

Song Genre Identification Tool

Thomas McCausland

2472525

Proposal

Motivation

[Clearly motivate the purpose of your project; why someone would care about what you are doing] Song genre is an important piece of metadata within MP3s. Sometimes tracks can be produced in a way that leads to the genre being absent from the MP3s metadata. This project will involve developing a machine learning model that can "listen" to an excerpt of a song and identify the genre(s) it belongs to. The model can be trained using a labelled dataset of song samples (e.g. GTZAN). With a suitably trained model develop a tool to identify the genre of a song and add the appropriate information to the metadata of the track.

Aims

[Clearly state what the project is intended to do. This should be something which is measurable; it should be possible to tell if you succeeded] This project aims to research and develop a machine learning model capable of identifying and classifying song genre, using image recognition. The minimum viable product is a model that has "better than random chance" testing accuracy.

Progress

[Briefly state your progress so far, as a bulleted list] - Developed a simple convolutional neural network model using pytorch for image recognition on the CIFAR-10 dataset - Researched convolutional neural networks and typical machine learning model layers - Improved the simple CNN model - Researched and developed script to convert song files into spectrograms - Researched and chose an appropriate existing song dataset labelled by genre: FMA dataset - Wrote script to generate custom image dataset of song spectrograms - Researched how to adapt the developed CNN model for use on custom image dataset - Adapted the CNN model for use on custom spectrogram dataset and ran tests to test its initial accuracy. Better than random chance testing accuracy has been achieved.

Problems and risks

Problems

[What problems have you had so far, that have held up the project?] - The first hurdle was learning how to use pytorch and teaching myself about Convolutional Neural Networks. - The next issue, once a simple example image classification model was made, was to improve the training time of the model by offloading tensor operations to the GPU and to learn about and implement hyperparameter training. - The greatest problem was figuring out what training dataset to use or whether to create my own. I first attempted to create my own but that quickly proved not feasible. I eventually settled on using the small FMA dataset after a lot of research. - There small problems with creating the spectrogram conversion script, as there is no clear best library to use to convert sound files to spectrograms. Additional research on spectrograms and the mel scale had to be done. The current approach is still flawed and needs refining - One particularly difficult issue to identify was that the FMA dataset is incomplete and erroneous. Some songs files are corrupted or missing, which caused runtime errors that were very difficult to locate. This was solved by removing the specific songs from the data. - Another issue with FMA was integrating it into the project. The repository it originates from contains code to help understand and run it, but a lot of that code is broken or incomplete. - Another small issue was using pandas to handle the metadata from the FMA dataset, which Chris helped solve. - An issue that also held up training on the spectrogram dataset was the computer running out of memory. This turned out to be a bug with the spectrogram generation. - The final issue was about how to load the data into the model for training. This was sorted using specific file structures that pytorch can use to read in custom datasets.

Risks

[What problems do you foresee in the future and how will you mitigate them?] - A potential important risk is that a simple CNN is not enough to categorise spectrograms by song genre. To mitigate this, I plan on testing the image classification problem with a more complex, off-the-shelf model and comparing the results. - Another potential issue is that the accuracy of the model is not

able to increase past a certain point. There are many reasons why this could happen and I plan on researching more about CNN layers to try and increase the accuracy of the existing model. - Another potential problem is that certain song genres might be easier to distinguish from each other than others. To study this potential behaviour, I plan on running a series of binary choice tests on the model and presenting the results as a confusion matrix.

Plan

[Time plan, in roughly weekly to monthly blocks, up until submission week]

This is a very rough plan, as I'll try to get most of the progress done as soon as possible so that I can focus on working on the dissertation. Of course, I'll try to be writing the dissertation continuously throughout the semester.

- January 2023:
 - Improve model through addition of layers and input tweaks
 - Set up experiments for binary choices and write code to generate confusion matrix
- February 2023:
 - Compare custom simple CNN model with an off-the-shelf model
 - Once model is relatively finalised, run binary choice experiments and adjust direction of work based on the results
- March 2023:
 - Work on any final, new issues and research directions that come up
 - Concentrate on dissertation writing