**Whisper model fine tuning**

Please down load all packages from download\_links.txt.

For example, download this model,

<https://huggingface.co/openai/whisper-tiny/resolve/main/model.safetensors?download=true>

and then copy this into whisper-finetune-master\models\_tiny.

For other models, it is same.

**֍ Preparing Development environment**

**✅ Install Cuda Toolkit 11.8**

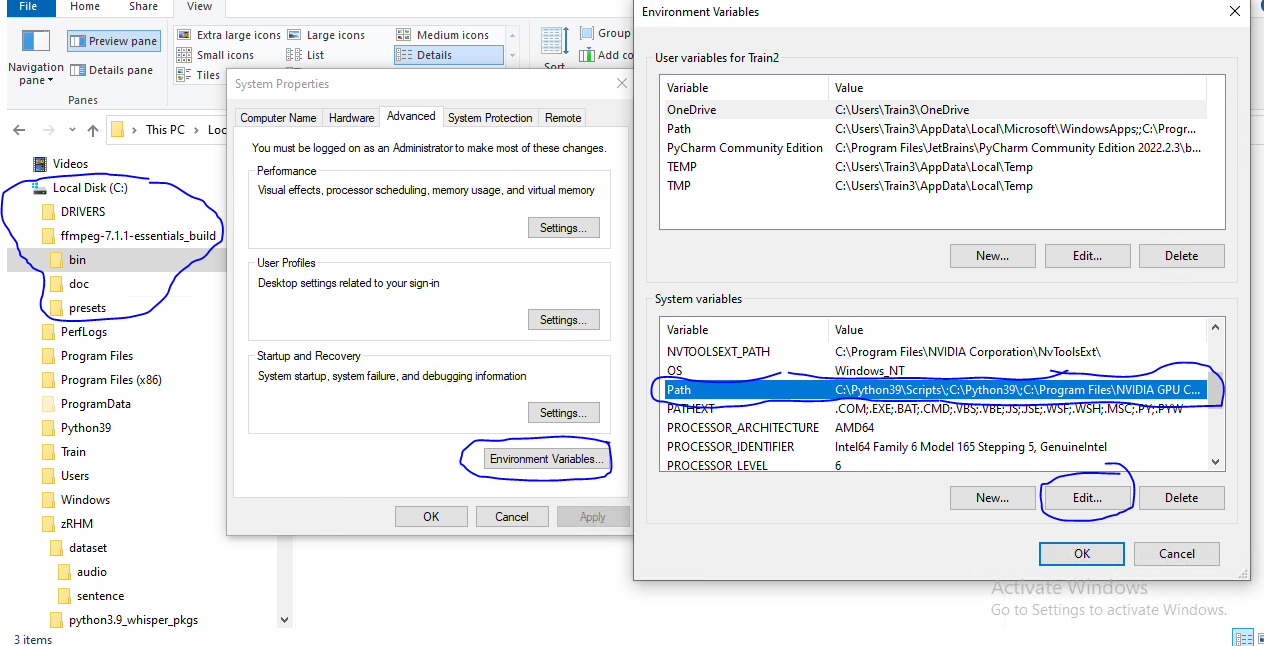
cuda\_11.8.0\_522.06\_windows.exe

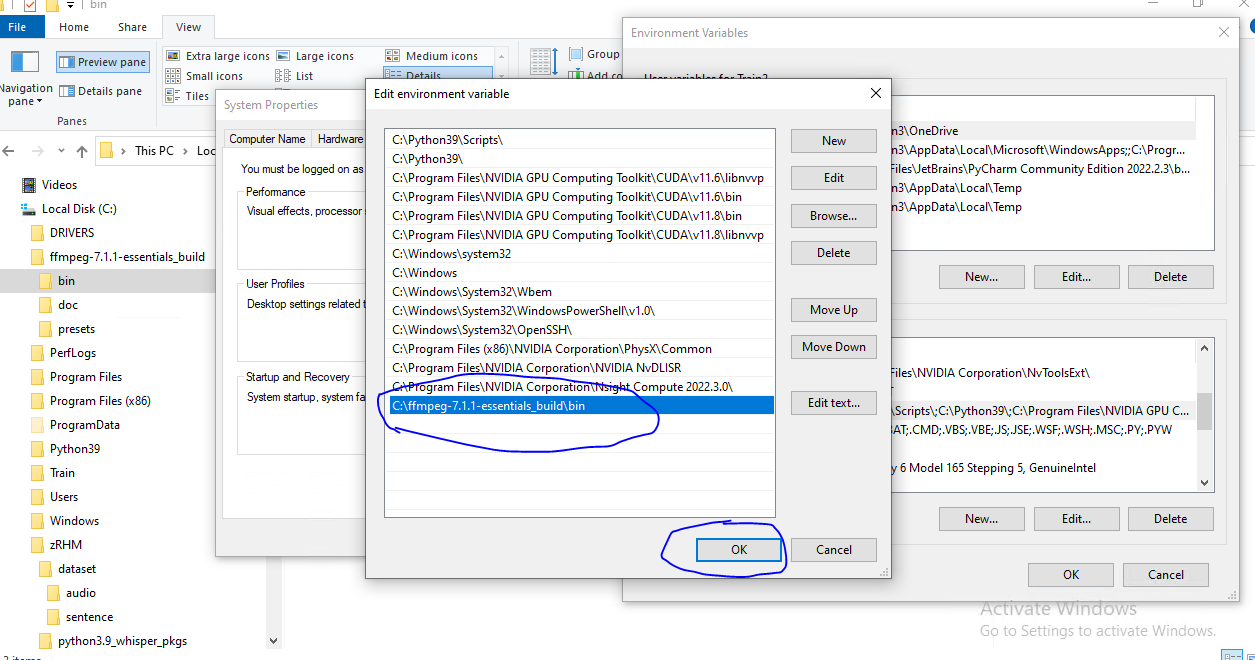
cudnn-windows-x86\_64-8.9.7.29\_cuda11-archive.zip

**✅ Install ffmpeg**

Unzip ffmpeg-7.1.1-essentials\_build.zip to “c:\\”.

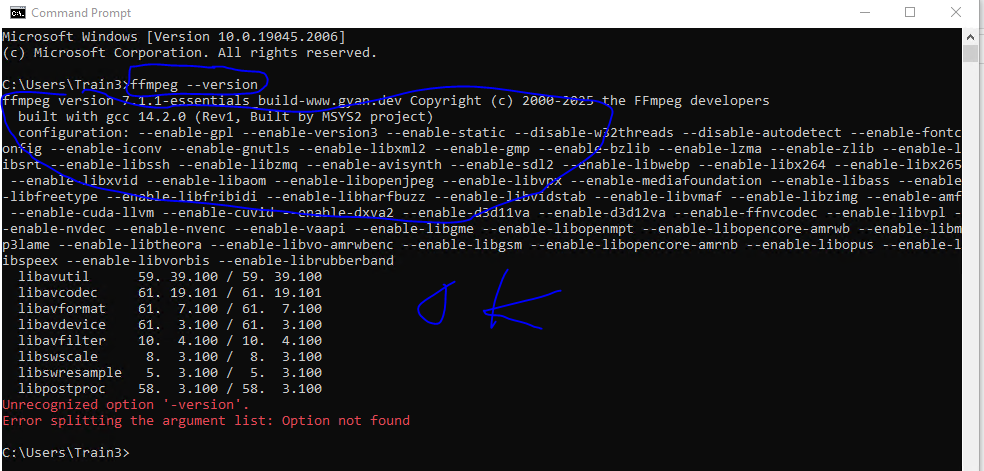
Add “C:\ffmpeg-7.1.1-essentials\_build\bin” to system path env.





After adding path, please check if it is correct.

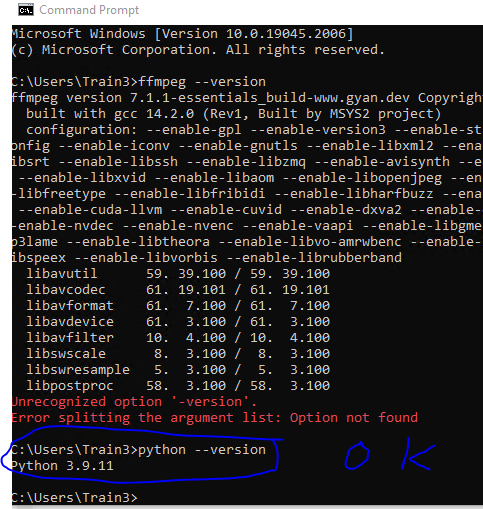
Open cmd and run ffmpeg –version



**✅ Install python**

Run python-3.9.11-amd64.exe.

Check python --version

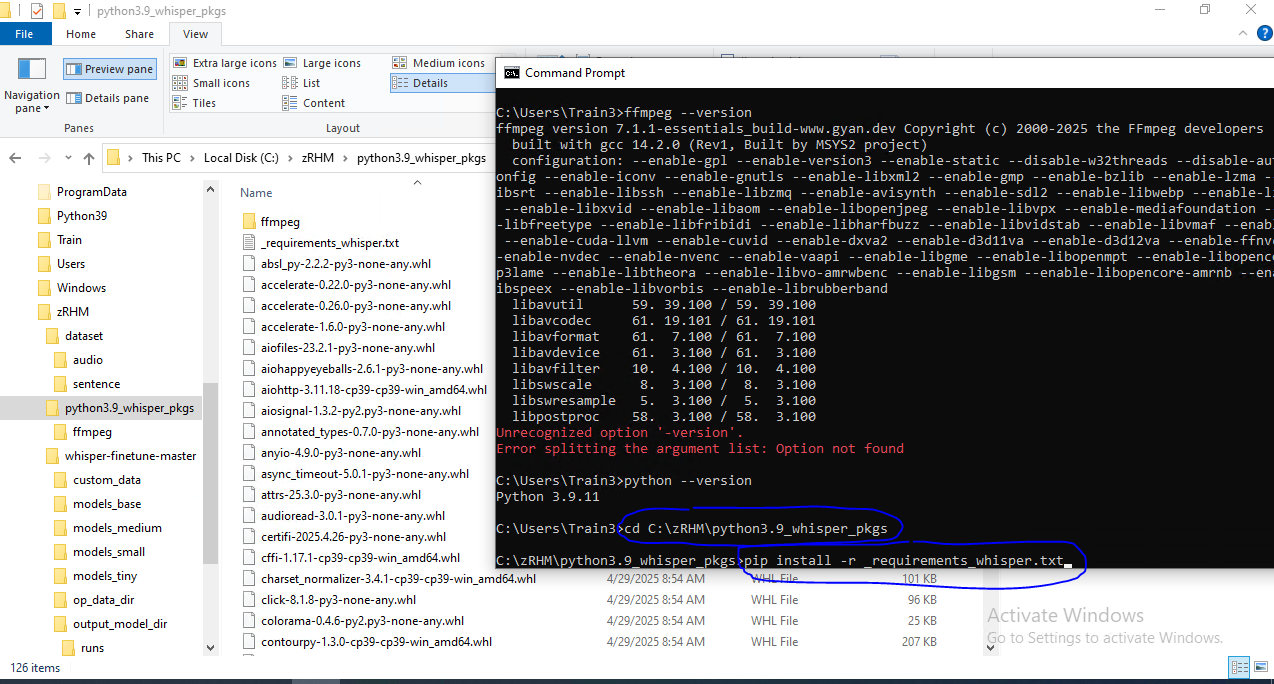


**✅ Install python packages for fine-tuning whisper models**

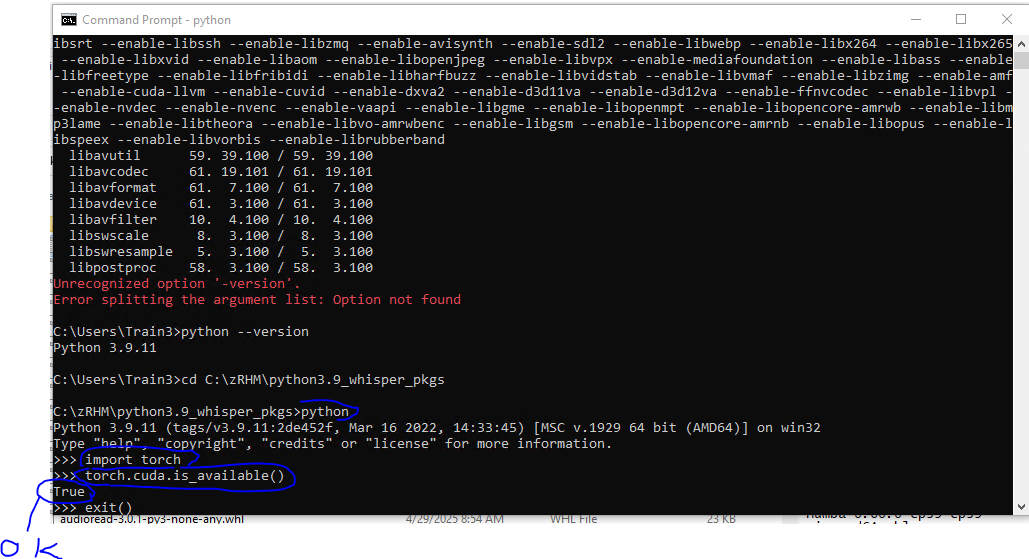
cd python3.9\_whisper\_pkgs

pip install -r \_requirements\_whisper.txt

\*\*\* pip install tensorflow-gpu --no-index --find-links F:/tensorflow/python\_package/1.15.2

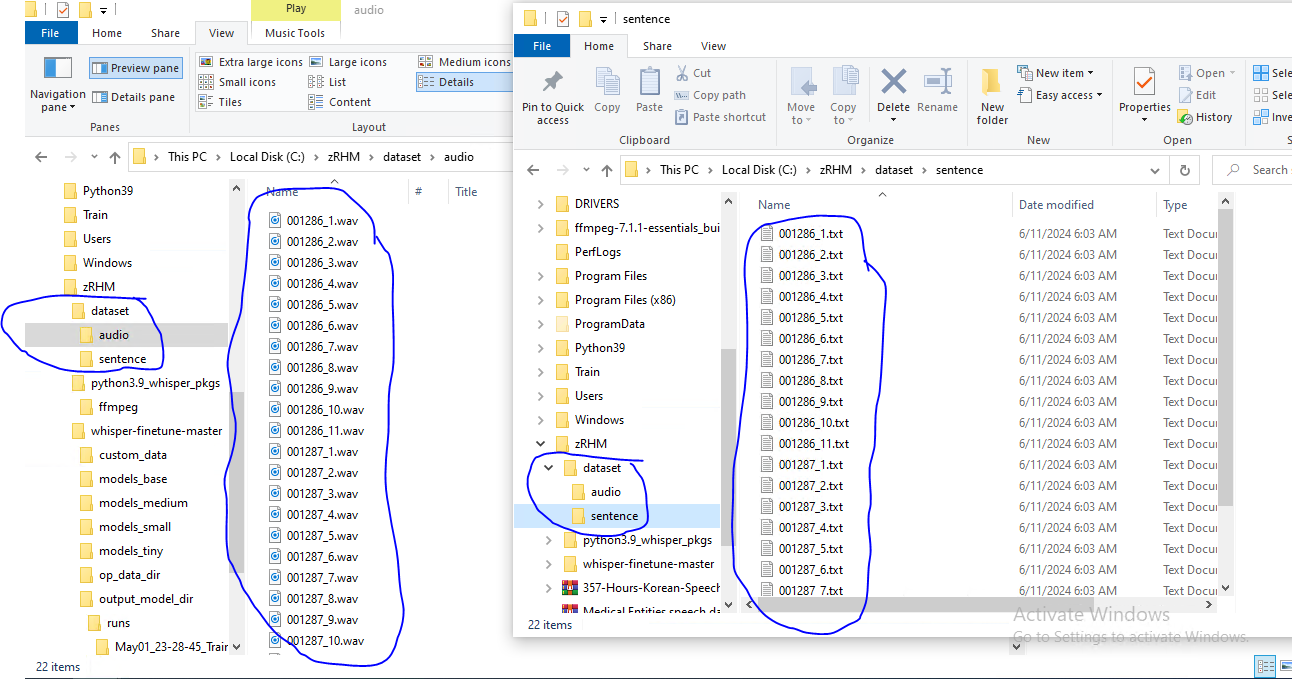


Once completed, check if torch is okay.



**֍ Preparing Dataset**

**✅ prepare your custom dataset structure as follows.**



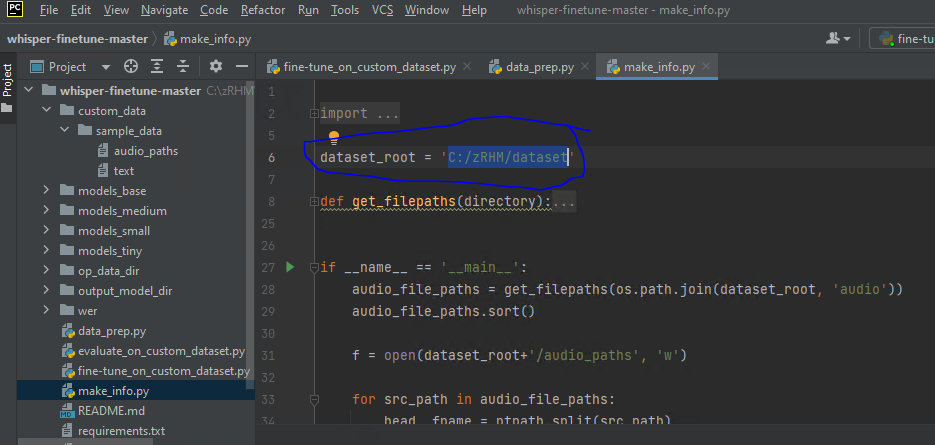
In “audio”, all wav files (16KHz 16bit mono, file name should be an unique ID, 30s or less)

In “sentence”, all text files (the same file name to wave file, transcription text, utf-8 format)

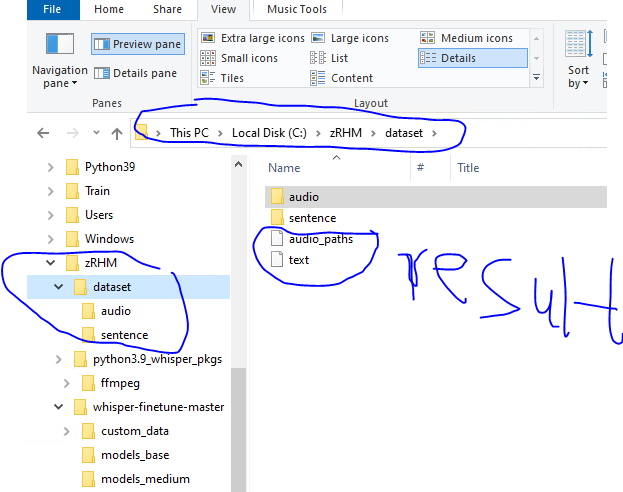
**✅ make meta info**

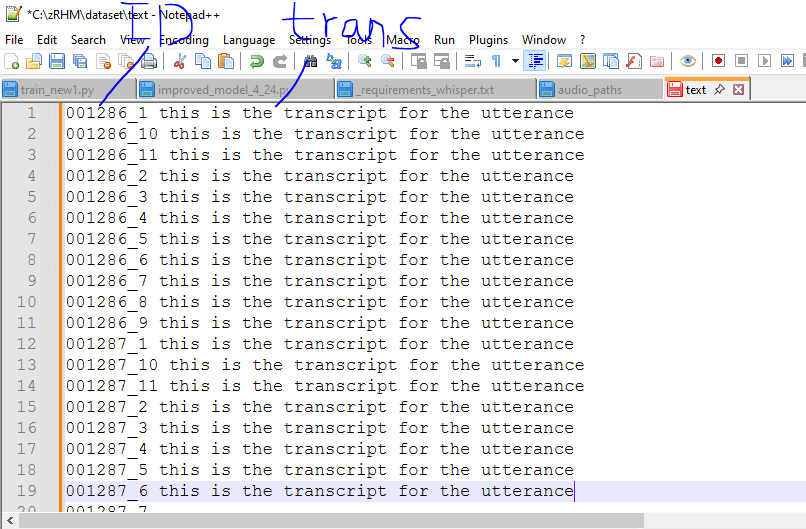
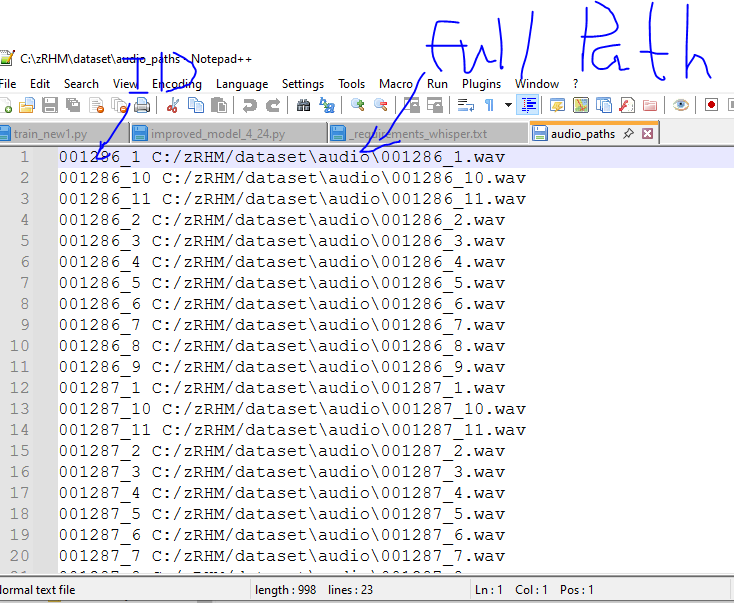
make\_info.py (in whisper-finetune-master)

make sure the dataset root directory.



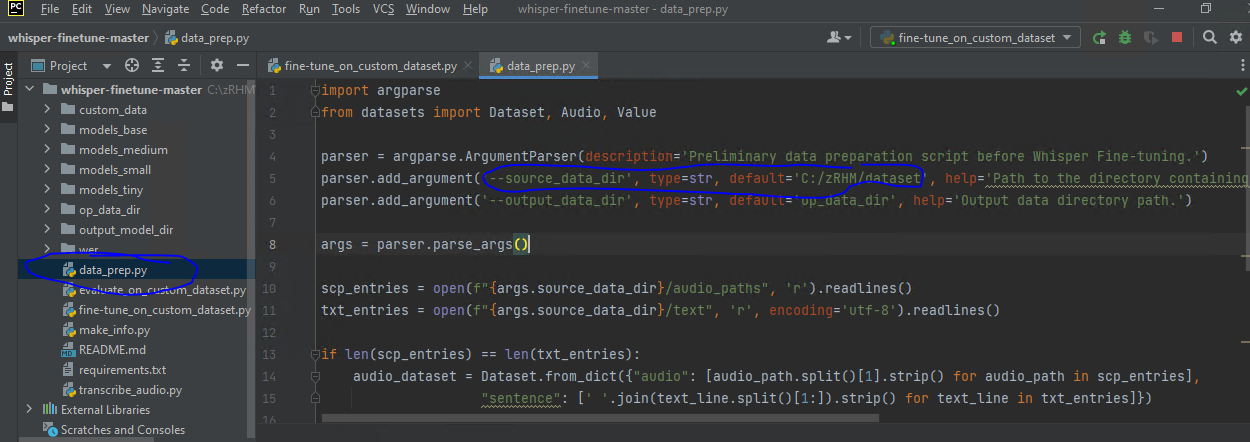
As result, you can see audio\_paths & text.



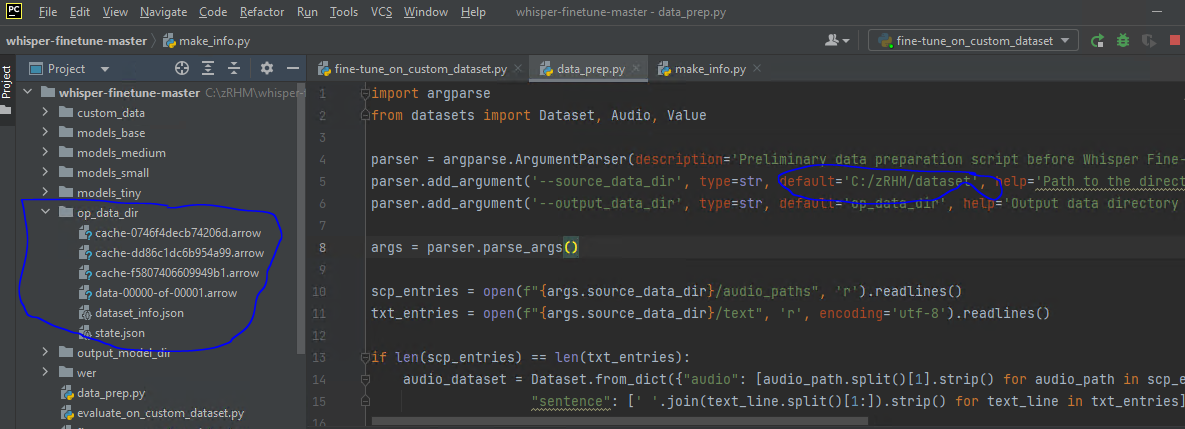


**✅ convert meta info to whisper dataset**

data\_prep.py (in whisper-finetune-master)



As result, ./op\_data\_dir



Until now, completed all preparation.

**֍ Fine tuning whisper model**

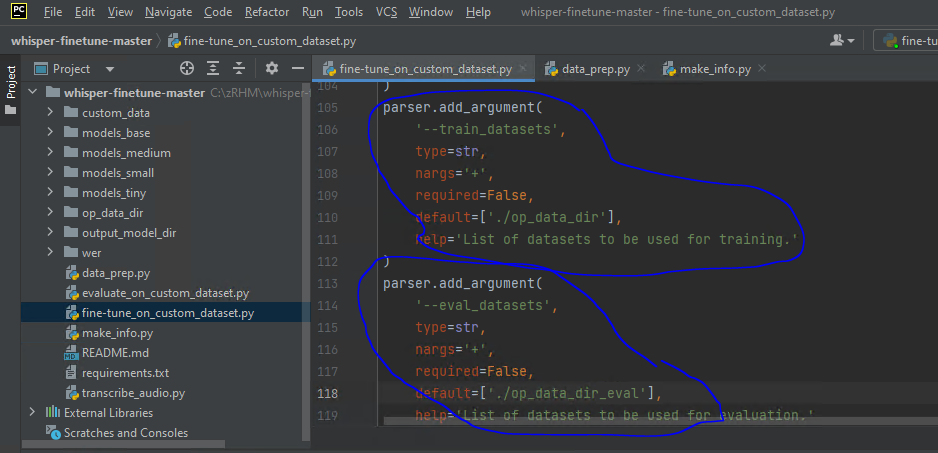
According above steps, you should make two datasets: train and eval.

For whole wave files, as ratio 8:2. If you have vast dataset, you can randomly choice a few ten thousands files for eval set. Make sure that there is no duplication between train and eval sets.

**✅ fine-tune\_on\_custom\_dataset.py** (in whisper-finetune-master)

Training Arguments.

--train\_datasets, --eval\_datasets



--model\_name

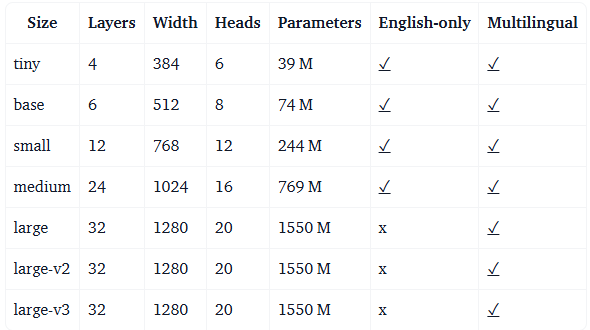
This is whisper model type. The following things are possible.

./models\_tiny

./models\_base

./models\_small

./models\_medium



On RTX3060 (vram 12GB), it is possible tiny/base/small.

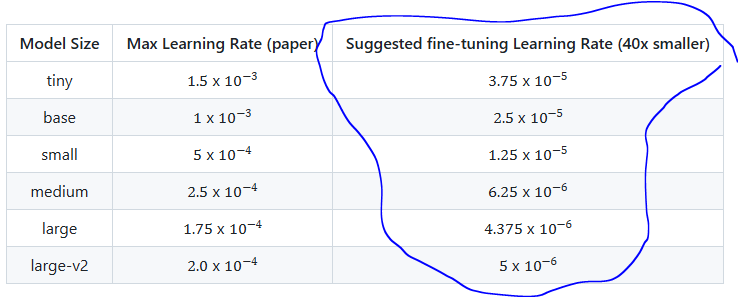
If you will use GPU version for inference, medium is a good candidate. But it is slow on CPU.

In my opinion, base or small is a good candidate for cpu.

--language

ko

--learning\_rate



The smaller, the more robust. But the smaller, the slower.

--train\_batchsize

Depending on GPU memory, you can select high value.

If occurring out of memory error, decrease this value.

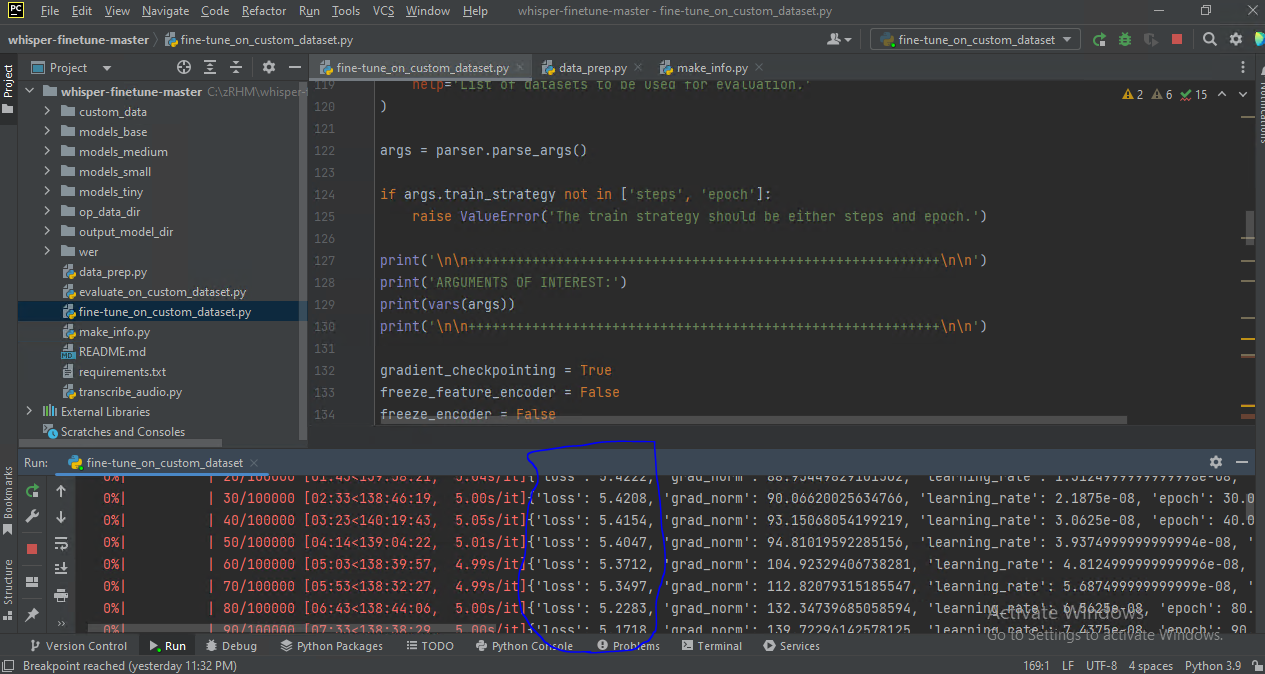
--num\_steps

Whole fine-tuning steps

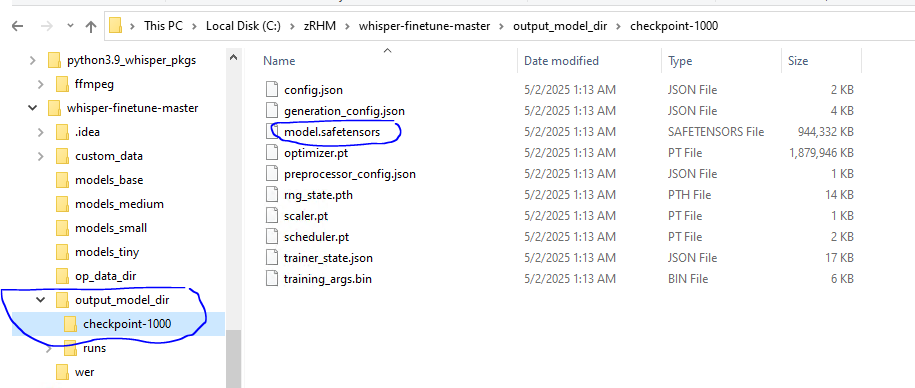
--output\_dir

Output model directory for the checkpoints generated.

After setting arguments, you can start the model fine-tuning.



You can see train loss is decreasing continually.



model.safetensors is a fine-tuned model file.

After reached --num\_steps, model fine tuning is completed.

From now, we will use this fine-tuned model to recognize wav files.

**֍ Recognizing WAV**

Whisper.cpp is a c++ version of whisper.

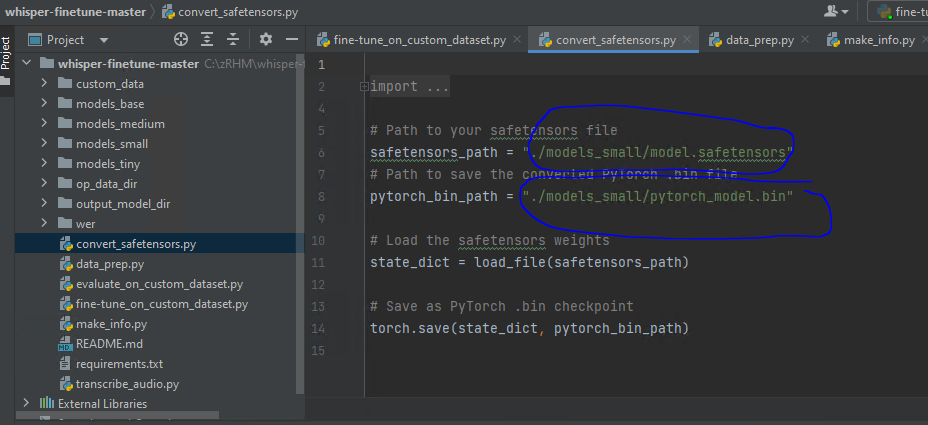
We will use this module in real environment for inference, not python.

Beforehand, you should install msvs2018.

**✅ converting model format**

First, you should convert saftensors format to PyTorch .bin.

convert\_saftensors.py (in whisper-finetune-master)



As result, you can see pytorch\_model.bin file.

**✅ converting pytorch\_model.bin to ggml format**

Whisper.cpp is using ggml model format.

For converting model format to ggml, will use module convert-h5-to-ggml.py of whisper.cpp/models.

cd whisper.cpp/models

python convert-h5-to-ggml.py fine\_tuned\_model\_dir\_path out\_dir\_path

For example,

In our case, fine\_tuned\_model\_dir\_path is “C:\zRHM\whisper-finetune-master\output\_model\_dir\checkpoint-2000”.

out\_dir\_path is “C:\zRHM\whisper-finetune-master\converted” .

✅ recognizing a wav file

Compile and build whisper\_asr project in vs2018.

whisper\_asr\_2.0\x64\Debug>audio\_recognizer.exe -m models\ggml-model\_small\_ko.bin -f 001286\_5.wav

