MODEL CARD

Keyword Spotting

Model Details

This model has approximately 39k parameters. The predetermined keywords include "你好方方", "Hi 方方", “打開屏幕”, “關閉屏幕”, “關閉防窺”, “打開防窺”, “關閉節能” and “打開節能”. It can be applied for real-time inference, with an inference time of approximately 21ms on the VA8801 platform.

Model Specifications

Model Type : Convolutional Neural Network

Model Architecture : A modification network based on ResNet for VA8801

Input : 1*88*24

Output : 17 classes

Application

Smart speakers, mobile phones.

Application example: Devices is able to recognize the keywords and then executes the follow-up instructions.

Limitation

(1) Hard to detect keywords in the low SNR environment.

(2) The length of keywords can’t be larger than 2 seconds.

Training Data

你好方方: Recordings from the Fitipower staffs

Subsets : "你好方方", "Hi 方方", “打開屏幕”, “關閉屏幕”, “關閉防窺”, “打開防窺”, “關閉節能” and “打開節能”

Total : 548 clips

Due to the limitation of EDM microphone, some clips are very low-SNR or low energy. Careful pre-processing is needed before training models.

Reference

Convolutional Recurrent Neural Networks for Small-Footprint Keyword Spotting