Music Genre Classification Using a Convolutional Neural Network – Report

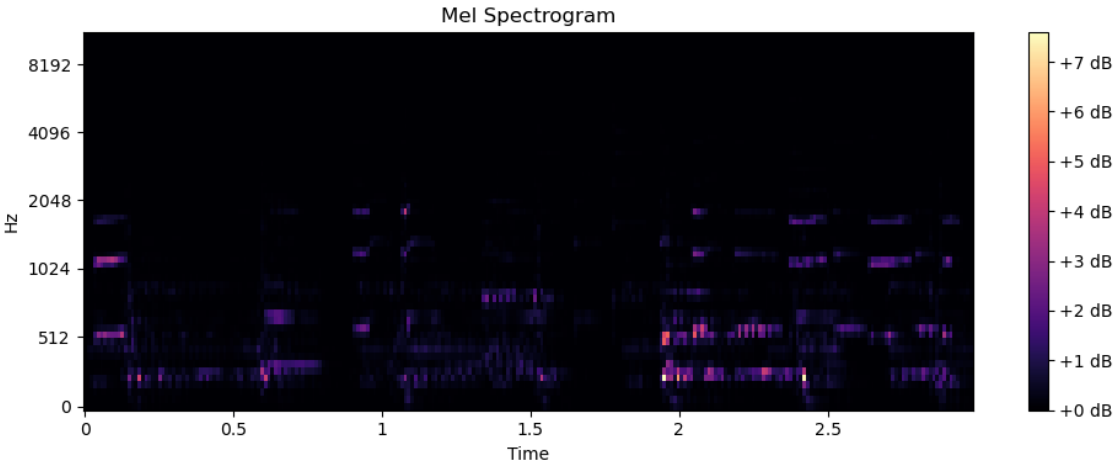
Lauri Kalliojärvi

1. Introduction

Music genre classification is determining the genre of a song by extracting features from it to learn genre-specific features. Music genre classification is used by streaming services to create song suggestions based on user preferences. In this project music genre classification is tackled by training a convolutional neural network model to learn genre-specific features and classify songs based on those features. This project and used methods are based on an article [1].

2. Data

The genre classification is performed on an open-source dataset GTZAN [2]. The dataset consists of 1000 song samples from 10 different music genres. The samples are divided in a way that for each of 10 genres there are 100 samples. For classification purposes the audio samples are divided into training, validation and testing samples. Each sample is converted from timescale to frequency scale, in the form of a Mel-spectrogram with 64 Mel bins. The spectrograms are then split into 3-second-long segments. The segments are then used for the genre classification.



The first 3 second segment from the Mel-spectrogram of blues.00004.wav.

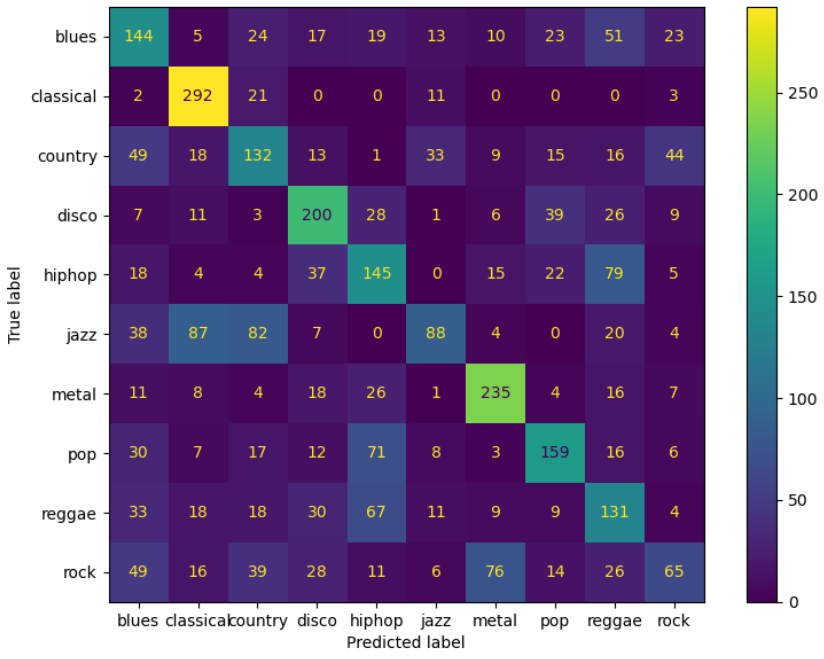
3. Methods

For this task the used classifier is a convolutional neural network with an architecture seen below. The used song samples are divided using split of 50% training, 20% validation and 30% testing. Cross-entropy is used to calculate loss during 30 epochs. Patience of 5 is used in aid of finding the best classification model.

```
MyModel(
  (block1): Sequential(
    (0): Conv2d(1, 64, kernel_size=(3, 3), stride=(1, 1))
    (1): ReLU()
    (2): MaxPool2d(kernel_size=(2, 4), stride=(2, 4), padding=0, dilation=1, ceil_mode=False)
    (3): Dropout(p=0.2, inplace=False)
  )
  (block2): Sequential(
    (0): Conv2d(64, 64, kernel_size=(3, 5), stride=(1, 1))
    (1): ReLU()
    (2): MaxPool2d(kernel_size=(2, 4), stride=(2, 4), padding=0, dilation=1, ceil_mode=False)
    (3): Dropout(p=0.2, inplace=False)
  )
  (fully_connected): Sequential(
    (0): Linear(in_features=13440, out_features=32, bias=True)
    (1): Dropout(p=0.2, inplace=False)
    (2): ReLU()
  )
  (output): Sequential(
    (0): Linear(in_features=32, out_features=10, bias=True)
  )
)
```

4. Results

The training exited due to early stopping during epoch 7 with loss of 1.7. The model reached 48% accuracy with the classification. The resulting confusion matrix of the predicted genres can be seen below.



5. Conclusion

Music genre classification is a difficult task for both the human ear and machines. The 48% accuracy of the used CNN model is adequate. The model classifies classical and metal music quite well, possible due to the distinct nature of the genres. The model performs especially poorly when classifying rock and jazz music, getting them mixed with other genres. This might be because there are no such solid boundaries for said genres with the influence of other genres being present. With the fine tuning of the hyperparameters of the used model, the classifying accuracy can be improved.

6. Sources

[1] [1802.09697] Convolutional Neural Network Achieves Human-level Accuracy in Music Genre Classification (arxiv.org)
[2] https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification

Notes:

jazz.00054.wav is faulty. Tried to replace it with another jazz snippet provided by Camilo Gomez O in the discussion section of the dataset page. Even the replaced file didn't work so deleted the file and replaced it with a copy of jazz.00053.wav.