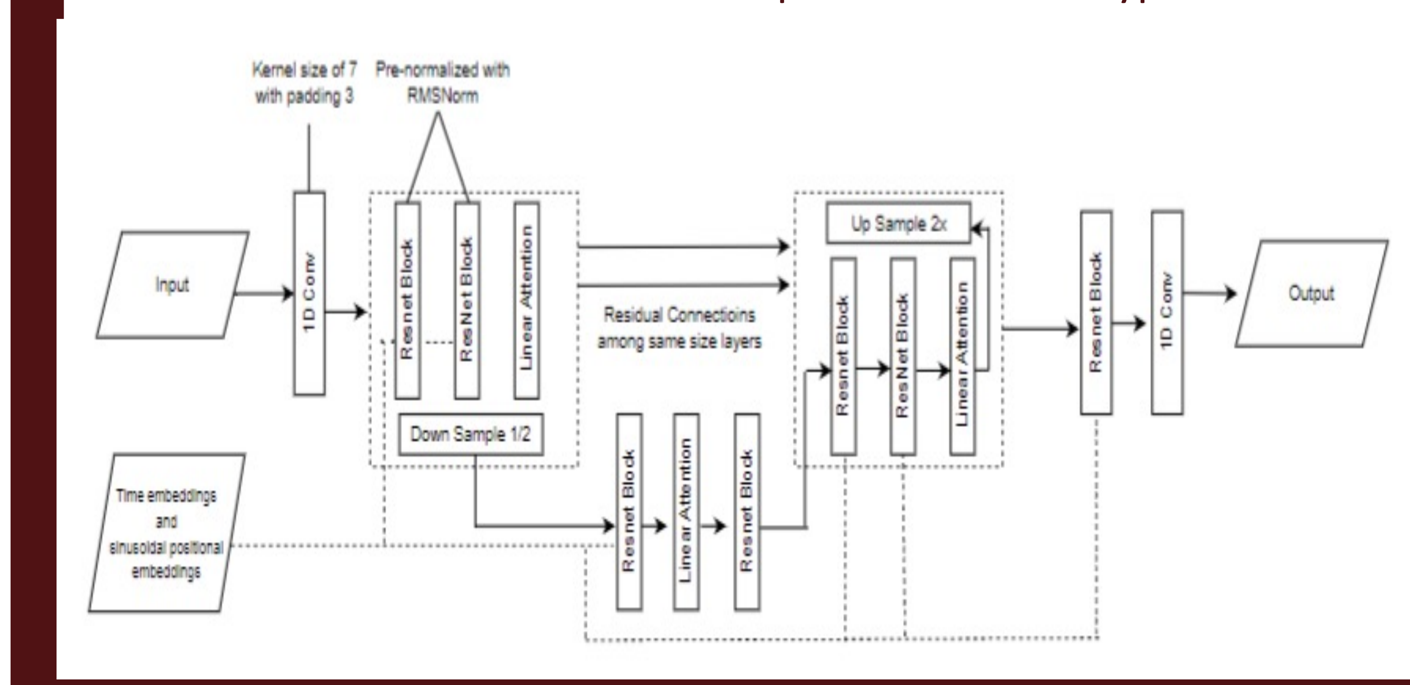

Figure 1: Schematic of the U-Net Architecture Adapted for Time-Series Data

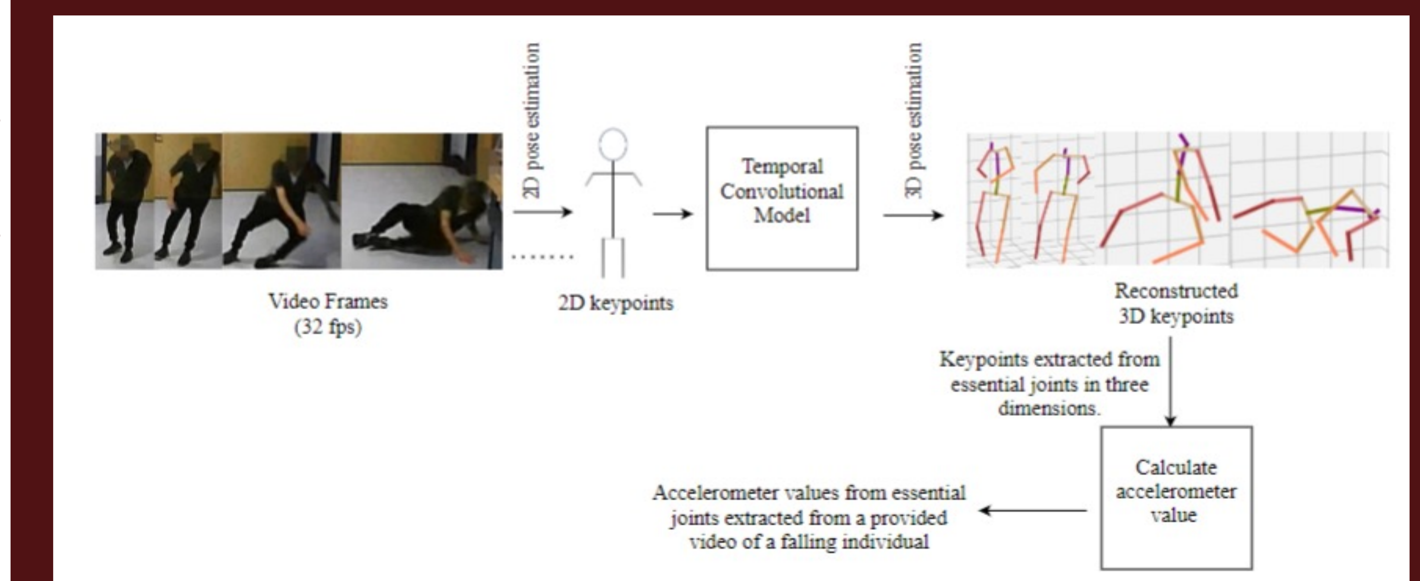

Figure 2: Accelerometer data extraction process from video

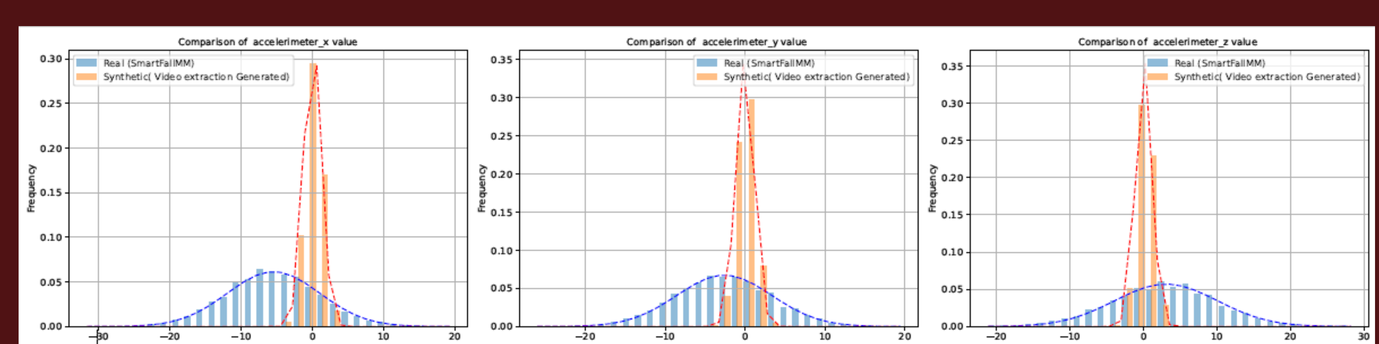

Figure 4: Comparison of SmartfallMM real fall data and synthetic fall data using video extraction method.

Results

- Tables 1 and 2 present a comparative analysis of fall detection performance on two datasets with and without synthetic data.
- The results demonstartes the critical role of meaningful data enrichment, as basic data augmentation techniques exhibit limited impact or even slight degradation in model performance compared to the significant enhancements observed with the diffusion method.

Results of Real and Synthetic data analysis

- Figures 3-6 depict the data distribution in both real and synthetic datasets, using mean and standard deviation as inputs.
- The diffusion method shows most similar distributions in both datasets, while video-extracted data is more tightly concentrated within a smaller range, typically between -0.5 and +0.5.

Dataset	Precision	Recall	F1 score	Accuracy
SmartfallMM w/o augmentation	.71	.88	.79	.78
SmartfallMM w diffusion	.79	.93	.85	.82
SmartfallMM w video extraction	.77	.97	.86	.81
SmartfallMM w jittering	.70	.79	.74	.76
SmartfallMM w magnitude warping	.71	.80	.85	.76
SmartfallMM w rotation	.69	.74	.71	.73

Dataset	Precision	Recall	F1 score	Accuracy
UniMiB w/o augmentation	.71	.88	.79	.78
UniMiB w diffusion	.79	.93	.85	**.82**
UniMiB w video extraction	.77	.97	.86	**.81**
UniMiB w jittering	.70	.79	.74	.76
UniMiB w magnitude warping	.71	.80	.85	.76
UniMiB w rotation	.69	.74	.71	.73

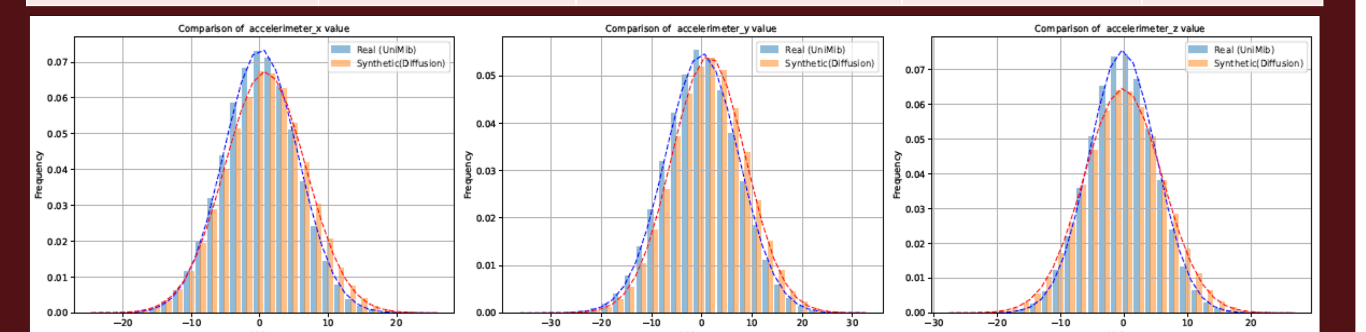

Figure 5: Comparison of UniMiB real fall data and synthetic fall data using diffusion method

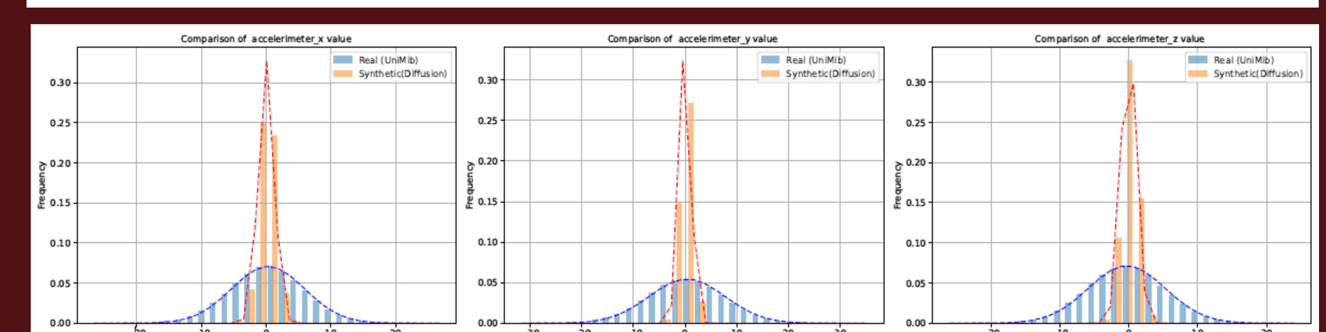

Figure 6: Comparison of UniMiB real fall data and synthetic fall data using video extraction method.

Results(Contd.)

Real time Fall detection application Result

- Initial real-time fall detection using basic LSTM achieved an F1 score of .63 with SmartFallMM data.
- Integration of diffusion model-generated watch data significantly improved the model's efficacy for real-time applications, **raising the F1 score to .85, an increase of nearly 25%.**
- incorporating video-extracted data also enhanced the performance, with the F1 score reaching 77%, nearly 15% higher.

Discussion

- The diffusion method generated data demonstrated an increase of 10-12% enhancement in the offline LSTM fall detection's F1 score compared to the baseline.
- The real-time test of the LSTM fall detection model with diffusion generated data resulted in almost 25-30% increase in accuracy.
- Video extraction data also showcased improved performance, but a bit less.
- Visual analysis of the generated data indicated a closer distribution resemblance between synthetic and real data with the diffusion method.
- **Generative AI Diffusion model** is a promising technique for solving the data scarcity problem in fall detection
- Future work includes ablation study on the optimal amount of generated data that will achieve the best result.

Acknowledgement

We thank the National Science Foundation for funding the research under the Smart and Connected Health Program (NSF-SCH-21223749) at Texas State University.

References

[1] Stephen Robinovitch. Falls experienced by older adult residents in long-term care homes, 2018.

[2]Daniela Micucci, Marco Mobilio, and Paolo Napoletano. Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. Applied Sciences, 7(10), 2017.

[3] Xiaomin Li. Mitigating Data Shortage in Biomedical Signal Analysis: An Investigation into Transfer Learning and Generative Models. Ph.D. dissertation, Texas State University, Texas, June 2023

[4] Junfa Liu, Juan Rojas, Yihui Li, Zhijun Liang, Yisheng Guan, Ning Xi, and Haifei Zhu. A graph attention spatio-temporal convolutional network for 3d human pose estimation in video. In 2021 IEEE International Conference on Robotics and Automation (ICRA), pages 3374–3380. IEEE, 2021.

TEXAS STATE UNIVERSITY®

MEMBER THE TEXAS STATE UNIVERSITY SYSTEM