

Pitch detection with Machine Learning – Project proposal

118339238 Young-Shiuan Hsu

1. Overview

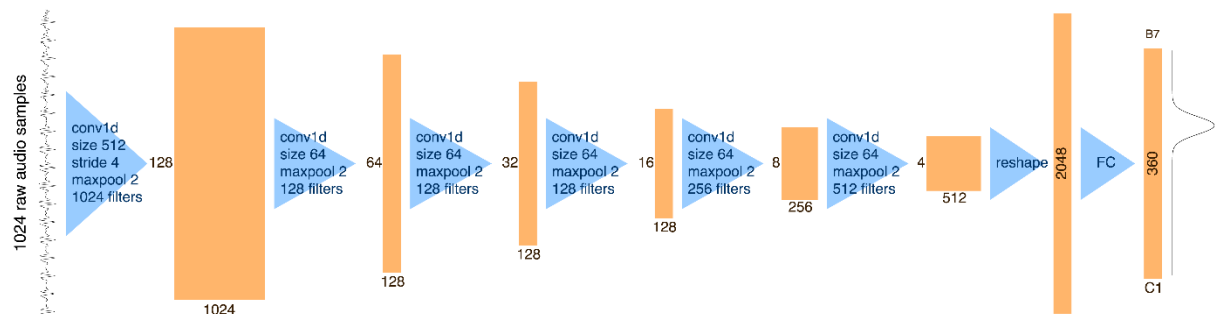
Pitch detection now has become a way for people to do speaker identification verification and has been used everywhere in our device now. There are various method and algorithms for pitch detection such as autocorrelation method, cepstrum and machine learning. In this project, I will be focusing on how machine learning models performs on pitch detection by changing the input of the model and also using different model.

2. System Design

Dataset to train the model: <https://www.kaggle.com/competitions/pitch/data>

Step1: Label the training dataset. With CREPE pitch tracker package in python, we are able to obtain the fundamental frequency of the training dataset.

The model of the CREPE pitch tracker:



Step2: Train the model.

The first model to train will be a 5-layer model (fully connected 1024->1024->2048->1024->360). The initial values are the same as CREPE model.

The second model I want to try is using transfer learning with VGG model. The input will be a 3*13 2D matrix which is the MFCC output.

The third model will be the substituting the first layer with a 39-node layer in the first model and see how it performs with feature extraction input.

Step3: Test data and compare with the results and accuracy of three different models.

3. Goals

We could see the accuracy and the output of the testing data with different data and also compare the result using feature extraction and raw data.

Insert my own audio into each model and do pitch detection for myself.

4. Expectations

The best performance model will be the VGG model. By feeding the feature extraction as an input will perform better than the model with raw data input.

5. Reference

<https://www.tensorflow.org/hub/tutorials/spice>

<https://pypi.org/project/crepe/>

[https://hal.archives-ouvertes.fr/hal-02439798/file/Fully-](https://hal.archives-ouvertes.fr/hal-02439798/file/Fully-Convolutional_Network_for_Pitch_Estimation_of_Speech_Signals.pdf)

[Convolutional_Network_for_Pitch_Estimation_of_Speech_Signals.pdf](https://hal.archives-ouvertes.fr/hal-02439798/file/Fully-Convolutional_Network_for_Pitch_Estimation_of_Speech_Signals.pdf)

<https://towardsdatascience.com/audio-classification-with-pre-trained-vgg-19-keras-bca55c2a0efe>

<https://www.arxiv-vanity.com/papers/1802.06182/>

<https://faroit.com/keras-docs/1.2.0/applications/>

<https://www.thepythoncode.com/article/gender-recognition-by-voice-using-tensorflow-in-python>