

Off-the-shelf HuggingFace models for audio deepfake detection

Adriana STAN

www.adrianastan.com



Cluj-Napoca, 29.10.2024

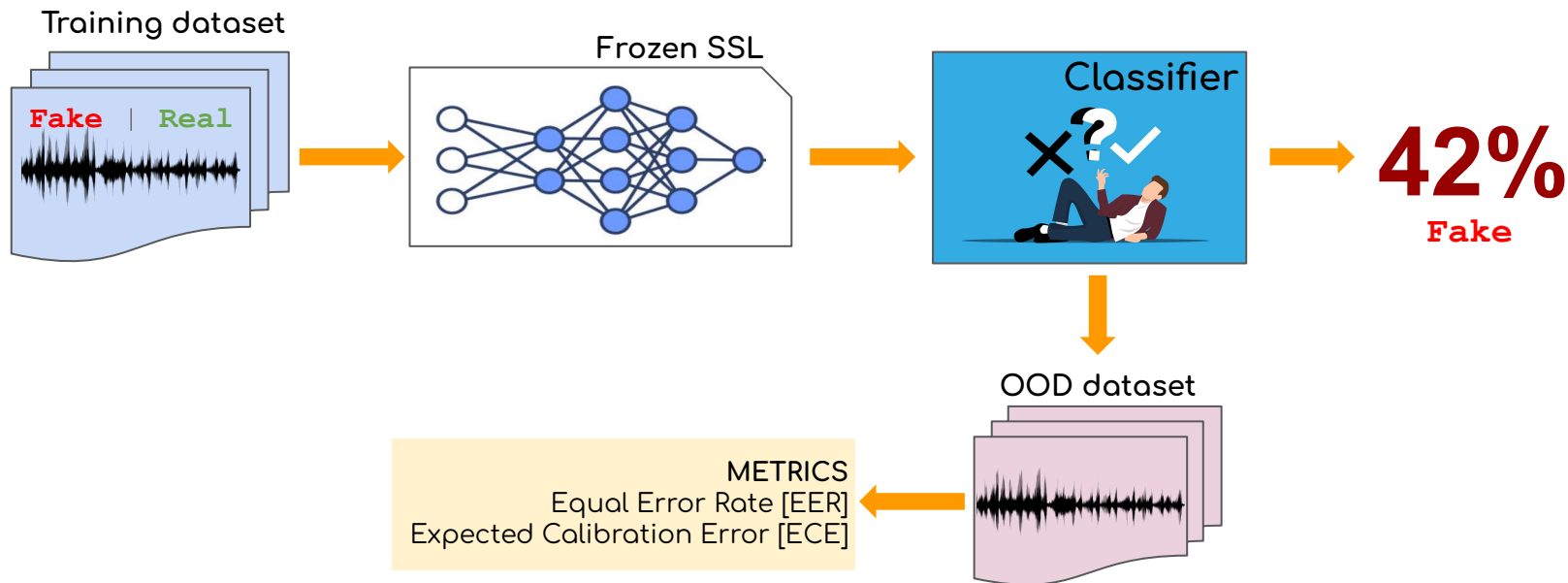
Audio deepfakes

Generating audio deepfakes

- Text-to-speech
- Voice cloning



Deepfake detection

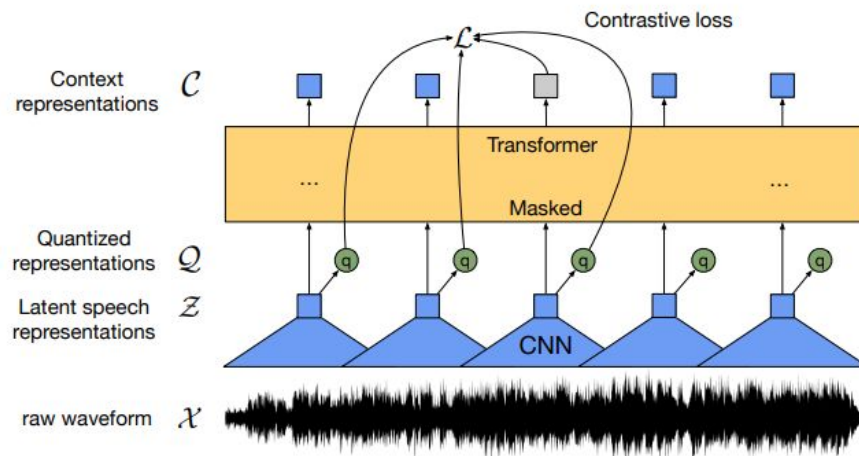


Q: SSL models?

Audio SSL: wav2vec2 and wavLM

wav2vec2

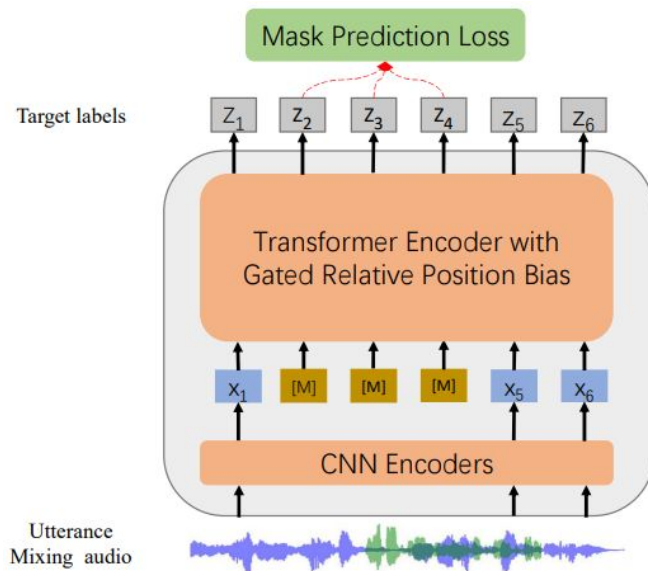
<https://huggingface.co/collections/facebook/wav2vec-20-651e865258e3dee2586c89f5>



<https://arxiv.org/abs/2006.11477>

wavLM

<https://huggingface.co/microsoft/wavlm-large>



<https://arxiv.org/abs/2110.13900>



Hugging Face



Hugging Face

- **Open-Source ML Hub:** Hugging Face is a popular platform providing an extensive repository of pre-trained machine learning models, particularly for Natural Language Processing (NLP), such as transformers for tasks like text generation, translation, and question-answering.
- **Transformers Library:** It offers the widely-used Transformers library, which makes it easy to integrate and fine-tune state-of-the-art transformer models from BERT, GPT, T5, and many others in a few lines of code.
- **Community and Collaboration:** Hugging Face fosters a collaborative environment where researchers and developers can share models and datasets, and it also provides tools for deploying, managing, and scaling ML models through their Inference API and Spaces platform.



Hugging Face

Models 1,144

audio

Full-text search

Sort: Trending

stabilityai/stable-audio-open-1.0

Text-to-Audio • Updated Jul 31 • 21.3k • 906

nvidia/audio-flamingo

Updated 13 days ago • 16

LeroyDyer/SpydazWeb_AI_Text_AudioVision_Project

Text Generation • Updated 4 days ago • 8 • 2

MIT/ast-finetuned-audioset-10-10-0.4593

Audio Classification • Updated Sep 6, 2023 • 75.9M • 257

MU-NLPC/whisper-tiny-audio-captioning

Updated Mar 12 • 141 • 11

descript/descript-audio-codec

Updated Jun 14, 2023 • 11

Jmica/audiobook_maker

Updated Jun 10 • 13

slseanwu/beats-conformer-bart-audio-captioner

Updated Jan 6 • 10 • 5

audio/stable-audio-open-1.0

Updated Jun 5 • 6

facebook/audiogen-medium

Updated Mar 12 • 3.38k • 115

Qwen/Qwen2-Audio-7B-Instruct

Text2Text Generation • Updated Aug 9 • 77.4k • 211

Talha/urdu-audio-emotions

Audio Classification • Updated Jul 2, 2022 • 27 • 11

bookbot/distil-ast-audioset

Audio Classification • Updated Sep 12, 2023 • 947 • 15

MU-NLPC/whisper-small-audio-captioning

Updated Mar 13 • 1.3k • 9

cvssp/audioldm2

Updated Apr 16 • 3.62k • 37

Qwen/Qwen-Audio-Chat

Text Generation • Updated Dec 8, 2023 • 32.2k • 71

hf-audio/wav2vec2-bert-CV16-en

Automatic Speech Recognition • Updated Jan 15 • 1.03k • 5

MelodyMachine/Deepfake-audio-detection-V2

Audio Classification • Updated Jun 18 • 617 • 6



Hugging Face

<https://huggingface.co/facebook>

Wav2Vec 2.0

A collection for the first release of Wav2Vec 2.0, a speech encoder that learns ...

- facebook/wav2vec2-large-960h-lv60-self
 - Automatic Speech Recognition · Updated May 23, 2022 · 29.5k · 125
- facebook/wav2vec2-large-960h
 - Automatic Speech Recognition · Updated Apr 5, 2022 · 49.4k · 24
- facebook/wav2vec2-base-960h
 - Automatic Speech Recognition · Updated Nov 14, 2022 · 1.35M · 289
- facebook/wav2vec2-base-100h

SeamlessM4T

SeamlessM4T is designed to provide high quality translation, allowing people ...

Seamless M4T 945

- facebook/hf-seamless-m4t-large
 - Text-to-Speech · Updated Dec 8, 2023 · 2.86k · 54
- facebook/hf-seamless-m4t-medium
 - Text-to-Speech · Updated Dec 8, 2023 · 8.14k · 28
- facebook/seamless-m4t-large

XLSR

A collection of multilingual Wav2Vec 2.0 checkpoints pre-trained on 53 langua...

- facebook/wav2vec2-large-xlsr-53
 - Updated Mar 18, 2022 · 120k · 111
- facebook/wav2vec2-xlsr-53-espeak-cv-ft
 - Automatic Speech Recognition · Updated Dec 10, 2021 · 354k · 23
- facebook/wav2vec2-large-xlsr-53-dutch
 - Automatic Speech Recognition · Updated Jul 6, 2021 · 47 · 2
- facebook/wav2vec2-large-xlsr-53-french

XLS-R

First release checkpoints for XLS-R, a large-scale model for cross-lingual spee...

- facebook/wav2vec2-xls-r-300m
 - Updated Aug 10, 2022 · 664k · 76
- facebook/wav2vec2-xls-r-1b
 - Updated Aug 10, 2022 · 9.68k · 27
- facebook/wav2vec2-xls-r-2b
 - Updated Aug 10, 2022 · 1.46k · 31
- facebook/wav2vec2-xls-r-300m-en-to-15

Robust Wav2Vec 2.0

A collection of "robust" Wav2Vec 2.0 checkpoints pre-trained on datasets fro...

- facebook/wav2vec2-large-robust
 - Updated Nov 5, 2021 · 2.09k · 30
- facebook/wav2vec2-large-robust-ft-libri-960h
 - Automatic Speech Recognition · Updated Jun 23, 2023 · 32.4k · 11
- facebook/wav2vec2-large-robust-ft-swbd-300h
 - Automatic Speech Recognition · Updated Apr 5, 2022 · 542 · 17

Robust wav2vec 2.0: Analyzing Domain Shift in Self-Supervised Pre-T...

VoxPopuli v2

A collection of checkpoints from the second VoxPopuli release.

VoxPopuli: A Large-Scale Multilingual Speech Corpus for Representat...

Paper · 2101.00390 · Published Jan 2, 2021 · 1

- facebook/wav2vec2-base-bg-voxpupuli-v2
 - Automatic Speech Recognition · Updated Feb 27, 2022 · 14
- facebook/wav2vec2-base-cs-voxpupuli-v2

VoxPopuli

A collection of open-source artefacts (datasets + checkpoints) from the first Vo...

- facebook/voxpupuli
 - Viewer · Updated Oct 14, 2022 · 169k · 7.86k · 84

VoxPopuli: A Large-Scale Multilingual Speech Corpus for Representat...

Paper · 2101.00390 · Published Jan 2, 2021 · 1

- facebook/wav2vec2-base-100k-voxpupuli
 - Automatic Speech Recognition · Updated Nov 5, 2021 · 192 · 3
- facebook/wav2vec2-base-10k-voxo0011-ft-cs

HuBERT

A collection of checkpoints from the HuBERT release, a speech encoder that le...

HuBERT: Self-Supervised Speech Representation Learning by Maske...

Paper · 2106.07447 · Published Jun 14, 2021 · 2

- facebook/hubert-base-ls960
 - Feature Extraction · Updated Nov 5, 2021 · 59.4k · 45
- facebook/hubert-large-l160k

Running the experiment...

<https://colab.research.google.com/drive/1qPq1QNMTLsCGgUhoa1EWL5jJV-KTAPrn?usp=sharing>

Thank you for your attention!

Questions?



Code: github.com/adrianastan/deepfake_pydata

adriana.stan@gmail.com
www.adrianastan.com