ECE4094 Design Specification

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……………………………………………………………………………….2

Pose/ joints estimation theory……………………………………………………………..2

Optical flow theory………………………………………………………………………..3

Design pipeline……………………………………………………………………………3

Design process…………………………………………………………………………… 4

Dataset considerations……………………………………………………………………5

Hardware specification……………………………………………………………………5

Timeline…………………………………………………………………………………. 6

Risk analysis revisit………………………………………………………………………6

Reference………………………………………………………………………………… 7

Introduction:

The initial intention for this project is to develop, implement, and test a deep learning method for the detection and classification of epileptic seizures using video data. This project is in collaboration with the Alfred Hospital and supervised by Dr Zongyuan Ge and Dr Deval Mehta. The main theory behind the scenes, potential design alternatives, data set preparation, equipment requirements, timeline and project management will be detailed in this report.

Pose/ joints estimation theory:

Knowing the pose of a human often has significant meaning in many applications. Pose estimation is a general technic that is commonly related to Computer Vision and Deep Learning. It is a skeleton in a 2D format that represents the orientation of a human. Once the individual joints are extracted from a picture/video, the image coordinates will be saved along with the joints type. And depending on the joints type, one can decide whether a 'valid connection' has been made. The linkage between two validly connected joints is known as a limb. Pose estimation shows its strength in myriad fields since the invention. Some representative applications are activity recognition, motion tracking/capture.

Figure 1. Human Pose Estimation [1]

Nowadays, human pose estimation has already become a well-developed and highly reliable field. There exists a wide range of applications related, most of them can be found as open source. It gives extra hand to help us go deeper in many fields concerning human activity.

Figure 2. Application of Pose Estimation [2]

Optical flow theory:

Optical flow, commonly defined as an apparent motion of pixels, is a concept used for objects motion detection. It was firstly introduced in the 1940s by James J. Gibson. It is a measure for relative motion between viewpoint and the observed scene. One of the major applications of optical flow is motion estimation algorithm. With the ability to deduce the regions of movement and the velocity of motion, optical flow can handle most of the fields with respect to motion analysis.

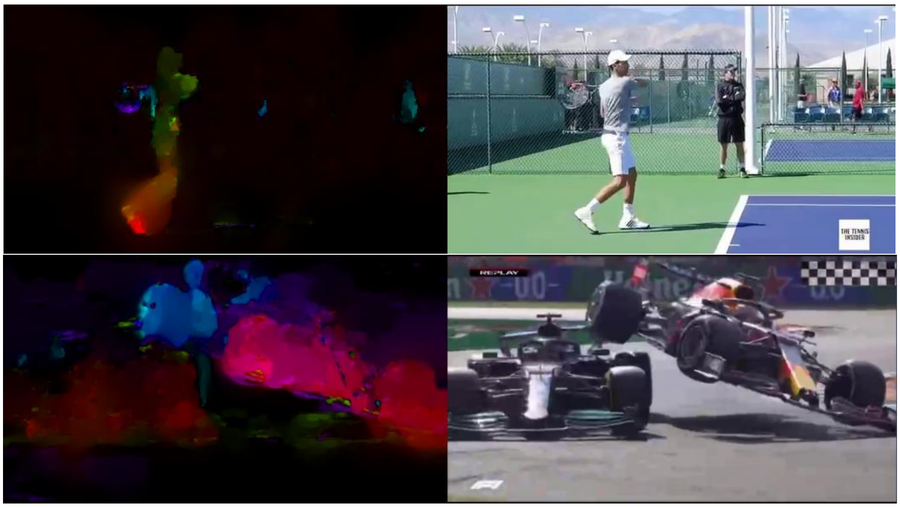

Figure 3. Examples of Optical Flow

Design pipeline:

Figure 4 illustrates the mechanism of the design pipeline from a higher hierarchy. The video data is on the bottom floor passed into pose estimation and optical flow software. The detection software is integrated with functions that can resize input videos, video jump-start, visualization display, and data set saving.

Once we have collected enormous data sets with high quality, they can be used as input to the deep learning model. To train a robust deep learning network, the model needs to be refined and tuned along with time. Time-consuming for this part can be expected to be huge. Finally, the application would be deployed to the clinical environment to detect and classify epilepsy seizures.

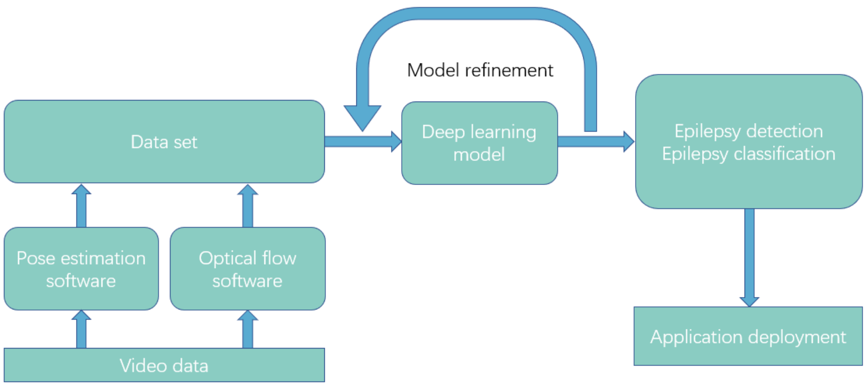

Figure 4. Pipeline of Design

Design process:

The project starts with the device/software setup. Due to the border restriction of Covid 19, I cannot attend the campus in person. The first design stage is going to be implemented on the Nvidia Jetson Xavier platform. In order to gain remote access to the device, we have built an ssh connection and set a VNC server. With VPN provided by Monash, the connection has high stability, the delay is about 2 seconds.

The joints movement of patients during an epileptic seizure strike is valuable to this project, it is looked upon as the main judging factor used in epilepsy classification. Therefore, the first stage of the project indeed was focus on implementing multiple pose estimation models on the device. Then make a comparison between them to find the optimal solution. Considering the comfortability of deployment in the clinical environment, the design is built on the Nvidia Jetson Xavier platform. The comparison can be roughly separated into two aspects: model accuracy and processing time.

The first pose estimation model implemented on the jetson device was 'lightweight-human-pose-estimation-3d' taken from GitHub. This repository performs human pose estimation in PyTorch and has an Nvidia TensorRT option for fast inference. There are 18 types of joints that can be detected by this model, which provides adequate information and satisfies our requirements. The detected joints will be reflected on 2D coordinates. The pre-trained model is available online with reasonable performance. One of the drawbacks of this model is the input video size need to be fixed when running with the TensorRT inference. This issue will be detailed in the later section. The processing performance of this model with TensorRT inference, tested on several sample videos on the jetson device, is about 17 frames per second. Under the same condition, the performance is about 13 frames per second without TensorRT inference.

Along with the 'lightweight-human-pose-estimation', I have also implemented retina face detector and dlib landmark estimator for the facial features collection. In the ideal condition, there will be 64 key points been detected on the face. These data have similar usage as the body joints.

Dr Deval first visits Alfred Hospital on 17/9 2021, the rudiment design has been tested in the real clinical environment. Based on the results of this visit, the findings are illustrated as follows:

1. Small objects (for instance bottles) sometimes are detected as part of the body, which causes the pose to become not continuous.
2. There is a lot of clutter in the background and the patients' position varies from different videos.
3. Video size could be different.
4. RetinaFace detection and landmarks estimation does not work as expected.
5. The lightweight pose information has expected accuracy, for half body detection it has fair enough performance.

Concerning the issues, we have made improvements and changes based on the early prototype. Firstly, Google MediaPipe has come into consideration. It is a pose estimation application provided by Google, which contains a holistic solution. According to the MediaPipe team, the human pose, hand landmarks and face landmarks will all be integrated into the holistic solution. They also provided several parameters to help with customization, for instance, there is a MIN_DETECTION_CONFIDENCE that can be used to tune the model.

For now, the overall performance of MediaPipe has not been tested on large video sets. But the detection accuracy and the detection consistency are still lower than expected based on the observation from sample videos. Major errors are many kinds of blocking distractors covers on the patients.
The video size problem is solved by normalizing input videos at the beginning of the program. Currently, the resolution is 640x480. A distortion problem brought should be considered in the future.

Dataset considerations:

Due to the specificity of the project, there is no open-source dataset available online. All the data will need to be collected in collaboration with the Alfred hospital. From the perspective of privacy, copying original videos is prohibited by the Alfred hospital. Therefore, the dataset will only contain joints coordinates/ optical transformation. The software is integrated with a section for saving pickle files and optical images, which automatically saves the legal information under folders based on the video name. Following this method, all the data is stored with the order, providing convenience for later work. Besides that, another consideration of using pickle files is to save disk space. A set of experiments have been made to test the skeleton reconstruction from pickle files.

Hardware Specification:

Local PC:
The local hardware has ability to handle/test most of the activity included in this project. However, the deep learning model may need to be trained on Monarch in later works.

Items	Technical Specification
GPU	Nvidia GeForce RTX 2070 with Max-Q Design
CPU	Intel Core i7-9750H
Memory	32GB

Table 1. Local Hardware Technical Specification

Nvidia Jetson Xavier:
In this project, we use the Nvidia Jetson Xavier as the platform. It has AI inferencing capabilities on edge devices. The Jetson device has excellent performance to handle tasks like visual odometry, mapping, object detection etc. Based on official data, the GPU performance is up to 32 TOPS of peak compute and 750 Gbps of high-speed I/O in a compact form factor. Besides the performance, another merit for the Jetson device is the flexibility. Not only it has a small size, but also the power profiles can be customized to a specific application.

Items	Technical Specification
GPU	512-core NVIDIA Volta™ GPU with 64 Tensor cores
CPU	8-core ARM® v8.2 64-bit CPU, 8MB L2 + 4MB L3
DL Accelerator	2x NVDLA
Memory	32GB 256-bit LPDDR4x | 137GB/s
Storage	32GB eMMC 5.1
Mechanical	105mm x 105mm x 65mm

Table 2. Nvidia Jetson Xavier Technical Specification

Timeline:

The timeline for the remaining part of this project is illustrated in table 3. Discussed with Dr Deval, we expect data collection will be officially started when we fix on a good model, for now, we have set up the lightweight-human-pose estimator and the MediaPipe pose estimator, but their performance should be able to further improve. The actual collection would not be time-consuming as the collection process is automated, I expect once it starts, the dataset can be built within one week. The training of the neural network will start at the beginning of the year 2022. Follows by experiments and tests on the model.

	2021-Nov	2021-Dec	2022-Jan	2022-Feb	2022-Mar	2022-Apr	2022-May	2022-Jun	2022-Jul
Repository set up									
Lwpose refine									
Mediapipe refine									
Other models									
Network traning									
Dataset collection									
Traning									
Tuning									
Validation									
Report writing									
Skeleton									
Finalise									
Video presentation									

Table 3. Timeline specification for the remaining project

Risk analysis revisit:

Currently, the risk factor of this project doesn't have a significant discrepancy with the initial version, however, it could be affected if the Covid-19 restriction is removed in the coming semester. The potential risk factor is still on electricity as the project involves activity on PC. The overall risk rating is still at a low level.

MONASH University

1. Activity or Task Based Risk Assessment [Ref Number: 31444]

Name	**ECSE_FYP_2021_S2_ZhaoningLu_Automated Video Based Epilepsy Detection and Classification Using Deep Learning**	**Current Rating**	**Residual Rating**
Location	**Off Campus: Jinan, Shandong, China** **The project will be completely done online in the home environment (due to covid border restriction I cannot go back to campus.)**	Low	
Business Unit		**Last Review Date**	**Risk Owner**
Dean Faculty of Engineering		14/08/2021	ZHAONING LU
Risk Assessment Team		**Risk Approver**	
Zhaoning Lu		Zongyuan Ge	

Additional Notes

Describe task / use

Using deep learning to detect and classify epilepsy.

The whole project is software-based design and will be finished fully online.
The major parts of the project have:
1. Extracting pose, facial, etc. information from patients.
2. Build the datasets based on the extracted information.
3. Train the Deep learning network.
4. Refine model.

Project Full Details
People with epilepsy need assistance and are at risk of sudden death when having convulsive seizures. Automated real-time seizure detection systems can help alert caregivers, but current automated methods for the detection of seizures rely on Electroencephalography and wearable sensors, which are not always tolerated by patients. Visual motion clues such as facial expression, hand joint stiffness, and body convolutions (pose) are natural semiology features which an epileptologist observes to detect epileptic seizures. Our goal is to breach this gap by developing a robust deep learning framework for classifying detecting and classifying seizures based on the video data.

Monash University

Reports identifying people are confidential documents.
Statistical information shall only be used for internal reporting purposes.

powered by riskware.com.au

Page 1 of 5

Table 4. Risk Analysis Revisit

Reference:

1. Ivan G, Valentin B. “MediaPipe Holistic — Simultaneous Face, Hand and Pose Prediction, on Device” Online. Google AI Blog. Dec 10, 2020.

2. Bharath R, Yoni O. “An Overview of Human Pose Estimation with Deep Learning” Online. Jun 2019.