ECE4094 Requirements Analysis

Project Title:

Automated Video-based Epilepsy Detection and Classification using Deep Learning

Student: Zhaoning Lu
Student ID: 29386993
Supervisor: Dr. Zongyuan Ge, Dr. Deval Mehta

Contents

Introduction

Epileptic seizure is very crucial, it may cause sudden death or other serious consequences. The traditional method to detect epileptic seizure attack is by Electroencephalography or wearable sensors. Those method makes inconvenience to patents every-day life. Also, because the seizure appearance is diverse for different patents, the clues such as countenance, pose movement has not been widely used in auto-detection. Given all these circumstances, the development of a robust deep learning frame work that can detect and classify epilepsy in real time will be imperative and meaningful. For the start of the project, Dr. Deval Mehta and I will collaborate to implement some environment settings, also some base ideas will be tested and put into practice, there will be more members joining the group as the project goes further and deeper.

Grading

The following grading table will be used as a marking reference for the student, which is both agreed by student and supervisors.

Grade	C	D	HD
Details	The student only understands the main point of the project topic. Developing a model which has some minor issues need to be fixed.	The student clearly understands the project topic. Developing a model which can be used to do epilepsy detection and classification.	The student well understands the project topic. Developing a model which can be used as a robust method to do epilepsy detection and classification.

Schedule & Milestones

There are two major milestones for this project:
1) The first stage of the project involves dataset collection, which is going to collaborate with Alfred Hospital. To be concrete, it needs set up the off-shelf models to run with "low time complexity" on Nvidia Jetson Xavier. Which involves implementation with TensorRT/ONNX, etc. And the models to be setup include - body, face, hand pose estimators and optical flow, etc.
2) The second milestone is to build machine learning / deep learning models for the motion data extracted from the Alfred hospital for epileptic seizure analysis. These will include training CNNs, LSTMs, GCNs, etc. and involve designing a combination / new model or a learning technique for learning from less data.

Schedule for milestones one	
Weeks	Tasks
Week1	Setup remote connection with the board; Read relative papers and materials.
Week2-3	Setup the python environment on the board; Build 'light-weight-human-pose-estimation' and 'pifpaf' on the board; Evaluate other potential methods.
Week4	Finish Requirement Analysis and Risk Analysis.
Week5-6	Build interface with TensorRT and optical flow; Do some test based on hospital environment, starting to collect sample data; Implement data saving method.
Week7-8	Decisions to be made.

Literature Review

Starting point and availability

Until now, there is a wide range of relative literature and research which are related to automated epilepsy detection and classification. Monitoring people with epilepsy is meaningful and necessary, this is well agreed by the society.

Facial expression (e.g., chewing, blinking), body joints movement, ictal head turning can be used as sign which potentially can be clues to detect and classify epilepsy. However, the study of these signs relies heavily on clinical experience and training. Automated analysis of semi logical patterns based on computer vision can support diagnosis by standard and objective assessment methods among evaluators, which is showed by prior work [1]. Given these circumstances, first aid will often be needed for patients who suffers seizures. However, current detection systems are always depending on sensors or wearable devices. Some patient groups such as children or people with intellectual disability may not tolerate wearable devices and may try to dislodge them. [2].

Epilepsy classification [3]

There are different types of epilepsy. And each of them has different symptoms. Here are the most common three types of them. There are also some unknown types for epilepsy, which always happens on a single patient or a small group of patients.

1. Generalized tonic-clonic: Patient will lose consciousness and is possible to fall. Then it follows generalized body stiffening. Patients may get injured when the seizure attacks. (e.g., may bit the tongue)
2. Absence: The patient will lose consciousness for a short amount of time (possibly a few seconds). For the most of the patient, they will have a feeling of “losing time”. Which means they didn’t aware the seizure has attack.
3. Myoclonic: Usually brief jerking will happen on both side of the body. Patients will have a feeling like being slightly shocked by electricity.

Deep Learning (DL) and Neural Networks

Deep learning is a subset of machine learning, most commonly, it is a neural network with multiple layers. By exploiting multiple layers with non-linearity added, processed by supervised or unsupervised learning it can do feature extraction, pattern analysis and classification. In short words, deep learning attempts to mimic the way human brain handles problems, as the name ‘neural networks’ suggested. With DL, we can implement a system to do classification or prediction with incredible accuracy and speed.

Convolutional Neural Networks (CNNs)

Convolutional Neural Networks is a class of neural network, which has a similar idea of the mathematical convolution operation. each module consists of a convolutional layer and a pooling layer. These modules are often stacked up with one on top of another. It is invented due to the restrictions of multilayer perceptions (MLP) (e.g., when vectorize the input function, information may lose. Also, MLP is hard to identify position transmission of a particular unit). Therefore, the Convolutional Neural Networks is widely used in field related to image analysing. By using techniques like pooling, padding, etc, the network can be trained to fit a specific task. Instead of manually manipulate and change parameters as we did in image processing algorithm traditionally, the Convolution Neural Networks use less pre-processing, it is an end-to-end learning process.

Long Short-Term Memory (LSTM)

LSTM is an architecture for recurrent neural network (RNN). The major difference of LSTM form traditional feedforward neural network is it has feedback connections. LSTM can process not only single data points (such as images), but also entire sequences of data (e.g., video).

Convolutional Pose Machines (CPMs)

CPMs approach is used by [4]. The CPMs is designed for tasks of pose estimation. It inherits the benefits of the pose machine architecture and combined with the advantages of convolutional architectures: the ability to learn feature representations for both image and spatial context directly from data; a differentiable architecture that allows for globally joint training with backpropagation; and the ability to efficiently handle large training datasets. CPMs consist of a sequence of convolutional networks, the output is learned end-to-end by backpropagation. At each stage in a CPM, image features and the belief maps produced by the previous stage are used as input. The belief maps provide the subsequent stage an expressive non-parametric encoding of the spatial uncertainty of location for each part, allowing the CPM to learn rich image-dependent spatial models of the relationships between parts [5].
In the specific task based on [4]: the joint position is labelled valid if the distance between the estimation is within a certain range when comparing with previous position. If not, the point will be rejected and the algorithm will search for the nearest neighbor within the area that satisfies the criteria, then it continuous the tracking to the next frame.

Head and upper limbs semiology

Recent deep learning networks for movement quantification and articulated pose estimation have shown remarkably robust performance and localization accuracy. There are some open-source libraries such as pifpaf, light-weight-human-pose-estimation could be used to trace the accurate position of people’s joints. From the experiment result based on [4]: The heat map shows the estimated joint positions for a sequence in fig 1.

Fig. 1. Heat maps of the joints detected in a sequence.

They have also used a method to consider whether a joint is detected: Only if the distance between the predicted location and the ground truth is within 8 pixels, the joint will be considered detected. From their result, the evaluation of the detected joints has an accuracy of 91%. And they found the most challenging part is the wrists and the elbows.

Design Prototype

The initial design for milestone one is based on 'lightweight-human-pose-estimation', which is an open source from GitHub. It detects a skeleton which consists of key points and connections between them to identify human body pose. There will be up to 18 key points included (eyes, nose, neck, shoulders, elbows, wrists, knees, etc.). The performance of the model was tested on a real sample video. From Fig2, one can see that the shoulders, elbows, wrists of the patient can be correctly detected. This screen shot was taken on local PC, the performance on the jetson board is about 12 FPS.

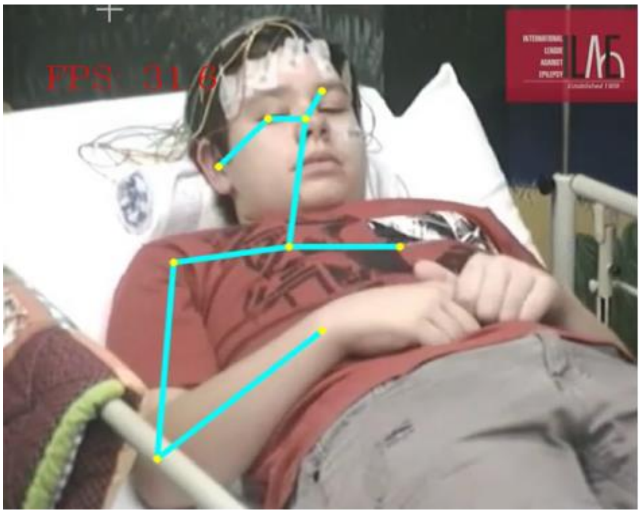

Fig 2. lightweight-human-pose-estimation tested

A data saving method was build based on the current model, which can extract the position of the joints split by frames and automatically save the data in pickle files under a nominated folder. Considering the amount of data will be huge in the future, we will also implement a way to save those pickle files directly to a Google Drive.

Potential Difficulties

Based on prior work [1][2][4][6], the difficulties may occur at following process:
1. Difficulty to extract robust features in the clinical environment, especially for elbows and wrists. For some videos, it is possible to extract only facial features due to occlusions from objects such as the blanket.
2. Lacking an integrated approach that can analyse facial and body motions simultaneously.
3. There is no public dataset for seizure semiology analysis, this makes it difficult to compare the methodology with other researcher's results. Due to ethical and legal considerations restrict the public release, researchers only use their own datasets.

Conclusion

The main objective of this project is to develop of a robust deep learning frame work that can detect and classify epilepsy in real time. The relative background and prior works have been discussed in this report. Also, the design prototype and the potential difficulties is shown. Overall, this project has high feasibility and will be meaningful.

References

1. David Ahmedt-Aristizabal, Clinton Fookes, et al. *Aberrant epileptic seizure identification: A computer vision perspective* Seizure: European Journal of Epilepsy 65 (2019) 65–71.
2. Evelien E. Geertsema, et al. *Automated video-based detection of nocturnal convulsive seizures in a residential care setting* Journal. 2018, 59(S1):53.
3. Cleveland Clinic *Epilepsy Types and their Symptoms* Medical website. Ava: link
4. David Ahmedt-Aristizabal, Kien Nguyen, et al. *Deep Motion Analysis for Epileptic Seizure Classification* IEEE publication. 2018, P2-3.
5. Shih-En Wei, Varun Ramakrishna, et al. *Convolutional Pose Machines* arXiv:1602.00134 2016, P1.
6. David Ahmedt-Aristizabal, Simon Denman, et al. *Understanding Patients' Behavior: Vision-Based Analysis of Seizure Disorders* IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 23, NO. 6, NOVEMBER 2019