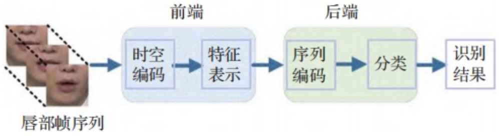
A Neural Network Accelerator for Lip Recognition Based on FPGA

INTRODUCTION

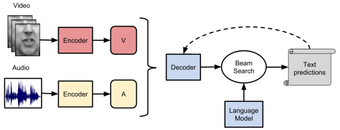
Lip recognition is a technology that integrates **computer version and natural language processing**, which can directly identify the content of speech from the image of someone speaking. Lip recognition transfer to new type since 2016 when **deep learning** technique was introduced to the filed. **Visual Speech Recognition (VSR),** is one of the most popular application direction. There are 2 main challenges for lip recognition: (1) when using independent frames for lip recognition, visual ambiguity is easy to cause due to insufficient use of lip information (2) the dataset of lip recognition is relatively scarce and cannot support the use of the entire language library. Therefore, it is difficult to unify the differences between regional language.



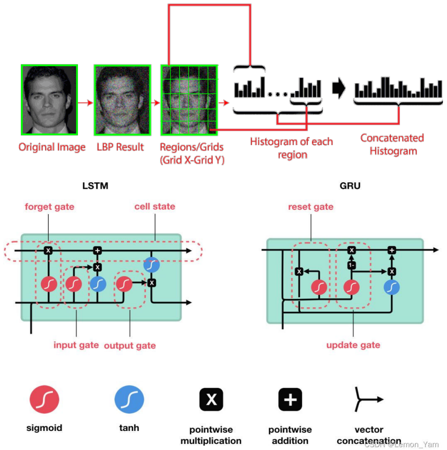
Process of visual speech recognition

CREATIVE DESING

Data flow in videos, audios, and the analysis system



Principles of HAAR and LSTM



PYNQ board, as the **sending end**, transmit the picture which captured by camera to the computer (**receiving end**), and display the picture in real time. **Haar Cascades** was used to extract the features of the tester's lip part, and the intercepted lip pictures were stitched into a long sequence of word pictures in the format of 36 28*28 . **LSTM neural network accelerator IP core**, which built in advance through **DMA,** was used to predict result.

Weilong Wu, Runkai Li, Hancheng Sun
Southeast University, Jiangsu Province

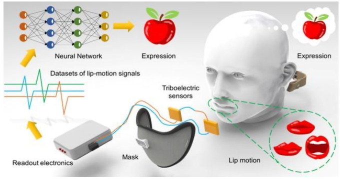
OpenHW2022

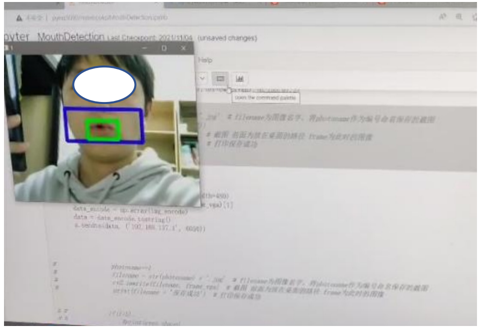
AMD XILINX

On board test by PYNQ-Z2

RESULT

Use Jupyter notebooks to drive the camera and lip-recognize the testers in real-time. The green boundary means the lip detected by cameras. In the future, we try to develop a portable device




pyter MouthDetection Last Checkpoint: 2021/11/04 (unsaved changes)
Help
Open the command palette

```
data_encode = np.array(img_encode)
data = data_encode.tostring()
s.sendto(data, ('192.168.137.1', 6000))
```

Demo of lip- recognition system and future design