

Readme

1.完成功能

本项目完成一个语音转换模型EVP

2.最终实现训练截图

```
(mbdf) [root@room-SYS-4028GR-TR:/data] /nlp1$ python run_talkingface.py --model=evp --dataset=evpDataset
<module 'talkingface.model.audio_driven_talkingface.evp' from '/data/tuluwei/nlp1/talkingface/model/audio_driven_talkingface/evp.py'> evp
22 Jan 23:27 INFO ['run_talkingface.py', '--model=evp', '--dataset=evpDataset']
22 Jan 23:27 INFO
General Hyper Parameters:
gpu_id = 3, 4, 5
use_gpu = True
seed = 2023
checkpoint_dir = saved/wav2lip
show_progress = True
log_wandb = False
reproducibility = True

Training Hyper Parameters:
epochs = 300
train_batch_size = 2048
learner = adam
learning_rate = 0.0001
eval_step = 1
stopping_step = 10
weight_decay = 0.0
```

```
    (2): ReLU()
  )
  (6): MaxPool2d(kernel_size=(1, 2), stride=(1, 2), padding=0, dilation=1, ceil_mode=False)
)
(emotion_encoder_fc): Sequential(
  (0): Linear(in_features=3072, out_features=2048, bias=True)
  (1): ReLU(inplace=True)
  (2): Linear(in_features=2048, out_features=128, bias=True)
  (3): ReLU(inplace=True)
)
(last_fc): Linear(in_features=128, out_features=8, bias=True)
)
(CroEn_loss): CrossEntropyLoss()
(triplet_loss): TripletMarginLoss()
)
Trainable parameters: 8107592
Trainable parameters: 8107592
train 0 {'loss': 2.1067702770233154}: 100%|
22 Jan 01:52 INFO epoch 0 training [time: 445.69s, loss: 2.0702]
Valid for 1 steps
Valid {'loss': 2.0629453659057617}: 100%|
22 Jan 01:53 INFO valid result:
loss : 2.0785579082215406
22 Jan 01:53 INFO Saving current: saved/wav2lip/evp-Jan-22-2024_01-45-22.pth
22 Jan 01:53 INFO Loading model structure and parameters from saved/wav2lip/evp-Jan-22-2024_01-45-22.pth
22 Jan 01:53 INFO {'lse': 0.01, 'ssim': 0.01}
```

```
Trainable parameters: 8107592
Trainable parameters: 8107592
train 0 {'loss': 2.106316328048706}: 100%| 29139/29139 [10:11<00:00, 47.68it/s]
22 Jan 23:57 INFO epoch 0 training [time: 611.18s, loss: 2.0703]
Valid for 1 steps
Valid {'loss': 2.0624935626983643}: 100%| 8937/8937 [01:37<00:00, 91.37it/s]
22 Jan 23:58 INFO valid result:
loss : 2.0787953066855214
22 Jan 23:58 INFO Saving current: saved/wav2lip/evp-Jan-22-2024_23-46-50.pth
22 Jan 23:58 INFO Loading model structure and parameters from saved/wav2lip/evp-Jan-22-2024_23-46-50.pth
22 Jan 23:58 INFO {'lse': 0.01, 'ssim': 0.01}
```

3.依赖安装

```
abs1-py==2.0.0
addict==2.4.0
aiosignal==1.3.1
appdirs==1.4.4
```

attrs==23.1.0
audioread==3.0.1
basicsr==1.3.4.7
cachetools==5.3.2
certifi==2020.12.5
cffi==1.16.0
charset-normalizer==3.3.2
click==8.1.7
cloudpickle==3.0.0
colorama==0.4.6
colorlog==6.7.0
contourpy==1.1.1
cycler==0.12.1
decorator==5.1.1
dlib==19.22.1
docker-pycreds==0.4.0
face-alignment==1.3.5
ffmpeg==1.4
filelock==3.13.1
fonttools==4.44.0
frozenlist==1.4.0
future==0.18.3
gitdb==4.0.11
GitPython==3.1.40
glob2==0.7
google-auth==2.23.4
google-auth-oauthlib==0.4.6
grpcio==1.59.2
hyperopt==0.2.5
idna==3.4
imageio==2.9.0
imageio-ffmpeg==0.4.5
importlib-metadata==6.8.0
importlib-resources==6.1.0
joblib==1.3.2
jsonschema==4.19.2
jsonschema-specifications==2023.7.1
kiwisolver==1.4.5
lazy_loader==0.3
librosa==0.10.1
llvmlite==0.37.0
lmdb==1.2.1
lws==1.2.7
Markdown==3.5.1
MarkupSafe==2.1.3
matplotlib==3.6.3
msgpack==1.0.7
networkx==3.1
numba==0.54.1
numpy==1.20.3
oauthlib==3.2.2
opencv-python==3.4.9.33
packaging==23.2
pandas==1.3.4
pathtools==0.1.2
Pillow==6.2.1

```
pkgutil_resolve_name==1.3.10
platformdirs==3.11.0
plotly==5.18.0
pooch==1.8.0
protobuf==4.25.0
psutil==5.9.6
pyasn1==0.5.0
pyasn1-modules==0.3.0
pycparser==2.21
pyparsing==3.1.1
python-dateutil==2.8.2
python-speech-features==0.6
pytorch-fid==0.3.0
pytz==2023.3.post1
PyWavelets==1.4.1
PyYAML==5.3.1
ray==2.6.3
referencing==0.30.2
requests==2.31.0
requests-oauthlib==1.3.1
rpds-py==0.12.0
rsa==4.9
scikit-image==0.16.2
scikit-learn==1.3.2
scipy==1.5.0
sentry-sdk==1.34.0
setproctitle==1.3.3
six==1.16.0
smmap==5.0.1
soundfile==0.12.1
soxr==0.3.7
tabulate==0.9.0
tb-nightly==2.12.0a20230126
tenacity==8.2.3
tensorboard==2.7.0
tensorboard-data-server==0.6.1
tensorboard-plugin-wit==1.8.1
texttable==1.7.0
thop==0.1.1.post2209072238
threadpoolctl==3.2.0
tomli==2.0.1
torch==1.13.1+cu116
torchaudio==0.13.1+cu116
torchvision==0.14.1+cu116
tqdm==4.66.1
trimesh==3.9.20
typing_extensions==4.8.0
tzdata==2023.3
urllib3==2.0.7
wandb==0.15.12
werkzeug==3.0.1
yapf==0.40.2
zipp==3.17.0
```

4.训练过程

1)首先将https://drive.google.com/file/d/1QjFo6oRu-PIIZII-6zPfnD_x4TW1iZ-3/view的文件下载下来放在项目中的dataset文件夹中

2)运行talkingface/data/dataset下的preprocess文件: `python preprocess.py`

3)运行 `python run_talkingface.py -model=evp -dataset=evpDataset`

5.实验中遇到的问题

1)由于是使用给定框架将evp模型嵌入,因此在实验过程中发现,有时候一些可能在框架中集成了但是需要补充的部分不容易找到,容易忽略到一部分内容。

2)一开始没有调整初始的epoch,导致在实验中可能eval的部分有问题,但是需要等很久,等到所有训练完成后才发现,这就导致时间上的浪费

3)在处理evp提供的数据集时也发现了问题,因为evp源代码提供的数据预处理部分本身是有问题的,这就导致需要自己对这部分的代码进行修改,但这样也很容易出错,因为得到的结果可能和作者期望的存在差距,这就需要仔细阅读论文以及github上的别的人提出的问题以及解答