Music_Informatics_Project

September 24, 2025

[2]:
```
!pip install torch torchaudio librosa soundfile tqdm

import os, glob, torch, torchaudio, librosa, numpy as np
from torch.utils.data import Dataset, DataLoader
import torch.nn as nn
import torch.nn.functional as F
from tqdm import tqdm
```

```
Requirement already satisfied: torch in /usr/local/lib/python3.12/dist-packages
(2.8.0+cu126)
Requirement already satisfied: torchaudio in /usr/local/lib/python3.12/dist-
packages (2.8.0+cu126)
Requirement already satisfied: librosa in /usr/local/lib/python3.12/dist-
packages (0.11.0)
Requirement already satisfied: soundfile in /usr/local/lib/python3.12/dist-
packages (0.13.1)
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-packages
(4.67.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-
packages (from torch) (3.19.1)
Requirement already satisfied: typing-extensions>=4.10.0 in
/usr/local/lib/python3.12/dist-packages (from torch) (4.15.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-
packages (from torch) (75.2.0)
Requirement already satisfied: sympy>=1.13.3 in /usr/local/lib/python3.12/dist-
packages (from torch) (1.13.3)
Requirement already satisfied: networkx in /usr/local/lib/python3.12/dist-
packages (from torch) (3.5)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packages
(from torch) (3.1.6)
Requirement already satisfied: fsspec in /usr/local/lib/python3.12/dist-packages
(from torch) (2025.3.0)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in
/usr/local/lib/python3.12/dist-packages (from torch) (12.6.77)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.6.77 in
/usr/local/lib/python3.12/dist-packages (from torch) (12.6.77)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.6.80 in
/usr/local/lib/python3.12/dist-packages (from torch) (12.6.80)
```

```
Requirement already satisfied: nvidia-cudnn-cu12==9.10.2.21 in
/usr/local/lib/python3.12/dist-packages (from torch) (9.10.2.21)
Requirement already satisfied: nvidia-cublas-cu12==12.6.4.1 in
/usr/local/lib/python3.12/dist-packages (from torch) (12.6.4.1)
Requirement already satisfied: nvidia-cufft-cu12==11.3.0.4 in
/usr/local/lib/python3.12/dist-packages (from torch) (11.3.0.4)
Requirement already satisfied: nvidia-curand-cu12==10.3.7.77 in
/usr/local/lib/python3.12/dist-packages (from torch) (10.3.7.77)
Requirement already satisfied: nvidia-cusolver-cu12==11.7.1.2 in
/usr/local/lib/python3.12/dist-packages (from torch) (11.7.1.2)
Requirement already satisfied: nvidia-cusparse-cu12==12.5.4.2 in
/usr/local/lib/python3.12/dist-packages (from torch) (12.5.4.2)
Requirement already satisfied: nvidia-cusparselt-cu12==0.7.1 in
/usr/local/lib/python3.12/dist-packages (from torch) (0.7.1)
Requirement already satisfied: nvidia-nccl-cu12==2.27.3 in
/usr/local/lib/python3.12/dist-packages (from torch) (2.27.3)
Requirement already satisfied: nvidia-nvtx-cu12==12.6.77 in
/usr/local/lib/python3.12/dist-packages (from torch) (12.6.77)
Requirement already satisfied: nvidia-nvjitlink-cu12==12.6.85 in
/usr/local/lib/python3.12/dist-packages (from torch) (12.6.85)
Requirement already satisfied: nvidia-cufile-cu12==1.11.1.6 in
/usr/local/lib/python3.12/dist-packages (from torch) (1.11.1.6)
Requirement already satisfied: triton==3.4.0 in /usr/local/lib/python3.12/dist-
packages (from torch) (3.4.0)
Requirement already satisfied: audioread>=2.1.9 in
/usr/local/lib/python3.12/dist-packages (from librosa) (3.0.1)
Requirement already satisfied: numba>=0.51.0 in /usr/local/lib/python3.12/dist-
packages (from librosa) (0.60.0)
Requirement already satisfied: numpy>=1.22.3 in /usr/local/lib/python3.12/dist-
packages (from librosa) (2.0.2)
Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.12/dist-
packages (from librosa) (1.16.2)
Requirement already satisfied: scikit-learn>=1.1.0 in
/usr/local/lib/python3.12/dist-packages (from librosa) (1.6.1)
Requirement already satisfied: joblib>=1.0 in /usr/local/lib/python3.12/dist-
packages (from librosa) (1.5.2)
Requirement already satisfied: decorator>=4.3.0 in
/usr/local/lib/python3.12/dist-packages (from librosa) (4.4.2)
Requirement already satisfied: pooch>=1.1 in /usr/local/lib/python3.12/dist-
packages (from librosa) (1.8.2)
Requirement already satisfied: soxr>=0.3.2 in /usr/local/lib/python3.12/dist-
packages (from librosa) (1.0.0)
Requirement already satisfied: lazy_loader>=0.1 in
/usr/local/lib/python3.12/dist-packages (from librosa) (0.4)
Requirement already satisfied: msgpack>=1.0 in /usr/local/lib/python3.12/dist-
packages (from librosa) (1.1.1)
Requirement already satisfied: cffi>=1.0 in /usr/local/lib/python3.12/dist-
packages (from soundfile) (2.0.0)
```

```
Requirement already satisfied: pycparser in /usr/local/lib/python3.12/dist-
packages (from cffi>=1.0->soundfile) (2.23)
Requirement already satisfied: packaging in /usr/local/lib/python3.12/dist-
packages (from lazy_loader>=0.1->librosa) (25.0)
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in
/usr/local/lib/python3.12/dist-packages (from numba>=0.51.0->librosa) (0.43.0)
Requirement already satisfied: platformdirs>=2.5.0 in
/usr/local/lib/python3.12/dist-packages (from pooch>=1.1->librosa) (4.4.0)
Requirement already satisfied: requests>=2.19.0 in
/usr/local/lib/python3.12/dist-packages (from pooch>=1.1->librosa) (2.32.4)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/usr/local/lib/python3.12/dist-packages (from scikit-learn>=1.1.0->librosa)
(3.6.0)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.12/dist-packages (from sympy>=1.13.3->torch) (1.3.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.12/dist-packages (from jinja2->torch) (3.0.2)
Requirement already satisfied: charset_normalizer<4,>=2 in
/usr/local/lib/python3.12/dist-packages (from
requests>=2.19.0->pooch>=1.1->librosa) (3.4.3)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-
packages (from requests>=2.19.0->pooch>=1.1->librosa) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.12/dist-packages (from
requests>=2.19.0->pooch>=1.1->librosa) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.12/dist-packages (from
requests>=2.19.0->pooch>=1.1->librosa) (2025.8.3)
```

[3]:
```
!wget -nc https://os.unil.cloud.switch.ch/fma/fma_small.zip
!unzip -q fma_small.zip
```

```
--2025-09-24 12:08:27--  https://os.unil.cloud.switch.ch/fma/fma_small.zip
Resolving os.unil.cloud.switch.ch (os.unil.cloud.switch.ch)… 86.119.28.16,
2001:620:5ca1:201::214
Connecting to os.unil.cloud.switch.ch
(os.unil.cloud.switch.ch)|86.119.28.16|:443… connected.
HTTP request sent, awaiting response… 200 OK
Length: 7679594875 (7.2G) [application/zip]
Saving to: ‘fma_small.zip’

fma_small.zip       100%[===================>]   7.15G  21.9MB/s    in 5m 31s

2025-09-24 12:13:59 (22.1 MB/s) - ‘fma_small.zip’ saved [7679594875/7679594875]
```

[4]:
```
SR = 22050
N_MELS = 128
```

```
N_FFT = 1024
HOP = 512
SEG_DUR = 5.0
SEG_SAMPLES = int(SR * SEG_DUR)

class FMADataset(Dataset):
    def __init__(self, root):
        self.files = glob.glob(os.path.join(root, "**/*.mp3"), recursive=True)
        print("Found", len(self.files), "files")
        self.mel = torchaudio.transforms.MelSpectrogram(
            sample_rate=SR,
            n_fft=N_FFT,
            hop_length=HOP,
            n_mels=N_MELS
        )

    def __len__(self):
        return len(self.files)

    def safe_load(self, path):
        try:
            wav, sr = torchaudio.load(path)
            if wav.shape[0] > 1:
                wav = wav.mean(dim=0, keepdim=True)
            if sr != SR:
                wav = torchaudio.functional.resample(wav, sr, SR)
            return wav.squeeze(0), True
        except Exception as e:
            print("Skipping:", path, e)
            return None, False

    def __getitem__(self, idx):
        path = self.files[idx]
        wav, ok = self.safe_load(path)
        if not ok:
            return self.__getitem__((idx+1) % len(self.files))

        if len(wav) < SEG_SAMPLES:
            wav = torch.nn.functional.pad(wav, (0, SEG_SAMPLES - len(wav)))
        else:
            start = np.random.randint(0, len(wav) - SEG_SAMPLES + 1)
            wav = wav[start:start+SEG_SAMPLES]

        mel = self.mel(wav.unsqueeze(0))
        logmel = torch.log(mel + 1e-6)
        return logmel, path
```

[5]:
```
class ConvAutoencoder(nn.Module):
    def __init__(self, latent_dim=128):
        super().__init__()
        self.enc = nn.Sequential(
            nn.Conv2d(1, 32, 3, 2, 1), nn.ReLU(),
            nn.Conv2d(32, 64, 3, 2, 1), nn.ReLU(),
            nn.Conv2d(64, 128, 3, 2, 1), nn.ReLU(),
            nn.Conv2d(128, 256, 3, 2, 1), nn.ReLU(),
        )
        self.fc_enc = nn.Linear(256*8*14, latent_dim)
        self.fc_dec = nn.Linear(latent_dim, 256*8*14)

        self.dec = nn.Sequential(
            nn.ConvTranspose2d(256, 128, 4, 2, 1), nn.ReLU(),
            nn.ConvTranspose2d(128, 64, 4, 2, 1), nn.ReLU(),
            nn.ConvTranspose2d(64, 32, 4, 2, 1), nn.ReLU(),
            nn.ConvTranspose2d(32, 1, 4, 2, 1),
        )

    def forward(self, x):
        B, C, H, W = x.shape
        z = self.enc(x)
        z_flat = z.view(B, -1)
        latent = self.fc_enc(z_flat)

        out = self.fc_dec(latent)
        out = out.view(B, 256, 8, 14)
        out = self.dec(out)

        # resize back to input shape
        out = F.interpolate(out, size=(H, W), mode="bilinear",␣
 ↪align_corners=False)
        return out, latent
```

[6]:
```
device = "cuda" if torch.cuda.is_available() else "cpu"
print("Using device:", device)

dataset = FMADataset("fma_small")
loader = DataLoader(dataset, batch_size=16, shuffle=True)

model = ConvAutoencoder(latent_dim=128).to(device)
opt = torch.optim.Adam(model.parameters(), lr=1e-3)
criterion = nn.L1Loss()


checkpoint_files = sorted(glob.glob("checkpoints/conv_autoencoder_epoch*.pth"))
if checkpoint_files:
```

```
    last_ckpt = checkpoint_files[-1]
    model.load_state_dict(torch.load(last_ckpt, map_location=device))
    print(f"  Resumed from {last_ckpt}")
else:
    print("No checkpoint found, starting fresh")

EPOCHS = 20

for epoch in range(EPOCHS):
    total_loss = 0
    for batch, _ in tqdm(loader, desc=f"Epoch {epoch+1}/{EPOCHS}"):
        batch = batch.to(device)
        out, _ = model(batch)
        loss = criterion(out, batch)

        opt.zero_grad()
        loss.backward()
        opt.step()

        total_loss += loss.item()

    print(f"Epoch {epoch+1}: {total_loss/len(loader):.4f}")

    torch.save(model.state_dict(), f"checkpoints/
 ↪conv_autoencoder_epoch{epoch+1}.pth")
    print(f"Checkpoint saved: epoch {epoch+1}")
```

```
Using device: cuda
Found 8000 files

Epoch 1/5:   0%|          | 0/500 [00:00<?,
?it/s]/usr/local/lib/python3.12/dist-packages/torchaudio/_backend/utils.py:213:
UserWarning: In 2.9, this function's implementation will be changed to use
torchaudio.load_with_torchcodec` under the hood. Some parameters like
``normalize``, ``format``, ``buffer_size``, and ``backend`` will be ignored. We
recommend that you port your code to rely directly on TorchCodec's decoder
instead: https://docs.pytorch.org/torchcodec/stable/generated/torchcodec.decoder
s.AudioDecoder.html#torchcodec.decoders.AudioDecoder.
  warnings.warn(
/usr/local/lib/python3.12/dist-packages/torchaudio/_backend/ffmpeg.py:88:
UserWarning: torio.io._streaming_media_decoder.StreamingMediaDecoder has been
deprecated. This deprecation is part of a large refactoring effort to transition
TorchAudio into a maintenance phase. The decoding and encoding capabilities of
PyTorch for both audio and video are being consolidated into TorchCodec. Please
see https://github.com/pytorch/audio/issues/3902 for more information. It will
be removed from the 2.9 release.
  s = torchaudio.io.StreamReader(src, format, None, buffer_size)
Epoch 1/5:  62%|       | 310/500 [07:54<04:42,  1.49s/it]
```

```
Skipping: fma_small/099/099134.mp3 Failed to open the input
"fma_small/099/099134.mp3" (Invalid argument).

Epoch 1/5:  81%|      | 406/500 [10:20<02:19,  1.49s/it]

Skipping: fma_small/108/108925.mp3 Failed to open the input
"fma_small/108/108925.mp3" (Invalid argument).

Epoch 1/5:  87%|      | 435/500 [11:03<01:37,  1.51s/it]

Skipping: fma_small/133/133297.mp3 Failed to open the input
"fma_small/133/133297.mp3" (Invalid argument).

Epoch 1/5: 100%|     | 500/500 [12:40<00:00,  1.52s/it]

Epoch 1: 1.6757

Epoch 2/5:   3%|          | 16/500 [00:24<12:51,  1.59s/it]

Skipping: fma_small/133/133297.mp3 Failed to open the input
"fma_small/133/133297.mp3" (Invalid argument).

Epoch 2/5:  44%|        | 222/500 [05:28<06:52,  1.48s/it]

Skipping: fma_small/099/099134.mp3 Failed to open the input
"fma_small/099/099134.mp3" (Invalid argument).

Epoch 2/5:  78%|      | 389/500 [09:36<02:47,  1.51s/it]

Skipping: fma_small/108/108925.mp3 Failed to open the input
"fma_small/108/108925.mp3" (Invalid argument).

Epoch 2/5: 100%|     | 500/500 [12:21<00:00,  1.48s/it]

Epoch 2: 1.4711

Epoch 3/5:  53%|       | 263/500 [06:24<06:02,  1.53s/it]

Skipping: fma_small/099/099134.mp3 Failed to open the input
"fma_small/099/099134.mp3" (Invalid argument).

Epoch 3/5:  65%|       | 324/500 [07:54<04:11,  1.43s/it]

Skipping: fma_small/108/108925.mp3 Failed to open the input
"fma_small/108/108925.mp3" (Invalid argument).

Epoch 3/5:  93%|     | 466/500 [11:22<00:51,  1.51s/it]

Skipping: fma_small/133/133297.mp3 Failed to open the input
"fma_small/133/133297.mp3" (Invalid argument).

Epoch 3/5: 100%|     | 500/500 [12:12<00:00,  1.47s/it]

Epoch 3: 1.4160

Epoch 4/5:   9%|          | 45/500 [01:25<10:25,  1.37s/it]

Skipping: fma_small/099/099134.mp3 Failed to open the input
"fma_small/099/099134.mp3" (Invalid argument).
```

```
Epoch 4/5:  29%|         | 145/500 [04:02<14:14,  2.41s/it]

Skipping: fma_small/108/108925.mp3 Failed to open the input
"fma_small/108/108925.mp3" (Invalid argument).

Epoch 4/5:  73%|       | 366/500 [09:38<03:17,  1.48s/it]

Skipping: fma_small/133/133297.mp3 Failed to open the input
"fma_small/133/133297.mp3" (Invalid argument).

Epoch 4/5: 100%|      | 500/500 [12:51<00:00,  1.54s/it]

Epoch 4: 1.3476

Epoch 5/5:  57%|        | 284/500 [06:48<05:04,  1.41s/it]

Skipping: fma_small/108/108925.mp3 Failed to open the input
"fma_small/108/108925.mp3" (Invalid argument).

Epoch 5/5:  59%|        | 293/500 [07:01<04:41,  1.36s/it]

Skipping: fma_small/133/133297.mp3 Failed to open the input
"fma_small/133/133297.mp3" (Invalid argument).

Epoch 5/5:  71%|       | 356/500 [08:31<03:24,  1.42s/it]

Skipping: fma_small/099/099134.mp3 Failed to open the input
"fma_small/099/099134.mp3" (Invalid argument).

Epoch 5/5: 100%|      | 500/500 [11:59<00:00,  1.44s/it]

Epoch 5: 1.3184
```

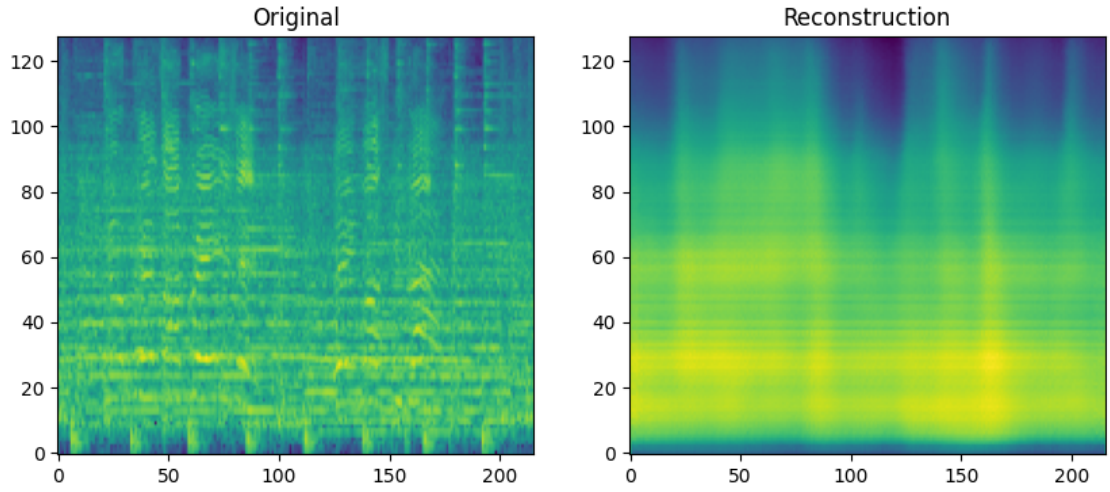
[7]:
```python
# Test the model on a few samples
model.eval()
with torch.no_grad():
    batch, paths = next(iter(loader))
    batch = batch.to(device)
    out, latents = model(batch)

import matplotlib.pyplot as plt

def show_reconstruction(original, reconstructed, idx=0):
    fig, axs = plt.subplots(1, 2, figsize=(10, 4))
    axs[0].imshow(original[idx][0].cpu().numpy(), aspect="auto", origin="lower")
    axs[0].set_title("Original")
    axs[1].imshow(reconstructed[idx][0].cpu().numpy(), aspect="auto",␣
 ↪origin="lower")
    axs[1].set_title("Reconstruction")
    plt.show()

show_reconstruction(batch, out)
```

```
# Save model
os.makedirs("checkpoints", exist_ok=True)
torch.save(model.state_dict(), "checkpoints/conv_autoencoder.pth")
print("Model saved.")
```




```
Model saved.
```