

# Note of the introductory sequence of ST7

Xinjian OUYANG

1900273

## 1. Description and understanding of the topic

This sequence mainly focuses on the image and sound processing, which is currently a rapidly growing field. Thanks to the remarkable progress of artificial intelligence (AI) and the improvement of computing power, it is no longer difficult to perform some complicated operations on images and sounds with the help of deep learning and big data.

From the academic aspect, the image processing, more precisely, the digital image processing (DIP), is the use of a digital computer to process digital images through an algorithm. The operations on images include segmentation, noise filtering etc. And the algorithms include Split & Merge, Watershed, Region growing, etc, which are discussed during the class.

Sound processing is often related to digital signal processing (DSP), since DSP is the common tool to analyze and operate audio signals. For instance, we can analyze an audio by its spectrogram or get the frequency representation of an audio through STFT(short time Fourier transform ). In other words, sound processing is the concrete application of DSP in audios.

Moreover, in recent years, the technologies of image and sound processing has given birth to several emerging disciplines , such as facial recognition, automatic drive, and speech recognition etc. These new applications change our life significantly as well as arise the public concern about data privacy.

## 2. Economic, social, industrial, current and future issues / challenges

Generally speaking, the progress of image and sound processing technology boosts economic growth, contributes to a better life , but brings some social issues as well.

On the one hand, the sound processing technology has been deployed into automatic speech recognition systems, such as in-car voice assistant and Siri in Apple devices. What's more, people with disabilities, especially those who are deaf or hard of hearing, can benefit from speech recognition systems by transforming audio/voice to text. When it comes to the (digital) image processing, there are various applications such as facial/expression recognition and medical imaging diagnosis. A common example is to unlock our mobile phone using facial recognition.

The new applications we mentioned above have created or accelerated a larger consumer demand for some technical products. That is to say, these new applications enriched the functions and the user's experience of some technical products, such as mobile phones, cars

etc, which could improve their sales, stimulate the consummation and finally boost the economy of a state.

On the other hand, as the popularization of image and sound processing technology in our daily life, concerns have been raised over data privacy and personal data protection. A typical example is facial recognition. Since individuals have limited ability to avoid face recognition tracking unless they hide their faces, there is a great risk of personal pictures being captured and shared. The good news is that governments have been aware of this problem and start to formulate laws and regulations to protect personal data.

### 3. Major scientific and technological challenges

Despite the power applications of image and sound processing technology, there are some non negligible challenges we have to face.

Firstly, **accuracy is always imperfect**, especially in the speech recognition and picture classification domains. Trustworthiness should be highly guaranteed in some occasions, such as self-driving technology, voice-to-text translation. However, background noise and loud environments may decrease the accuracy dramatically.

Furthermore, as mentioned in the introductory courses, **recognition of tones** in a language (such as Chinese) remains as an interesting challenge for researchers. Last but not least, the non-interpretability of parameters in neural networks might be a challenging issue in the near future. As we all know, AI is almost a “**black box**” model and most of its parameters are not interpretable, which means it is hard to give semantic interpretation to parameters of an AI neural network.

### 4. Key areas of innovation for engineers

As a rapid growing field, there are some critical areas of innovation in image and sound processing technology. The most interesting one is automatic driving technology in the background of 5G.

As the progress of digital image processing and 5G, automatic driving technology has made a lot of achievements theoretically and practically. However, there is a long way to go for the commercialization of this technology. For instance, it is not easy to determine the intentions of pedestrians, bicyclists, and animals, and models of behavior must be programmed into driving algorithms.

Moreover, considering the high safety of automatic driving in the road, a very accurate classification of objects is critical. It also should be noticed that expression recognition is helpful in determining and predicting the behaviors of pedestrians. In brief, how to improve the trustworthiness of object classification and the predictability of expression recognition, in chaotic inner-city environments, requires innovation and efforts of engineers.

In fact, there are lots of unsolved challenges which need the innovations of engineers, such as the mechanism of depression which is mentioned in the introductory section. In a word, image and sound processing is such a hot and promising discipline which is extremely powered by AI and 5G. However, there is a long way left to go.