
Tensorflow Speech Recognition

Speech recognition using google's tensorflow deep learning framework, sequence-to-sequence neural networks.

Replaces caffe-speech-recognition, see there for some background.

Update 2024: Use Whisper !

This (relatively) old project is NO LONGER UP TO DATE.

The tensorflow 1.0 used is not compatible anymore and the theory is no longer state of the art either.

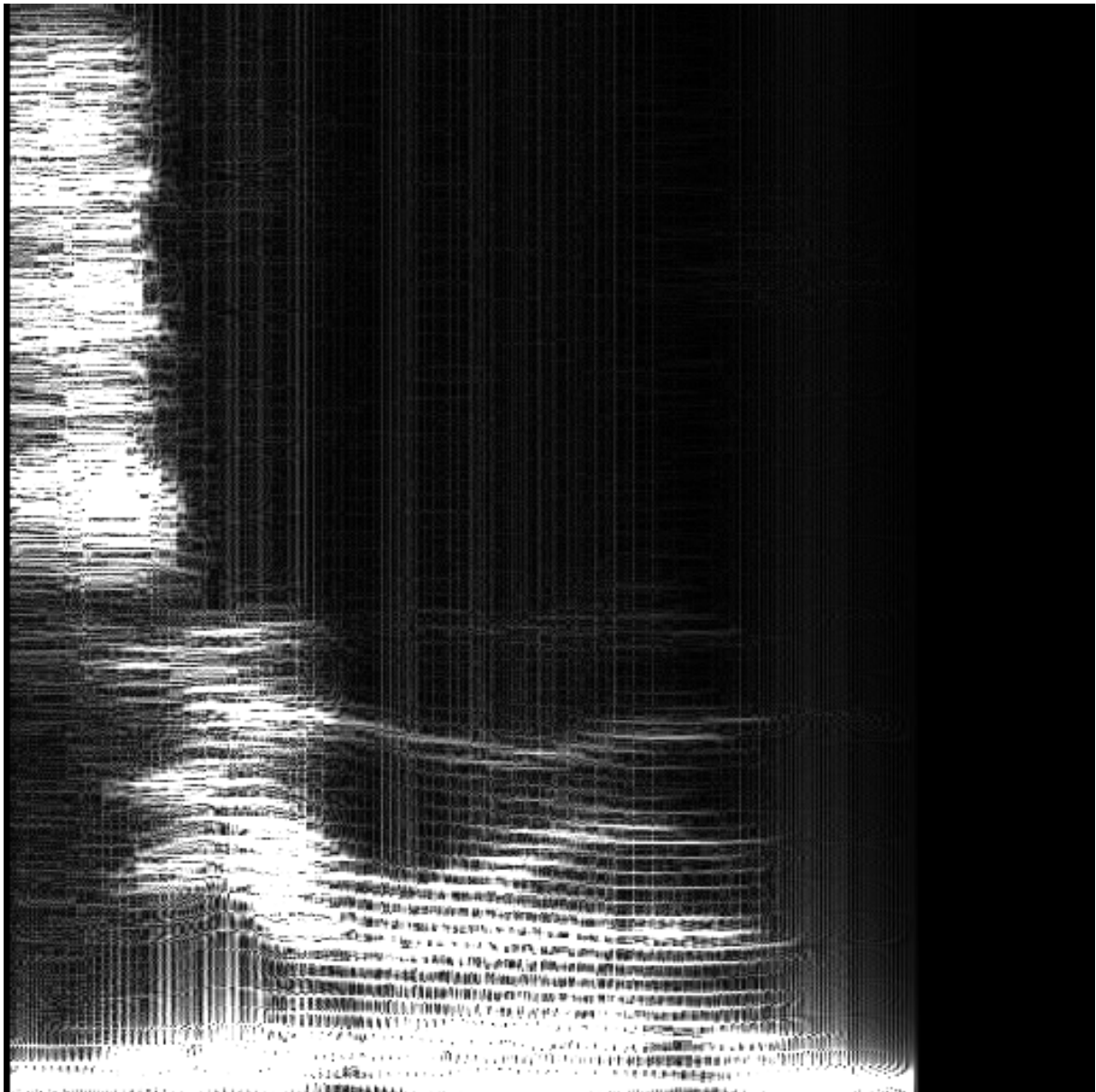
We highly recommend you check out and use whisper

Update 2020: Mozilla released DeepSpeech

They achieve good error rates. Free Speech is in good hands, go *there* if you are an end user. For now *this* project is only maintained for educational purposes.

Ultimate goal

Create a decent standalone speech recognition for Linux etc. Some people say we have the models but not enough training data. We disagree: There is plenty of training data (100GB here and 21GB here on openslr.org , synthetic Text to Speech snippets, Movies with transcripts, Gutenberg, YouTube with captions etc etc) we just need a simple yet powerful model. It's only a question of time...



Sample spectrogram, Karen uttering 'zero' with 160 words per minute. ## Installation ### clone code

```
1 git clone https://github.com/pannous/tensorflow-speech-recognition
2 cd tensorflow-speech-recognition
3 git clone https://github.com/pannous/layer.git
4 git clone https://github.com/pannous/tensorpeers.git
```

pyaudio

requirements portaudio from <http://www.portaudio.com/>

```
1 git clone https://git.assembla.com/portaudio.git
2 ./configure --prefix=/path/to/your/local
3 make
4 make install
5 export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/path/to/your/local/lib
6 export LIBRARY_PATH=$LIBRARY_PATH:/path/to/your/local/lib
7 export CPATH=$CPATH:/path/to/your/local/include
8 source ~/.bashrc
```

install pyaudio

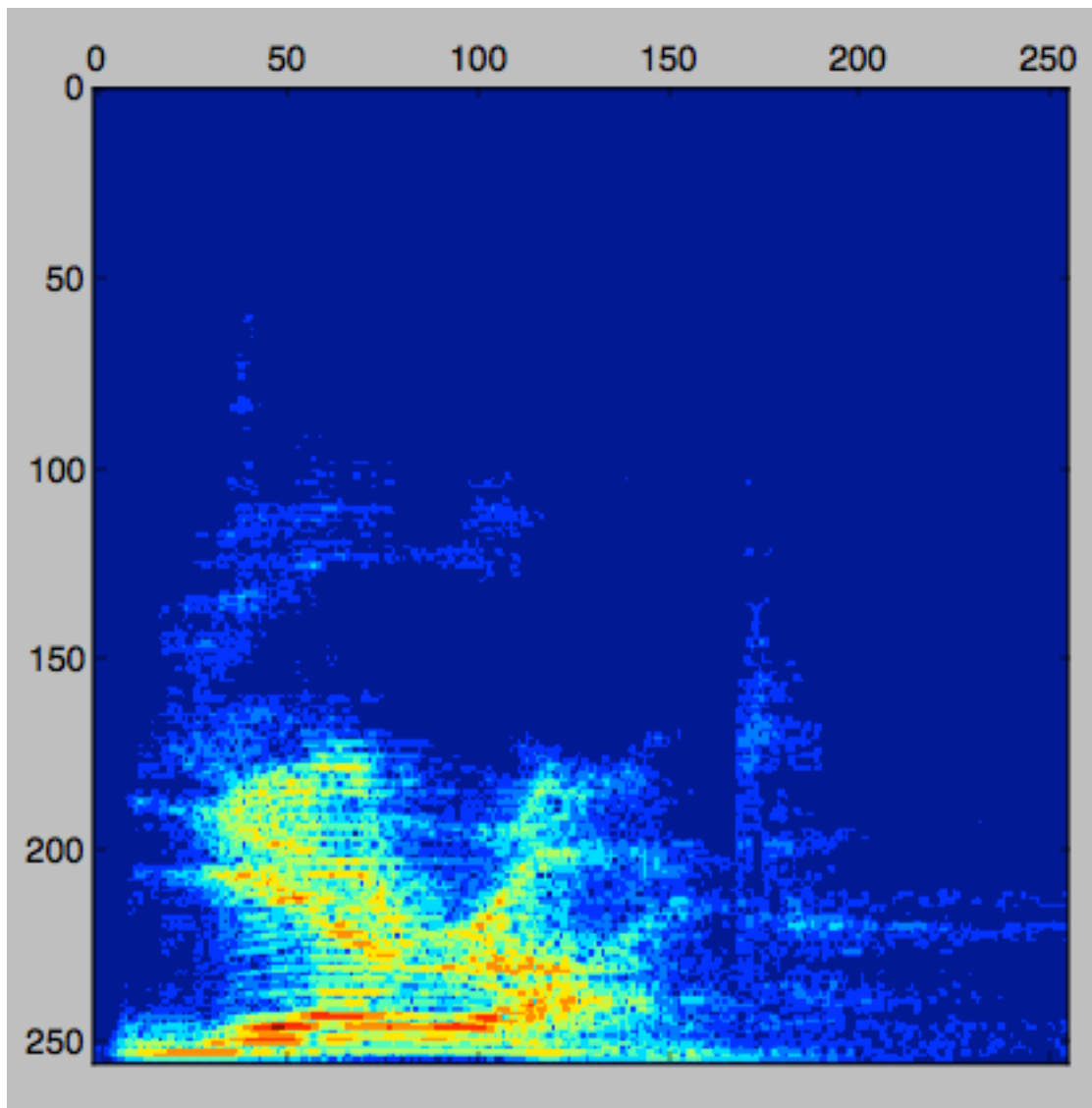
```
1 pip install pyaudio
```

Getting started

Toy examples: `./number_classifier_tflearn.py` `./speaker_classifier_tflearn.py`

Some less trivial architectures: `./densenet_layer.py`

Later: `./train.sh` `./record.py`



Update: Nervana demonstrated that it is possible for 'independents' to build speech recognizers that are state of the art.

Fun tasks for newcomers

- Watch video : <https://www.youtube.com/watch?v=u9FPqkuoEJ8>
- Understand and correct the corresponding code: `lstm-tflearn.py`
- Data Augmentation : create on-the-fly modulation of the data: increase the speech frequency, add background noise, alter the pitch etc,...

Extensions

Extensions to current tensorflow which are probably needed: * WarpCTC on the GPU see issue * Incremental collaborative snapshots ('P2P learning') ! * Modular graphs/models + persistence

Even though this project is far from finished we hope it gives you some starting points.

Looking for a tensorflow collaboration / consultant / deep learning contractor? Reach out to info@pannous.com