Load Wav2Vec

```
import torch
import librosa
import numpy as np
from tqdm import tqdm
from transformers import Wav2Vec2Model, Wav2Vec2Processor
```

```
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
print("Device:", device)
```

```
Device: cuda
```

```
processor = Wav2Vec2Processor.from_pretrained("facebook/wav2vec2-base")
wav2vec = Wav2Vec2Model.from_pretrained("facebook/wav2vec2-base")

wav2vec = wav2vec.to(device)

wav2vec.eval()
```

```
Loading weights: 100%                                                                211/211 [00:00<00:00, 685.35it/s, Materializing param=masked_spec_embed]
Wav2Vec2Model LOAD REPORT from: facebook/wav2vec2-base
Key                           | Status     |  |
------------------------------+------------+--+-
quantizer.weight_proj.weight | UNEXPECTED |  |
project_hid.weight           | UNEXPECTED |  |
project_q.weight             | UNEXPECTED |  |
project_q.bias               | UNEXPECTED |  |
project_hid.bias             | UNEXPECTED |  |
quantizer.weight_proj.bias   | UNEXPECTED |  |
quantizer.codevectors        | UNEXPECTED |  |

Notes:
- UNEXPECTED    :can be ignored when loading from different task/architecture; not ok if you expect identical arch.
Wav2Vec2Model(
  (feature_extractor): Wav2Vec2FeatureEncoder(
    (conv_layers): ModuleList(
      (0): Wav2Vec2GroupNormConvLayer(
        (conv): Conv1d(1, 512, kernel_size=(10,), stride=(5,), bias=False)
        (activation): GELUActivation()
        (layer_norm): GroupNorm(512, 512, eps=1e-05, affine=True)
      )
      (1-4): 4 x Wav2Vec2NoLayerNormConvLayer(
        (conv): Conv1d(512, 512, kernel_size=(3,), stride=(2,), bias=False)
        (activation): GELUActivation()
      )
      (5-6): 2 x Wav2Vec2NoLayerNormConvLayer(
        (conv): Conv1d(512, 512, kernel_size=(2,), stride=(2,), bias=False)
        (activation): GELUActivation()
      )
    )
  )
  (feature_projection): Wav2Vec2FeatureProjection(
    (layer_norm): LayerNorm((512,), eps=1e-05, elementwise_affine=True)
    (projection): Linear(in_features=512, out_features=768, bias=True)
    (dropout): Dropout(p=0.1, inplace=False)
  )
  (encoder): Wav2Vec2Encoder(
    (pos_conv_embed): Wav2Vec2PositionalConvEmbedding(
      (conv): ParametrizedConv1d(
        768, 768, kernel_size=(128,), stride=(1,), padding=(64,), groups=16
```