The Faculty of Engineering, The Chinese University of Hong Kong

The Faculty is committed to fostering the advancement of innovative technologies to benefit society. With strong support from the university, the Faculty has built up an inter-disciplinary research team since its inception in 1991. After two decades of development, we have developed core competencies in various engineering disciplines.

Computational LEGO Technic Design

We developed an AI programme to automatically design LEGO Technic models. Even for professional designers, the LEGO Technic design process usually requires several weeks or even months of trial and error. With our AI programme, users now just need to draw a sketch, then our programme can automatically design a beautiful and stable LEGO Technic model that resembles the sketch in less than one minute. Moreover, the programme will also generate the assembly instruction to assist users in the assembly. We invited several volunteers to draw sketches and our programme generated the LEGO Technic models for them (see picture). This project will be submitted to SIGGRAPH 2019 and also demonstrated in HK SciFest 2019.

AI System for Computer-aided Diagnosis

A research team led by Prof. Heng Pheng Ann, Department of Computer Science and Engineering, has developed an automated image processing technology using AI that is able to offer efficient and accurate diagnosis using CT scan and histopathological images. The technology has been tested on two of Hong Kong’s most prevalent cancers – lung cancer and breast cancer – achieving diagnostic accuracies of 91% and 99% respectively. The tests, which take between 30 seconds to 10 minutes, demonstrate that the technology not only boosts efficiency in clinical diagnosis, but also reduces misdiagnosis. The automated screening and analysis technology is expected to be widely adopted.

Automatic Recognition of Disordered Speech

Speech disorders such as dysarthria are commonly found among the elderly population, hindering their verbal communication with the outside world. Speech disorders not only introduce a negative impact on their quality of life but also increase the cost of care. As Hong Kong is ageing rapidly, the number of people with speech disorders will further increase. Due to the great mismatch between normal and disordered speech, state-of-the-art speech recognition systems designed for normal speech often produce very low recognition accuracy when applied to disordered speech.

Hence, there is a pressing need to develop new technologies that perform disordered speech recognition with high accuracy. Recent research conducted at the Faculty and department led to a highly accurate automatic speech recognition system that produced the best performance on disordered speech published to date. The CUHK research team is currently developing more advanced AI technologies to further enhance the performance of such systems and allow them to be more widely usable for multiple languages including Cantonese, English and Putonghua.

This research will allow easier and more fluent dialogue between Cantonese speakers with such disorders, in Hong Kong and the Greater Bay Area, and the outside world. It also forms a strong basis for future research on disordered speech recognition for Putonghua and other Chinese dialects to help a much larger number of similarly afflicted people in China.

Voice Conversion

Voice Conversion (VC) is a technique aiming at transforming one person’s speech such that it sounds as if it were spoken by another person. The Human-Computer Communications Lab (HCCCL) of CUHK recently proposed the Phonetic Posteriorgrams (PPGs)-based VC technique, which received the only Best Paper Award of ICMC 2016 and has since then been adopted by other research groups in the VC community. The proposed VC system first recognises the speech into a speaker-and-language-independent representation known as “Phonetic Posteriorgram”. Through speech synthesis this representation is then outputted as the ‘speech’ of the target speaker. The Automatic Speech Recogniser and the Speech Synthesiser are trained separately by deep learning approaches using a large amount of speech data from the target speaker and many other speakers.
Visual Question Answering
Prof. Li Hongsheng’s group has recently developed an AI program that can answer questions related to an input image. Such a challenging task is called Visual Question Answering (VQA). The AI program is required to understand and reason the contents of the image as well as the question to predict the correct answer. The developed AI program detects objects in the input image and infers the relationships between them. The cross-modal relations between objects in the image and words in the question are also automatically modeled by the AI program. The answers can be correctly predicted based on the cross-modal reasoning. The proposed AI program currently ranks 1st in the public VQA benchmark VQA 2.0.

Q: How many vehicles are in the street? A: 1
Q: Are there any people on the street? A: No
Q: Where will the bus go? A: School

Intelligent Vision Systems
Leading by Prof. LIN Dahua, the current AI research in the Multimedia Lab has three main themes. Firstly, by applying the technology of machine learning and deep learning, we aim to build an intelligent system that is able to extract useful information from our daily videos (bottom figure). For example, our system can automatically recognize the persons, objects as well as the mood and actions in videos. Secondly, our research endeavor also lies at the interaction of images and languages (top right). Human intelligence has a remarkable ability of using languages to describe the visual world; we would like to equip machines with a similar capacity. Lastly, we teach creativity and aesthetics to computers (top left). Our newly developed AI agent is able to create its own oil paintings.

Considering the Motion Speed and Orientation in 3D Human Action Recognition Using Two-stream CNN
Prof. LIU Yunhui’s group recently published paper about the 3D action recognition problem. Based on human skeleton data, the skeletal kinematics features are calculated first, i.e., linear velocity and orientation displacement, to capture the action variation along time. Then a novel image encoding method is introduced to encode the proposed kinematics features into images. Key frame selection scheme is also proposed to guarantee the same image input size. When training the CNN, the popular two-stream CNN architecture is adopted, incorporating spatial and temporal networks. The spatial ConvNet is trained on still RGB images, while the temporal ConvNet is trained on the proposed encoded kinematics features. This method is evaluated on a challenging multi-view dataset and the experiment results show that the proposed method is fast to train and outperforms many handcrafted features. (Published as conference paper “Kinematics Features for 3D Action Recognition Using Two-Stream CNN” on WCICA 2018)

A Robot for Baby-caring: Detecting the Dangerous Behaviors Conducted by Young Children in Daily Life
Prof. LIU Yunhu’s group recently developed a child caring robot, for detecting some dangerous behavior performed by child in the domestic environment based on the human action recognition and object recognition technologies. A human behavior is an interactive process between human and objects. Therefore, three factors need to be considered: the engaged objects, human actions and the relationship between human and the engaged objects. For the human action recognition, a new motion encoding is introduced and a convolutional neural network is utilized. Evaluation on the Northwestern-UCLA dataset verified the effectiveness of this method when action categories are small. The proposed action recognition method is simple and efficient, which is crucial for online behavior detection. Extensive experiments in the real physical world for detecting the behavior of eating allergic fruit and touching/playing electrical socket have achieved good performance. (Published as Conference paper “A child caring robot for the dangerous behavior detection based on the object recognition and human action recognition” on IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND BIOMIMETICS (ROBIO) 2018)