AI-Generated voice Detection

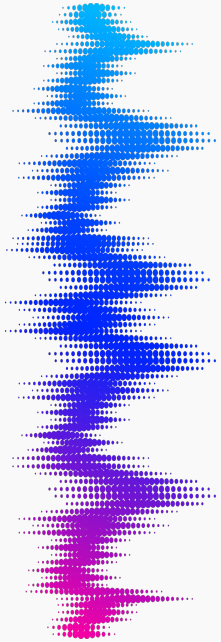
CSCE460402 - Advanced Machine Learning (2023 Fall)
Course Project Final Presentation

Ahmed Mohammed Bahssain ID: 900196218
Mokhtar Salem Ba Wahal ID: 900196209

Problem statement:

- Our project aims to enhance Deepfake audio detection algorithms due to the serious impact on people's opinions and finances.

- Simply the model will receive an audio file and return whether it is real or fake.



Baseline Model

RawNet 2, a CNN-GRU hybrid model

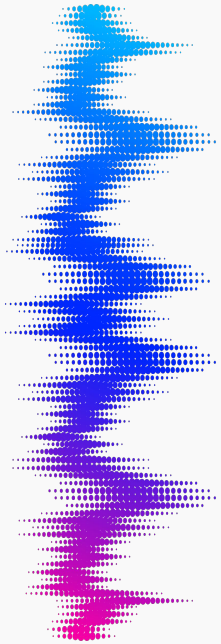
Layer	**Input : 64000 samples**	Output Shape
Since Filters	Conv (129,1,128) Maxpooling (3) BN & LeakyReLU	(21290,128)
Res Block **(X2)**	BN & LeakyRelu Conv (3,1,20) BN & LeakyReLU Conv(3,1,128) Maxpooling(3) FMS	(2365,128)
Res Block **(X4)**	BN & LeakyRelu Conv (3,1,128) BN & LeakyReLU Conv(3,1,128) Maxpooling(3) FMS	(29,128)
GRU	GRU(1024)	1024
FC	1024	1024
Output	1024	1

Updates on the Baseline

- **Adding 4 New Residual Blocks :**
  - Deeper is better :)
  - RawNet2 is an improvement of RawNet1 (By adding 2 ResBlocks)

- **Survey and fuse more reference audio data to increase versatility of the real dataset**
  - WaveFake have real audio for **only 2** people
  - Fused a new dataset (in_the_wild), with 38 hrs & 58 celebrities

- **Cross Validation :** Experiment to fine-tune the hyperparameters of the baseline for more generalization improvements, by excluding subset of the dataset



Final RawNet2

Layer	**Input : 64000 samples**	Output Shape
Since Filters	Conv (129,1,128) Maxpooling (3) BN & LeakyReLU	(21290,128)
Res Block **(X2)**	BN & LeakyRelu Conv (3,1,20) BN & LeakyReLU Conv(3,1,128) Maxpooling(2) FMS	(5298,128)
Res Block **(X4)**	BN & LeakyRelu Conv (3,1,128) BN & LeakyReLU Conv(3,1,128) Maxpooling(2) FMS	(331,128)
Res Block **(X4)**	BN & LeakyRelu Conv (3,1,128) BN & LeakyReLU Conv(3,1,128) Maxpooling(2) FMS	(20,128)
GRU	GRU(1024)	1024
FC	1024	1024
Output	1024	1

Training the models

- Both the baseline and modified models trained on each of the seven datasets
- For In distribution training, we used batch-size = 32, #samples= 3200

```
(torch) g9@csep072178g9-Alienware-Aurora-R11:~/ahmed/WaveFake$ CUDA_VISIBLE_DEVICES=1 python train_models.py dataset/LJSpeech-1.1/wa
vs dataset/generated_audio --raw_net -c -v  -b 32 --epochs 10  -a 2000 --ckpt oldModel
2023-11-11 15:20:59,703 - INFO - Loading data...
2023-11-11 15:20:59,703 - INFO - Loading data...
number of fake_training_distributions is 7
2023-11-11 15:20:59,792 - INFO - Training /home/g9/ahmed/WaveFake/dataset/generated_audio/ljspeech_multi_band_melgan
2023-11-11 15:20:59,792 - INFO - Training /home/g9/ahmed/WaveFake/dataset/generated_audio/ljspeech_multi_band_melgan
2023-11-11 15:21:01,793 - INFO - Training rawnet model on (3200,3) audio files.
2023-11-11 15:21:01,793 - INFO - Training rawnet model on (3200,3) audio files.
done init
2023-11-11 15:21:01,794 - DEBUG - Starting training for 10 epochs!
2023-11-11 15:21:01,794 - DEBUG - Starting training for 10 epochs!
2023-11-11 15:22:10,238 - INFO - [0001/0010]: 0.5832135647535324 - train acc: 69.9609375 - test_acc: 53.90625
2023-11-11 15:22:10,238 - INFO - [0001/0010]: 0.5832135647535324 - train acc: 69.9609375 - test_acc: 53.90625
2023-11-11 15:23:18,594 - INFO - [0002/0010]: 0.5053420819342136 - train acc: 76.40625 - test_acc: 72.96875
2023-11-11 15:23:18,594 - INFO - [0002/0010]: 0.5053420819342136 - train acc: 76.40625 - test_acc: 72.96875
2023-11-11 15:24:26,074 - INFO - [0003/0010]: 0.4505730792880058 3 - train acc: 79.6484375 - test_acc: 72.1875
2023-11-11 15:24:26,074 - INFO - [0003/0010]: 0.4505730792880058 3 - train acc: 79.6484375 - test_acc: 72.1875
2023-11-11 15:25:35,166 - INFO - [0004/0010]: 0.3826569659635425 - train acc: 83.3984375 - test_acc: 50.78125
2023-11-11 15:25:35,166 - INFO - [0004/0010]: 0.3826569659635425 - train acc: 83.3984375 - test_acc: 50.78125
2023-11-11 15:26:43,521 - INFO - [0005/0010]: 0.2983633386595174 7 - train acc: 87.8515625 - test_acc: 81.5625
```

```
(torch) g9@csep072178g9-Alienware-Aurora-R11:~/ahmed/WaveFake$ python train_models.py dataset/LJSpeech-1.1/wavs dataset/generated_
dio --raw_net -c -v  -b 32 --epochs 10  -a 2000 --ckpt newModel
2023-11-11 15:18:55,007 - INFO - Loading data...
2023-11-11 15:18:55,007 - INFO - Loading data...
number of fake_training_distributions is 7
2023-11-11 15:18:55,096 - INFO - Training /home/g9/ahmed/WaveFake/dataset/generated_audio/ljspeech_multi_band_melgan
2023-11-11 15:18:55,096 - INFO - Training /home/g9/ahmed/WaveFake/dataset/generated_audio/ljspeech_multi_band_melgan
2023-11-11 15:18:57,184 - INFO - Training rawnet model on (3200,3) audio files.
2023-11-11 15:18:57,184 - INFO - Training rawnet model on (3200,3) audio files.
done init
2023-11-11 15:18:57,184 - DEBUG - Starting training for 10 epochs!
2023-11-11 15:18:57,184 - DEBUG - Starting training for 10 epochs!
2023-11-11 15:20:10,675 - INFO - [0001/0010]: 0.5782105106860399 - train acc: 70.9765625 - test_acc: 70.78125
2023-11-11 15:20:10,675 - INFO - [0001/0010]: 0.5782105106860399 - train acc: 70.9765625 - test_acc: 70.78125
2023-11-11 15:21:20,860 - INFO - [0002/0010]: 0.5303499672561884 - train acc: 74.5703125 - test_acc: 65.15625
2023-11-11 15:21:20,860 - INFO - [0002/0010]: 0.5303499672561884 - train acc: 74.5703125 - test_acc: 65.15625
2023-11-11 15:22:32,351 - INFO - [0003/0010]: 0.4709834836423397 - train acc: 79.0625 - test_acc: 62.65625
2023-11-11 15:22:32,351 - INFO - [0003/0010]: 0.4709834836423397 - train acc: 79.0625 - test_acc: 62.65625
2023-11-11 15:23:42,305 - INFO - [0004/0010]: 0.3801120653748512 - train acc: 83.2421875 - test_acc: 70.625
2023-11-11 15:23:42,305 - INFO - [0004/0010]: 0.3801120653748512 - train acc: 83.2421875 - test_acc: 70.625
2023-11-11 15:24:53,336 - INFO - [0005/0010]: 0.2982563217170536 3 - train acc: 87.734375 - test_acc: 82.65625
```

Training the models (Leave one out experiment)

- Both the baseline and modified models trained on six of the seven datasets EXCEPT for one Dataset (To test Generalization of the Model on Unseed Generative Models)
- Bach Size= 32, #samples = 6160

```
(torch) g9@csep072178g9-Alienware-Aurora-R11:~/ahmed/WaveFake$ CUDA_VISIBLE_DEVICES=1 python train_models.py dataset/LJSpeech-1.1/wa
vs dataset/generated_audio --raw_net -c -v  -b 32 --epochs 10  -a 1100 --ckpt oldModel
2023-11-11 16:38:27,679 - INFO - Loading data...
2023-11-11 16:38:27,679 - INFO - Loading data...
2023-11-11 16:38:27,791 - INFO - Training out-of-distribution models!
2023-11-11 16:38:27,791 - INFO - Training out-of-distribution models!
2023-11-11 16:38:27,791 - INFO - Training all but /home/g9/ahmed/WaveFake/dataset/generated_audio/ljspeech_multi_band_melgan
2023-11-11 16:38:27,791 - INFO - Training all but /home/g9/ahmed/WaveFake/dataset/generated_audio/ljspeech_multi_band_melgan
2023-11-11 16:38:30,270 - INFO - Training rawnet model on 6160 audio files.
2023-11-11 16:38:30,270 - INFO - Training rawnet model on 6160 audio files.
done init
2023-11-11 16:38:30,271 - DEBUG - Starting training for 10 epochs!
2023-11-11 16:38:30,271 - DEBUG - Starting training for 10 epochs!
2023-11-11 16:40:41,898 - INFO - [0001/0010]: 1.131470904334799 - train acc: 58.94886363636363 - test_acc: 69.4078947368421
2023-11-11 16:40:41,898 - INFO - [0001/0010]: 1.131470904334799 - train acc: 58.94886363636363 - test_acc: 69.4078947368421
2023-11-11 16:42:55,692 - INFO - [0002/0010]: 1.0826553783633492 - train acc: 64.40746753246754 - test_acc: 30.427631578947366
2023-11-11 16:42:55,692 - INFO - [0002/0010]: 1.0826553783633492 - train acc: 64.40746753246754 - test_acc: 30.427631578947366
2023-11-11 16:45:05,696 - INFO - [0003/0010]: 0.9709476576997088 - train acc: 68.62824675324676 - test_acc: 79.52302631578947
2023-11-11 16:45:05,696 - INFO - [0003/0010]: 0.9709476576997088 - train acc: 68.62824675324676 - test_acc: 79.52302631578947
2023-11-11 16:47:14,344 - INFO - [0004/0010]: 0.8580387004397132 - train acc: 73.09253246753246 - test_acc: 84.62171052631578
2023-11-11 16:47:14,344 - INFO - [0004/0010]: 0.8580387004397132 - train acc: 73.09253246753246 - test_acc: 84.62171052631578
2023-11-11 16:49:23,028 - INFO - [0005/0010]: 0.7212283907385616 - train acc: 77.55681818181817 - test_acc: 87.17105263157895
2023-11-11 16:49:23,028 - INFO - [0005/0010]: 0.7212283907385616 - train acc: 77.55681818181817 - test_acc: 87.17105263157895
2023-11-11 16:51:32,452 - INFO - [0006/0010]: 0.6803509974247449 - train acc: 79.28165584415584 - test_acc: 86.67763157894737
2023-11-11 16:51:32,452 - INFO - [0006/0010]: 0.6803509974247449 - train acc: 79.28165584415584 - test_acc: 86.67763157894737
2023-11-11 16:53:42,604 - INFO - [0007/0010]: 0.6046076734344681 - train acc: 82.73133116883116 - test_acc: 61.01973684210527
2023-11-11 16:53:42,604 - INFO - [0007/0010]: 0.6046076734344681 - train acc: 82.73133116883116 - test_acc: 61.01973684210527
2023-11-11 16:55:52,601 - INFO - [0008/0010]: 0.5197660670458496 - train acc: 84.07061688311688 - test_acc: 80.50986842105263
2023-11-11 16:55:52,601 - INFO - [0008/0010]: 0.5197660670458496 - train acc: 84.07061688311688 - test_acc: 80.50986842105263
```

```
(torch) g9@csep072178g9-Alienware-Aurora-R11:~/ahmed/WaveFake$ python train_models.py dataset/LJSpeech-1.1/wavs dataset/gen
dio --raw_net -c -v  -b 32 --epochs 10  -a 1100 --ckpt newModel
2023-11-11 16:39:35,755 - INFO - Loading data...
2023-11-11 16:39:35,755 - INFO - Loading data...
2023-11-11 16:39:35,841 - INFO - Training out-of-distribution models!
2023-11-11 16:39:35,841 - INFO - Training out-of-distribution models!
2023-11-11 16:39:35,841 - INFO - Training all but /home/g9/ahmed/WaveFake/dataset/generated_audio/ljspeech_multi_band_melga
2023-11-11 16:39:35,841 - INFO - Training all but /home/g9/ahmed/WaveFake/dataset/generated_audio/ljspeech_multi_band_melga
2023-11-11 16:39:38,240 - INFO - Training rawnet model on 6160 audio files.
2023-11-11 16:39:38,240 - INFO - Training rawnet model on 6160 audio files.
done init
2023-11-11 16:39:38,241 - DEBUG - Starting training for 10 epochs!
2023-11-11 16:39:38,241 - DEBUG - Starting training for 10 epochs!
2023-11-11 16:41:56,027 - INFO - [0001/0010]: 1.1446207734671505 - train acc: 58.80681818181818 - test_acc: 60.115131578947
2023-11-11 16:41:56,027 - INFO - [0001/0010]: 1.1446207734671505 - train acc: 58.80681818181818 - test_acc: 60.115131578947
2023-11-11 16:44:12,575 - INFO - [0002/0010]: 1.0825966933330933 - train acc: 60.4099025974026 - test_acc: 69.736842105263
2023-11-11 16:44:12,575 - INFO - [0002/0010]: 1.0825966933330933 - train acc: 60.4099025974026 - test_acc: 69.736842105263
2023-11-11 16:46:26,732 - INFO - [0003/0010]: 0.985130048804469 - train acc: 66.37581168831169 - test_acc: 87.253289473684
2023-11-11 16:46:26,732 - INFO - [0003/0010]: 0.985130048804469 - train acc: 66.37581168831169 - test_acc: 87.253289473684
2023-11-11 16:48:40,932 - INFO - [0004/0010]: 0.8713517080653798 - train acc: 72.07792207792207 - test_acc: 68.421052631578
2023-11-11 16:48:40,932 - INFO - [0004/0010]: 0.8713517080653798 - train acc: 72.07792207792207 - test_acc: 68.421052631578
2023-11-11 16:50:55,258 - INFO - [0005/0010]: 0.8180463695293897 - train acc: 73.27516233766234 - test_acc: 67.269736842105
2023-11-11 16:50:55,258 - INFO - [0005/0010]: 0.8180463695293897 - train acc: 73.27516233766234 - test_acc: 67.269736842105
2023-11-11 16:53:10,714 - INFO - [0006/0010]: 0.7033039331436157 - train acc: 76.66396103896103 - test_acc: 83.141447368421
2023-11-11 16:53:10,714 - INFO - [0006/0010]: 0.7033039331436157 - train acc: 76.66396103896103 - test_acc: 83.141447368421
2023-11-11 16:55:25,663 - INFO - [0007/0010]: 0.6441707542383825 - train acc: 79.62662337662337 - test_acc: 80.263157894736
2023-11-11 16:55:25,663 - INFO - [0007/0010]: 0.6441707542383825 - train acc: 79.62662337662337 - test_acc: 80.263157894736
2023-11-11 16:57:41,788 - INFO - [0008/0010]: 0.5924798145696715 - train acc: 81.37175324675324 - test_acc: 87.664473684210
```

Comparing aEER of BaseLine & Updated RawNet2 Model

	All_But_Full_Band_Melgan		All_But_Melgan		All_But_MB_MelGan		All_But_HiFI_GAN		All_But_PWG	
	BaseLine	Updated	Base Line	Updated	BaseLine	Updated	BaseLine	Updated	Base Line	Updated
Full_Band_Melgan	0.345	0.335	0.344	0.2549	0.260	0.3349	0.2299	0.18	0.325	0.1749
HiFi-GAN	0.285	0.2899	0.34	0.29	0.245	0.33	0.29	0.25	0.3249	0.215
MelGan	0.15	0.1249	0.255	0.1348	0.125	0.14	0.07	0.05	0.0949	0.0649
MelGAN (L)	0.14	0.1149	0.1849	0.115	0.1149	0.145	0.069	0.069	0.1049	0.04
PWG	0.17	0.17	0.194	0.135	0.115	0.195	0.11	0.104	0.195	0.15
WaveGlow	0.10	0.13	0.1999	0.1249	0.11	0.1449	0.08	0.05	0.075	0.02
MB-MelGAN	0.205	0.1555	0.22	0.125	0.205	0.289	0.1349	0.065	0.21	0.1349

Fusing the new Dataset

- Test Accuracy after fusing the new dataset (~91%)
-
- EER is lower when we train on the new dataset separately (0.095)

- After fusing WaveFake & in_the_wild datasets : The model performs much better on the old dataset
  - On the old dataset, EER ~ 0.008
  - On the new dataset, EER ~ 0.36 :(

```
(myenv) group09-f2023@group09f2023:~/WaveFake/DeepMl_FakeDetectv$ python eva_new_dat
ta/generated fffff/mfcc/raw_net/model_with_new/ --raw_net -c -a 1200 --output result
Model: {'data_set_name': 'fffff/mfcc/raw_net/model_with_new/ckpt.pth'}
2023-12-05 21:03:14,719 - INFO - Evaluating /home/group09-f2023/WaveFake/DeepMl_Fake
ated/p1...
Size: 240
before path fffff/mfcc/raw_net/model_with_new/ckpt.pth
Model path fffff/mfcc/raw_net/model_with_new/ckpt.pth
2023-12-05 21:03:16,660 - INFO - calculatnig EER for model trained on new dataset
2023-12-05 21:03:18,608 - INFO - /home/group09-f2023/WaveFake/DeepMl_FakeDetectv/dat
        EER: 0.09583333333333365 Thresh: -0.0051432182081052
```

```
2023-12-05 21:08:04,077 - INFO - calculatnig EER for model trained on new dataset
2023-12-05 21:08:09,948 - INFO - /home/group09-f2023/WaveFake/DeepMl_FakeDetectv/dataset/generated_aud
io/ljspeech_melgan_large:
        EER: 0.008333333334626257 Thresh: -1.5095536168565852e-06
2023-12-05 21:08:15,473 - INFO - /home/group09-f2023/WaveFake/DeepMl_FakeDetectv/dataset/generated_aud
io/ljspeech_parallel_wavegan:
        EER: 0.008333333334654 Thresh: -1.3852443502440743e-06
2023-12-05 21:08:17,327 - INFO - /home/group09-f2023/WaveFake/DeepMl_FakeDetectv/dataset/generated_aud
io/p1:
        EER: 0.36249999999882226 Thresh: -0.9999972581863403
2023-12-05 21:08:23,031 - INFO - /home/group09-f2023/WaveFake/DeepMl_FakeDetectv/dataset/generated_aud
io/ljspeech_hifiGAN:
        EER: 0.008333333333334072 Thresh: -1.6472110928623211e-06
2023-12-05 21:08:28,714 - INFO - /home/group09-f2023/WaveFake/DeepMl_FakeDetectv/dataset/generated_aud
io/ljspeech_melgan:
        EER: 0.008333333333334119 Thresh: -1.4908500816101545e-06
2023-12-05 21:08:34,285 - INFO - /home/group09-f2023/WaveFake/DeepMl_FakeDetectv/dataset/generated_aud
io/ljspeech_full_band_melgan:
        EER: 0.008333333333333335 Thresh: -1.5679636362620467e-06
2023-12-05 21:08:39,769 - INFO - /home/group09-f2023/WaveFake/DeepMl_FakeDetectv/dataset/generated_aud
io/ljspeech_multi_band_melgan:
        EER: 0.00833333333333406 Thresh: -1.5121239584924042e-06
2023-12-05 21:08:45,299 - INFO - /home/group09-f2023/WaveFake/DeepMl_FakeDetectv/dataset/generated_aud
io/ljspeech_waveglow:
        EER: 0.008333333334654 Thresh: -1.5904138350158405e-06
```

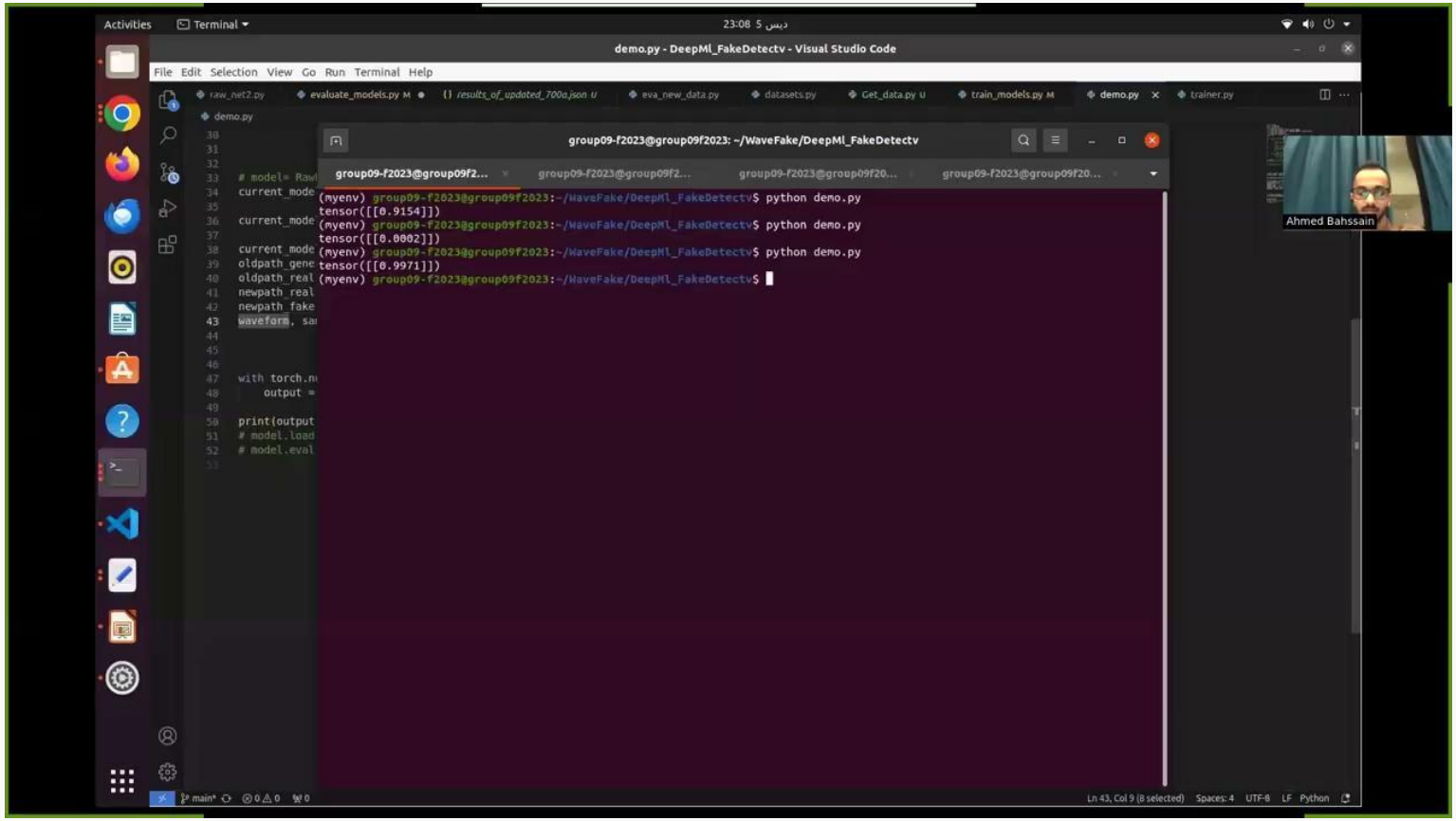
Conclusion

- Deep-Fakes is a real threat to our society especially with the current political situation in the region.
- Error is harmful even if it is small.
- Going deeper is usually a good idea :)
- Hardware limitations should be considered beforehand

Lessons learned:

- PyTorch is fun, cleaner, and object-oriented, we enjoyed it.
- Choosing a hard problem and complicated code helped us learn many thing about design machine learning systems that includes dealing with the infrastructure and hardware.
- Never stack a sigmoid activation after a softmax :)
- This was a major problem we face, softmax in the model file, sigomd in the trainer.

Demo


Activities
Terminal
23:08 5 ديس
demo.py - DeepMl_FakeDetectv - Visual Studio Code
File Edit Selection View Go Run Terminal Help
raw_net2.py
evaluate_models.py M
results_of_updated_700a.json U
eva_new_data.py
datasets.py
Get_data.py U
train_models.py M
demo.py
trainer.py
group09-f2023@group09f2023: ~/WaveFake/DeepMl_FakeDetectv
(myenv) group09-f2023@group09f2023:~/WaveFake/DeepMl_FakeDetectv$ python demo.py
tensor([[0.9154]])
(myenv) group09-f2023@group09f2023:~/WaveFake/DeepMl_FakeDetectv$ python demo.py
tensor([[0.0002]])
(myenv) group09-f2023@group09f2023:~/WaveFake/DeepMl_FakeDetectv$ python demo.py
tensor([[0.9971]])
(myenv) group09-f2023@group09f2023:~/WaveFake/DeepMl_FakeDetectv$
Ln 43, Col 9 (8 selected)
Spaces: 4
UTF-8
LF
Python
Ahmed Bahssain

Member's contribution

- Ahmed:
    - Fixed the GPU Drivers to train the model.
    - Updated the architecture of the RawNet2
    - Added the new dataset
- MokhtaR:
    - Traced the Baseline code to fix the no learning problem
    - Fixed the errors in the evaluation and adapted trainer of the model to train both the baseline and updated models

- Both trained the models and made observations on the conducted experiments

References

- Frank, J., & Schönherr, L. (2021). Wavefake: A data set to facilitate audio deepfake detection. arXiv Preprint. https://doi.org/arXiv:2111.02813
- Hemlata Tak, Jose Patino, Massimiliano Todisco, Andreas Nautsch, Nicholas Evans, and Anthony Larcher. End-to-End anti-spoofing with RawNet2. In International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021.
- https://www.kaggle.com/datasets/andreadiubaldo/wavefake-test
- https://deepfake-demo.aisec.fraunhofer.de/in_the_wild