Project : Staying Connected: Personalising stroke recovery and rehabilitation using AI

Project's Primary Objective: The main goal of this project is to enhance the quality and productivity of current stroke recovery and rehabilitation by integrating technologies like accelerometry, computer vision, and mobile communication platforms. The focus of this project is specifically on upper limb rehabilitation and assessment methods.

Role of Computer Vision: Within the scope of this project, we work with videos depicting stroke patients with upper limb impairments. Our task involves utilizing computer vision techniques to evaluate these patients. We aim to discern finer details and deviations in their activity performance when compared to individuals without impairments. This process includes a longitudinal analysis of their progress throughout their therapy journey.

Currently, we are in the process of collecting clinical data, a task that is expected to take an additional 6-8 months to gather a substantial amount of data. Consequently, I have initiated work with existing datasets to conduct preliminary algorithmic developments pertaining to human activity recognition using computer vision.

Topic: Context-Aware Multimodal Network for Human Activity Recognition

Objective: The main objective of this model is to accurately classify given videos into their respective activity classes.

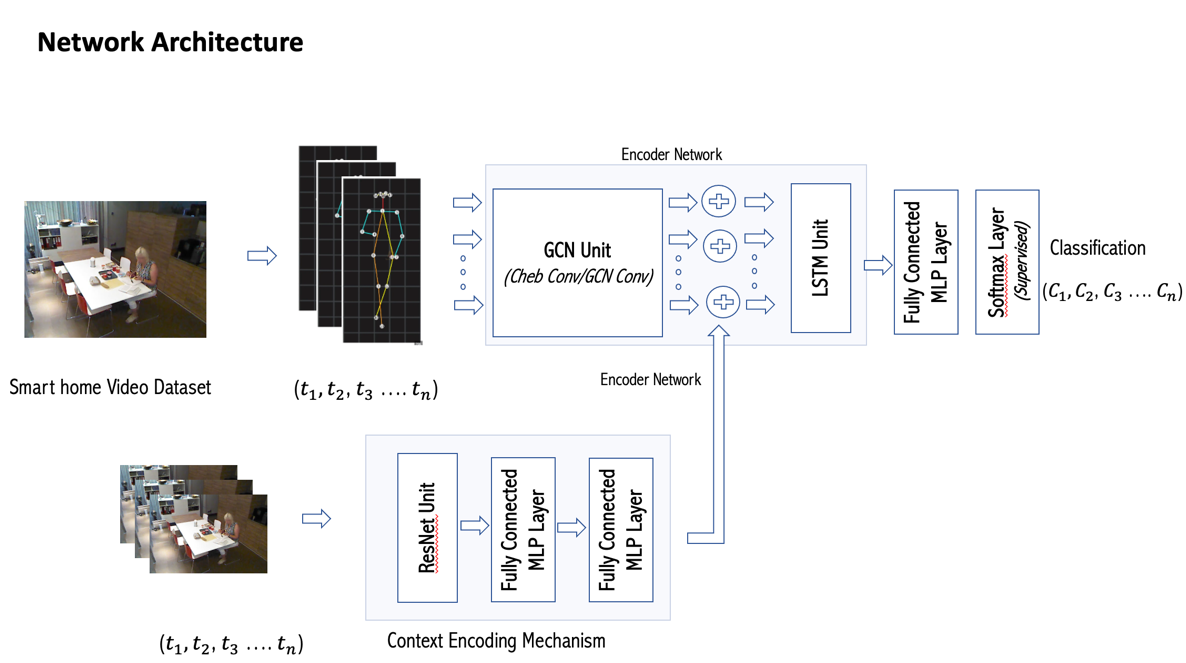
Dataset: Toyota Smarthome dataset (https://project.inria.fr/toyotasmarthome/)

This dataset comprises video clips pertaining to 31 distinct activity classes and offers three modalities: RGB video, depth, and 3D skeleton.

Proposed Approach: The core idea behind this approach is to leverage the skeleton structure of the human subjects present in the videos for analysis. As the human skeleton inherently follows a graph structure, we intend to employ Graph Neural Networks (GNN) for analysis. By utilizing GNN, we anticipate the capture of significant spatial features related to human movements. To address temporal aspects, we incorporate an LSTM module.

One observed drawback of this model is the potential loss of crucial context-related features from the background when focusing solely on the skeleton information. To account for this, we propose an additional module designed to extract each frame and generate ResNet encodings. These encodings will then be concatenated with the output from the Graph Convolutional layer. The combined output will subsequently be inputted into the LSTM module to capture temporal relationships.

Here is the overall model architecture.



I would greatly value your feedback regarding this approach. Are there any specific areas where you believe substantial modifications or enhancements are necessary? Alternatively, do you have any suggestions that could contribute to its improvement? Your insights would be highly valuable.

References :

Das, Srijan, et al. "Toyota smarthome: Real-world activities of daily living." Proceedings of the IEEE/CVF international conference on computer vision. 2019.

Shi, Lei, et al. "Two-stream adaptive graph convolutional networks for skeleton-based action recognition." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

Bruce, X. B., et al. "Mmnet: A model-based multimodal network for human action recognition in rgb-d videos." IEEE Transactions on Pattern Analysis and Machine Intelligence 45.3 (2022): 3522-3538.

Yan, Sijie, Yuanjun Xiong, and Dahua Lin. "Spatial temporal graph convolutional networks for skeleton-based action recognition." Proceedings of the AAAI conference on artificial intelligence. Vol. 32. No. 1. 2018.