“Speech resynthesis from discrete disentangled self-supervised representations” online resource

[1] A. Polyak et al., “Speech resynthesis from discrete disentangled self-supervised representations,” *arXiv preprint arXiv:2104.00355*, Jul. 2021.

Author : Sian-Yi Chen

Advisor : Tay-Jyi Lin and Chingwei Yeh

Outline

- Action item
  - 找 disentangled 相關研究：使用 [1] 所提供的 open online resource
- Status report
  - 實作步驟分為以下四點：
    1. **Environment setup**

       此專案預設使用 GPU 執行，目前**執行方案二**
       - 方案一：使用 AC922 (GPU) 執行，但在 PPC64el 架構下，Anaconda 不提供作者要求所使用套件 torch=1.8.0 的版本，需自行從 source code 建構 (附錄 p.4)
       - 方案二：將程式中所使用到 GPU 的指令更改為 CPU 執行，並架設 Google Cloud Platform (GCP) 中免費版的 Virtual Machine (VM) (附錄 p.5)
    2. **Dataset preprocess**
       - 作者提供的連結為 flac 檔，需自行將 flac 轉換為 wav 檔，才能繼續 audio file downsampling
    3. **Network training**
       A. F0 Quantizer Model (**訓練中**)

       B. Resynthesis Model，需完成 A 才能繼續訓練
    4. **Generate Audio file**
  - 後續規劃
    1. (使用 VM) 等待 F0 Quantizer Model 訓練完接著訓練 Resynthesis Model
    2. (使用本機) 將目前 F0 Quantizer Model 訓練到一半的 checkpoint 拿去訓練 Resynthesis Model，之後與 VM 完整訓練完結果比較
    3. 再次嘗試使用 AC922 進行實作 (在 Environment setup 中，從 source code 建構所需環境)

[1] A. Polyak et al., “Speech resynthesis from discrete disentangled self-supervised representations,” *arXiv preprint arXiv:2104.00355*, Jul. 2021.

附錄

Pytorch don't have ppc64le binaries

intel / packages / pytorch 1.8.0

0

PyTorch is an optimized tensor library for deep learning using GPUs and CPUs.

Conda | Files | Labels | Badges

License: BSD 3-Clause
Home: http://pytorch.org/
1446 total downloads
Last upload: 3 days and 16 hours ago

Installers

Info: This package contains files in non-standard labels.

conda install

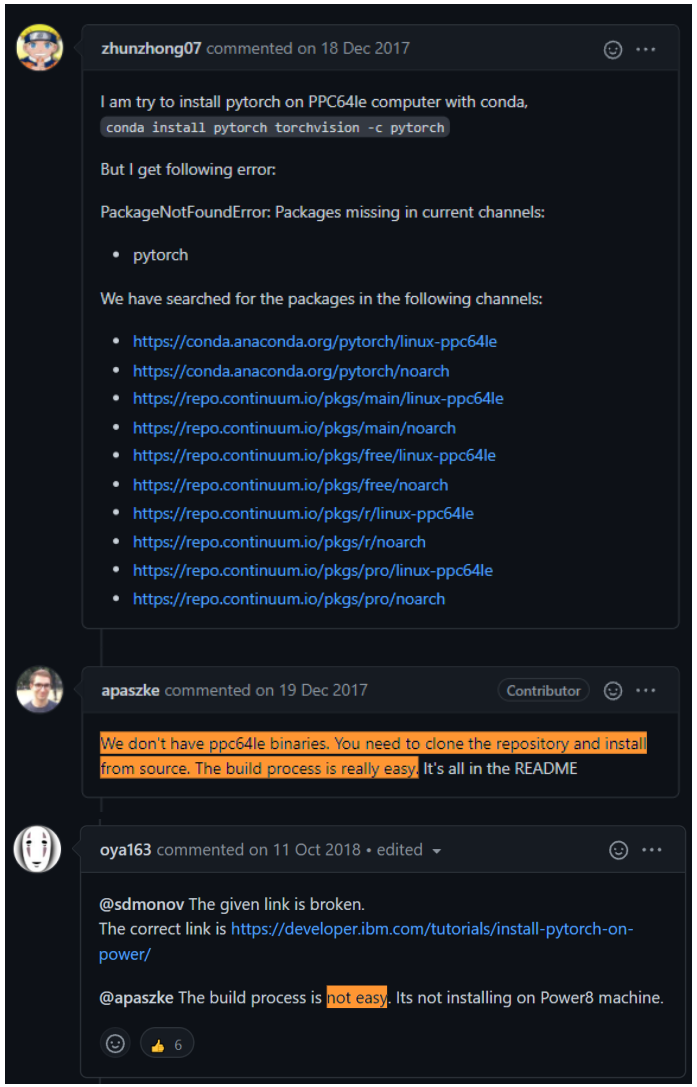
linux-64 v1.8.0

To install this package with conda run one of the following:

```
conda install -c intel pytorch
conda install -c intel/label/oneapibeta pytorch
```

Description

[Pytorch :: Anaconda.org]



zhunzhong07 commented on 18 Dec 2017

I am try to install pytorch on PPC64le computer with conda,

```
conda install pytorch torchvision -c pytorch
```

But I get following error:

PackageNotFoundError: Packages missing in current channels:

- pytorch

We have searched for the packages in the following channels:

- https://conda.anaconda.org/pytorch/linux-ppc64le
- https://conda.anaconda.org/pytorch/noarch
- https://repo.continuum.io/pkgs/main/linux-ppc64le
- https://repo.continuum.io/pkgs/main/noarch
- https://repo.continuum.io/pkgs/free/linux-ppc64le
- https://repo.continuum.io/pkgs/free/noarch
- https://repo.continuum.io/pkgs/r/linux-ppc64le
- https://repo.continuum.io/pkgs/r/noarch
- https://repo.continuum.io/pkgs/pro/linux-ppc64le
- https://repo.continuum.io/pkgs/pro/noarch

apaszke commented on 19 Dec 2017 Contributor

We don't have ppc64le binaries. You need to clone the repository and install from source. The build process is really easy. It's all in the README

oya163 commented on 11 Oct 2018 • edited

@sdmonov The given link is broken.
The correct link is https://developer.ibm.com/tutorials/install-pytorch-on-power/

@apaszke The build process is not easy. Its not installing on Power8 machine.

6

Using GCP's VM

```
Epoch: 37
Steps : 25020, Gen Loss Total : 0.054, s/b : 0.648
Steps : 25025, Gen Loss Total : 0.061, s/b : 0.489
Steps : 25030, Gen Loss Total : 0.063, s/b : 0.512
Steps : 25035, Gen Loss Total : 0.062, s/b : 0.429
Steps : 25040, Gen Loss Total : 0.036, s/b : 0.473
Steps : 25045, Gen Loss Total : 0.040, s/b : 0.462
Steps : 25050, Gen Loss Total : 0.030, s/b : 0.495
Steps : 25055, Gen Loss Total : 0.049, s/b : 0.470
Steps : 25060, Gen Loss Total : 0.048, s/b : 0.471
Steps : 25065, Gen Loss Total : 0.044, s/b : 0.497
Steps : 25070, Gen Loss Total : 0.054, s/b : 0.464
Steps : 25075, Gen Loss Total : 0.047, s/b : 0.528
Steps : 25080, Gen Loss Total : 0.055, s/b : 0.484
Steps : 25085, Gen Loss Total : 0.032, s/b : 0.470
Steps : 25090, Gen Loss Total : 0.044, s/b : 0.470
Steps : 25095, Gen Loss Total : 0.040, s/b : 0.476
Steps : 25100, Gen Loss Total : 0.044, s/b : 0.451
Steps : 25105, Gen Loss Total : 0.057, s/b : 0.443
Steps : 25110, Gen Loss Total : 0.039, s/b : 0.449
Steps : 25115, Gen Loss Total : 0.061, s/b : 0.481
Steps : 25120, Gen Loss Total : 0.039, s/b : 0.415
Steps : 25125, Gen Loss Total : 0.042, s/b : 0.417
Steps : 25130, Gen Loss Total : 0.029, s/b : 0.486
Steps : 25135, Gen Loss Total : 0.038, s/b : 0.398
/home/fg6ts15test/.local/lib/python3.8/site-packages/scipy/signal/signaltools.p
y:1531: UserWarning: kernel_size exceeds volume extent: the volume will be zero
-padded.
  warnings.warn('kernel_size exceeds volume extent: the volume will be '
Steps : 25140, Gen Loss Total : 0.038, s/b : 0.489
Steps : 25145, Gen Loss Total : 0.034, s/b : 0.428
Steps : 25150, Gen Loss Total : 0.034, s/b : 0.415
Steps : 25155, Gen Loss Total : 0.036, s/b : 0.507
Steps : 25160, Gen Loss Total : 0.038, s/b : 0.389
Steps : 25165, Gen Loss Total : 0.033, s/b : 0.460
Steps : 25170, Gen Loss Total : 0.046, s/b : 0.510
Steps : 25175, Gen Loss Total : 0.049, s/b : 0.435

[0] 0:bash  1:bash!  2:bash  3:bash  4:unzip LJSpeech-  5:training*
```

目前執行到 Steps：25175
最終　　Steps：40000

訓練完 F0 Quantizer Model 會在下方資料夾 (lj_f0_vq)
看到 g_00040000 的 checkpoint 檔

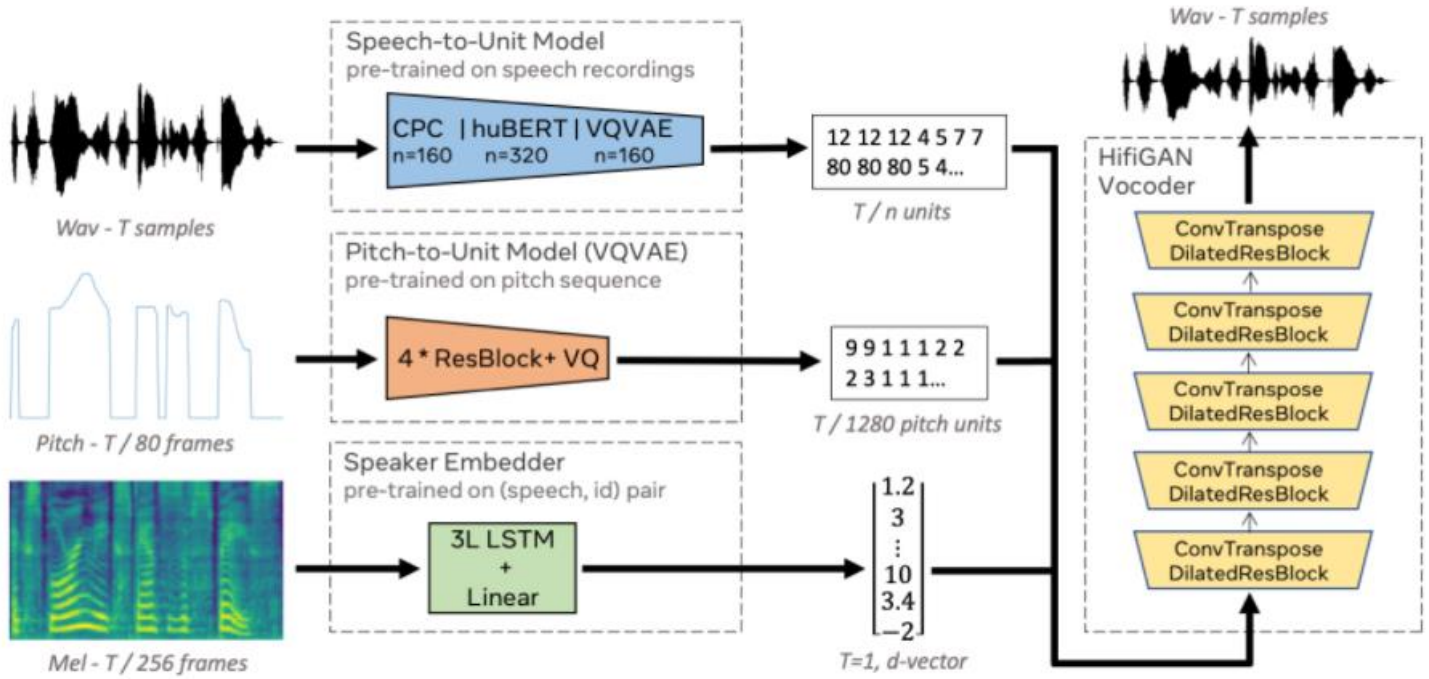
```
fg6ts15test@ubuntu-focal-2:/mnt/disks/nvmessd/speech-resynthesis/data$ c
d ..
fg6ts15test@ubuntu-focal-2:/mnt/disks/nvmessd/speech-resynthesis$ cd c
checkpoints/ configs/
fg6ts15test@ubuntu-focal-2:/mnt/disks/nvmessd/speech-resynthesis$ cd c
checkpoints/ configs/
fg6ts15test@ubuntu-focal-2:/mnt/disks/nvmessd/speech-resynthesis$ cd che
ckpoints
fg6ts15test@ubuntu-focal-2:/mnt/disks/nvmessd/speech-resynthesis/checkpo
ints$ ls
lj_f0_vq  lj_vqvae
fg6ts15test@ubuntu-focal-2:/mnt/disks/nvmessd/speech-resynthesis/checkpo
ints$

fg6ts15test@ubuntu-focal-2:/mnt/disks/nvmessd/speech-resynthesis/checkpo
ints/lj_f0_vq$ ls
config.json  g_00010000  g_00020000  logs
```

並使用 g_00040000 checkpoint 繼續訓練第二個 Resynthesis Model

Paper Demo Link

[Speech Resynthesis from Discrete Disentangled Self-Supervised Representations]



[HiFi-GAN: High-Fidelity Denoising and Dereverberation Based on Speech Deep Features in Adversarial Networks]