Glasgow College, UESTC

Project Specifications and Preliminary Report on UESTC4006P(BEng) Final Year Project

Please start by saving this file with the name: GUID_Surname_UESTC4006P_Preliminary_year
**** *Please add appropriate course code***

Student Name	Zilai Wei
Student Matriculation Number	2429587W
UESTC Student Number	2018190504035
Degree programme	Bachelor
Academic year	2022

Placement Company (if appropriate)	
Working Title of Project	Deep person detection in video
Name of First Supervisor	Jin Qi
Name of Second Supervisor	Joao Ponciano
Declaration of Originality and Submission Information	*I affirm that this submission is all my own work in accordance with the University of Glasgow Regulations and the School of Engineering requirements* Signed (Student) : Zilai Wei

Your report should be NO more than 6 pages in length and include the below subject headings and incorporated within this document:

Project Description (no more than half a page)
Technical Background (no more than one page)
Main tasks and targets (no more than half a page)
Measureable Outcomes (no more than half a page): Tangible outcomes (Hardware, Software, Hardware & Software, Theoretical research)
Project outline (no more than one page)
Work plan (no more than one page)

Resources: Complete the component request form and email the form to your 1st supervisor separately
Rist Assessment Form: You may be asked to submit a Risk Assessment Form if your project includes hazardous activities. Please check with your supervisor if you need to submit it.
Ethical Consideration Form: You must submit an Ethical Consideration Form. Please have it signed with both your supervisors.

Deadlines for submission of this report: Please upload this report via the Moodle page by the deadline mentioned in Table 1 of your project handbook.

Comments from your Second Supervisor will be made via Moodle or via email.

Project Description:

Face recognition is a biometric technology for identity recognition based on human face feature information. The portrait recognition and face recognition technology is capturing images or video streams containing human faces with cameras, then automatically detecting and tracking the face in the image, and carring out face recognition on the detected face.

In this project, I will use deep learning and neural network algorithm to train the artificial intelligence to learn how to extract the human face feature in the vedio automaticly, and achieve the effect of real-time tracking the face in the video, which can label all faces in the video. If possible, I will try to train the artificial intelligence to remember some specifc people, and then recogonices them and labels their name in the tested vedio.

Technical Background:

Deep learning and has gradually became a hot topic and research area in recent years. Deep learning is a general term for a class of pattern analysis methods. It usually containes three types of methods: (1) The neural network system, usually based on convolution operation, such as convolution neural network (CNN). (2) The self-encoding neural network based on multi-layer neurons includes automatic encoding and sparse encoding, and has received extensive attention in recent years. (3) Pre-training the multi-layer self-encoding neural network, and then combine the identification information to further optimize the deep belief network (DBN) of the neural network weight.

Neural networks are generally used in deep learning. The structure of neural networks can be roughly divided into four parts, including convolutional layer, pooling layer, activation function and fully connected layer. Among them, the main purpose of the convolutional layer is to extract different features of the input. When a new image is given, CNN cannot know exactly which parts of the original image these features should match, so it will try every possible position in the original image, which is equivalent to turning this feature into With a filter, this process is a convolution operation[1]. For the pooling layer, the pooling area is usually 2x2 in size, and then converted to the corresponding value according to certain rules, such as taking the maximum value (max-pooling) or average value (mean-pooling) in the pooling area, and take this as the resulting pixel value. In addition, in the convolutional neural network, the activation function generally uses the ReLU function, which is characterized by fast convergence and simpler gradient calculation. In addition, the fully connected layer functions as a "classifier" in the entire neural network, that is, after convolution, activation function, pooling and other deep networks[2], the results are classified and recognized through the fully connected layer.

The research of face recognition technology generally began in the 1950s-1960s and continued in the 1980s. This is the early research stage of the development of face recognition. The vigorous development of face recognition technology mainly appeared in the 1990s. Although the overall duration is not very long, face recognition technology has made a substantial breakthrough. Since the 21st century, face recognition technology has been widely used in many subdivided industries and fields, which greatly facilitates our daily work and life.

The wide application of face recognition technology mainly benefits from the following aspects: First, the improvement of face image modeling methods, including nonlinear modeling methods represented by the Kernel method, and linear modeling represented by linear discriminant analysis Modeling methods and 3D face modeling methods based on 3D information, including several different face space model construction methods; Second, more in-depth research on factors affecting face recognition accuracy and recognition speed, including changes in ambient light, changes in human posture, and changes in facial expressions; The third is the research and analysis of data sources of facial images. Previously, it was mainly for image recognition under normal conditions, but now it is based on face recognition in video and face recognition based on sketches and near-infrared images (where the wide use of near-infrared image technology can effectively detect whether a face image is a living body); The fourth is the study of feature extraction methods for face images Analysis, the current prevailing methods mainly include face recognition methods based on geometric features, face recognition methods based on subspace, face recognition methods based on statistics, and face recognition methods based on neural networks.

Main tasks and targets:

The main tasks contained in this project mainly fall into the parts below:

(1). Collect the dataset. It should contain several videos of different faces in different scenes, which could let the neural network extract more obvious facial features without overfitting.

(2). Extract every frame of the video, and preprocess the image, like resize the image, add gauss blur or other tools.

(3). Locate the face in the image. I will manually mark some faces in the picture, and let the computer automaticaly extract the features of the face. I will build a deep learning neural network model, such as CNN, to learn the feature of human face, and achieve the goal that can accurately locate the position of the face in the test video.

(4). If possible, I will try to use a classifier to learn the different features from some different specific people, and let the neural network recogonize them, that achieve the improvement from detection to recogonition.

In conclusion, I will build a neural network model to detect the face in a video fastly and accurately.

Measureable Outcomes:

The main measurement and evaluation criteria are mainly in three aspects: Precision, Recall and Speed.

Precision: It is a software outcomes. The precision is based on our prediction results, and it indicates how many of the samples whose predictions are positive are truly positive samples. The detector can also make mistakes and may consider other things to be human faces. Precision = the relative number of detected faces by the system / the total number of detected faces by the system. The fewer false detections, the better.

Recall: It is a software outcomes. The recall rate is for our original sample. It indicates how many positive examples in the sample are predicted correctly. The more the number of faces that the detector can detect, the better. Recall = the number of detected faces/the total number of faces in the image[3].

Speed: It is a hardware outcomes. The less time the detector takes to detect an image, the better, usually expressed in frame-per-second (FPS). Many detectors are that the smaller the image, the fewer faces in the image, the larger the smallest face detected, and the faster the detection speed. However, the speed of the detectors will be easily influenced by the hardware, like the model, main frequency and other indicators of the CPU or GPU, and we would better test some outcome of the GPU or CPU when running face detection, like the FPS and so on, and we will easily know the performance of this neural network.

Project outline:

Analysis

The project is mainly theoratic research and comparison under more specific situation. The data set will mainly based on the online public database, and some test set may collected by myself. Of course in the process of building the neural network model and training, due to the huge data set, some technical support will be given from some other students in the advisor's lab, like running the program on the school serve. Then I will acquire the compare the data and adjust parameters until achieving high correctness and accuracy in face dectecting.

Algorithms

Most of the algorithms are based on the research done by former scientists, their papers will give the brief introduction to the algorithms they used. There are two common algorithms:

1. Regression algorithm

Regression algorithm is a kind of algorithm that attempts to explore the relationship between variables by measuring the error. Regression algorithm is a powerful tool for statistical machine learning. Common regression algorithms include least squares, logistic regression, stepwise regression, multivariate adaptive regression spline and local scatter smoothing estimation.

2. Case-based algorithm

Case-based algorithm is often used to establish a model for decision-making problems. Such a model often selects a batch of sample data, and then compares the new data with the sample data according to some approximation, so as to find the best match in this way.

Specification&Verification

In this project, as is mentioned, there will be fairly less physical interaction. Most of the work will be done on the PC or the servers, only a few are connected with physical adjustment when the data collection devices are under construction. And all the verification will be done with the network models constructed from those papers.

Language&Software

The most often used software will be Pycharm, it is a powerful Python editor, a lot of programming will be done on this software, and the model training and testing will also be done on it. Sometimes I will also use Matlab to assist me in data processing. Before starting, a software Anaconda will play an important role in environmental configuration, such as Pytorch, it contains almost all tool installation packages I need. For the language, Python will be most commonly used, sometimes C language will also be used. Also some radar and its related software will be helpful when collecting the data. Office Word, Excel, and Power point will offer me the platform to show the outcome.

Background material

[1]. Shipu Xu, Yong Liu, Nondestructive detection of yellow peach quality parameters based on 3D-CNN and hyperspectral images, Conference Series, Volume 1682, 2020 International Conference on Machine Learning and Computer Application 11-13 September 2020, Shangri-La, China

[2]. Qi Ailing; Tian Ning, Fine-grained vehicle recognition method based on improved ResNet, 2020 2nd International Conference on Information Technology and Computer Application (ITCA)

[3]. Shiqing Fan, Ye Luo Deblurring, Processor for Motion-Blurred Faces Based on Generative Adversarial NetworksICDSP 2021: 2021 5th International Conference on Digital Signal ProcessingFebruary 2021 Pages 272–277

[4]. Xiaojun, L., et al., Feature Extraction and Fusion Using Deep Convolutional Neural Networks for Face Detection. Mathematical Problems in Engineering, 2017. 2017.

[5]. Huang, G.B., et al., Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments, in Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition. 2008: Marseille, France.

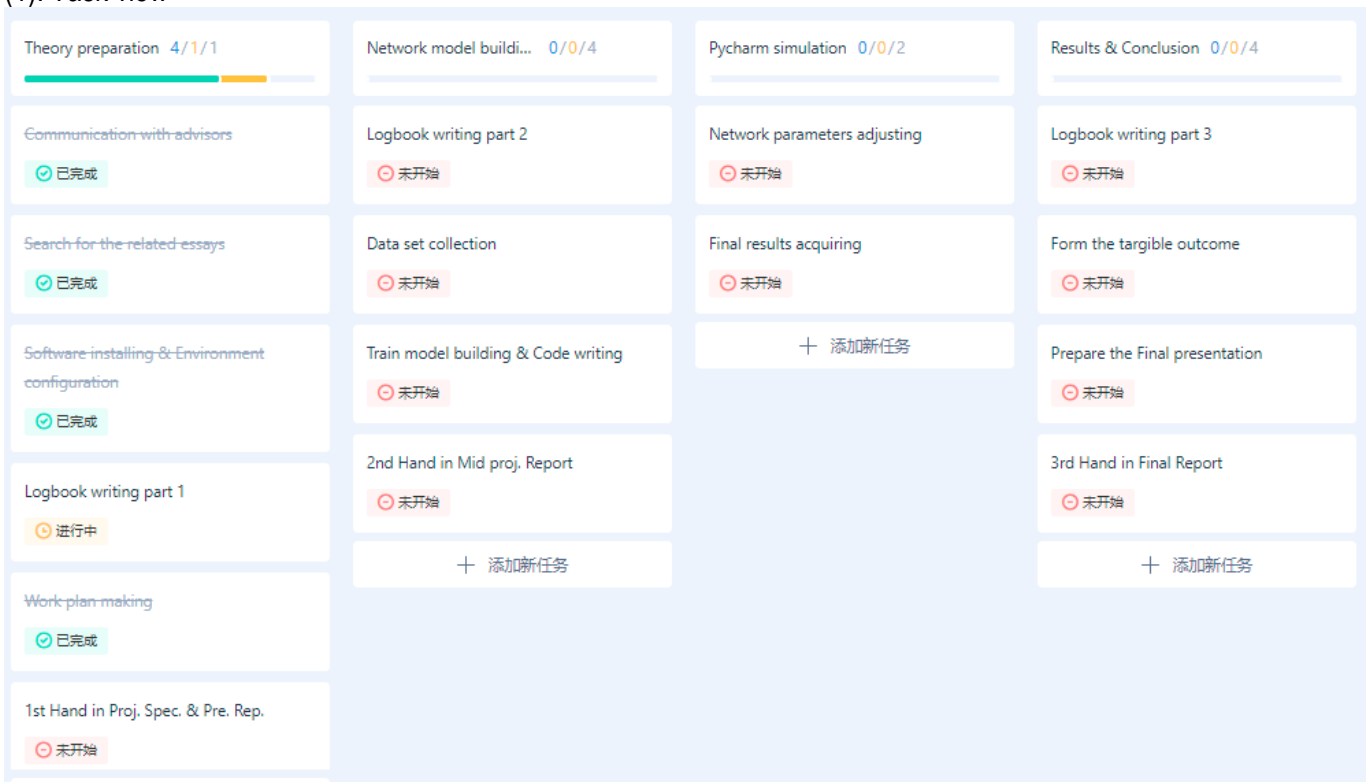
Work plan:

In this part, I used a software named *Worktile* to assist me to build the task plan and Gantt Chart. Using a software is more convenient and seens more clearly than hand painting

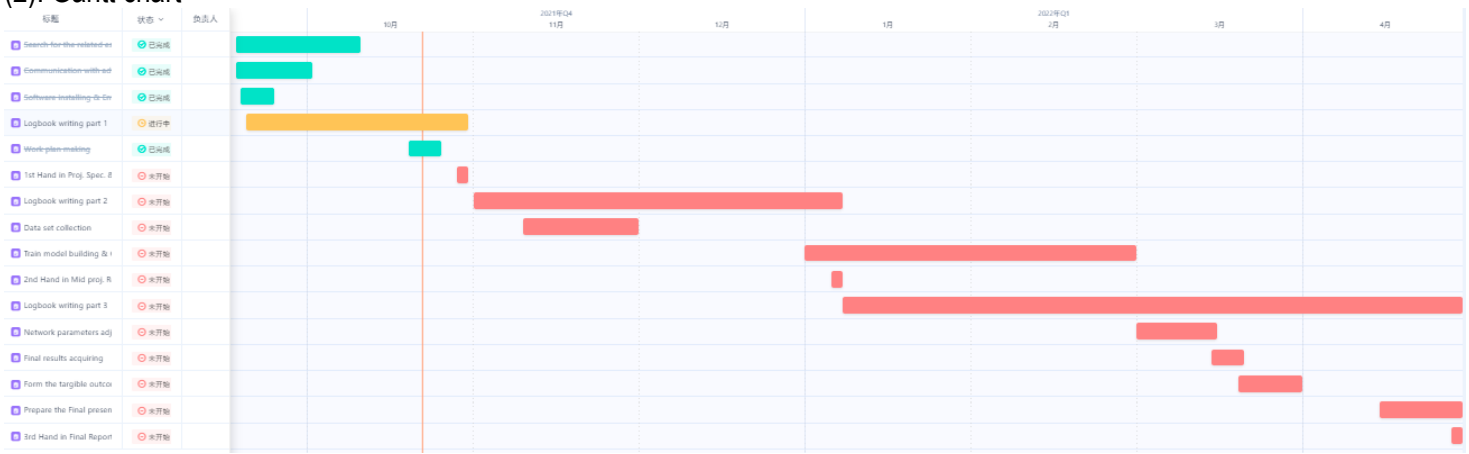
The work will be broken down to four big parts and some subtask are hidden in the chart but can be seen in the Gantt chart. The four big parts are *Theory preparation*, *Network model building*, *Pycharm simulation* and *Results & Conclusion*.

Both in the Task view and the Gantt chart, the colour Green means the task has already done, Yellow means it is in progress, and Red shows that I have not begin to do it yet. When the report is submitted, the task **1st Hand** in should be have done(in the colour of Green). The two diagrams is shown below:

(1). Task view



(2). Gantt chart



Project Specifications and Preliminary Report on UESTC4006P(BEng) Final Year Project

Feedback from Second Supervisors: Second supervisors may provide their feedback by adding comments directly on Moodle taking into account the questionnaire below **or** by filling out the below form and uploading it to Moodle.

Name of Second Supervisor	

Was the report satisfactory?	Yes		No	
Is project suitable for an BEng project?	Yes		No	
Is the project plan feasible?	Yes		No	
Would you like to give any suggestions/recommendations?	Yes		No	

Please write your comments in the space provided below:

Signature: **Date:**