Graph Convolutional Networks for Assessment of Physical Rehabilitation Exercises

Swakshar Deb[1], Md Fokhrul Islam[1] Shafin Rahman[2], Sejuti Rahman*[1]

[1] Department of Robotics and Mechatronics Engineering, University of Dhaka, Dhaka, Bangladesh

[2]Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Summary

Goal: Developing a novel end-to-end model for assessing rehabilitation exercises and providing guidance on improving assessment scores.

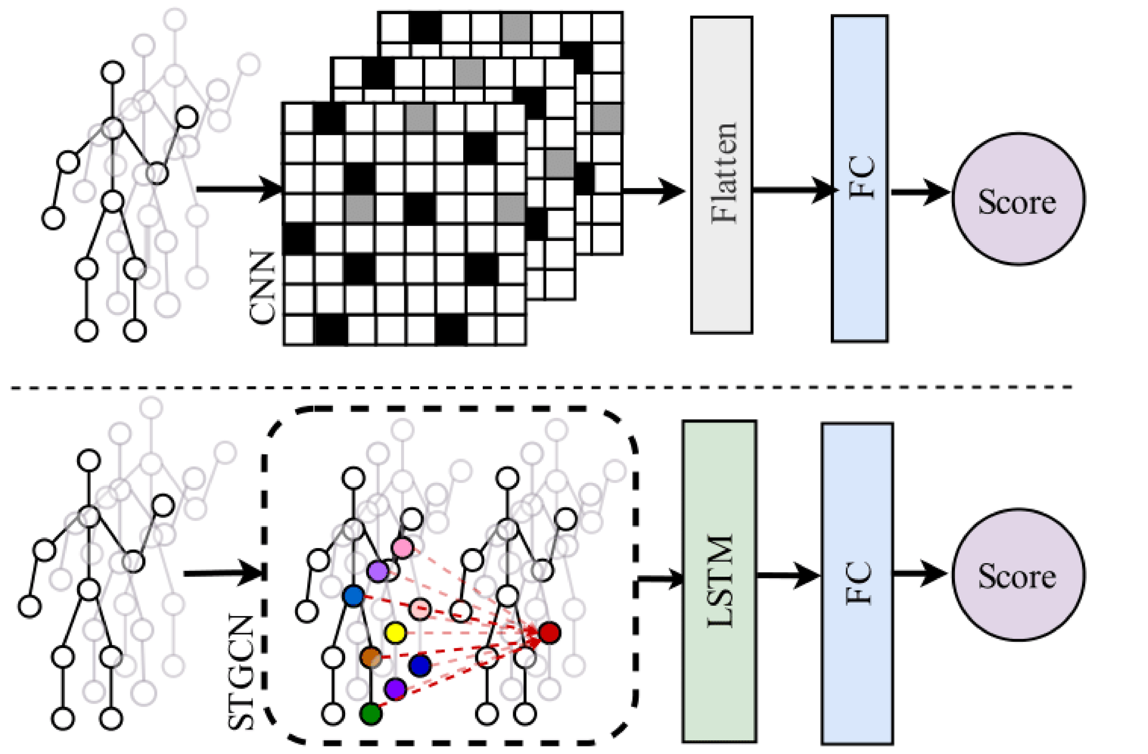
Motivation: Patients in home rehabilitation lose motivation and confidence, which worsens their medical condition. An automated assessment system could keep track of a patient's progress and provide feedback to both the patient and the physician.

Limitations:

- Recent approaches [1] use CNN to extract features while ignoring the topological structure of the human skeleton.
- Current methods use fixed duration input videos and reduce model resilience.

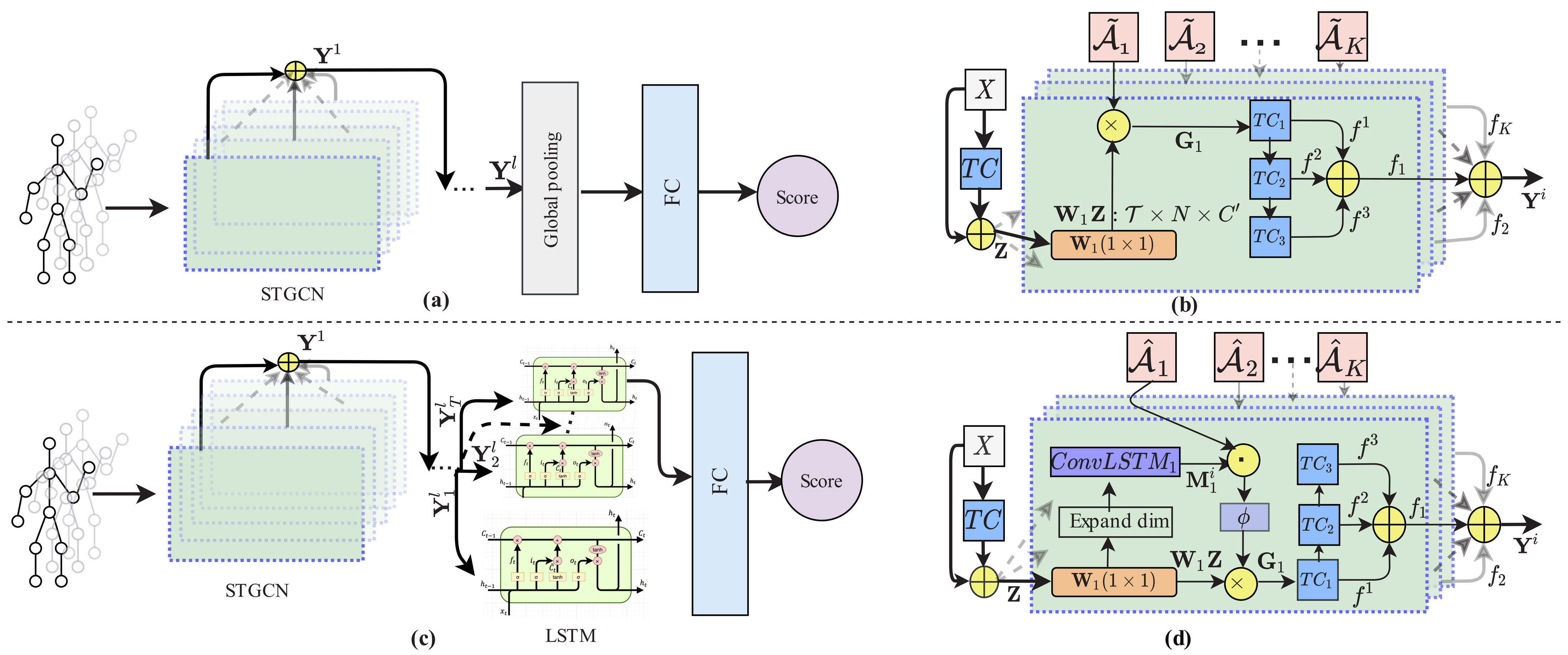
Our Approach: We propose dynamic attention-guided graph convolution, which focuses on the most essential joints and an LSTM layer to extract subtle movement information, given an RGB-D video. We identify how body-joints contribute to the final score by analyzing the attention amount, which eventually guides the user to perform better in subsequent trials.

Contributions



- We extend the popular STGCN [2] to adapt it for the assessment of rehabilitation exercises in an end-to-end manner.
- Our proposed model supports variable-length exercise input considering any number of repetitions of a given exercise during training and testing.
- We also offer a self-attention mechanism to guide users by highlighting body-joints contributing more to the prediction.

Method



- The model takes RGB-D data as input.
- Passes it through the STGN layers to extract spatio-temporal features.
- Vanilla STGCN ignores sequential dependencies using several pooling layers (Figure a).
- Vanilla STGCN ignores joints importance while treating every joint equally (Figure b): $\mathbf{G}_k = \sigma(\tilde{A}_k \mathbf{Z} \mathbf{W}_k)$.
- We provide input flexibility (Figure c) that supports variable length input and also considers sequential dependencies using LSTM.
- We consider the role of body joints (also providing guidance, Figure d) using a dynamic attention module, $\mathbf{G}_k = \sigma(\phi(\hat{A}_k \odot \mathbf{M}_k)\mathbf{Z}\mathbf{W}_k)$.
- Get the score (0-50) as output from the Fully Connected (FC) layer.

Experiments

Results on KIMORE and UI-PRMD datasets (lower values indicate better result):

Metric	Ex	Ours	Song *et al.[3]*	Zhang *et al.*[4]	Liao *et al.*[1]	Yan *et al.*[2]	Li *et al.*[5]	Du *et al.*[]7]
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

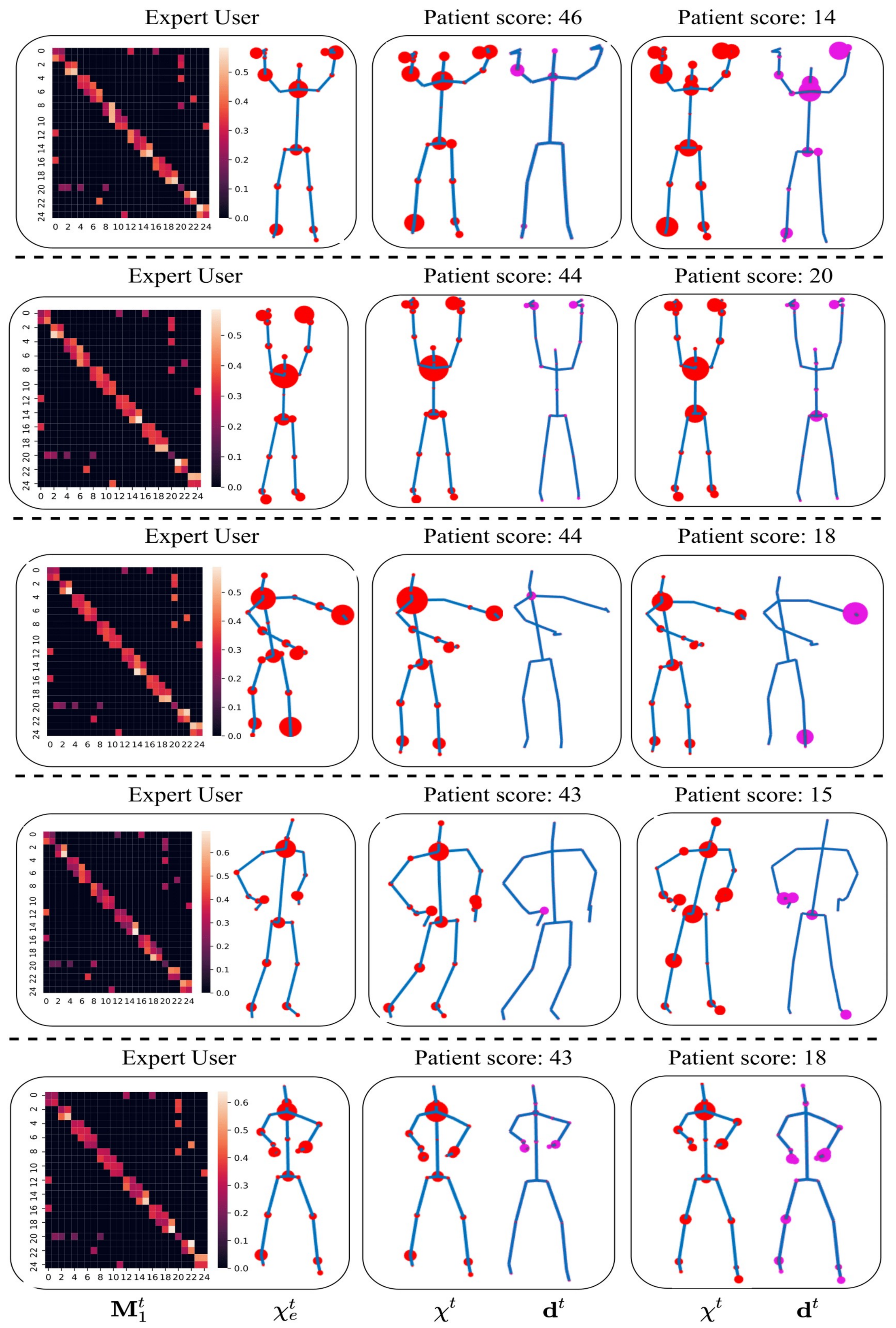
(a) KIMORE RGB-D

Ex	Ours	Song *et al.*[3]	Zhang *et al.*[4]	Liao *et al.*[1]	Li *et al.*[5]	Shahroudy *et al.*[6]	Du *et al.*[7]
Ex1	**0.009**	0.011	0.022	0.011	0.011	0.018	0.030
Ex2	**0.006**	**0.006**	0.008	0.028	0.029	0.044	0.077
Ex3	0.013	**0.010**	0.016	0.039	0.056	0.081	0.137
Ex4	**0.006**	0.014	0.016	0.012	0.014	0.024	0.036
Ex5	**0.008**	0.013	0.008	0.019	0.017	0.032	0.064
Ex6	**0.006**	0.009	0.008	0.018	0.019	0.034	0.047
Ex7	**0.011**	0.017	0.021	0.038	0.027	0.049	0.193
Ex8	**0.016**	0.017	0.025	0.023	0.025	0.051	0.073
Ex9	**0.008**	**0.008**	0.027	0.023	0.027	0.043	0.065
Ex10	**0.031**	0.038	0.066	0.042	0.047	0.077	0.160

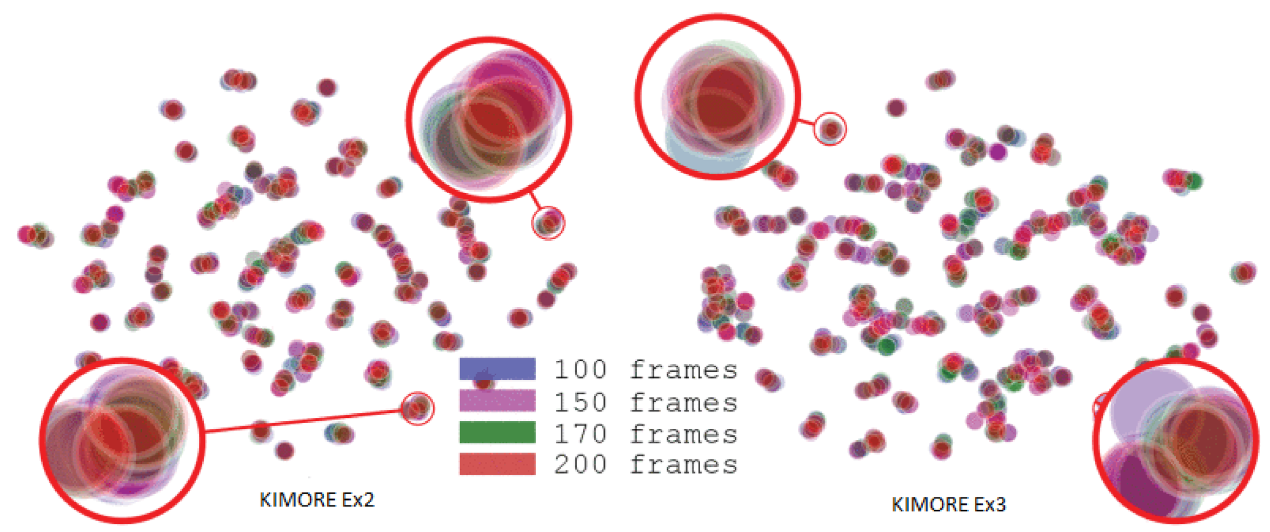
(b) UI-PRMD RGB-D

Metric	Algorithm	Ours	Liao *et al.*[1]	Yan *et al.*[2]	Li *et al.*[5]	Du *et al.*[7]
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

(c) KIMORE RGB video (Exercise 5)



Role of body joints



Visualization on different length of videos

References:

[1] Liao *et al.* A deep learning framework for assessing physical rehabilitation exercises. IEEE TNSRE, vol. 28, no. 2, pp. 468–477, 2020.

[2] Yan *et al.* Spatial temporal graph convolutional networks for skeleton-based action recognition. In AAAI, 2018.

[3] Song *et al.* Richly activated graph convolutional network for robust skeleton-based action recognition. IEEE Trans. Circuits Syst. Video Technol.,vol. 31, no. 5, pp. 1915–1925, 2021.

[4] Zhang *et al.* Semantics guided neural networks for efficient skeleton-based human action recognition. In CVPR, 2020.

[5] Li *et al.* Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation. In IJCAI, 2018.

[6] Shahroudy *et al.* Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In CVPR, 2016.

[7] Du *et al.* Hierarchical recurrent neural network for skeleton based action recognition. In CVPR, 2015.