Music generation models finetuning for Peña music

METAIS Robin, LANIER Théo

2025-12-22

Big intro

$$f(x) = x^2$$
$$g(x) = \frac{1}{x}$$
$$F(x) = \int_b^a \frac{1}{3}x^3$$

1 Models description

Hello World!

1.1 MusicGen

MusicGen's description

1.2 Riffusion

Riffusion's description

2 Models comparaison

In order to compare models between each other, and to evaluate how close from the wanted results their generations are, we cant evaluate their individuals latents representation.
Therefore we decided to create a small Convolutional Neural Network (CNN) which will generate latents representations of spectograms. This CNN will be train on the same dataset that the other models, and for the inference, it will therefore receive the spectrogram computed from the generations of the chosen models (MusicGen, MusicLM, Riffusion). Then, the latents representations of the generated musics will be compared to the representations of the ones from the dataset by using the followings metrics :

- Fréchet Audio Distance (FAD)
- Maximum Mean Discrepancy (MMD)
- Cosine Similarity

2.1 Metrics

Fréchet Audio Distance (FAD) : The fréchet Audio distance is a

Maximum Mean Discrepancy (MMD) : The Maximum Mean Discrepancy is a

Cosine Similarity : Cosine Similarity

2.2 Models results

3 title

4 title