StarGAN-VC

This is a pytorch implementation of the paper: StarGAN-VC: Non-parallel many-to-many voice conversion with star generative adversarial networks.

The converted voice examples are in samples directory

Dependencies

- Python 3.6
- pytorch 1.0
- librosa
- pyworld
- tensorboardX
- scikit-learn

Usage

Download dataset

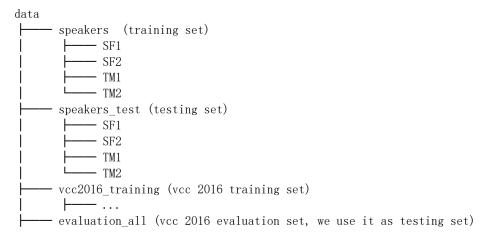
Download the vcc 2016 dataset to the current directory

python download.py

The downloaded zip files are extracted to ./data/vcc2016_training and ./data/evaluation_all.

- training set: In the paper, the author choose four speakers from ./data/vcc2016_training. So we move the corresponding folder(eg. SF1, SF2, TM1, TM2) to ./data/speakers.
- 2. testing set In the paper, the author choose four speakers from ./data/evaluation_all. So we move the corresponding folder(eg. SF1, SF2, TM1, TM2) to ./data/speakers_test.

The data directory now looks like this:



```
Preprocess
```

Extract features (mcep, f0, ap) from each speech clip. The features are stored as npy files. We also calculate the statistical characteristics for each speaker.

python preprocess.py

This process may take minutes!

Train

python main.py

Convert

python main.py --mode test --test_iters 200000 --src_speaker TM1 --trg_speaker "['TM1','SF1']"

Network structure

Snip20181102_2

Note: Our implementation follows the original paper's network structure, while pytorch StarGAN-VC code use StarGAN's network. Both can generate good audio quality.

Reference

tensorflow StarGAN-VC code

StarGAN code

CycleGAN-VC code

pytorch-StarGAN-VC code

StarGAN-VC paper

StarGAN paper

CycleGAN paper

If you feel this repo is good, please star!

Your encouragement is my biggest motivation!