

1. 安装 ctcdecode

神坑 ctcdecode，只适用于 Linux 下，大部分安装都是 gcc 的问题，本人再 gcc11.2 和 gcc9.4 都安装成功过，torch 版本 1.8.1、1.10.1、1.12.1、1.13.1、python 都是 3.8 版本，都成功安装上，记得要安装 cuda 和 wget，同时本人环境是 A100 显卡，最大支持 cuda12.2 的版本，目前安装的 torch 版本都是 cuda11.1 和 11.6 的，都成功安装。如果报错请查看自己 Linux 的 gcc 版本问题。在运行作者的代码中，碰到两个关于 ctcdecode 的问题

一个是 not enough space，另外一个是在出现当时已经成功安装 ctcdecode 0.4.0，自己是刚开始是根据某 DN 中网盘下载的 ctc 源码进行安装的，经过自己的尝试，发现还是 ctcdecode 的问题，自己从 github 下载了 master 分支下的 1.0 版本，同时将 third_party 直接替换到下载的 1.0 版本当中，然后也是成功安装了，没想到这次运行也没有报这两个错误了，虽然分支是 1.0 版本的，但是最后安装成功还是 0.4 版本的 ctcdecode，报错信息如下图：

运行可能还会出现未定义的符号，应该也是该 ctcdecode 版本的问题

```
Traceback (most recent call last):
  File "main.py", line 255, in <module>
    processor.start()
  File "main.py", line 70, in start
    dev_wer = seq_eval(self.arg, self.data_loader['dev'], self.model, self.device,
  File "/data/liujing_data/CorrNet-main/seq_scripts.py", line 65, in seq_eval
    ret_dict = model(vid, vid_lgt, label=label, label_lgt=label_lgt)
  File "/data/Anaconda3/envs/ljtest/lib/python3.8/site-packages/torch/nn/modules/module.py",
    result = self.forward(*input, **kwargs)
  File "/data/liujing_data/CorrNet-main/slr_network.py", line 100, in forward
    else self.decoder.decode(outputs, lgt, batch_first=False, probs=False)
  File "/data/liujing_data/CorrNet-main/utils/decode.py", line 28, in decode
    return self.BeamSearch(nn_output, vid_lgt, probs)
  File "/data/liujing_data/CorrNet-main/utils/decode.py", line 42, in BeamSearch
    beam_result, beam_scores, timesteps, out_seq_len = self.ctc_decoder.decode(nn_output, vid_lgt)
  File "/data/Anaconda3/envs/ljtest/lib/python3.8/site-packages/ctcdecode/__init__.py", line
    ctc_decode.paddle_beam_decode(probs, seq_lens, self._labels, self._num_labels, self._bea
RuntimeError: Not enough space
```

```
| 0/270 [00:20<?, ?it/s]
Traceback (most recent call last):
  File "main.py", line 254, in <module>
    processor.start()
  File "main.py", line 71, in start
    'dev', epoch, self.arg.work_dir, self.recoder, self.arg.evaluate_tool)
  File "/data/liujing_data/CorrNet-main/seq_scripts.py", line 65, in seq_eval
    ret_dict = model(vid, vid_lgt, label=label, label_lgt=label_lgt)
  File "/data/Anaconda3/envs/py3.7/lib/python3.7/site-packages/torch/nn/modules/module.py", line 1102, in _ca
    return forward_call(*input, **kwargs)
  File "/data/liujing_data/CorrNet-main/slr_network.py", line 100, in forward
    else self.decoder.decode(outputs, lgt, batch_first=False, probs=False)
  File "/data/liujing_data/CorrNet-main/utils/decode.py", line 28, in decode
    return self.BeamSearch(nn_output, vid_lgt, probs)
  File "/data/liujing_data/CorrNet-main/utils/decode.py", line 42, in BeamSearch
    beam_result, beam_scores, timesteps, out_seq_len = self.ctc_decoder.decode(nn_output, vid_lgt)
  File "/data/Anaconda3/envs/py3.7/lib/python3.7/site-packages/ctcdecode/__init__.py", line 40, in decode
    output, timesteps, scores, out_seq_len)
RuntimeError: Invalid UTF-8
```

在碰到空间不足这个问题时，首先怀疑是否内存不足，寻找了好久，在 main.py 当中有一个 pin_memory 参数，好像是锁定内存的，详细请到 pytorch 官网查询，最后设置为 False，还是报错，无果，最后将该参数删掉，batchsize 设置为 8，成功运行第一轮训练后，报错 UTF-8 问题，最后尝试还是因为 ctcdecode 1.0（实际还是 0.4）的问题。最后从新装 ctcdecode（实际还是 0.4）解决该问题。在作者 github 讨论区中就有人说是升级 torch 版本解决空间不足问题，应该正是如此，那位询问的人也说是下载的是 1.0.3 的版本最后是 0.4 的 ctcdecode。后续将 pin_memory 从新加上，发现不影响，因此还是 ctcdecode 版本的问题。最后运行的环境时 Linux gcc9.4.0，显卡 cuda 版本是 12.0，pytorch 版本是 1.10.1+cuda11.1，python3.7.1 运行成功，目前训练稳定。

同时在安装 ctc (例如修改 build.py 中 c++11 替换成 c++14 等) 中报的一些错, 在其他博客都有, 可以进行参考。

此外 我的 pytorch 都是通过本地安装的 然后安装 pip install torchvision==相对应 torch 版本的

然后安装 conda install cudatoolkit=版本号也是对应的 -c nvidia

```
torch-1.0.1-cp36-cp36m-linux_x86_64.whl      torch-1.8.0+cull1-cp36-cp36m-linux_x86_64.whl
torch-1.10.1+cull1-cp37-cp37m-linux_x86_64.whl  torch-1.8.0+cull1-cp38-cp38-linux_x86_64.whl
torch-1.12.0+cull16-cp38-cp38-linux_x86_64.whl
torch-1.13.0+cull16-cp38-cp38-linux_x86_64.whl
```

附上自己的 ctcdecode 链接: 链接:

<https://pan.baidu.com/s/1pdNAJUQcg6h8Eymx2s61LQ?pwd=nlgm>

提取码: nlgm

Ctc 安装就此完成,

2. kaldi 安装: 网上教程都有, 相对来说还比较简单, 主要是 gcc (目前 gcc9.4) 的问题感觉, 报错参考, 最后成功运行测试一般就没问题了。(源码直接在 github 进行下载的)

同时作者中进行软连接当中 tools 下没有 sctk-2.4.10 安装好的目录如下

- scilite [\[kaldi-asr/kaldi\]](#), install kaldi tool to get scilite for evaluation. After installation, create a soft link toward the scilite: `mkdir ./software ln -s PATH_TO_KALDI/tools/sctk-2.4.10/bin/scilite ./software/scilite`

```
total 4900
drwxrwxr-x 11 ubuntu ubuntu 4096 Nov 15 12:05 ./
drwxrwxr-x 11 ubuntu ubuntu 4096 Nov 13 04:13 ../
drwxrwxr-x 3  ubuntu ubuntu 4096 Nov 13 04:13 ATLAS_headers/
drwxrwxr-x 2  ubuntu ubuntu 4096 Nov 13 04:13 CLAPACK/
drwxrwxr-x 2  ubuntu ubuntu 4096 Nov 13 04:13 config/
lrwxrwxrwx 1  ubuntu ubuntu   9 Nov 15 09:20 cub -> cub-1.8.0/
drwxrwxr-x 8  ubuntu ubuntu 4096 Nov 15 09:20 cub-1.8.0/
-rw-rw-r-- 1  ubuntu ubuntu 384644 Nov 15 09:20 cub-1.8.0.tar.gz
-rw-rw-r-- 1  ubuntu ubuntu   65 Nov 15 09:18 env.sh
drwxrwxr-x 2  ubuntu ubuntu 4096 Nov 13 04:13 extras/
-rw-rw-r-- 1  ubuntu ubuntu 505 Nov 13 04:13 .gitignore
-rw-rw-r-- 1  ubuntu ubuntu 1206 Nov 13 04:13 INSTALL
lrwxrwxrwx 1  ubuntu ubuntu  29 Nov 15 09:18 install_pfile_utils.sh -> extras/install_pfile_utils.sh
lrwxrwxrwx 1  ubuntu ubuntu  27 Nov 15 09:18 install_portaudio.sh -> extras/install_portaudio.sh*
lrwxrwxrwx 1  ubuntu ubuntu  23 Nov 15 09:18 install_speech.sh -> extras/install_speech.sh*
lrwxrwxrwx 1  ubuntu ubuntu  23 Nov 15 09:18 install_srilm.sh -> extras/install_srilm.sh*
-rw-rw-r-- 1  ubuntu ubuntu 6970 Nov 13 04:13 Makefile
lrwxrwxrwx 1  ubuntu ubuntu  13 Nov 15 12:05 openfst -> openfst-1.7.2/
drwxr-xr-x 7  ubuntu ubuntu 4096 Nov 15 12:03 openfst-1.7.2/
-rw-rw-r-- 1  ubuntu ubuntu 1269292 Jul 17 2019 openfst-1.7.2.tar.gz
drwxrwxr-x 2  ubuntu ubuntu  4096 Nov 15 09:18 python/
lrwxrwxrwx 1  ubuntu ubuntu  12 Nov 15 09:20 sctk -> sctk-20159b5/
drwxrwxr-x 5  ubuntu ubuntu  4096 Nov 15 09:20 sctk-20159b5/
-rw-rw-r-- 1  ubuntu ubuntu 3067724 Nov 15 09:20 sctk-20159b5.tar.gz
lrwxrwxrwx 1  ubuntu ubuntu  13 Nov 15 09:20 sph2pipe -> sph2pipe_v2.5/
-rw-rw-r-- 1  ubuntu ubuntu 290693 Nov 15 09:20 sph2pipe-2.5.tar.gz
drwxrwxr-x 3  ubuntu ubuntu  4096 Nov 15 09:20 sph2pipe_v2.5/
(list) ubuntu@localhost:~/data/liujing_data/kaldi-master/tools
```

因此软链接需要修改 `ln -s PATH_TO_KALDI/tools/sctk/bin/scilite ./software/scilite` 同时下图时一般出现的问题可以作为参考

```
1、OpenFst问题
Checking compiler g++ ...
Checking OpenFst library in ...
***configure failed: Could not find file /include/fst/fst.h:
you may not have installed OpenFst. See ../tools/INSTALL ***
1
2
3

2、cub
Checking compiler c++ ...
Checking OpenFst library in /root/kaldi/tools/openfst-1.7.2 ...
Checking cub library in ...
***configure failed: Could not find file /cub/cub.cuh:
you may not have installed cub. Go to ../tools/ and type
'make cub' to download and unpack it. We'll detect it then. ***
从网上下载一个cub-1.8.0.tar.gz, 然后将这个文件拷贝到kaldi/tools目录下, 最后make cub.
```

3. warp-ctc

安装教程网上都有可以参考

在安装这个时出现错误

```
-- Build files have been written to: /home/warp-ctc/build
[ 11%] Building NVCC (Device) object CMakeFiles/warpctc.dir/src/warpctc_generated_reduce.cu.o
nvcc fatal : Unsupported gpu architecture 'compute_30'
CMake Error at CMakeLists.txt:100 (add_executable):
```

Compute30

只需修改 CMakeList 文件中这几行代码即可安装

```
Processing /data/liujing_data/corrNet-main/warp-ctc/pytorch_binding
Preparing metadata (setup.py) ... done
Building wheels for collected packages: warpctc-pytorch
Building wheel for warpctc-pytorch (setup.py) ... done
Created wheel for warpctc-pytorch: filename=warpctc_pytorch-0.1-cp37-cp37m-linux_x86_64.whl size=2462517 sha256=9e0e45ba48580258842082b23136931e838546016d551aedb5e52009c166cfd
Stored in directory: /tmp/pip-ephem-wheel-cache-w0bpo5g/wheels/b2/d5/dd/cde0c41c25c02bd2b44886d12b3f1aa89b253bc694799bb882
Successfully built warpctc-pytorch
Installing collected packages: warpctc-pytorch
Successfully installed warpctc-pytorch-0.1
py3-7@ubuntu@localhost:~/data/liujing_data/corrNet-main/warp-ctc/pytorch_binding$ pip list
Package Version
```

```
CMakeLists.txt - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

include_directories(include)

FIND_PACKAGE(CUDA 6.5)
MESSAGE(STATUS "cuda found ${CUDA_FOUND}")

option(WITH_GPU "compile warp-ctc with cuda." ${CUDA_FOUND})
option(WITH_OMP "compile warp-ctc with openmp." ON)

if(NOT WITH_OMP)
    add_definitions(-DCTC_DISABLE_OMP)
endif()
if(WITH_OMP)
    set(CMAKE_CXX_FLAGS "${CMAKE_CXX_FLAGS} -fopenmp")
    set(CUDA_NVCC_FLAGS "${CUDA_NVCC_FLAGS} -Xcompiler -fopenmp")
endif()

# need to be at least 30 or _shfl_down in reduce wont compile
set(CUDA_NVCC_FLAGS "${CUDA_NVCC_FLAGS} -gencode arch=compute_30,code=sm_30 -O2")
set(CUDA_NVCC_FLAGS "${CUDA_NVCC_FLAGS} -gencode arch=compute_35,code=sm_35")

set(CUDA_NVCC_FLAGS "${CUDA_NVCC_FLAGS} -gencode arch=compute_50,code=sm_50")
set(CUDA_NVCC_FLAGS "${CUDA_NVCC_FLAGS} -gencode arch=compute_52,code=sm_52")
IF(CMAKE_CXX_COMPILER_VERSION VERSION GREATER 5)
    SET(CUDA_NVCC_FLAGS "${CUDA_NVCC_FLAGS} -D_MWAITXINTRIN_H_INCLUDED -D_FORCE_INLINES")
ENDIF()

IF(CUDA_VERSION GREATER 7.6)
    set(CUDA_NVCC_FLAGS "${CUDA_NVCC_FLAGS} -gencode arch=compute_60,code=sm_60")
    set(CUDA_NVCC_FLAGS "${CUDA_NVCC_FLAGS} -gencode arch=compute_61,code=sm_61")
    set(CUDA_NVCC_FLAGS "${CUDA_NVCC_FLAGS} -gencode arch=compute_62,code=sm_62")
```

这是我目前碰到的一些问题 以及解决办法，能力不行，口才有限，局限于此。希望可以帮助到其他人。（也懒的排版了哈哈哈）

整体来说 应该时 gcc 的问题（我这样认为，有大佬可以在给出意见!!! 供我学习）