# What are the applications using AI?

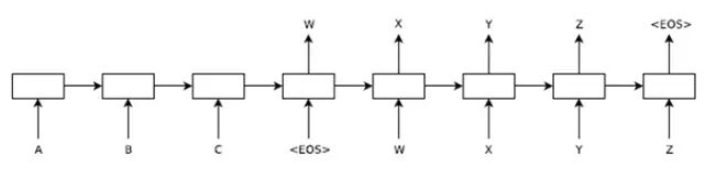
**Guo Zikun**

**2021225559**

1. **Neural Machine Translation**

Neural network machine translation is the use of neural networks to transform a language sequence into a target language sequence.

Encoder-Decoder:



The encoding-decoding framework is shown in the figure above, where the left side of <EOS>-W is the encoder, and the right side is the decoder. A, B, C, and <EOS> represent the source language input sequence, and X, Y, Z, and <EOS> represent the target language output sequence given by the translation machine. <EOS> represents the terminator of a sentence. W is the encoding vector of the input language sequence A, B, C, and <EOS> by the encoder. Each box in the figure represents a RNN (Recurrent Neural Network or LSTM) neural network that is unfolded at all times.

This encoding-decoding structure simulates the translation process of the human brain, which is to store the heard language in the brain first, and then give the output of the target language based on the understanding in the brain. Here, the W vector simulates the vector corresponding to the read source language stored in the brain.

This architecture also unites language understanding and language models, and finally realizes end-to-end machine translation. In addition, this coding-decoding structure is extremely flexible, and it can be applied to tasks such as image labeling, video, and words. In addition, this architecture can also be well integrated with external corpus and has good scalability.

1. **Face recognition**

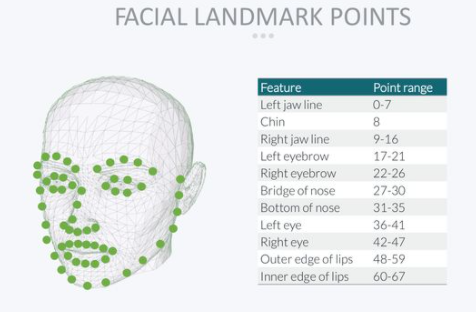
**Face Detection**

The histogram of orientation gradients (HOG) can be used to detect the position of the face. First, the picture is grayed out, because the color has no obvious effect on finding the position of the face, and then the gradient of each pixel in the image is calculated.

**Face alignment**

The face in a picture may be slanted, or just a side face. In order to facilitate the coding of the human face, the human face needs to be aligned into the same standard shape.

The first step in face alignment is to estimate that the face is a feature point. Dlib has special functions and models, which can realize the positioning of 68 feature points of the face.



After finding the feature points, align each feature point (displace the eyes, mouth, etc. to the same position) through geometric transformation of the image (affine, rotation, zoom).

**Face encoding**

Train a neural network to generate a 128-dimensional prediction value from the input face image. The general process of training is: feed two different photos of the same person and another person's photos into the neural network, and continuously iteratively train, so that the predicted values of the two photos of the same person after encoding are close, and the predicted values of photos of different people are increased. Far. That is to reduce the distance within the class and increase the distance between the classes. The specific algorithm refers to facenet[3].

**Identification**

Put everyone's connections into the face database in advance, and use the above-mentioned neural network to encode them into 128 dimensions and save them. During recognition, after predicting the face as a 128-dimensional vector, it is compared with the data in the face database. There are many comparison methods. You can directly find the face with the smallest Euclidean distance within the threshold range, or train a terminal SVM or knn classifier to directly generate the person's code (identity).

1. **Speech Recognition**

**Speech recognition concept**

Voice recognition technology is to make smart devices understand human voices. It is a multidisciplinary science involving digital signal processing, artificial intelligence, linguistics, mathematical statistics, acoustics, emotions, and psychology. This technology can provide multiple applications such as automatic customer service, automatic voice translation, command control, and voice verification codes. In recent years, with the rise of artificial intelligence, speech recognition technology has made great breakthroughs in both theory and application. It has begun to move from the laboratory to the market, and has gradually entered our daily lives. Now voice recognition has been used in many fields, including voice recognition dictation, voice paging and answering platforms, independent advertising platforms, intelligent customer service.

**Principles of Speech Recognition**

The essence of speech recognition is a pattern recognition based on speech characteristic parameters, that is, through learning, the system can classify the input speech according to a certain pattern, and then find the best matching result according to the judgment criterion. At present, the principle of pattern matching has been applied to most speech recognition systems. Figure 1 is a block diagram of a speech recognition system based on the principle of pattern matching.

General pattern recognition includes basic modules such as preprocessing, feature extraction, and pattern matching. As shown in the figure, the input speech is preprocessed first. The preprocessing includes framing, windowing, and pre-emphasis. The second is feature extraction, so it is particularly important to select appropriate feature parameters. Commonly used characteristic parameters include: pitch period, formant, short-term average energy or amplitude, linear prediction coefficient (LPC), perceptually weighted prediction coefficient (PLP), short-term average zero-crossing rate, linear prediction cepstrum coefficient (LPCC), Autocorrelation function, Mel cepstrum coefficient (MFCC), wavelet transform coefficient, empirical mode decomposition coefficient (EMD), gamma pass filter coefficient (GFCC), etc. In actual recognition, a template is generated for the test speech according to the training process, and finally the recognition is performed according to the distortion judgment criterion. Commonly used distortion judgment criteria are Euclidean distance, covariance matrix and Bayes distance.