MODEL CARD

Keyword Spotting

Model Details

This model has approximately 6.2 thousand parameters. The predetermined keywords include "Hi fiti", "貨物訊息", “barcode 辨識” and “貨物辨識”. It can be applied for real-time inference, with an inference time of approximately 225ms on the VA8801 platform.

Model Specifications

Model Type : Convolutional Neural Network

Model Architecture : A modification network based on ResNet for VA8801

Input : 1*192*40

Output : 16 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

Hi-Fit: Recordings from the Fitipower staffs

Subsets : Hi fi, 貨物訊息, barcode 辨識, 物件辨識,

Total : 553 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

基于 Pytorch 实现的声音分类系统