An Intelligent Agent for Evaluating and Guiding the Post-Stroke Rehabilitation Exercises

Presented by
Md Fokhrul Islam (Exam Roll : 1206)
Swakshar Deb (Exam Roll : 1219)

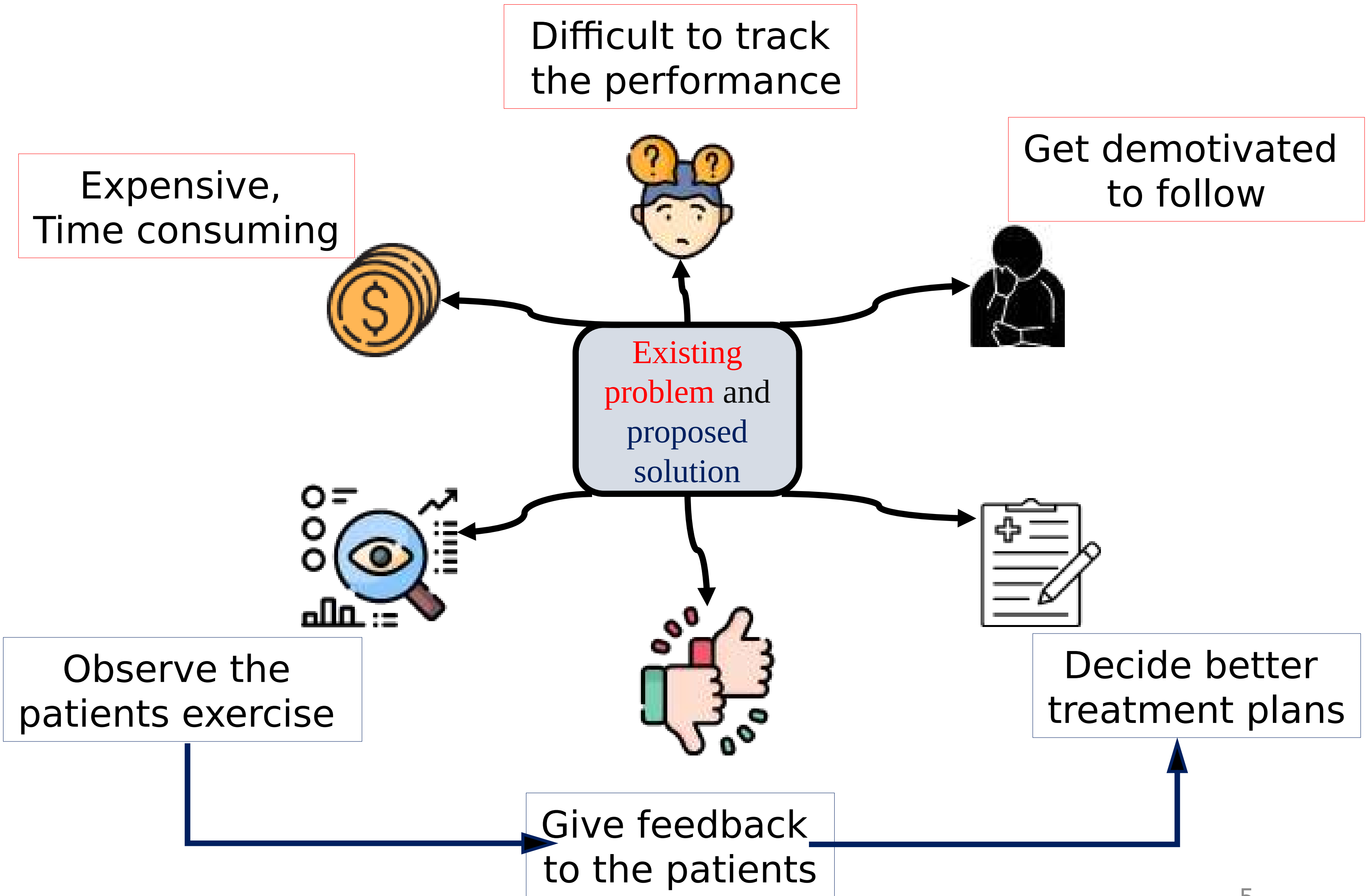
Supervised by
Dr. Sejuti Rahman
Assistant Professor
Department of Robotics and Mechatronics Engineering
University of Dhaka

Agenda

- Introduction
  - Motivation
- Preliminary
  - Data Collection, Graph, Skeleton as a Graph
  - Demo of Pose Estimation
- Novelty
  - Limitation of Existing Literature
  - Ours' Novelty
- Solution Framework
  - STGCN
- Proposed Approach
  - Role of body joint
  - Guidance
  - Varying Length
- Experiments & Results
  - Dataset Description
  - Overall Result
  - Analysis of Handcrafted and Featuring Learning
- Assessment on RGB camera
  - Overall Result
  - Trade off
- Conclusion

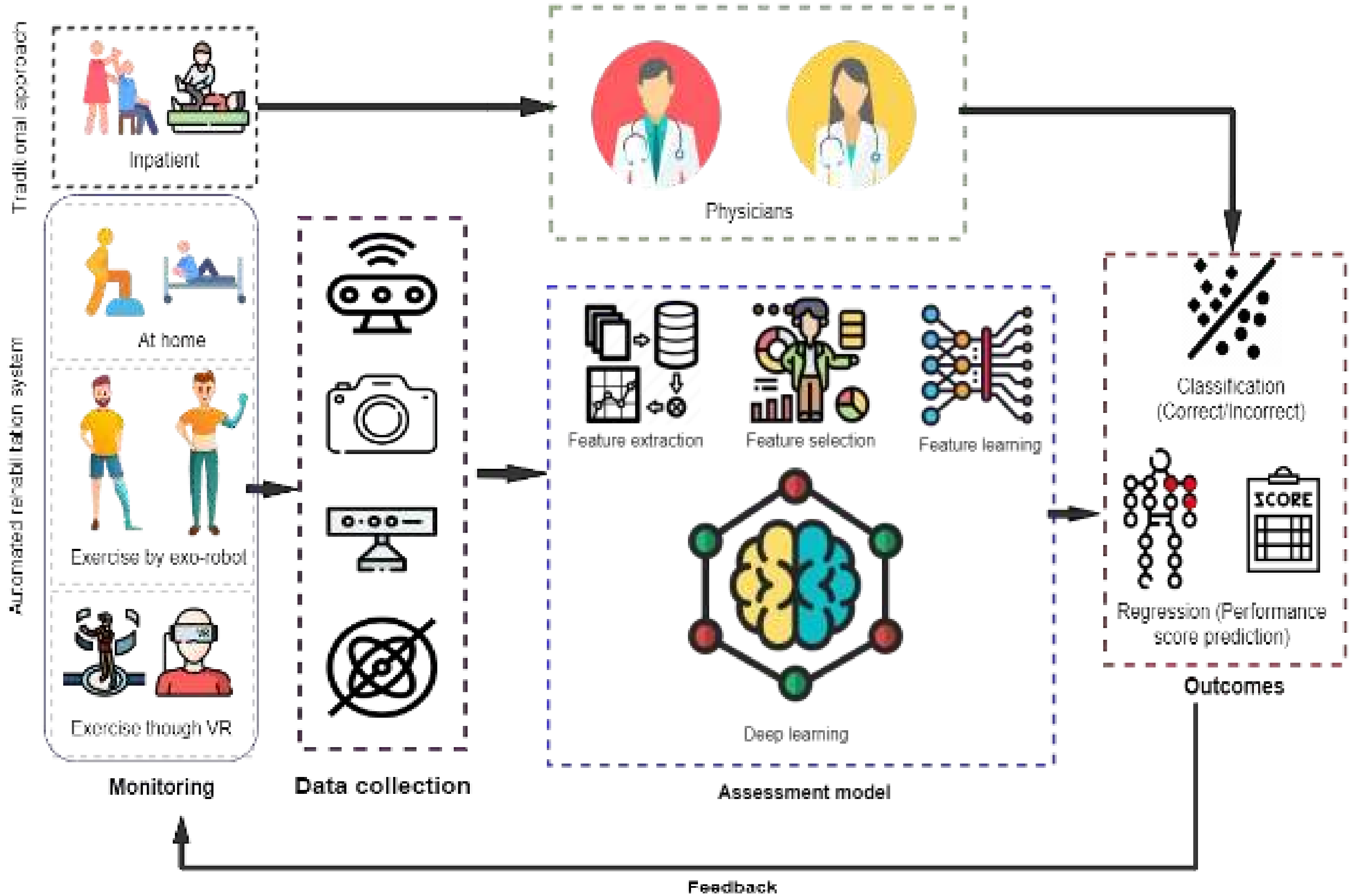
Introduction

Introduction



Introduction

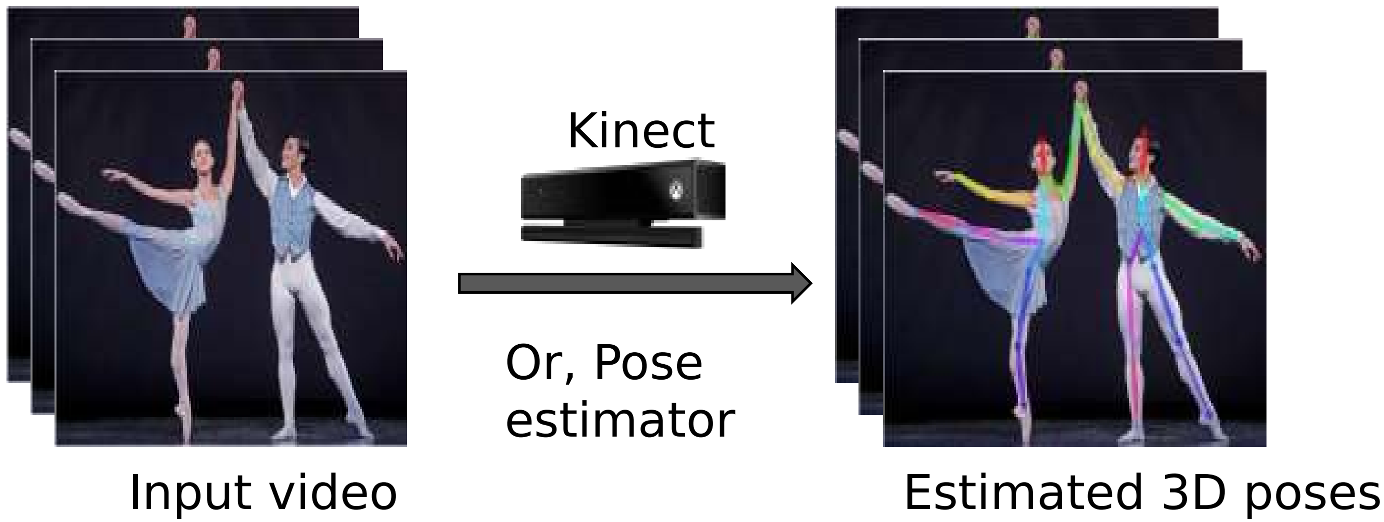
Motivation



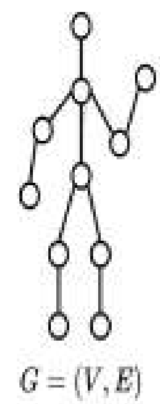
Preliminaries

Preliminaries

Data collection, Graph, Skeleton as Graph



- Human body can efficiently represent as skeleton[1]
- Joints → **nodes(V)**, bones → **edges(E)**
- **Graph (G)** naturally captures the structure of human body



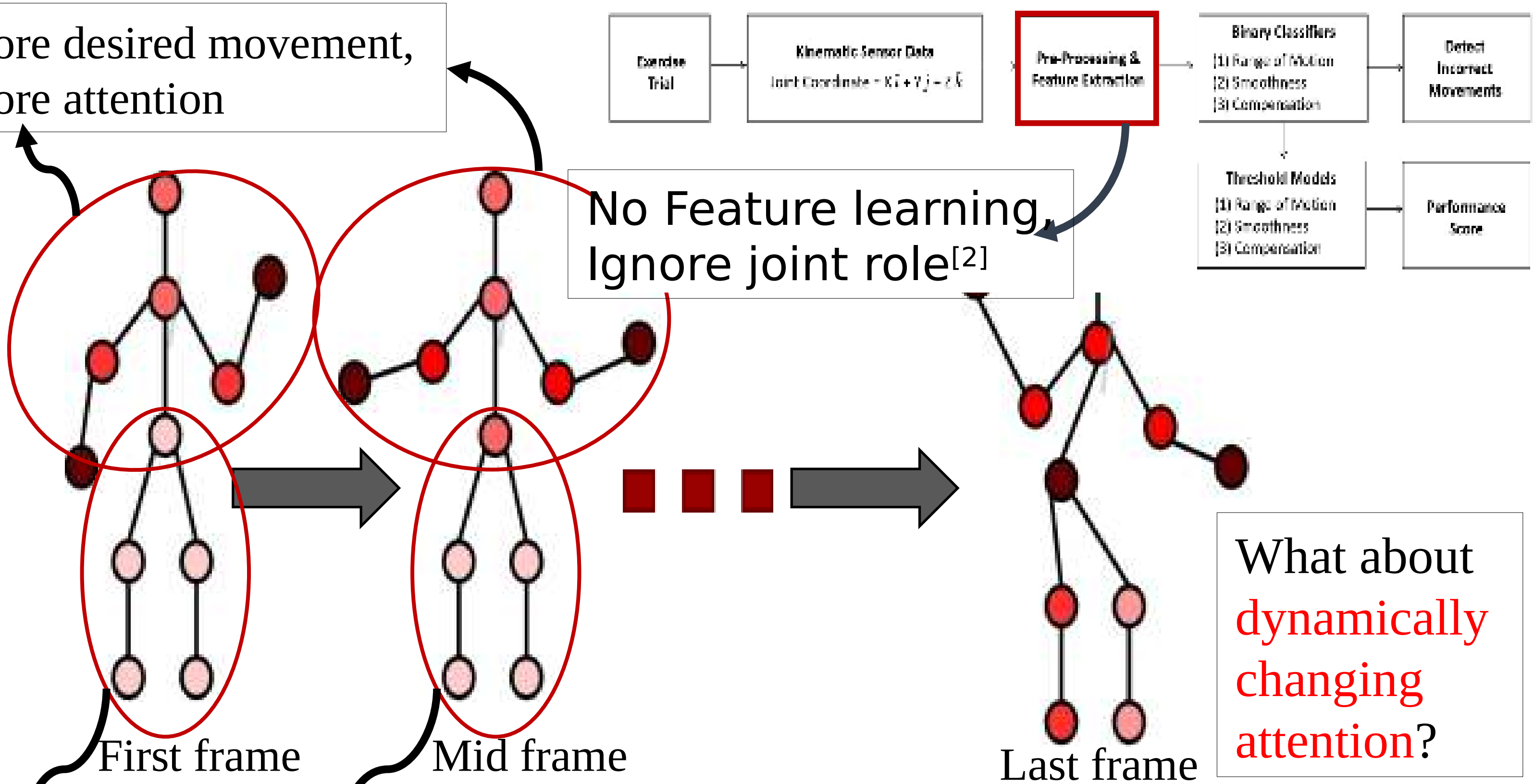
[1] Cao, Zhe, et al. "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields.", CVPR 2019.

Novelty

Novelty

Existing Literature & Ours' Improvement

More desired movement,
More attention



No Feature learning,
Ignore joint role[2]

What about dynamically changing attention?

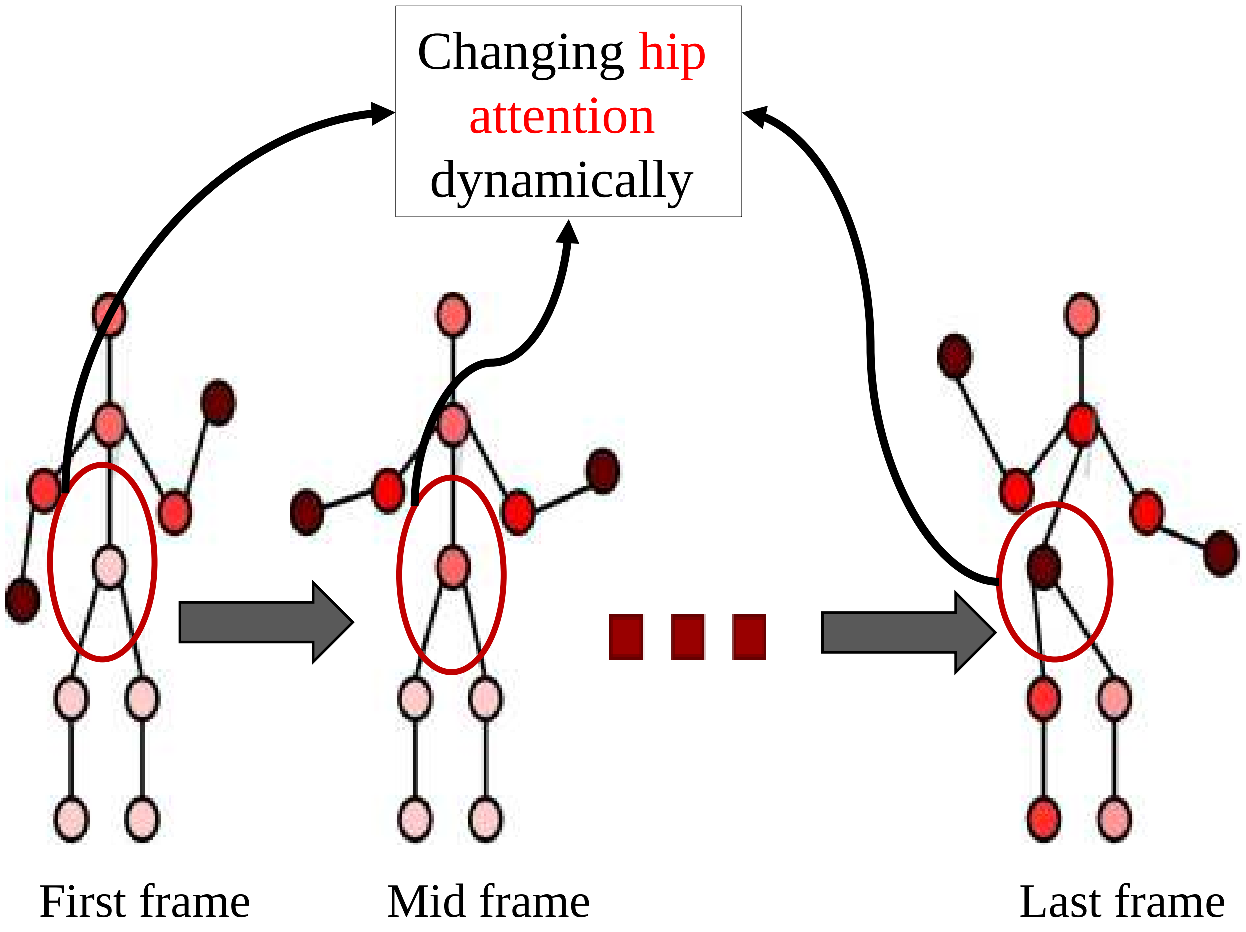
Less desired movement,
Less attention

Attention value changing spatially

[2] Lee et al., "Learning to assess the quality of stroke rehabilitation exercises," in Proceedings of the 24th International Conference on Intelligent User Interfaces, 2019

Novelty

Existing Literature & Ours' Improvement



Novelty

Existing Literature & Ours' Improvement

Subject 01

Model should support varying length Video

Subject 02

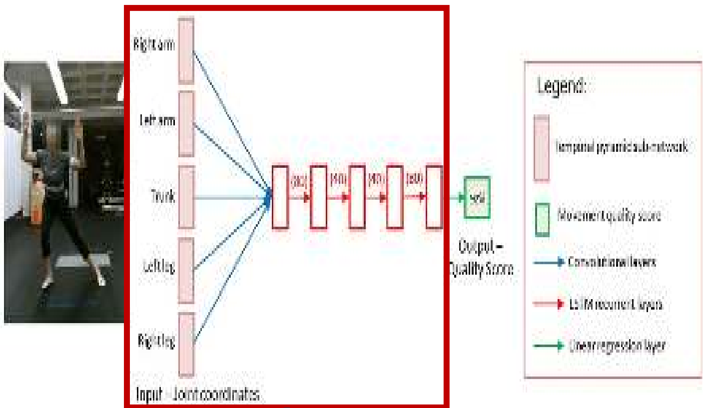
Increase robustness

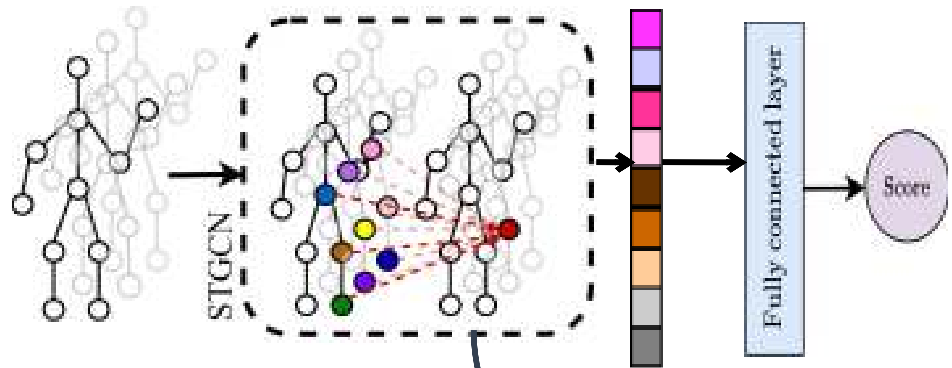
Subject 01 need less time

Subject 02 need more time

Novelty

Existing Literature & Ours’ Improvement





- Ignore Structural information,
- Don’t support varying length video[3]

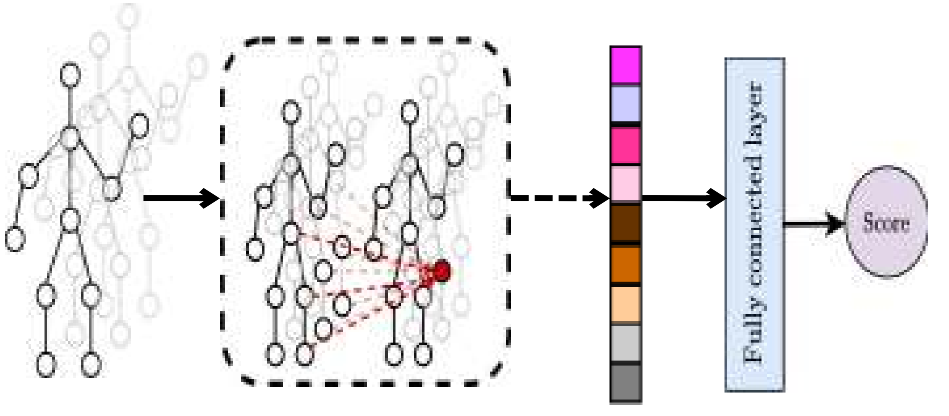
- Preserve structure while extracting features
- Support varying length video

[3] Liao et al., “A deep learning framework for assessing physical rehabilitation exercises,” IEEE TSNRE 2020.

Solution Framework

Solution Framework

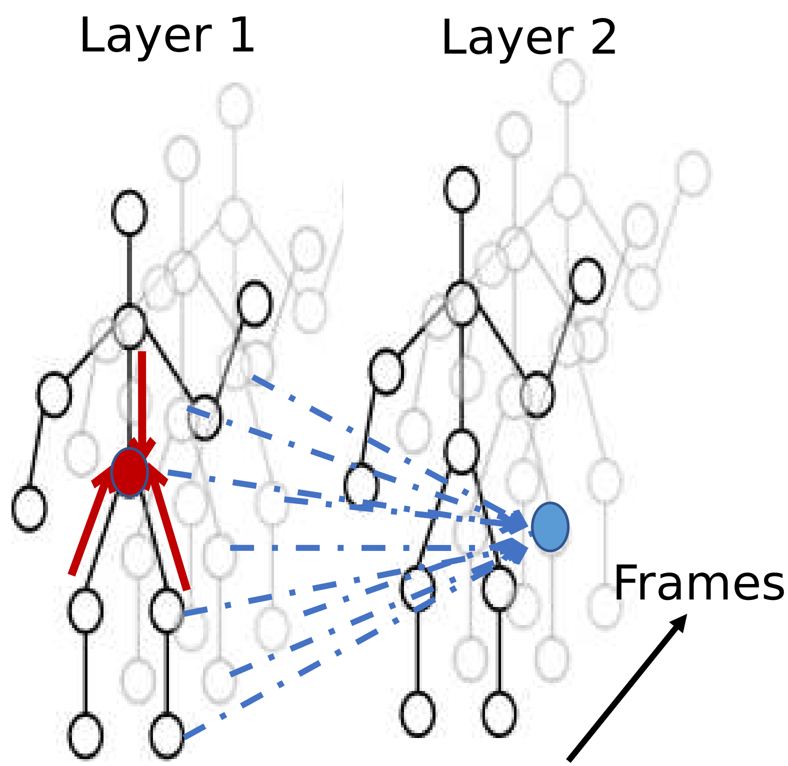
- Take the RGBD data as input.
- Pass it through the STGN layers to extract spatio-temporal features.
- Perform global pooling to convert it to a vector representation.
- Get the score output from the fully connected layer.



Solution Framework

ST-GCN

- Extract spatio-temporal features through successive STGCN layer[4].
- Each STGCN layer has two component:
  - **Spatial layer:** Extract spatial features performing Graph Convolution.
  - **Temporal layer(TC):** Extract temporal features.



Feature Update

$$\mathbf{G}_k = \sigma(\boxed{\tilde{A}_k \mathbf{Z}}\boxed{\mathbf{W}_k})$$

K-th layer

Neighborhood Aggregation

Structure: $\tilde{A}_k \in \mathbb{R}^{N\times N}$

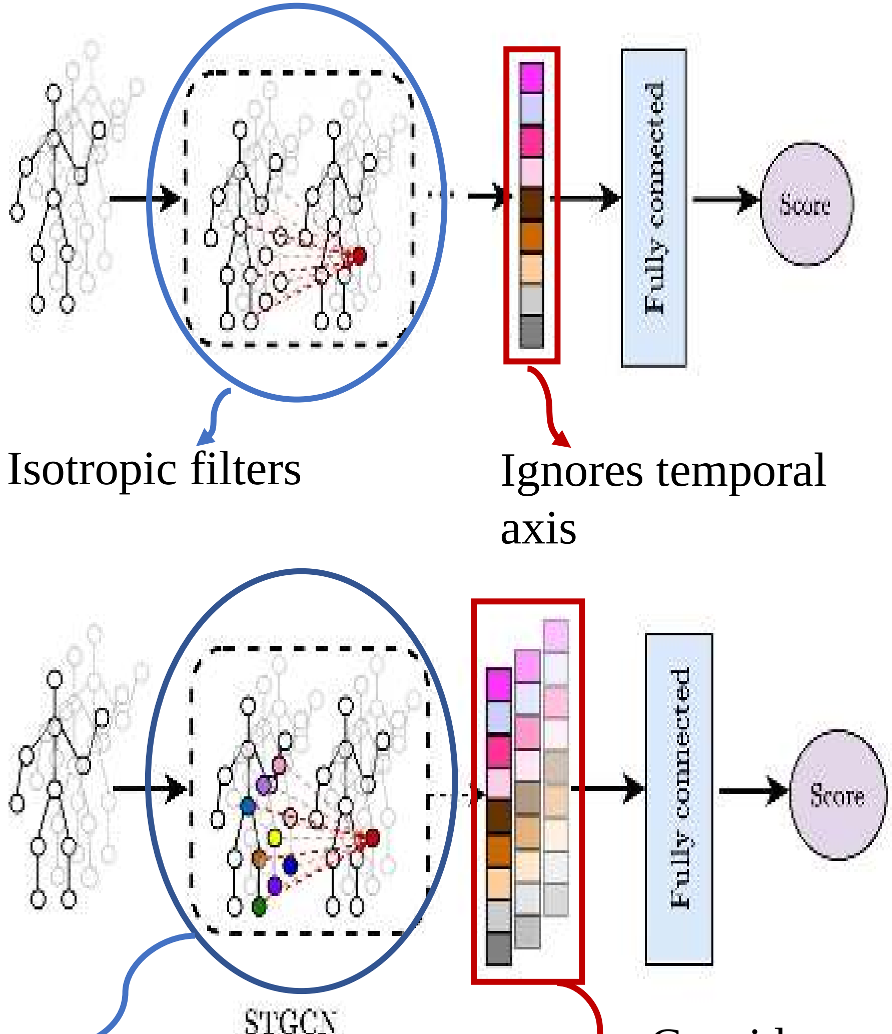
Feature: $Z \in R^{T\times N\times C}$

[4] Yan et. al., “Spatial temporal graph convolutional networksfor skeleton-based action recognition” AAAI 2018

Proposed Approach

Proposed Approach

- Role of body joints.
  - Dynamic attention in body joints.
- Input flexible.
  - Support variable length input.
  - Vanilla approach ignores sequential dependences reside in the features.



Isotropic filters

Ignores temporal axis

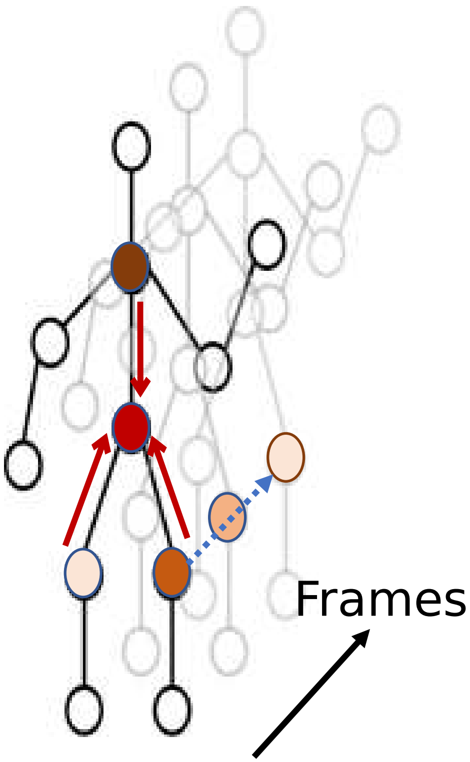
Dynamic attention

Attention weight indicated by colors

Consider sequential dependencies

Proposed Approach

Role of Body Joints



- Produce anisotropic filters that are more powerful than isotropic GCN.
- Treat spatial and temporal axis separate.

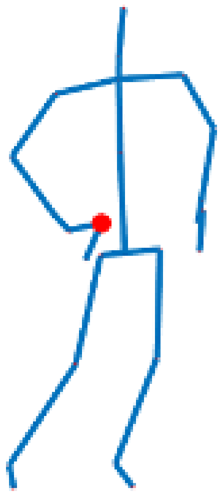
Feature Update

$$\mathbf{G}_k = \sigma(\boxed{\phi(\hat{A}_k \odot \mathbf{M}_k)\mathbf{Z}}\boxed{\mathbf{W}_k})$$

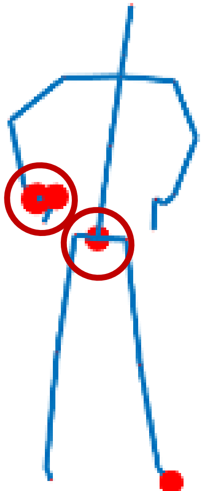
Attention based Aggregation

Proposed Approach

Guidence



Patient score: 43



Patient score: 18

Experiments

Experiments

Dataset Description

Feature	UI-PRMD	KIMORE
Reference	Vakanski et. al. (2018)	capecci et. al. (2019)
Year	2018	2019
Sensor	Kinect v2 + Vicon	Kinect v2
Modality	Skeleton data	RGB-D and skeleton data
No. of Subjects	10	78
No. of Exercises	10	5
Score range	0 - 1	0 - 50

- KIMORE[5] and UI-PRMD[6] are well established dataset on rehabilitation exercises.
- **Data is captured by professional physio therapists in their respective field and the labels are also given by them.**

[5] Capecci et. al. The kimore dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation," IEEE Transactions on Neural Systems and Rehabilitation Engineering,

[6] Vakanski el. al. A data set of human body movements for physical rehabilitation exercises, Data, vol. 3.

Experiments

Evaluation metrics

- In our task the evaluation criteria is mean absolute deviation(MAD), Mean absolute percentage error(MAPE), Root mean square error.
- Calculate deviation from the actual value.
- The lower the values better the performance.

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|y - \hat{y}|$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y - \hat{y}}{y}| \times 100$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y - \hat{y})^2}$$

n = sample size, y = label, y'=prediction

Experiments

Overall Result (UI-PRMD)

Ex	Ours	Song et al.[4]	Zhang et al.[5]	Liao et al. [3]	Li et al. [6]	Shahroudy et al. [7]	Du et al.[8]	Deep CNN	Deep LSTM
Ex1	0.009	0.011	0.022	0.011	0.011	0.018	0.030	0.013	0.016
Ex2	0.006	0.006	0.008	0.028	0.029	0.044	0.077	0.02	0.049
Ex3	0.013	0.010	0.016	0.039	0.056	0.081	0.137	0.041	0.093
Ex4	0.006	0.014	0.016	0.012	0.014	0.024	0.036	0.016	0.016
Ex5	0.008	0.013	0.008	0.019	0.017	0.032	0.064	0.013	0.025
Ex6	0.006	0.009	0.008	0.018	0.019	0.034	0.047	0.023	0.021
Ex7	0.011	0.017	0.021	0.038	0.027	0.049	0.193	0.033	0.040
Ex8	0.016	0.017	0.025	0.023	0.025	0.051	0.073	0.029	0.045
Ex9	0.008	0.008	0.027	0.023	0.027	0.043	0.065	0.024	0.044
Ex10	0.031	0.038	0.066	0.042	0.047	0.077	0.160	0.036	0.051

Table 1: Results on the UI-PRMD dataset

[3] Liao et al., “A deep learning framework for assessing physical rehabilitation exercises,” IEEE TSNRE 2020. [4] Song et. al., “Richly activated graph convolutional network for robust skeleton-based action recognition,” IEEE TCSVT, 2021. [5] Zhang et. al., “Semantics guided neural networks for efficient skeleton-based human action recognition,” CVPR 2020.[7]LI et. al., Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation,” AAAI 2018.[8] Du et al., “Hierarchical recurrent neural network for skeleton based action recognition,” CVPR 2015[6] Yanet al., “Spatial temporal graph convolutional networks for skeleton-based action recognition,” AAAI2018.

Experiments

Overall Result (KIMORE)

Metric	Exercise	Ours	Song et al.[4]	Zhang et al.[5]	Liao et al.[3]	Yan et al.[6]	Li et al.[7]	Du et al.[8]
MAD	Ex1	0.799	0.977	1.757	1.141	0.889	1.378	1.271
	Ex2	0.774	1.282	3.139	1.528	2.096	1.877	2.199
	Ex3	0.369	1.105	1.737	0.845	0.604	1.452	1.123
	Ex4	0.347	0.715	1.202	0.468	0.842	0.675	0.880
	Ex5	0.621	1.536	1.853	0.847	1.2184	1.662	1.864
RMS	Ex1	2.024	2.165	2.916	2.534	2.017	2.344	2.440
	Ex2	2.120	3.345	4.140	3.738	3.262	2.823	4.297
	Ex3	0.556	1.929	2.615	1.561	0.799	2.004	1.925
	Ex4	0.644	2.018	1.836	0.792	1.331	1.078	1.676
	Ex5	1.181	3.198	2.916	1.914	1.951	2.575	3.158
MAPE	Ex1	1.926	2.605	5.054	2.589	2.339	3.491	3.228
	Ex2	1.272	3.296	10.436	3.976	6.136	5.298	6.001
	Ex3	0.728	2.968	5.774	2.023	1.727	4.188	3.421
	Ex4	0.824	2.152	3.901	2.333	2.325	1.976	2.584
	Ex5	1.591	4.959	6.531	2.312	3.802	5.752	5.620

Table 2: Results on KIMORE dataset

[3] Liao et al., “A deep learning framework for assessing physical rehabilitation exercises,” IEEE TSNRE 2020. [4] Song et. al., “Richly activated graph convolutional network for robust skeleton-based action recognition,” IEEE TCSVT, 2021. [5] Zhang et. al., “Semantics guided neural networks for efficient skeleton-based human action recognition,” CVPR 2020.[7]LI et. al., Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation,” AAAI 2018.[8] Du et al., “Hierarchical recurrent neural network for skeleton based action recognition,” CVPR 2015[6] Yanet al., “Spatial temporal graph convolutional networks for skeleton-based action recognition,” AAAI2018.

Experiments

Ablation Study

Is-Stacked	Aggregation Style	Has-TCN Concatenated	Has-self attention	MAD	RMSE	MAPE
No	Global Pool	No	No	2.585	3.795	8.920
Yes	Global Pool	No	No	1.472	2.560	4.878
Yes	Global Pool	Yes	No	1.365	2.184	4.320
Yes	LSTM	Yes	No	0.767	1.484	2.340
Yes	LSTM	Yes	Yes	0.478	0.981	1.516

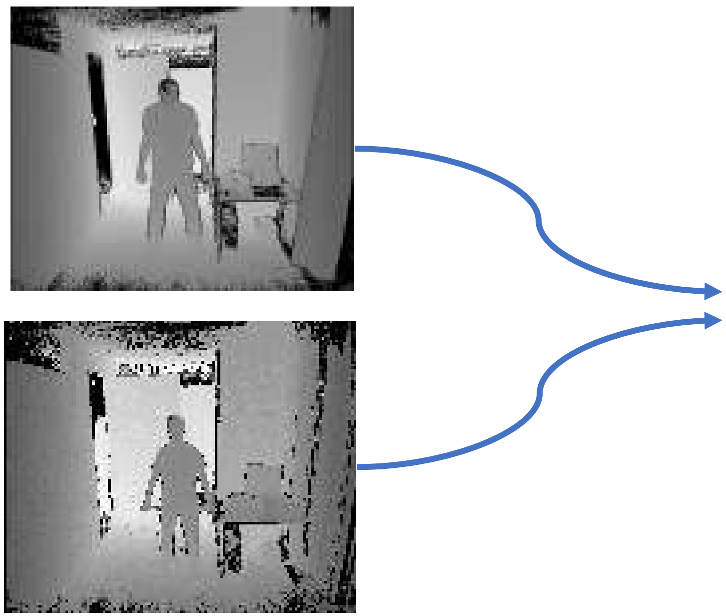
Considering sequential dependences of the features

Producing anisotropic filers

Experiments

Robustness Analysis

- Does speed matter for an exercise? No



Same correctness score

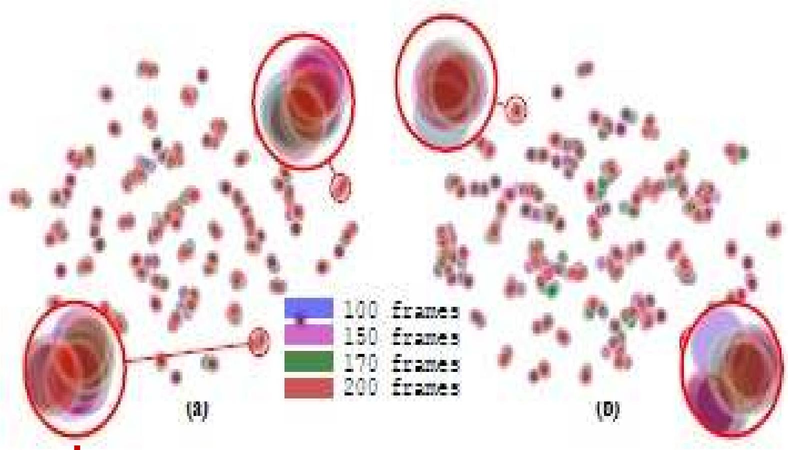
- Independent of peace, flexible model detect similar features.

Experiments

Robustness Analysis

- 2-D representation of the extracted spatio temporal features.
- Closer points indicate similar features

T-SNE plot



Detect similar features independent of input length

Handcrafted vs Feature Learning

Importance

- Benchmarking
- Scarcity of Benchmarking
- Benchmarking on Gold Standard Datasets

Handcrafted vs Feature Learning

Benchmark on UI-PRIMD

Ex	Deep Learning approach					Handcrafted features [5]			
	Ours	Du et al.[8]	Li et al. [7]	Yan et al. [4]	Liao et al. [3]	K-NN	RF	SVM	NN
Ex1	0.009	0.042	0.040	0.041	0.039	0.030	0.010	0.018	0.010
Ex2	0.006	0.008	0.008	0.009	0.018	0.077	0.029	0.044	0.028
Ex3	0.013	0.015	0.018	0.016	0.023	0.137	0.055	0.080	0.039
Ex4	0.006	0.018	0.021	0.021	0.025	0.035	0.013	0.023	0.011
Ex5	0.008	0.019	0.019	0.018	0.019	0.063	0.016	0.031	0.016
Ex6	0.006	0.014	0.014	0.014	0.014	0.046	0.018	0.034	0.017
Ex7	0.011	0.016	0.018	0.019	0.020	0.192	0.027	0.049	0.038
Ex8	0.016	0.014	0.020	0.016	0.022	0.072	0.024	0.050	0.023
Ex9	0.008	0.016	0.016	0.017	0.018	0.065	0.027	0.043	0.022
Ex10	0.031	0.076	0.091	0.085	0.095	0.160	0.046	0.077	0.041

Table 4: Results on the UI-PRMD dataset

[3] Liao et al., “A deep learning framework for assessing physical rehabilitation exercises,” IEEE TSNRE 2020 [7] Li et. al., Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation,” AAAI 2018. [8] Du et al., “Hierarchical recurrent neural network for skeleton based action recognition,” CVPR 2015[6] Yanet al., “Spatial temporal graph convolutional networks for skeleton-based action recognition,” AAAI2018.

Handcrafted vs Feature Learning

Benchmark on KIMORE

Ex	Matric	Deep Learning approach					Handcrafted features [5]			
		Ours	Du et al.[8]	Li et al. [7]	Yan et al. [4]	Liao et al. [3]	K-NN	RF	SVM	NN
Ex1	MAD	0.79	1.27	1.37	0.88	1.14	1.38	4.77	3.12	3.08
	RMS	2.02	2.44	2.34	2.01	2.53	2.33	5.93	4.34	4.35
	MAPE	1.92	3.22	3.49	2.33	2.58	3.90	21.34	23.82	8.51
Ex2	MAD	0.774	2.19	1.87	2.09	1.52	3.19	4.82	2.83	3.97
	RMS	2.12	4.29	2.82	3.26	3.73	4.22	5.71	3.68	5.22
	MAPE	1.27	6.00	5.29	6.13	3.97	8.65	21.54	23.53	10.50
Ex3	MAD	0.36	1.12	1.45	0.60	0.84	3.38	5.08	3.94	4.53
	RMS	0.55	1.92	2.00	0.79	1.56	5.23	6.42	5.40	6.55
	MAPE	0.72	3.42	4.18	1.72	2.02	9.67	19.23	20.89	12.26
Ex4	MAD	0.34	0.88	0.67	0.84	0.46	2.41	4.86	3.34	4.76
	RMS	0.64	1.67	1.07	1.33	0.79	3.73	5.99	4.52	6.94
	MAPE	0.84	2.58	1.97	2.32	2.33	5.95	19.33	21.36	12.58
Ex5	MAD	0.62	1.86	1.66	1.21	0.84	2.71	4.80	3.04	2.34
	RMS	1.18	3.15	2.57	1.95	1.91	4.28	6.07	4.91	3.15
	MAPE	1.59	5.62	5.75	3.80	2.31	6.84	16.74	18.04	5.41

Table 5: Results on the KIMORE dataset

[3] Liao et al., "A deep learning framework for assessing physical rehabilitation exercises," IEEE TSNRE 2020 [7] Li et. al., Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation," AAAI 2018. [8] Du et al., "Hierarchical recurrent neural network for skeleton based action recognition," CVPR 2015[6] Yanet al., "Spatial temporal graph convolutional networks for skeleton-based action recognition," AAAI2018.

Assessment with RGB Camera

Assessment on RGB Data

Overall Result

Successfully outlook unnecessary joints

GCN

Unnecessary joint degrade performance(BlazePose)

Metric	Algorithm	Ours	Liao et al. [3]	Yan et al. [6]	Li et al. [7]	Du et al. [8]
MAD	BlazePose	0.971	4.043	3.709	4.548	6.309
	VideoPose3D	1.855	2.554	3.084	3.546	4.669
	Kinectv2	0.621	0.847	1.218	1.663	1.864
RMS	BlazePose	1.993	5.991	5.657	7.194	8.681
	VideoPose3D	3.822	3.908	4.943	5.202	6.012
	Kinectv2	1.180	1.914	1.951	2.575	3.158
MAPE	BlazePose	3.081	15.618	15.917	20.897	25.816
	VideoPose3D	6.810	8.102	10.790	11.964	14.750
	Kinectv2	1.591	2.312	3.802	5.752	5.620

Table 6: Results on KIMORE RGB Exercise 5

[3] Liao et al., "A deep learning framework for assessing physical rehabilitation exercises," IEEE TSNRE 2020 [7] Li et. al., Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation," AAAI 2018. [8] Du et al., "Hierarchical recurrent neural network for skeleton based action recognition," CVPR 2015[6] Yanet al., "Spatial

BlazePose: 33 joints
VideoPose3D : 17

Assessment on RGB Data

- Pros
  - Cost Efficient
  - Easily Available
  - No Need Extra Sensors

- Cons
  - Decrease Performance
  - Time Consuming
  - Depth Ambiguity

Trade off between price and performance

Conclusion

- Propose attention guided GCN for assessing physical rehabilitation exercises.
- Leverage attention mechanism (dynamic), flexible to input length and guidance system.
- Future direction
    - ➢ Improve the guidance system using supervision of therapist.
    - ➢ Take input as cheaply available RGB data.

Publications, Code and Other Resources

1. S. Deb, M. F. Islam, S. Rahman and S. Rahman, "Graph Convolutional Networks for Assessment of Physical Rehabilitation Exercises," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, , vol. 30, pp. 410-419, 2022, DOI: 10.1109/TNSRE.2022.3150392. (SJR rank: Q1, IF= 4.528)

2. S. Rahman, S. Sarker, A. K. M. Nadimul Haque, M. M. Uttsha, M. F. Islam and S. Deb, "AI-driven Stroke Rehabilitation Systems and Assessment: A Systematic Review," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022, DOI: 10.1109/TNSRE.2022.3219085. (SJR rank: Q1, IF= 4.528)

Check code and other resources @

Reference

Cao, Zhe, et al. "OpenPose: realtime multi-person 2D pose estimation using
art Affinity Fields.“, CVPR 2019.

Liao et. al., “A deep learning framework for assessingphysical rehabilitation
ercises,” IEEE Trans. Neural Syst. Rehabilitation Eng, 2020

Yan et. al., “Spatial temporal graph convolutional networksfor skeleton-based
ction recognition” AAAI 2018

] Song et. al., “Richly activated graph convolutional network for robust skeleton-
ction recognition,” IEEE TCSVT, 2021.

] Zhang et. al., “Semantics guided neural networks for efficient skeleton-based
ction recognition,” CVPR 2020.

]LI et. al., Co-occurrence feature learning from skeleton data for action recogniti
nd detection with hierarchical aggregation,” AAAI 2018.

Reference

[8] Du et al., “Hierarchical recurrent neural network for skeleton based action recognition,” CVPR 2015

Thank you