CP-AGCN: Pytorch-based Attention Informed Graph Convolutional Network for Identifying Infants at Risk of Cerebral Palsy

Haozheng Zhang (Durham University, haozheng.zhang@durham.ac.uk),

Edmond S. L. Ho (University of Glasgow, Shu-Lim.Ho@glasgow.ac.uk),

Hubert P. H. Shum (Durham University, hubert.shum@durham.ac.uk, corresponding author)

Abstract

Early prediction is clinically considered one of the essential parts of cerebral palsy (CP) treatment. We propose to implement a low-cost and interpretable classification system for supporting CP prediction based on General Movement Assessment (GMA). We design a Pytorch-based attention-informed graph convolutional network to early identify infants at risk of CP from skeletal data extracted from RGB videos. We also design a frequency-binning module for learning the CP movements in the frequency domain while filtering noise. Our system only requires consumer-grade RGB videos for training to support interactive-time CP prediction by providing an interpretable CP classification result.

Keywords

Code metadata

Nr.	Code metadata description	Please fill in this column
C1	Current code version	v1
C2	Permanent link to code/repository used for this code version	https://github.com/zhz95/CP-AGCN
C3	Permanent link to Reproducible Capsule	https://codeocean.com/capsule/6073072/tree/v1
C4	Legal Code License	MIT License
C5	Code versioning system used	git
C6	Software code languages, tools, and services used	Python
C7	Compilation requirements, operating environments & dependencies	Python 3.7, PyTorth 1.8.10, OpenPose 1.7.0
C8	If available Link to developer documentation/manual	https://codeocean.com/capsule/6073072/tree/v1
C9	Support email for questions	haozheng.zhang@durham.ac.uk

1. Introduction

Due to the abnormal development or damage to parts of the brain, cerebral palsy (CP) appears in the early stages of the patient's childhood and permanently affects the patient's quality of life. There are about $2-3$ CP patients in 1000 children in the UK [1], which is similar to other developed countries. Although CP cannot be completely cured at present, early prediction of CP and intervention are considered as a paramount part of the treatment.

Current clinical early prediction of CP is investigated by General Movement Assessment (GMA) [2]. GMA can be done in person by GM assessors to assess an infant, or it can be done via watching an RGB video that has recorded the general movements of the infant. However, the GMA training is time- and resource-consuming, making it challenging to cope with the high demand for CP prediction. To tackle this problem, we propose automating this process by analyzing the general movements of infants from RGB videos. This allows the early prediction to cover even the lower-risk population. Motivated by the encouraging results reported in recent research based on skeletal data [3, 4, 5, 6, 7, 8, 9, 10], the 2D joint locations of the infant are extracted from RGB videos as the input of the system for CP prediction. The computational intelligence of our system is implemented with a graph convolution network, a kind of deep artificial neural network that models relational data very well, making it suitable for skeleton data. Our graph convolution operation encodes the input of human joint movement as frequency features in the graph structure. In addition, we employed the attention mechanism to enhance the learning performance with better interpretability. As a result, the implemented system can model the graph structure of the human pose features and amplify the influence of the most important parts of the data that are considered by the neural network.

We implement our system in PyTorch due to the availability of compatible open-source resources and its good coverage of the required deep learning functionalities. Firstly, we adapt OpenPose [11], a pose estimation software, to extract the human skeletal pose features from only the consumer-level RGB videos. Then, we implement the Fast Fourier Transform (FFT) algorithm with a tailor-made binning operation to analyse the human joint movement frequencies from the pose features. With the frequency operations, we can reduce the noise in the data and filter the high-frequency movement features that are less relevant for the CP classification [12, 13, 9]. We implement our CP classification GCN network based on the backbone (ST-GCN) framework from Shi *et al.* [14], because it shows promising classification performance and is widely identified as the benchmark in human pose-based studies. We validate our system in two datasets, showcasing that our system achieves state-of-the-art performance. The theoretical background of the system has been published in [13]. In this paper, we explain the software implementation details and discuss the impact of our software.
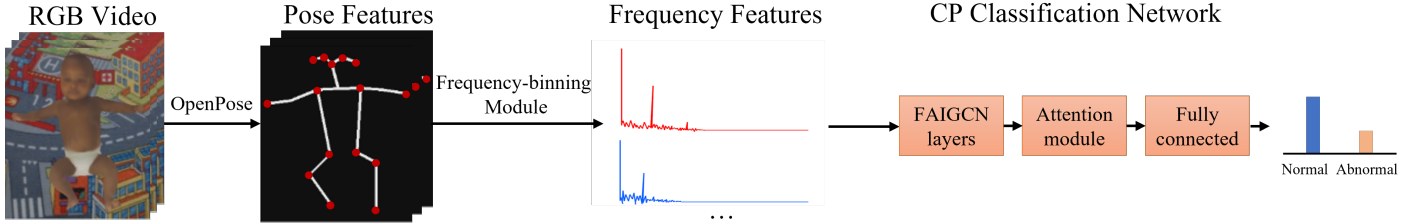
Figure 1: The architectures of proposed CP classification system.

2. System description

Algorithm 1 Attention GCN for CP classification

Inputs: $\mathbf{H}$: the set of joint features (dim: $B \times N$); A: the adjacency matrix of the graph distance between joints (dim: $N \times N$); An one-hot classification label embedding.

Output: A 0/1 classification result.

1: $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_L$, $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$.

2: **if** s $<$ S **then**:

3:    Take the graph convolution operation by Eq. 2, where $\mathbf{W}$ is updated by Eq. 3 and 4.

4:    s $\leftarrow$ s +1.

5: Encode the output to a 0/1 classification result by concatenation.

As shown in Figure 1, our proposed CP classification system first extracts infants' pose features (i.e., joint positions) by the pose estimation algorithm OpenPose [11], due to its robust performance. The latest version of OpenPose (1.7.0) comes with the pre-trained model, such that one can use the network effectively without training. Then, we implement a frequency-binning module to convert the pose features into the Fast Fourier Transform (FFT) coefficients for learning CP in the frequency domain. Both pose features or frequency features are fed into our designed PyTorch-based attention-informed graph convolutional network (CP-AGCN) for classifying normal and abnormal CP labels.

In particular, the frequency-binning module adapts Bluestein's FFT algorithm [15] for feature conversion. We tailor-made a binning method to filter high-frequency components, which mainly consist of data noise and high-frequency movement features:

$$f_n = \begin{cases} \text{Round}(f_0 \cdot c^n), & \text{if } f_n \cdot c^n < 3, \\ \text{Ceiling}(f_0 \cdot c^n), & \text{if } f_n \cdot c^n \geq 3, \end{cases} \tag{1}$$

where f_n ($f_0 = 1$) is the width of the nth bin, and c is a parameter that control the bin width.

The design in Eq. 1 enables the classification network to focus on learning the low-to-mid band movement frequency features. To retain the low-to-mid frequency human movement information, this function takes the round for widths less than three units. This function can distinguish the middle frequency and high-frequency bands based on the exponential growth of c and the ceiling function. Since we have not included any learnable parameter in this module, such a low computational cost module is adaptable for both machine learning-based or deep neural network (DNN) based classification systems with 24-60 FPS input videos.

For CP classification, the algorithm flow is shown in Algorithm 1. We employ the GCN [16] as the benchmark model to learn the joint dependencies from the skeleton graph. The graph convolutional propagation is demonstrated by:

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right) \tag{2}$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_L$ is the adjacency matrix. $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$. $\mathbf{I}_L$ is an identity matrix with L dimensions. $\mathbf{W}^{(l)}$ is the learnable weight matrix of layer l. $\sigma(\cdot)$ is the $ReLU$ nonlinear activation function.

By using the architecture backbone from [14], we apply the human skeleton graph $G = (V, E)$ for interpreting the importance ranking of joint's features CP classification. In this graph, $\{V = v_{bi} | b = 1, ..., B; i = 1, ..., N\}$ is the set of joint features, where v_{bi} is the b-th bin of i-th joint. The edge set E consists of: (1) the intra-skeleton connection, $\{v_{bi}v_{bj} | (i, j) \in K\}$, where K is the amount of the natural connections of skeleton joints. (2) the inter-feature edges which connect the bins of a joint from low to high, $\{v_{bi}v_{(b+1)i}\}$.

We propose to use attention-informed mechanism to learn the weight of features. We aggregate the network input features $\{\mathbf{h}_{1,i}, \mathbf{h}_{2,i}, \ldots, \mathbf{h}_{B,i}\}$ with attentions $\alpha_{b,i}$ by:

$$\mathbf{v}_k = \sum_{b=1}^{B} \alpha_{b,i}\mathbf{h}_{b,i} \tag{3}$$

and the attention weight $\alpha_{b,i}$ is calculated by:

$$\alpha_{b,i} = \frac{\exp\left(\sigma_n'\left(\mathbf{w}_\alpha^\mathsf{T}, \mathbf{z}_{b,i}\right)\right)}{\sum_b \exp\left(\sigma'\left(\mathbf{w}_\alpha^\mathsf{T}, \mathbf{z}_{b,i}\right)\right)}, \qquad \mathbf{z}_{b,i} = tanh\left(\mathbf{W}_z\mathbf{h}_{b,i}\right) \tag{4}$$

where σ_n' is an designable activation function $\sigma_n' = \mathbf{w}_\alpha^\mathsf{T}\mathbf{z}_\alpha$, and $\mathbf{w}_\alpha$ and $\mathbf{W}_z$ are learnable parameters.

The introduced system above is programmed under the PyTorch framework with packages including numpy 1.20.3, torch 1.8.1 and torchvision 0.3.0. We conduct comprehensive experiments on a synthetic dataset (MINI-RGBD) [17] with annotation provided by [5] and RVI-38 dataset [9], the result shows our system achieves state-of-the-art performance on both datasets [13].

Regarding computational costs, although we have applied the attention mechanism, our low-cost system can achieve a training speed of around 4 frames/second with an NVIDIA GeForce RTX 3080, with only 0.3 frames/second drop compared with our variant without the attention module. It means the total system training time on a 12 video sequences (1000 frames each) dataset is only about 50 minutes, including OpenPose pose estimation. This is considered to be very efficient in deep learning algorithms. During run-time, it only needs about 45s for the CP classification of a 33.3s 30FPS video with an Intel Core i7 CPU (i.e., no GPU needed). It shows that our system is employable for interactive-time prediction in a daily hospital environment with normal computer equipment setting, and is a feasible solution to support CP early prediction.

3. Impact overview

Our proposed system is expected to provide lower-risk patients with low-cost, non-intrusive CP abnormal movement classification results as a warning sign, with its accuracy supported by two datasets [13]. This provides a way for supporting the early prediction of CP in the clinical resource-limited areas, and relieving the labour stress of expert-based GMA [2]. In addition, our system can provide clinicians with information about the importance-ranking of joints' movement or frequency features in CP classification by the attention weights.

Apart from the early prediction of CP, we expect this system can be extended for a number of human motor disorders. For example, Parkinson's disease patients usually have motor dysfunction symptoms like tremors, unpaired gait patterns and limb stiffness [18]. The movement frequency of patients is highly likely to be different from that of healthy people, particularly due to tremors, and the use of frequency features may inform deep learning based Parkinson's disease diagnostic systems [19].

This easy-to-use system also benefits researchers in different fields like body motion and hand gesture analysis since it provides an alternative way for learning these features in the frequency domain. Recent work has shown that hand tremor can be modelled and reduced by deep learning as a denoising process [20], but the use of frequency is still yet to be considered. Similarly, frequency information is likely to be useful for sports motion analysis and visualisation [21]. In addition, the users can extend our system for multi-class classification tasks by minor modifications on the final fully connected layer. Existing work already shows that frequency information is useful for action recognition [22] and gesture recognition [23]. Our frequency binning idea and its uses on a graph neural network may inform such applications.

4. Conclusion and future improvements

In this work, we implement a Pytorch-based attention-informed graph convolutional network to classify cerebral palsy infants. We propose a frequency binning module for CP joint position features to increase to improve the classification performance, which is adaptable for other deep learning or traditional machine learning based classification models. Besides, we design an attention module to both improve the classification performance and interpret the joint features' importance ranking. Our system is validated on two datasets in [13] and achieves interactive speed on a consumer-grade computer.

In future versions, we consider integrating our system into an embedded system, such as the pixel processor array that integrates convolutional neural network operations [24], for autonomous disease monitoring and real-time decision-making. In addition, the performance of our system relies on the accuracy of the pose estimation process. Future versions will focus on improving the pose estimation part by adapting more advanced systems such as [25]. Finally, we wish to improve hardware compatibility by generalising the system with different cameras (e.g., unfocused camera, blurry camera). This can be implemented with a hardware abstraction layer that deals with device-specific procedures [26], which connects to a unified prediction algorithm.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Carter, C. V. Bennett, J. Bethel, H. M. Jones, T. Wang, and A. Kemp, "Identifying cerebral palsy from routinely-collected data in england and wales," *Clinical epidemiology*, vol. 11, pp. 457–468, 2019.

[2] C. Einspieler and H. F. R. Prechtl, "Prechtl's assessment of general movements: a diagnostic tool for the functional assessment of the young nervous system," *Mental retardation and developmental disabilities research reviews*, vol. 11, no. 1, p. 61—67, 2005.

[3] K. D. McCay, E. S. L. Ho, H. P. H. Shum, G. Fehringer, C. Marcroft, and N. D. Embleton, "Abnormal infant movements classification with deep learning on pose-based features," *IEEE Access*, vol. 8, pp. 51 582–51 592, 2020.

[4] D. Sakkos, K. D. Mccay, C. Marcroft, N. D. Embleton, S. Chattopadhyay, and E. S. L. Ho, "Identification of abnormal movements in infants: A deep neural network for body part-based prediction of cerebral palsy," *IEEE Access*, vol. 9, pp. 94 281–94 292, 2021.

[5] K. D. McCay, E. S. L. Ho, C. Marcroft, and N. D. Embleton, "Establishing pose based features using histograms for the detection of abnormal infant movements," in *IEEE EMBC*, 2019, pp. 5469–5472.

[6] B. Nguyen-Thai, V. Le, C. Morgan, N. Badawi, T. Tran, and S. Venkatesh, "A spatio-temporal attention-based model for infant movement assessment from videos," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2021.

[7] K. D. McCay, E. S. L. Ho, D. Sakkos, W. L. Woo, C. Marcroft, P. Dulson, and N. D. Embleton, "Towards explainable abnormal infant movements identification: A body-part based prediction and visualisation framework," in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2021, pp. 1–4.

[8] M. Zhu, Q. Men, E. S. L. Ho, H. Leung, and H. P. H. Shum, "Interpreting deep learning based cerebral palsy prediction with channel attention," in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2021, pp. 1–4.

[9] K. D. McCay, P. Hu, H. P. H. Shum, W. L. Woo, C. Marcroft, N. D. Embleton, A. Munteanu, and E. S. L. Ho, "A pose-based feature fusion and classification framework for the early prediction of cerebral palsy in infants," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 8–19, 2022.

[10] E. S. Ho, K. D. McCay, C. Marcroft, and N. D. Embleton, "PCPP: A matlab application for abnormal infant movement detection from video," *Software Impacts*, p. 100412, 2022.

[11] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *arXiv e-prints*, p. arXiv:1812.08008, 2018.

[12] H. Rahmati, H. Martens, O. M. Aamo, O. Stavdahl, R. Stoen, and L. Adde, "Frequency analysis and feature reduction method for prediction of cerebral palsy in young infants." *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 11, pp. 1225–1234, 2016.

[13] H. Zhang, H. P. H. Shum, and E. S. L. Ho, "Cerebral palsy prediction with frequency attention informed graph convolutional networks," in *Proceedings of the 2022 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, ser. EMBC '22. IEEE, 2022.

[14] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," *AAAI Conference on Artificial Intelligence*, 2018.

[15] L. Bluestein, "A linear filtering approach to the computation of discrete fourier transform," *IEEE Transactions on Audio and Electroacoustics*, vol. 18, no. 4, pp. 451–455, 1970.

[16] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *ICLR*, 2017.

[17] N. Hesse, C. Bodensteiner, M. Arens, U. G. Hofmann, R. Weinberger, and A. S. Schroeder, "Computer vision for medical infant motion analysis: State of the art and RGB-D data set," in *ECCV 2018 Workshops*, 2018.

[18] W. Poewe, K. Seppi, C. M. Tanner, G. M. Halliday, P. Brundin, J. Volkmann, A.-E. Schrag, and A. E. Lang, "Parkinson disease," *Nature Reviews Disease Primers*, vol. 3, 2017.

[19] H. Zhang, E. S. L. Ho, X. Zhang, and H. P. H. Shum, "Pose-based tremor classification for parkinson's disease diagnosis from video," in *Proceedings of the 2022 International Conference on Medical Image Computing and Computer Assisted Intervention*, ser. MICCAI '22, 2022.

[20] Z. Leng, J. Chen, H. P. H. Shum, F. W. B. Li, and X. Liang, "Stable hand pose estimation under tremor via graph neural network," in *Proceedings of the 2021 IEEE Conference on Virtual Reality and 3D User Interfaces*, ser. VR '21. IEEE, Mar 2021, pp. 226–234.

[21] Y. Shen, H. Wang, E. S. L. Ho, L. Yang, and H. P. H. Shum, "Posture-based and action-based graphs for boxing skill visualization," *Computers and Graphics*, vol. 69, no. Supplement C, pp. 104–115, 2017.

[22] G. Hu, B. Cui, and S. Yu, "Joint learning in the spatio-temporal and frequency domains for skeleton-based action recognition," *IEEE Transactions on Multimedia*, vol. 22, no. 9, pp. 2207–2220, 2020.

[23] R. Cutler and M. Turk, "View-based interpretation of real-time optical flow for gesture recognition," in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 416–421.

[24] L. Bose, P. Dudek, J. Chen, S. Carey, and W. Mayol-Cuevas, "A camera that cnns: Towards embedded neural networks on pixel processor arrays," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 1335–1344.

[25] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: Regional multi-person pose estimation," in *ICCV*, 2017.

[26] I. Abdelkader, Y. El-Sonbaty, and M. El-Habrouk, "Openmv: A python powered, extensible machine vision camera," *arXiv preprint arXiv:1711.10464*, 2017.