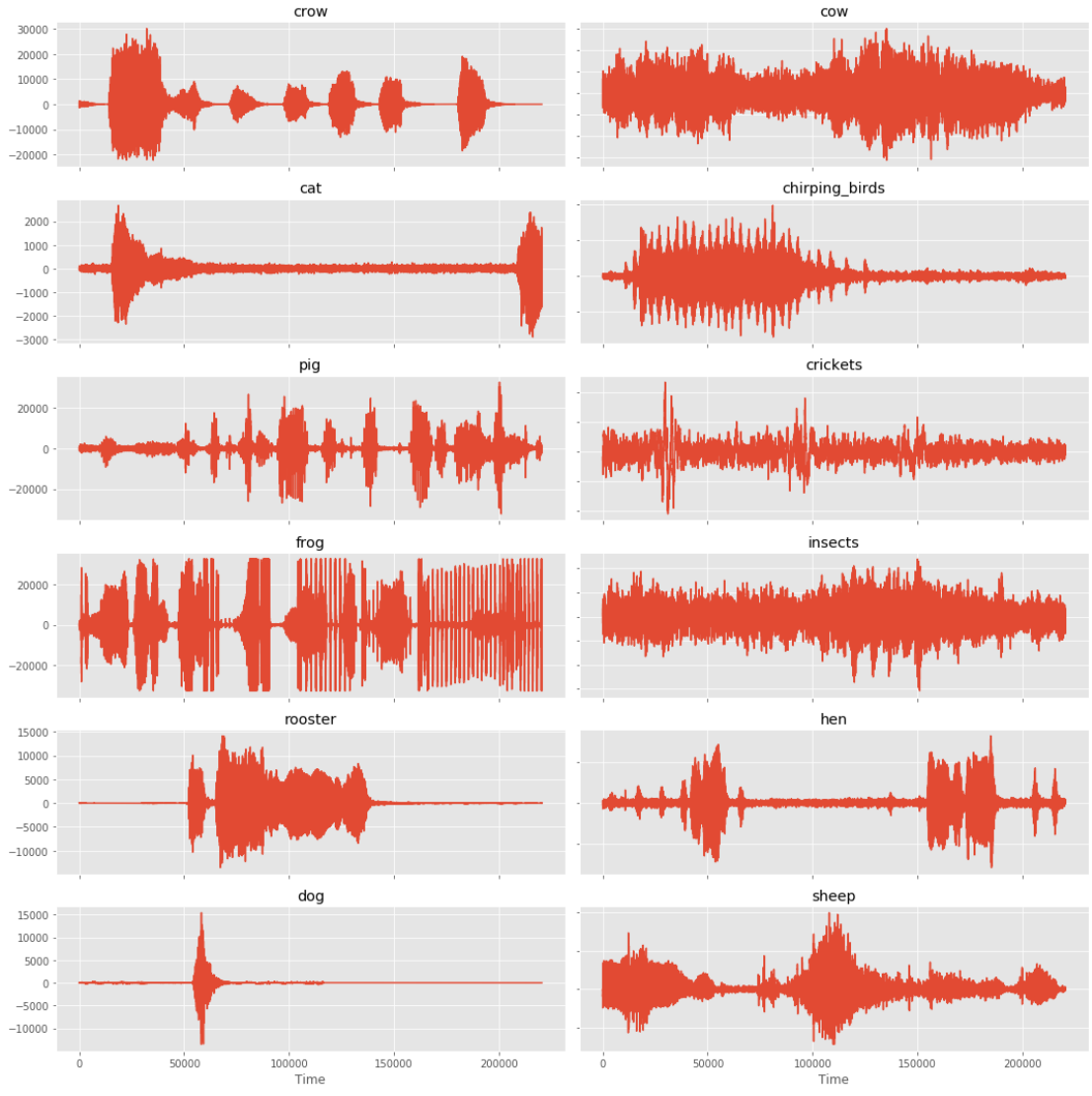
Sound Classification by Using Convolutional Neural Network



B34064061 賴宥瑜(Yo-Yu,Lai)

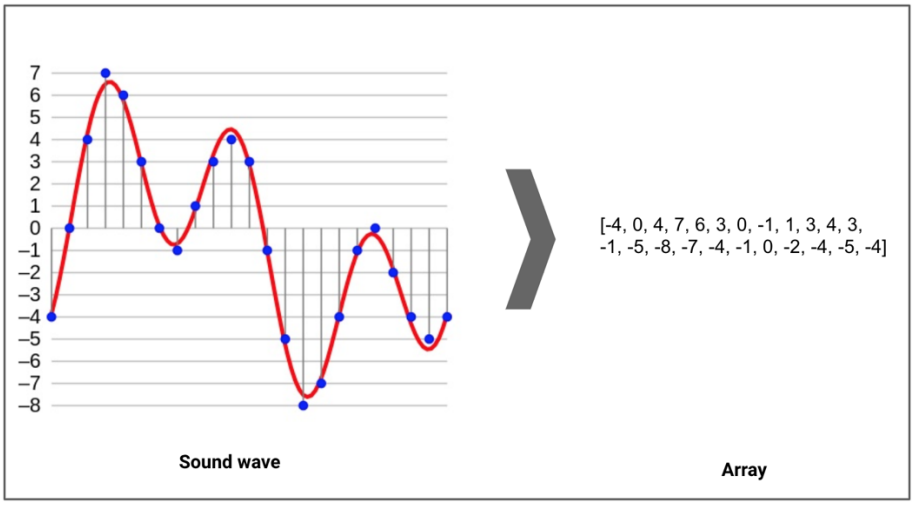
Motivation

Voiceprint recognition has been popular in recent years. Everyone has different kinds of voice, frequency and waveform. By voiceprint recognition system we can recognize people from different. It’s can be use on identify system. I use Convolution Neural Network trying to classify different kinds of animal, and this the try before I use CNN to identify human voice. I think it’s helpful for biology research, in wild by this system we can know what creature are in this ecosystem.

Experiment Process

using python enviroment

1.Data preprocess

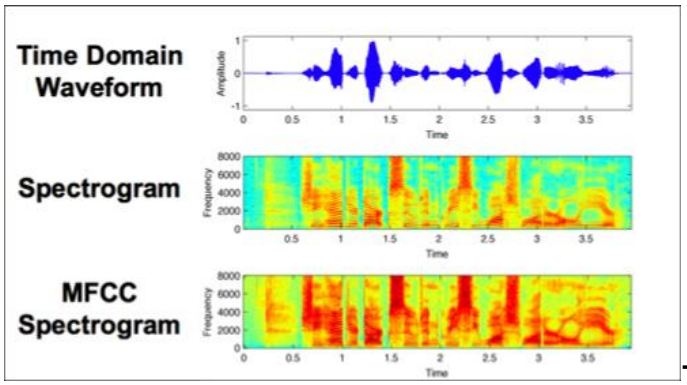


Continuous signal turn into discrete signal by sampling , after sampling we can represent signal by number ,then we can analyze it on the computer , if the sample rate is too low, the signal will be distortion. The data I used is from Kaggle opensource data. For the consistent of data I sampled in same rate and normalized the data.

Then we do Mel-frequency cepstral coefficients (MFCCs) signal transfer

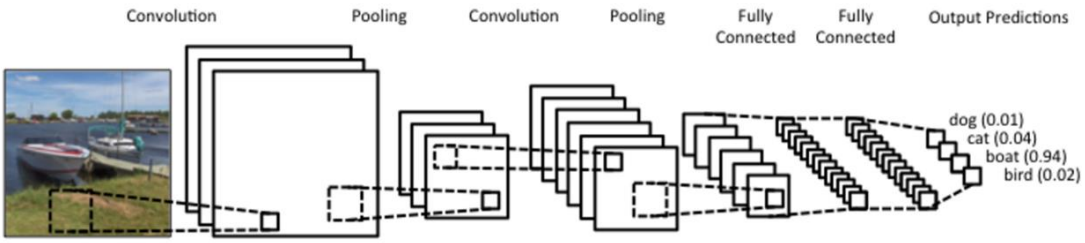
1. Take the Fourier transform of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows.
3. Take the logs of the powers at each of the mel frequencies.
4. Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the resulting spectrum.

By MFCCs we can avoid noise and compensate high frequency part. Compare to directly Fourier Transfer we get more detail about the signal



+

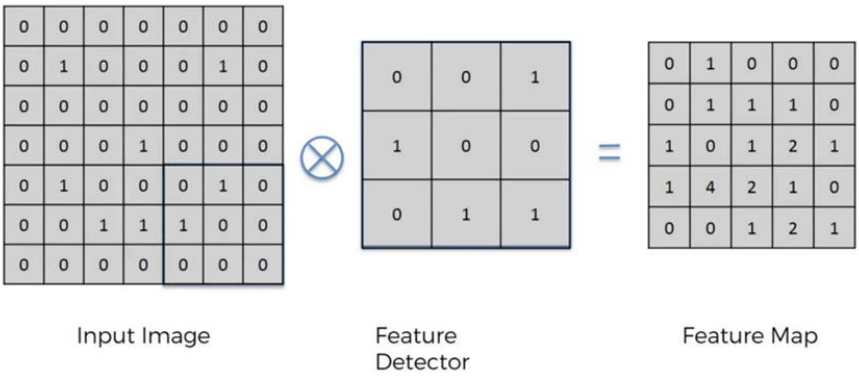
2.Build training model



Convolution Neural Network (CNN) is combined with Convolution Layer, Pooling Layer, Fully Connected Layer, and CNN is good at image recognition.

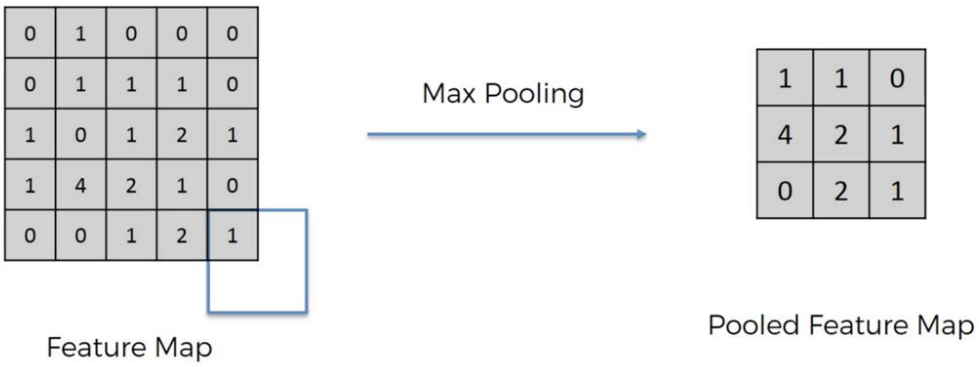
Convolution Layer:

extract the features from the image



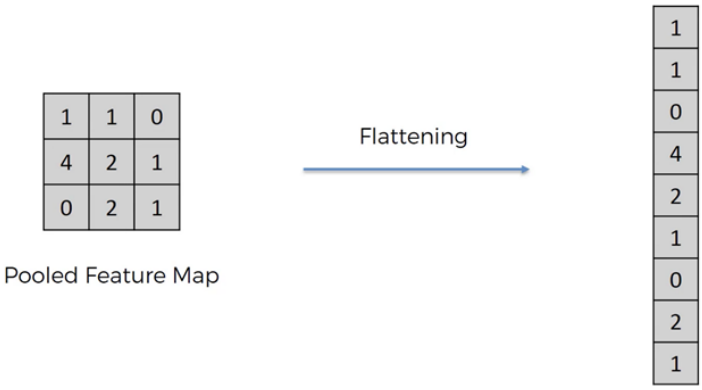
Pooling Layer:

maximum the value in the corner to avoid noise

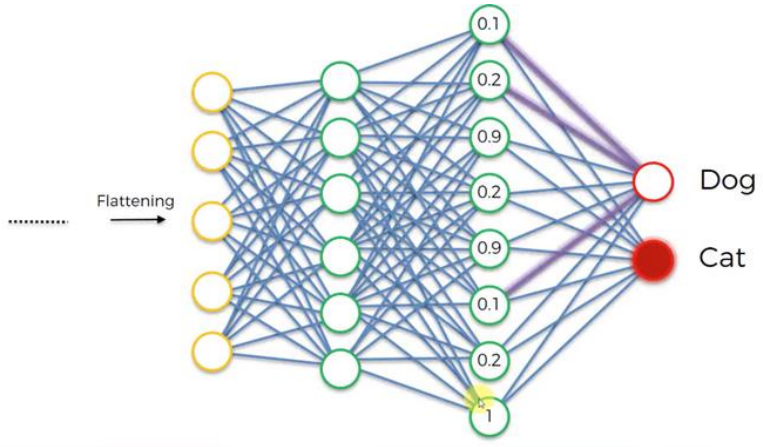


Fully Connected Layer:

flatten the result into neural network



neural network has a lots of nodes connect together, input value through activate function pass to next layer until able to predict the result



Experiment Result

There are 520 samples used in experiment are, 13 different kinds of animals ,every animal with 40 samples. There are cat, bird, cow, cricket, dog, frog, hen, insect, mouse, pig ,rooster, and sheep.

```
468/468 [==============================] - 1s 1ms/step - loss: 0.1222
Epoch 56/60
468/468 [==============================] - 1s 1ms/step - loss: 0.0691
Epoch 57/60
468/468 [==============================] - 1s 1ms/step - loss: 0.1618
Epoch 58/60
468/468 [==============================] - 1s 1ms/step - loss: 0.0835
Epoch 59/60
468/468 [==============================] - 1s 1ms/step - loss: 0.0625
Epoch 60/60
468/468 [==============================] - 1s 1ms/step - loss: 0.0719
52/52 [==============================] - 0s 154us/step
Test:
Loss: 0.3753053362552936
Accuracy: 0.9038461446762085
```

Randomly picks 468 samples for training, 52 for testing, after long train and modify. The final model structure using

4 convolution layer, 4 pooling layer, and 4 layer of neural network. And got 90.38% accuracy of classification animals in the experiment.

Conclusion

After MFCCs transfer we get better predict accuracy then directly Fourier Transfer. CNN again shows how powerful it can do on image recognition, and prove voiceprint recognition on CNN is possible. Machine Learning is a useful tool, after this project I am more skillful on it. In this experiment, I spent lots of time collecting data, and I realize how important to data preprocess, if you got a lots of data but useless, it may damage your training. Finally, I got 90% of accuracy, I think the data not more enough for higher accuracy.40 for each animal maybe come to the limit of model training. If I make the project next time, the data quality and quantity are important factor need to be consider.

Reference

Pandeya, Y. R., Kim, D., & Lee, J. (2018). Domestic cat sound classification using learned features from deep neural nets. *Applied Sciences*, *8*(10), 1949.

Nanni, L., Maguolo, G., & Paci, M. (2020). Data augmentation approaches for improving animal audio classification. *Ecological Informatics*, 101084.

UDACITY MACHINE LEARNING NANODEGREE 2018
Classifying Urban sounds using Deep Learning

Mel-frequency cepstrum- Wikipedia

Dataset

Environmental Sound Classification 50
raw audio classification of environmental sounds

URBANSOUND8K DATASET