

# Automatic Speech recognition on EC2- Habana Gaudi processors

habanalabs-base-ubuntu20.04-1.3.0-499-1644739541-9a75c51a-a4d1-4470-884f-6be27933fcc8

Mohamed A. Bencherif,  
Center of Smart Robotics Research-CS2R  
College of Computer & Information Sciences -CCIS  
King Saud University-KSU  
Saudi Arabia





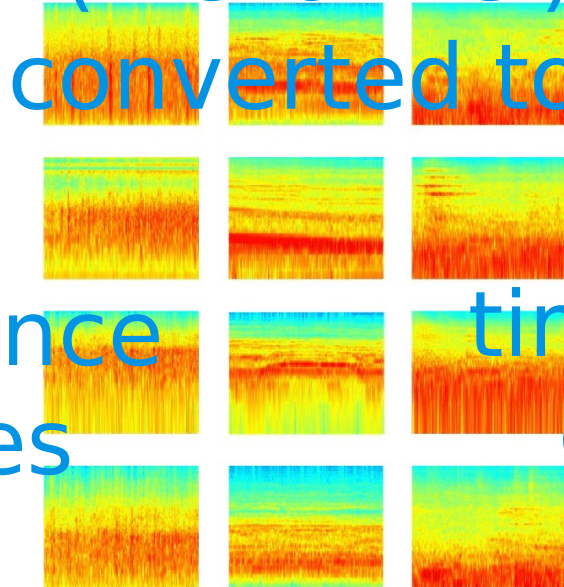
THANKS TO :  
AWS, DEVOST, HABANA TEAM  
ALL WHO HELPED....

# Problem Definition

How to train an ASR in the Habana Gaudi processors on the AWS Cloud ?

Raw speech (1 channel) is in general converted to :

2D  
time Sequence  
of Matrices



3D  
time Sequence  
of Matrices

# Candidate Networks

Name	Type	Params	In sizes	Out sizes
<hr/>				
0   cnn	Conv2d	320	[8, 1, 128, 1151]	[8, 32, 64, 576]
1   rescnn_layers	Sequential	56.3 K	[8, 32, 64, 576]	[8, 32, 64, 576]
2   fully_connected	Linear	1.0 M	[8, 576, 2048]	[8, 576, 512]
3   birnn_layers	Sequential	22.1 M	[8, 576, 512]	[8, 576, 1024]
4   classifier	Sequential	539 K	[8, 576, 1024]	[8, 576, 29]
5   criterion	CTCLoss	0	?	?
<hr/>				

<https://github.com/SeanNaren/deepspeech.pytorch>

<https://github.com/jiwidi/DeepSpeech-pytorch>

<https://github.com/mozilla/DeepSpeech>  
in tensorflow

# First Aws Credit 87\$ spent on compiling and .....

mozilla / DeepSpeech Public

Notifications Fork 3.4k Star 19.1k

<> Code Issues 106 Pull requests 15 Actions Projects 1 Wiki Security

master Go to file Code

About

DeepSpeech is an open source embedded (offline, on-device) speech-to-text engine which can run in real time on devices ranging from a Raspberry Pi 4 to high power GPU servers.

machine-learning embedded deep-learning offline tensorflow speech-recognition neural-networks

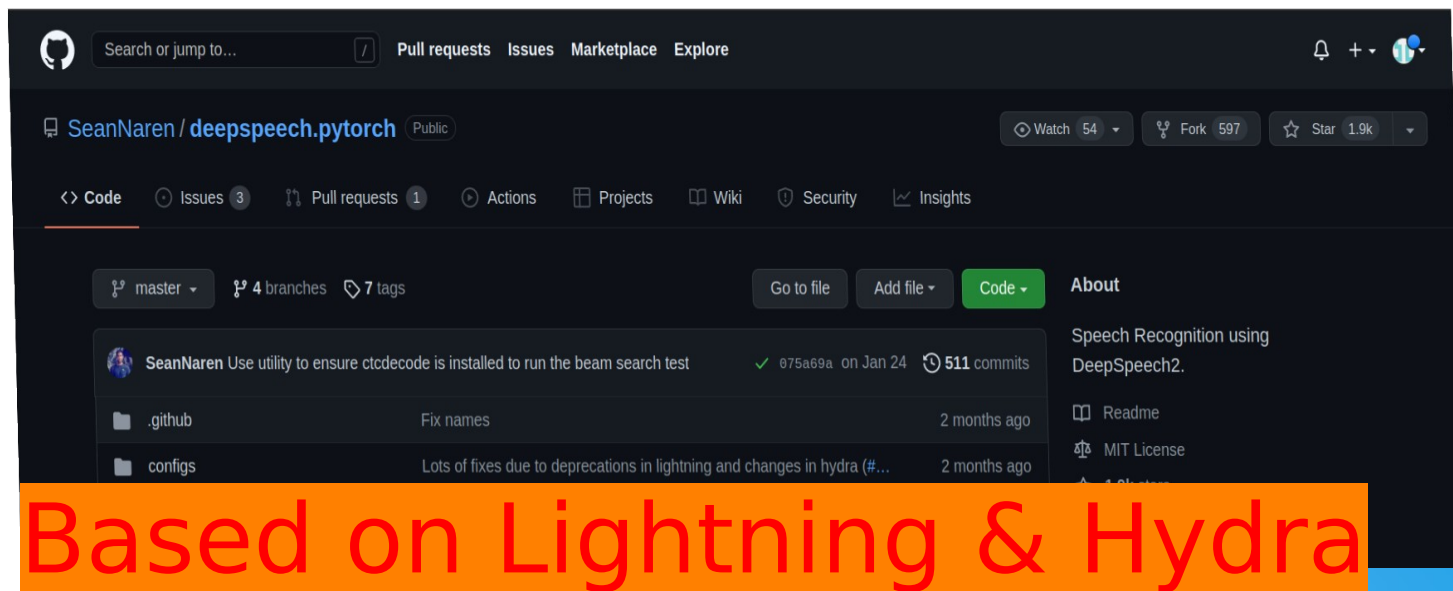
Readme MPL-2.0 License Code of conduct 19.1k stars 652 watching

File	Commit Message	Time Ago
.github	Add NodeJS 16.0.0	10 months ago
bin	Fixed M-AIILABS broken link	3 months ago
ci_scripts	CI: Linux ARMv7 / Aarch64	10 months ago
data	Add small bytes output mode scorer for tests	17 months ago
doc	Update conf.py	7 months ago
examples	Revert "Merge branch 'tensorflow-real'"	2 years ago
images	Updating Geometry	2 years ago
kenlm @ 0c4dd4e	MSVC doesn't like const Proxy operator*() c...	12 months ago
native_client	Add NodeJS 16.0.0	10 months ago
taskcluster	Fix #3608: Remove code refs to TaskCluster	11 months ago
tensorflow @ 23ad9...	Updating commit of submodule	2 years ago

Could not compile, neither work with

<https://github.com/mozilla/DeepSpeech> in tensorflow

# Second Aws Credit ...

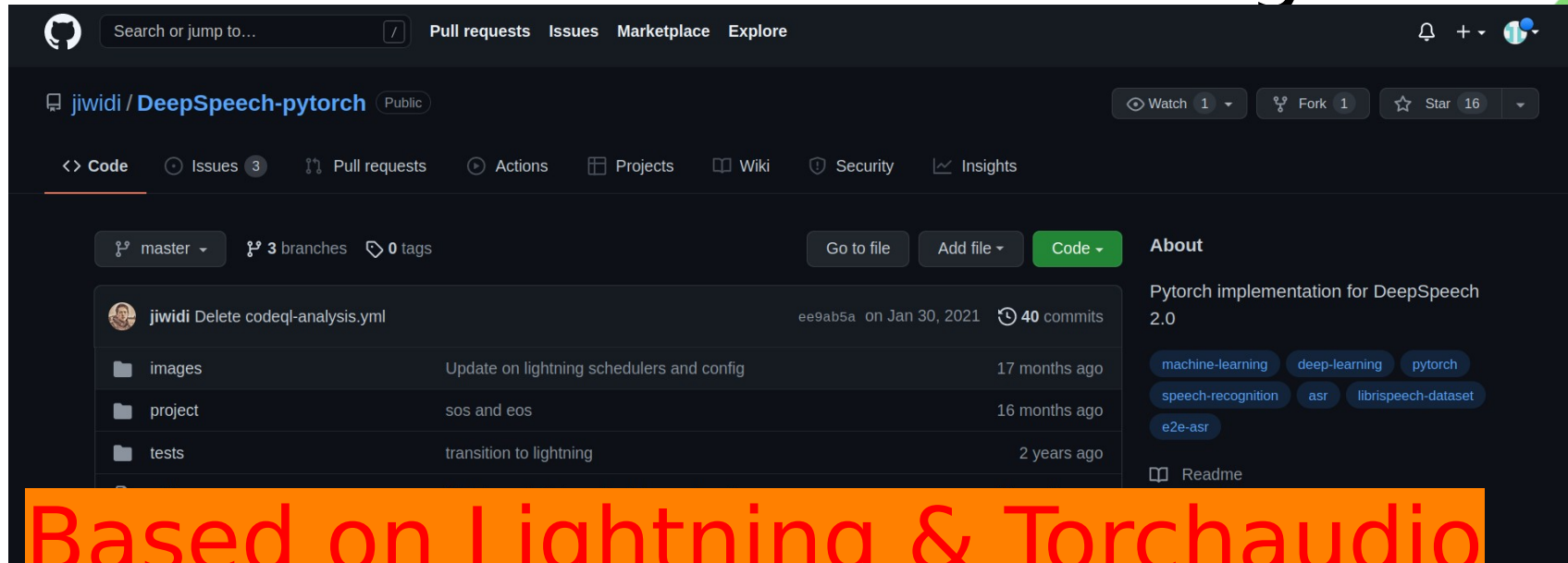


Did not find a way to port hydra style to Habana

<https://github.com/SeanNaren/deepspeech.pytorch>



# Second Aws Credit ... Still being Used



Based on Lightning & Torchaudio

<https://github.com/jjiwidi/DeepSpeech-pytorch>

Worked fine on my local machine, ported to habana with torchaudio tweeking,  
but some errors in negative dimension of the Tensors??  
+ I did a toy example for the deepspeech in the same MNIST Style



# Code Constraints

- Do not break the Code of the habana framework
- Code in tensorflow or pytorch need to be adjusted to work on the habana machines.
- Tensorflow codes are restricted to  $tf > 2.6$
- **Torchaudio does not work on Habana machine as it breaks the torch version of the machine**
- Librosa can not be installed as it uses numba and this later requires numpy  $< 1.21$ , but numpy used in habana is  $> 1.22$
- Working with PytorchLightning
  - Which trainer ?? Model-References/Pytorch/central
  - Trainer on local machine and Ec2-HB

# Platform Constraints

- **Compiling from source requires complex procedure as CUDA is not supported and many available codes are CUDA dependent.**
- **Due to the hourly cost of the Ec2 instance, one needs to work on a local and a distant machine.**
- **But the Model-References Code needs to be run on the EC2.**

# What went wrong?

- **Could not find the documentation on code adaptation easily.**
- **Workshops not working for every one as in the in the demo file,**
- **I could not benefit from it.**
- **Time Delay between question and answers, due to different time zones.**

# Unet Example : my short nightmare...:)

- **One of the recommended code was the Unet Example.**
- **So much switches not find the documentation on code adaptation easily.**
- **Too much files to check and understand...**
- **Code should be improved and librairies should be better organized**
- **Workshop not working for every one in the demo file.**
-

# Improvements

- - Used tweaked torchaudio on Gaudi Processors

- - Read flac files : replaced by soundfile

- Transformations

- - Melspectrum:
  - - Time Masking
- Frequency Masking

•

•

# Gentle recommendations

- **Afford two types of machines, a low cost or free tiers machine containing 1 Gaudi processor for learning purposes and testing issues.**
- **A second machine for code deployment.**
- **An elastic IP for the Gaudi machine would have solved a lot copy-paste code, mainly when stopping and restarting the machine, each time a new allocated ip and new ec2-name, as well as file transfer via filezilla or aws cli commands, and ease the work with**

# Thank you!