Documentation for Spectrogram Image Diffusion Model Fine-Tuning

Overview:

This script fine-tunes a diffusion model on a custom dataset of spectrogram images using PyTorch and the Diffusers library. The model is trained to generate images based on given prompts, and the training is performed on a GPU using mixed precision.

Step 1:

Data Preparation:

- we need spectrograms of the audio files and corresponding text description for trainig and fine-tuning both.
- At present for POC, we used text prompt as “blues” the genre of the dataset, as dataset containes all the audio files for the blues genre.

Step 2:

Fine-Tuning:

- The riffusion model is based on diffusion process, in which it learn to generate the spectrograms of the audio files.
- A custom dataset class SpectrogramDataset is defined to load spectrogram images and associated prompts.
- Then, image transformations are applied.
- Data set and dataloader for batch processing
- Load pre-trained riffusion model.
- Set optimisers and scheduler for fine-tuning.
- In training loop: spectrogram image with text prompts will be passed .
- Loss is generated using generated spectrogram image and the original image.
- Save the fine-tuned pipeline.

Step 3:

Inferencing:

- Load the fine tuned model pipelne.
- Enter prompt (how you like your melody to be generated) and negative-prompt(things which you don’t want in your melody).
- Genearte melody and enjoy!!!