seungwonpark / melgan

Code Issues 31 Pull requests 3 Actions Sec

MelGAN vocoder (compatible with NVIDIA/tacotron2)

swpark.me/melgan/

BSD-3-Clause license

639 stars 116 forks 29 watching Branches Activity
Tags

Public repository

6 Branches 3 Tags Go to file t Go

seungwonpark add resource from reproducibilit...

aca5990 · 4 years ago

assets	update README,...	5 years ago
config	add description ...	5 years ago
datasets	zero-pad short a...	5 years ago
model	fix remove_weig...	5 years ago
utils	remove unneces...	5 years ago
.gitignore	fix noise of unvoi...	5 years ago
	Initial commit	5 years ago

LICENSE		
README....	add resource fro...	4 years ago
hubconf.py	fix noise of unvoi...	5 years ago
inference....	add pretrained ...	5 years ago
preproces...	zero-pad short a...	5 years ago
requireme...	torch doesn't ne...	5 years ago
trainer.py	now training / va...	5 years ago

MelGAN

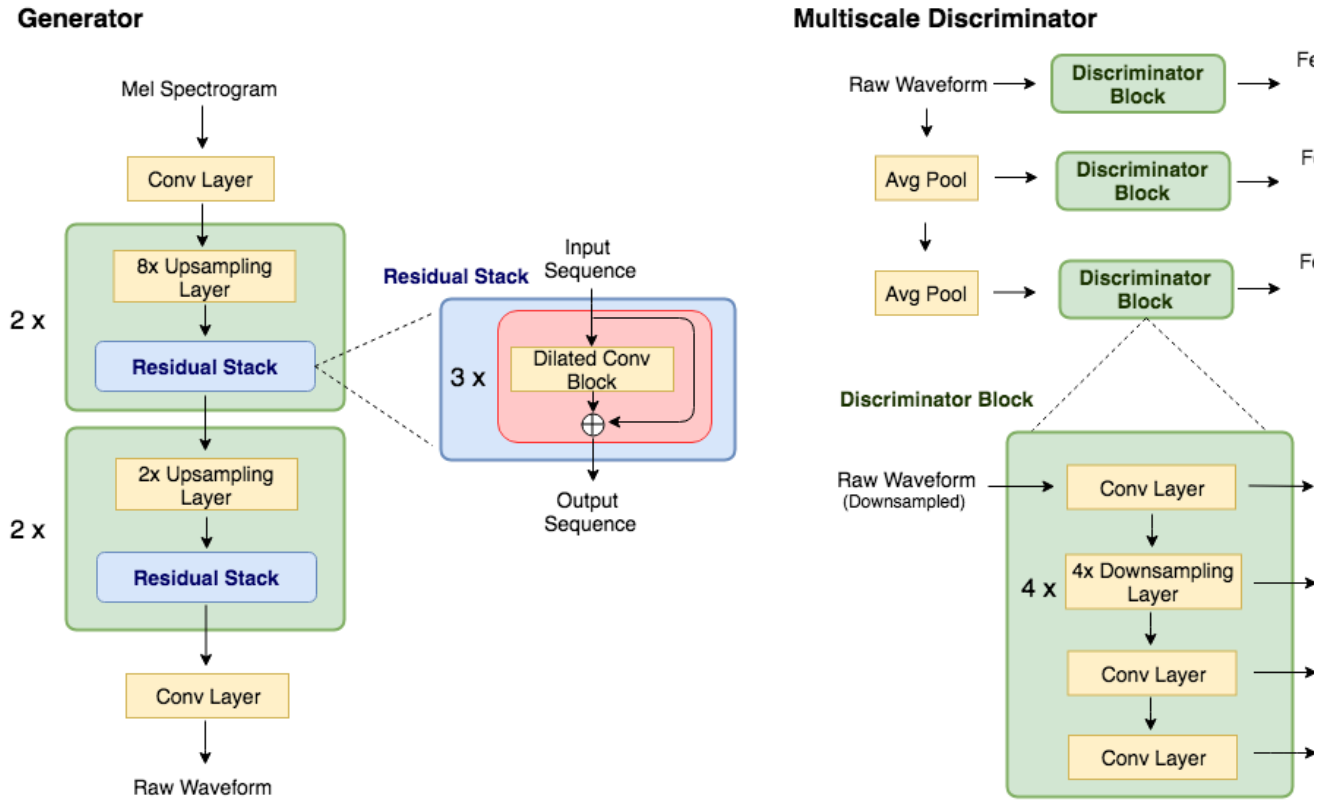
Unofficial PyTorch implementation of MelGAN vocoder

Key Features

- MelGAN is lighter, faster, and better at generalizing to unseen sp than WaveGlow.
- This repository use identical mel-spectrogram function from

README BSD-3-Clause license

- Pretrained model on LJSpeech-1.1 via PyTorch Hub.



Prerequisites

Tested on Python 3.6

```
pip install -r requirements.txt
```

Prepare Dataset

- Download dataset for training. This can be any wav files with sam 22050Hz. (e.g. LJSpeech was used in paper)
- preprocess: `python preprocess.py -c config/default.yaml -d [d root path]`
- Edit configuration `yaml` file

Train & Tensorboard

- `python trainer.py -c [config yaml file] -n [name of the run]`

    - `cp config/default.yaml config/config.yaml` and then edit `config.yaml`
    - Write down the root path of train/validation files to 2nd/3rd
    - Each path should contain pairs of `*.wav` with corresponding (preprocessed) `*.mel` file.
    - The data loader parses list of files within the path recursively
- `tensorboard --logdir logs/`

Pretrained model

Try with Google Colab: TODO

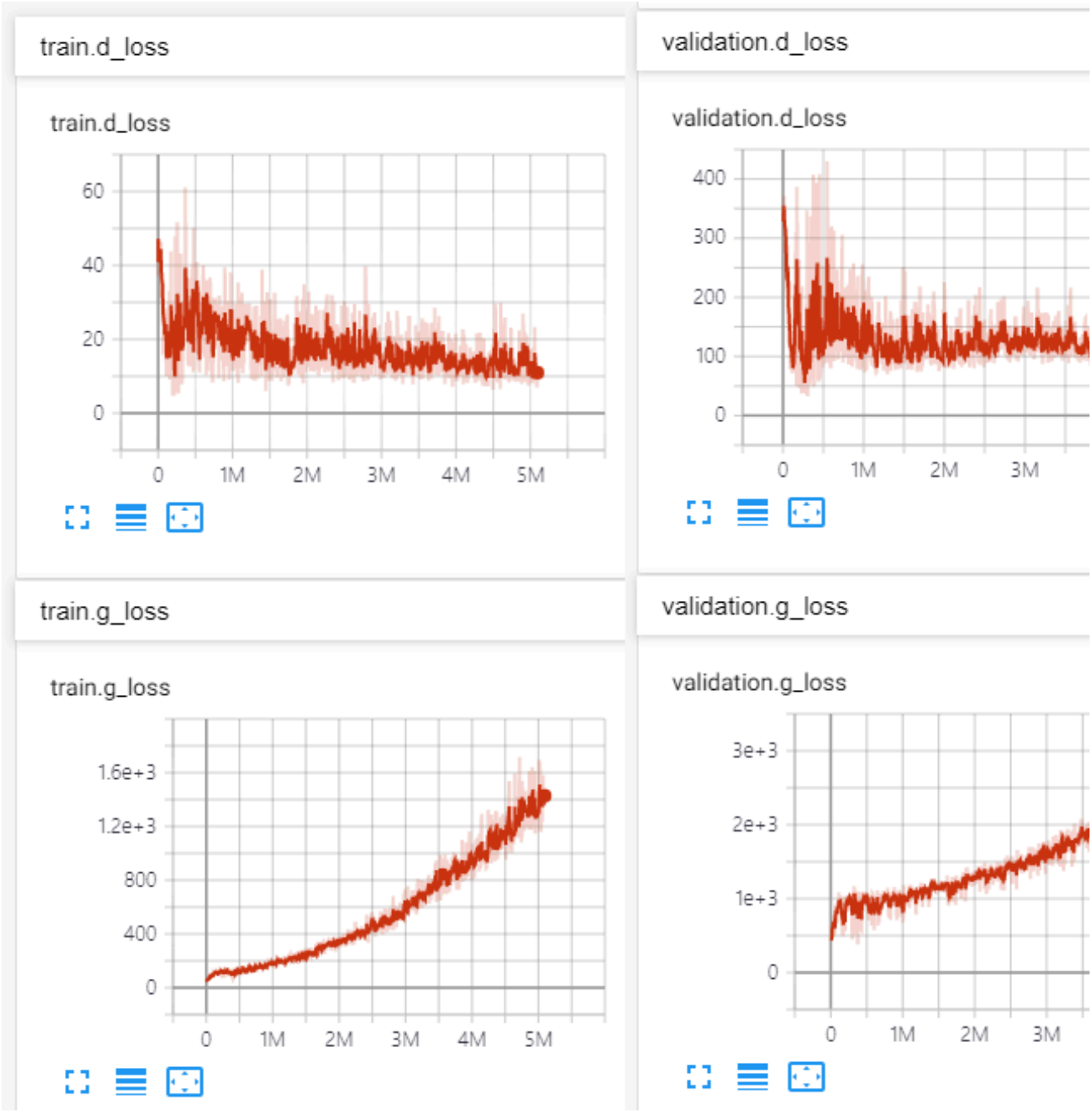
```
import torch
vocoder = torch.hub.load('seungwonpark/melgan', 'melgan')
vocoder.eval()
mel = torch.randn(1, 80, 234) # use your own mel-spectrogram her

if torch.cuda.is_available():
    vocoder = vocoder.cuda()
    mel = mel.cuda()

with torch.no_grad():
    audio = vocoder.inference(mel)
```

Inference

- `python inference.py -p [checkpoint path] -i [input mel path]`

Results

See audio samples at: http://swpark.me/melgan/. Model was trained a GPU for 14 days using LJSpeech-1.1.



Implementation Authors

- Seungwon Park @ MINDsLab Inc. (yyyyy@snu.ac.kr, swpark@min
- Myunchul Joe @ MINDsLab Inc.
- Rishikesh @ DeepSync Technologies Pvt Ltd.

License

BSD 3-Clause License.

- utils/stft.py by Prem Seetharaman (BSD 3-Clause License)
- datasets/mel2samp.py from https://github.com/NVIDIA/waveglo Clause License)
- utils/hparams.py from https://github.com/HarryVolek/PyTorch_Speaker_Verification (No specified)

Useful resources

- How to Train a GAN? Tips and tricks to make GANs work by Soun Chintala
- Official MelGAN implementation by original authors
- Reproduction of MelGAN - NeurIPS 2019 Reproducibility Challen (Ablation Track) by Yifei Zhao, Yichao Yang, and Yang Gao
  - "replacing the average pooling layer with max pooling layer replacing reflection padding with replication padding improv performance significantly, while combining them produces w

Releases 3

fixed noise of unvoiced segment (fix #30, #17) Latest
on Dec 2, 2019

+ 2 releases

Packages

No packages published

Contributors 2

seungwonpark Seung-won Park

benwu95

Deployments 16

github-pages 5 years ago

+ 15 deployments

Languages

- **Python** 100.0%