Attention-enhanced graph convolutional network for assessing rehabilitation exercises

Smita Priyadarshani and Hiroshi Watanabe

***Abstract*—An automated vision-based rehabilitation exercise assessment is a portable and cost-effective way of evaluating the patients' performance in a home-based setting and predicting a performance score by analyzing the correct and incorrect exercise sequences performed by the patient. Recent works have shown that exploring spatial and temporal features of the skeleton data is vital for this task. However, most of the methods treat all body joints equally and fail to capture the correlation information between all joints. Hence, to address this limitation, we have proposed an attention-enhanced graph convolutional network (ST-GCN) that captures the spatial and temporal dependencies among the body joints incorporating an attention mechanism to learn global information for the joint-specific roles to provide better assessment results.**

I. Introduction

Current deep learning approaches extract features but fail to capture the patterns embedded in the spatial configuration of the joints as well as their temporal dynamics. Thus, we incorporate a non-local attention module [1] into the spatial temporal graph convolutional network to capture long-range dependencies directly by computing interactions between any two positions, regardless of their positional distances.

II. Proposed Method

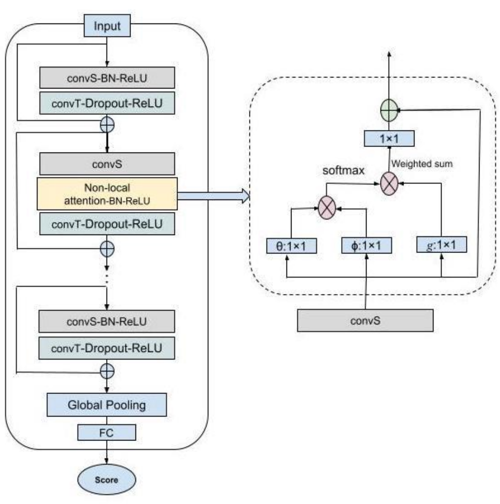

Fig.1. Network architecture of the proposed attention-enhanced spatial temporal graph convolutional network (left). Non-local attention block (right).

First, we start with constructing graph and performing spatial and temporal convolutions on the input data. This processed input is fed into the ST-GCN which includes the non-local attention module [1] in layer 2 and 3 between the spatial and temporal convolution layer. Inside this attention module, the first 1×1 linear mapping is performed on the feature maps of spatial convolution. Second, a matrix point multiplication operation is performed on θ and ϕ to calculate the autocorrelation in the feature, followed by SoftMax to obtain the self-attention coefficient. Lastly, the attention coefficient is multiplied by the feature matrix g and a residual connection is added with the original input feature map. After processing through 7 layers of spatial and temporal blocks along with a non-local attention mechanism, a global pooling layer aggregates the features across the layers to predict the final assessment score.

III. Experimental Results

We evaluated the model using PyTorch and two Nvidia RTX 3060 GPUs on two publicly available rehabilitation exercise datasets: (1) UI-PRMD and (2) KIMORE. We use mean absolute error (MAE) score as the evaluation metrics. Table 1. below summarizes the quantitative results.

TABLE I. Average mean absolute error (MAE) scores of exercise assessment

Avg. MAE score	Our method	GCN + self-supervision [2]	Deep CNN[3]
UI-PRMD	**0.0219**	0.0221	0.0252
KIMORE	**0.7840**	-	0.0966

Conclusion

In this paper we present an attention-enhanced spatial temporal graph convolutional network that efficiently learns the spatial and temporal relationships, incorporating an attention mechanism that computes the response at a position as a weighted sum of the features at all positions in the input feature maps. This helps the network learn long-range dependencies between the joints that might not be directly connected to a specific joint in the graph but play an important role in a particular exercise.

References

[1] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He, "Non-local neural networks," Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR2018), pp.7794-7803, Jun. 2018.

[2] Chen Du,Sarah Graham, Colin Depp, and Truong Nguyen, "Assessing physical rehabilitation exercises using graph convolutional network with self-supervised regularization," 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp.281-286, Oct. 2021.

[3] Yalin Liao, Aleksandar Vakanski, and Min Xian, "A deep learning framework for assessing physical rehabilitation exercises," IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 28, No. 2, pp.468-477, Feb. 2020.

Computer Science and Communications Engineering, Waseda Universiy, Tokyo, Japan (e-mail: smitapriya@toki.waseda.jp).