188.498 Similarity Modeling 1/2

Intermediate Project Report Hand-in

Craig Lincoln (11828331) and Gent Rexha (11832486)

18 11 2019

Approach

Out approach is to use a pre-trained neural net and feed both audio and image features into the same model. The proposed neural net would be ImageAI's ResNet which has an input size of 224,244,3. Obviously, feeding images is no problem. For the audio features we plan to add them to the images. Currently, the audio features are MFCC and Chroma from Librosa and will have dimensions, ~20 and ~12 respectively. This maybe something we vary as well add more features.

Tasks & Open Questions

☒ Research on the neural network framework possibilities.
☒ Created a GUI application for labeling image data.
☐ How will we divide training, test and validation sets?
☐ Do we want to add more audio features and which?
☐ How do we add the audio features explicitly, eg, just to the first color channel or the same features vector to all three and what do we do with in the case of augmentation (allow the audio features to be modified or always include them last?)

Current Timesheet

Date	Time	Description	Person responsible
19/10/2019	16:00-19:00	Initialized repository and initial research.	Craig
23/10/2019	14:00-15:00	Watched and got Kermit times from episode 02-01-01.	Craig
04/11/2019	20:00-22:00	Set project structure and moved some stuff around for the meeting.	Gent
07/11/2019	20:00-24:00	Started working on some dataset preparation.	Gent
08/11/2019	10:00-11:00	Sync Meeting.	Craig & Gent
17/11/2019	14:00-18:00	Look into Librosa and audio features.	Craig
18/11/2019	14:00-17:00	Finished labeling GUI application.	Gent