How to: for Windows 11 environment

**MER2025 – Track 1**

**Setup & Execution Guide – Condensed Version**

This guide explains step-by-step how to run the **Multimodal Emotion Recognition** project for **MER2025 Track 1**, including environment setup, data preparation, feature extraction, unimodal and multimodal runs, and submission generation.

**Step 1 – Create Anaconda Environment**

1. Open **Anaconda PowerShell**:

bash

CopyEdit

conda env create -f E:\BentziAddOn\Win\_Env\_vllm2MERMER\_.yml (create -f E:\UserName\Win\_Env\_vllm2MERMER\_.yml)

conda activate vllm2MER

1. Install PyTorch with CUDA:

bash

CopyEdit

pip install torch==2.2.2+cu118 torchvision==0.17.2+cu118 torchaudio==2.2.2 --index-url https://download.pytorch.org/whl/cu118

*(or cu121 if needed do not install other versions for compatibility)*

**Step 2 – Clone MER2025 Tools**

bash

CopyEdit

git clone https://github.com/zeroQiaoba/MERTools.git

cd E:\MERTools\MER2025\MER2025\_Track1

**Step 3 – Download Raw Dataset**

Source: [HuggingFace MER2025](https://huggingface.co/datasets/MERChallenge/MER2025) (HuggingFace payment token is needed)  
Only **Track 1** (~200GB) is required.

* **OpenFace** → run OpenFaceRawToTrack1.ipynb
* **Video** → run VideoRawToTrack1.ipynb
* **Audio** → unzip audio.zip into Track 1 folder

Delete the original ZIP and extracted full dataset to save disk space.

**Step 4 – Install Required Transformers**

In:

makefile

CopyEdit

E:\MERTools\MER2025\MER2025\_Track1\tools\transformers

Install:

* Audio:

bash

CopyEdit

git clone https://huggingface.co/TencentGameMate/chinese-hubert-large

* Text:

bash

CopyEdit

git clone https://huggingface.co/hfl/chinese-roberta-wwm-ext-large

* OpenFace: Download from Google Drive → tools/openface\_win\_x64

**Step 5 – Update config.py**

Example:

python

CopyEdit

PATH\_TO\_PRETRAINED\_MODELS = "E:\\MERTools\\MER2025\\MER2025\_Track1\\tools"

DATA\_DIR = {

"MER2025Raw": "E:\\Track1Raw",

"MER2025": "E:\\Track1Processed\\dataset\\mer2025-dataset-process",

}

Ensure **27,369 samples** exist for each modality.

**Step 6 – Dataset Preprocessing**

bash

CopyEdit

cd E:\MERTools\MER2025\MER2025\_Track1

python -m toolkit.preprocess.mer2025

This creates processed data in Track1Processed and the filtered subtitle CSV.

**Step 7 – Feature Extraction**

All features must be (1,1024) vectors.

**Audio (HuBERT):**

bash

CopyEdit

cd feature\_extraction/audio

python extract\_audio\_huggingface.py --dataset=MER2025 --feature\_level='UTTERANCE' --model\_name='chinese-hubert-large' --gpu=0

**Text (RoBERTa):**

bash

CopyEdit

cd feature\_extraction/text

python extract\_text\_huggingface.py --dataset=MER2025 --feature\_level='UTTERANCE' --model\_name='chinese-roberta-wwm-ext-large' --gpu=0

**Video (OpenFace → ViT):**

* Convert OpenFace vectors to 1024-dim vectors using a mean + padding script.

**Step 8 – Run Unimodal**

Test each modality separately with the Attention model:

bash

CopyEdit

python -u main-release.py --model='attention' --feat\_type='utt' --dataset=MER2025 --audio\_feature chinese-hubert-large-UTT --text\_feature chinese-roberta-wwm-ext-large --video\_feature clip-vit-large-patch14-UTT --gpu 0

**Step 9 – Run Multimodal (Top-N Fusion)**

Control script: RunMultiModal\_Ben.py  
Set parameters:

python

CopyEdit

names = ['AVT'] # Modalities combination

topn = [1, 2] # Number of top features per modality

epocs = 5 # Or 100 for full run

Run:

bash

CopyEdit

python RunMultiModal\_Ben.py

**Top-N Fusion** selects the N best features from each modality and passes them to **LLaMA 2** for fusion & reasoning.

**Step 10 – Prepare Submission**

1. Install OpenCV:

bash

CopyEdit

pip install opencv-contrib-python

1. Generate submission CSV according to competition format.

**Key Notes**

* All features must be aligned in (1,1024) dimensions.
* Remove unnecessary ZIPs to save disk space.
* Feature extraction can be parallelized across modalities.
* Top-N Fusion improves accuracy by keeping only the most relevant features.