

James Dyson Foundation Undergraduate Bursary

Supporting IIB projects in problem solving and design

As part of the application process for a James Dyson Foundation Undergraduate Bursary, students are required to complete all three parts of this application form and upload it using the online submission system. Completed application forms will be used by the Foundation to decide whether an award should be made.

Part I: Applicant information

Student name: Junxiao Shen

Student CRSID: js2283

Supervisor's name: Per Ola Kristensson

Supervisor's CRSID: pok21

Project code: C-POK21-1

Project title: a deep learning approach for gesture-to-text in Augmented Reality

Part II: Questions to applicants

Each question should be answered in no more than 250 words unless otherwise stated.

1) Why did you choose to study engineering at university?

Engineering is a subject which solves real-world problems. I have always wanted to solve real problems and then give back to society by inventing new technologies or by solving difficult technical problems. When I was in high school, I was the captain of the math competition team of my school. Engineering is, therefore, an area where I can utilize my math skills to fulfil my potential. Engineering course, especially in Oxbridge, is a general engineering course in the first two years and I truly think that having a mixed knowledge of different discipline can be very useful in terms of tackling problems in the 21st century. This is because nowadays, one problem can no longer be just put into one specific category in engineering and instead, it usually requires a team or a person that knows several engineering areas to solve. Therefore, I have chosen engineering and more specifically, engineering in Cambridge.

2) Where would you like to be in 10 years?

I would like to become an entrepreneur but meanwhile, I would like to stay closely connected to academia. This is not only because I always loved pure academics but also I think that in the future, one company needs

to have a barrier and that barrier is most likely to be technology barrier, Augmented reality is an industry that has great potential since it can not only be popular among consumers but also be of great importance in the industry such as design and construction. I would like to start a company that focus on Augmented Reality control systems such as text input and image segmentation. Augmented reality, nowadays, can be immature due to the constraints both on the hardware and software, However, I think that, which is also what I am doing for my fourth-year project, deep learning can break the constraints on the software and then can ignore the restrictions on the hardware to some extent. Therefore, as a conclusion, I would like to develop an efficient and accurate control system in AR in 10 years.

3) What invention do you wish you had thought of?

I wish I had thought of inventing the mouse. This is one of the greatest inventions in the 20th century. The mouse seems to be a very intuitive and obvious design to us now. However, it was not the case before it was invented since no one would know how a computer user interface looks like. The invention of the mouse has greatly improved people's efficiency in terms of the humane-computer-interaction. Furthermore, I think the ability to jump out of the box is a very strong skill that one should possess and I am trying to achieve that skill as well. This is because only a mind which can jump out the box can build disruptive technologies.

4) Who do you look up to?

According to Question 3, I am very interested in designing the next stage control system in Augmented Reality and it requires not only a solid mathematical and coding skills but also the ability to make the abstract thing into real. Therefore, I have been learning machine learning and its applications such as computer vision and speech recognition. This is to strengthen my skills for computing the data and the algorithm required in the control system of augmented reality.

5) What advice would you give a young person considering engineering as a career?

Choose what you like and focus on it, then not only you will enjoy it but also you will make something big because of being focused.

6) Engineering is... (in one sentence)

Engineering is a methodology to solve real-world problems.

7) My hidden talent is...

My hidden talent is creativity. In the user interface design, being creative is very important to the final results. I can make myself like a kernel within a Gaussian process to predict what people desire.

Part III: Project information

Please describe the proposed project, including (a) what the project will involve, (b) why you chose it (if relevant), (c) an indication of how the bursary would be used, and (d) an indication of what outreach materials will be produced (maximum 500 words). You are encouraged to add images to illustrate your ideas.

A deep learning approach for gesture-to-text in Augmented Reality

Junxiao Shen^{a,b}

^a*Department of Engineering, University of Cambridge, UK*

^b*js2283@cam.ac.uk*

Keyword: Human-Computer Interaction, Deep Learning, Machine Learning, Gesture Text Input, Augmented Reality

1. Research Goal

Due to the hardware inability to accurately sense the trace of the figure, there is a big displacement between the trace of the figure tips we see in the user interface in HoloLens and the actual data input of the trace. Therefore, a recognition model to map the trace to word or sentences should be invariant of the starting position of the keyboard. Therefore, Deep learning can be used to train this recognition model. Innovated from speech recognition, since speech is a 1D temporal data and the trace is a temporal and spatial data, therefore, the network models and state-of-art algorithms for speech recognition can be used to innovate us to design and train the neural networks for the trace recognition.

2. Research Plan

Figure 1 shows the overall plan. it is a component diagram and it is used to visualise and help us to have a better idea of the working process.

2.1. Data Synthesis

We will thus use the data we collected from the users' experiment to synthesis the data together with the data for the gaze collected from the experiment since we also believe that the trace of the gaze can be relevant to the final output as well. And the synthesised data will be from used to train the neural network and the trained model will thus replace the recognition component in Figure 1. The reason to synthesis the data is that unlike speech recognition, the recorded labelled data can be accessed quite easily nowadays since people can pay others to read and label whereas in our project. The dataset is limited due to /the lack of funding to hire a lot of people to generate the traces and due to the short time frame considering it is an only one-year project.

2.2. Data Training

We will do another users experiment of the established system and test the results and compare them with the state-of-art text input system. The user experience will still be carried out in the OptiTrack system.

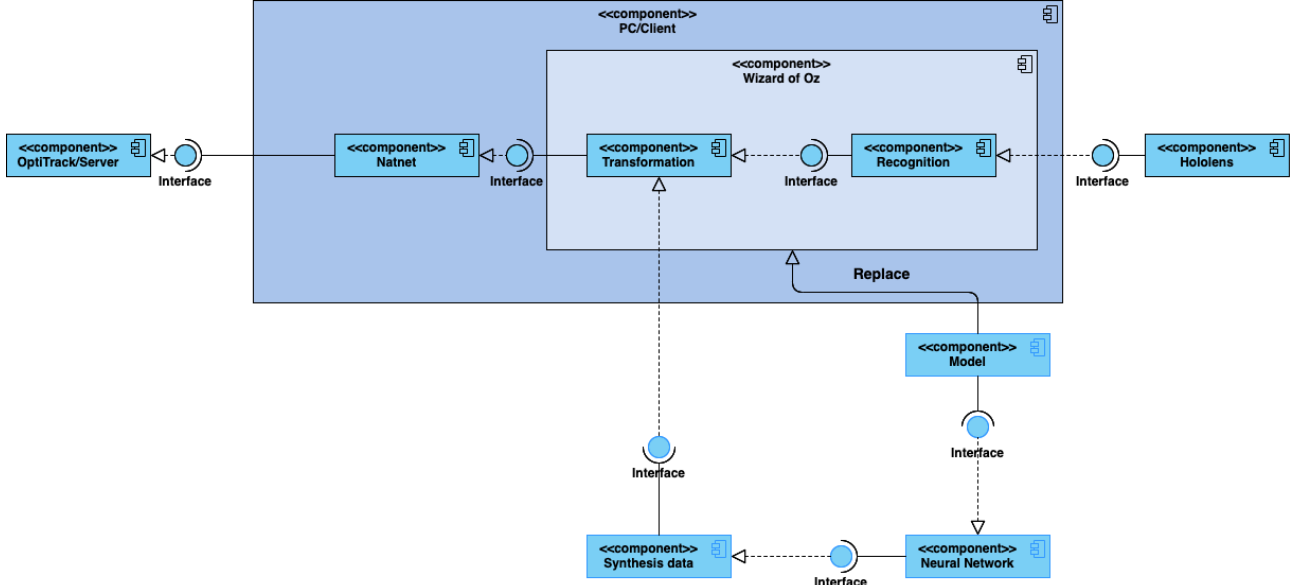


Figure 1: Component diagram for the development

Different neural networks can be used. I would start from Recurrent Neural Networks (RNN) first since the trace of fingertips is a 3d spatial and temporal signal and this can be in analogy to speech recognition. Furthermore, we could also try to combine CNN and RNN to see whether it will lead to better results. Transfer learning can also be used since the fine-tuning train ml model with synthetic data in first iterations and the with real data in further iterations. We should also be careful that we shouldn't use synthetic data in a test set since it can have a corresponding pair of real data which may lead to a dangerous result.

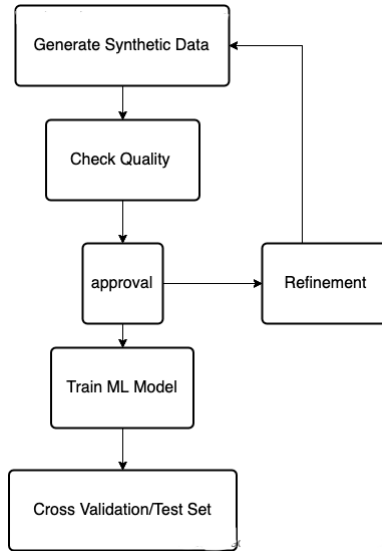


Figure 2: work flow for second and third stage

3. Cost

The majority of the cost here will be for the computational power used to generate the data and to train the deep neural networks. We are going to use Google Cloud to run the neural network and doing prototyping and large scale production code on this platform consistently can be very expensive. An alternative is to buy a strong GPU straight away and then the front cost is higher but the marginal costs is much lower. However, regardless of which way to use, 1 thousands pounds will be the starting point for this machine learning intensive project. Therefore the cost in our case is simple and straight away.