Shazzzam Cam

Angela Akoto | Wonyoung Lee | Serena Martin | Kira Teramoto

Motivation

Music, and their associated artwork, tell a story about the zeitgeist – artistic, cultural, and political… Think about the White Album, the Nirvana cover, Pink Floyd…

Define the problem

So our **question** was: are there discernable patterns in the features of the artwork that can be extracted to identify the genre of the music from its associated artwork?

Goal

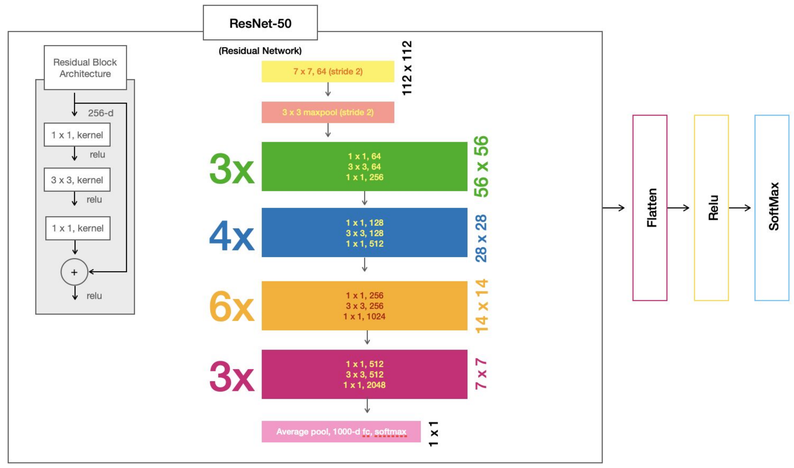
classifying the main genre of an album based on the cover art

Practically:

1. Using CNNs, and in particular a modified ResNet-50, to achieve an above-chance classification of a main genre that is associated with the album/song artwork
2. Exploring other models to find the best solution

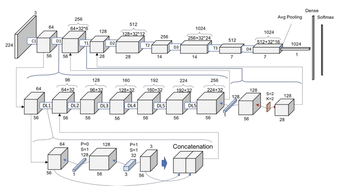
Methodology

1. We used the Million Song Dataset[1] which contained 30,713 tracks and their related album cover images, each annotated with a unique genre label among 15 classes.
   - Divided into 70\% (training), 15\% (validation), and 15\% (test), with no artist and album overlap across these sets to avoid overfitting.
2. We tested our program with a modified **ResNet-50**, which we altered by adding the Adam optimizer (learning rate 0.01), adding flatten and dense layers after initial calls, and customizing hyperparameters & final classes to match our data
3. We then tested other models to see which alternative solutions would provide us with the best results.



Alternative solutions

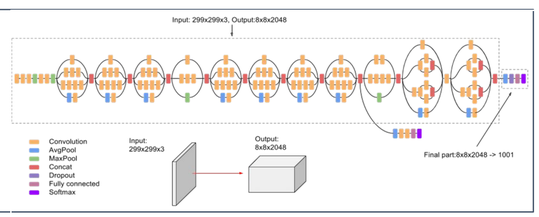
DenseNet 169



This CNN has 169 layers connected by Dense Blocks. Unlike ResNet, DenseNet concatenates the output of the previous layer with the next layer in a feed-forward fashion. Where traditional models would have 169 connections, this has 14,365 direct connections

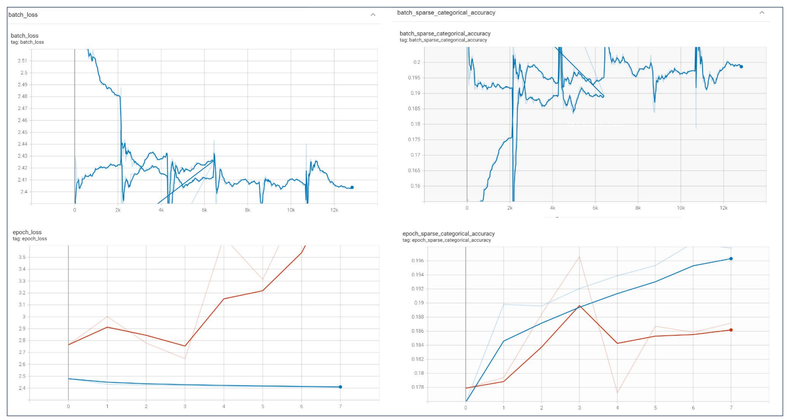
Inception Net V3

This CNN is made up of symmetric and asymmetric building blocks (convolutions, average pooling, max pooling, concatenations, dropouts, fully connected layers). Batch normalization is applied to activation inputs. Loss is computed using Softmax.

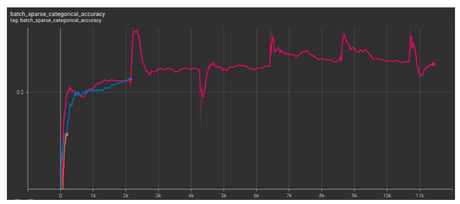


Results

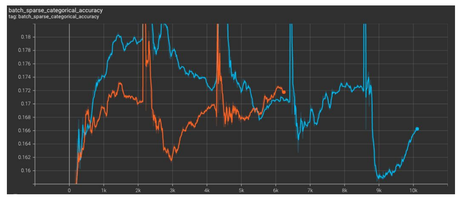
1. We found that the DenseNet169 model gave us the best prediction accuracies, with our strongest accuracy being 22%.
2. The ResNet50 model had an accuracy of around 20%
3. The InceptionV3 model was the next best, with the highest accuracy of 19%



ResNet-50 Results
(clockwise from top left: batch loss, categorical accuracy, epoch loss, epoch sparse categorical accuracy)



DenseNet169 Categorical Accuracy



InceptionV3 Categorical Accuracy

Some Image Classification Results

References

[1] Sergio Oramas et al. MSD-I: Million Song Dataset with Images for Multimodal Genre Classification, May 2018.
[2] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
[3] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016 pp. 2818-2826.

Acknowledgements

We would like to thank Will Chen, our project mentor, for their dedicated support throughout the process. Thank you also to all the TAs and Professor Tompkin.