A Diffusion Model for Video Background Music Generation

October 22, 2024

1. Brief Overview

The paper titled **Diff-BGM: A Diffusion Model for Video Background Music Generation** presents a novel approach for automatically generating background music for videos using diffusion models. The model is designed to control various aspects of music generation using visual and semantic features extracted from videos. The original repository for the project is available at: https://github.com/sizhelee/Diff-BGM.

2. Implementation Details

The official GitHub repo is very much incomplete and it doesn't have various files including weight file, train-split-pnt, feature extraction of video dataset, requirement version, and various other things. So in the requirement file, I changed the version for torch and torchvision. The Python version used is 3.10.14. A mir-eval file present in the installation command is being downloaded from polyfussion github repo. For training command is : python diffbgm/main.py –model ldmchd8bar –output (can see in readme file) The code is not working in the original repo, I have resolved multiple errors in my local system and local environment. I made the code work but in the first epoch it is throwing an error related to ran out of input (while loading the pickle file, which is again not present in the original repo, I have added it in my own repo). Though this has generated a log file and params.json file which has the model information and architecture, it had around 4.6 M parameters. Now the inference command is :

python diffbgm/inferencesdf.py –modeldir=[modeldir] –uncondscale=5. (can see in readme file)

This code is working but it is running over multiple folders, so it is giving inference for only files present in repo (code is set also for other folders but it is not present in repo or any other drive of the author), but for present folders, I am getting the inference result in diffbgm/exp folder.

So the command is working and takes about 4 hours to run.

3. Dataset Information

The paper introduces a new dataset named **BGM909**, which consists of 909 video-music pairs. The dataset is publicly available at the provided GitHub link in the project repository. I have included the link in the form. The model is based upon another work of which uses a dataset named pop909, again a video dataset from which music scores are generated, so it use it's pre-trained weight and fine tune the model with some modification on BGM909 dataset.

4. Results

The training code wasn't completed due to a pickle file loading error but the result generated (midi file) from the inference code is present in the repository in the directory diffbgm/exp .