Music Genre Classification Using CNN-hybrid Architectures

Avital Vetshchaizer
avetshch@sfu.ca

Henry Yip
hya59@sfu.ca

Andy Liu
ala158@sfu.ca

Abstract

In this paper we seek to improve the accuracy and runtime performance of the Convolutional Neural Networks (CNN) by introducing multiple modified architectures to achieve better runtime and accuracy. Our proposed models are known as a Convolutional Recurrent Neural Network (CRNN), Duplicated CNN, and Duplicated CRNN. Using the Free Music Archive (FMA) dataset, we investigated different methods of improving the performance of our models by measuring the efficiency, error, and training speed of the final models. As a result, our paper has provided firm evidence that both the duplicated CNN and CRNN have achieved an overall higher test accuracy in a shorter training period, thus proving the applicability and potential usage of these architectures.

Introduction

Signal decoding and classification plays a large role in the execution of everyday tasks in the modern world. A large subset of these signals, also known as music, has accompanied humans throughout the eras by evolving to act as symbols and records of the important events at the time. Different styles of music have helped inspire individuals and played large roles in people's self-perception and personal identity. Thus, in this research we attempt to quantify these genres to gain a measurable understanding of their similarities and differences. Our data consists of the spectrum of frequencies and their magnitudes for each song. Once processed, our input data will consist of a spectrogram in an image format. During this project, we experimented with four classification models: CNNs, CRNNs, Duplicated CRNNs (DCRNN), and Duplicated CNNs (DCNN) to attempt to gain a higher confidence rate above the standard CNN. To achieve the best results, we traversed through a series of data augmentations and model optimizations.

Related work

Duplicated Convolution Layers

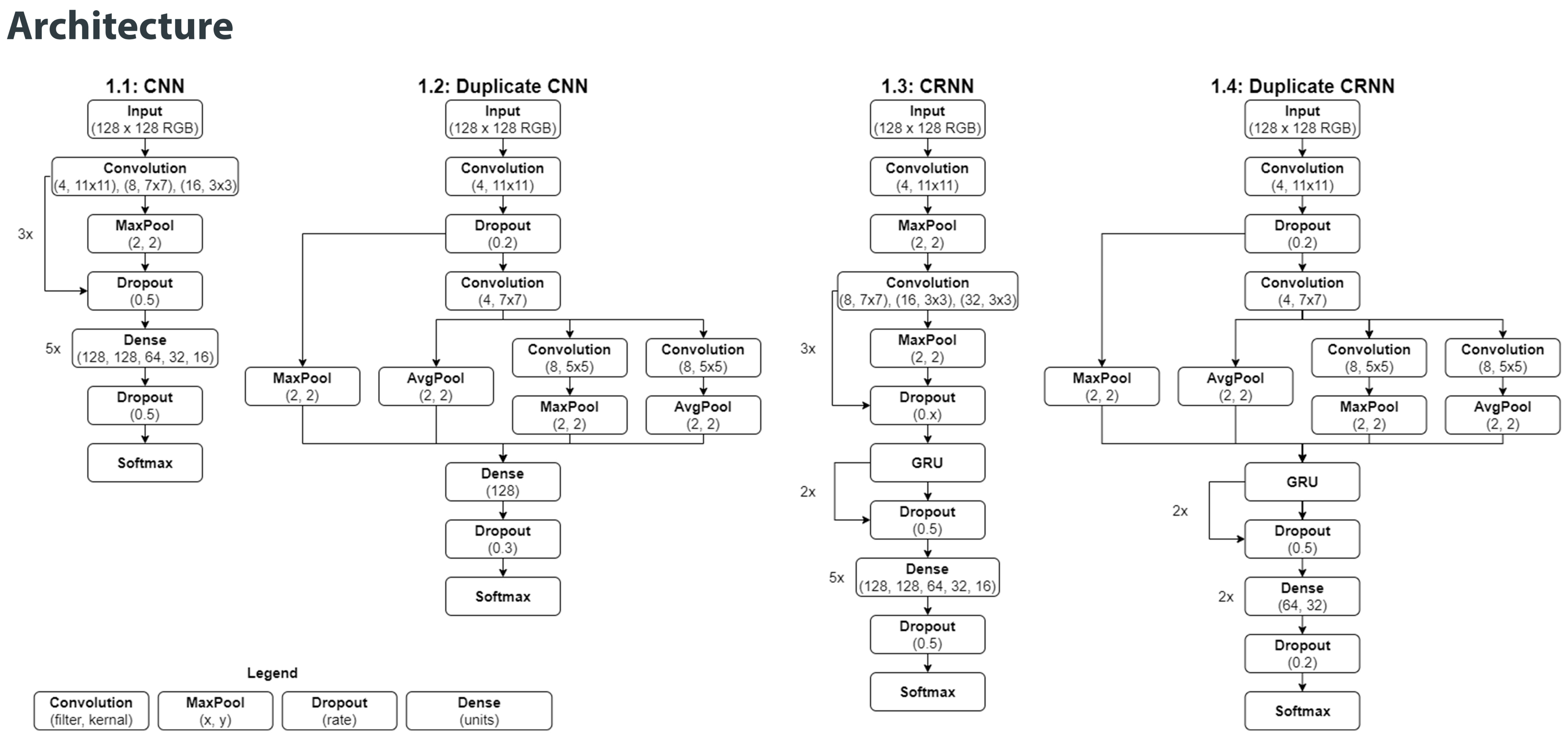
In their research paper, Yang and Zhang [1] used a Duplicated CNN to increase the predictive accuracy of their network in respective genres. Their network consisted of four parallel pooling layers, each with different types of layers (Max Pooling, Convolution, Average Pooling, or Dense layers). Using the GTZAN dataset, they discovered that each pooling layer best predicted a specific genre, thus allowing the collective network to achieve higher predictive accuracy.

Recurrent Neural Networks

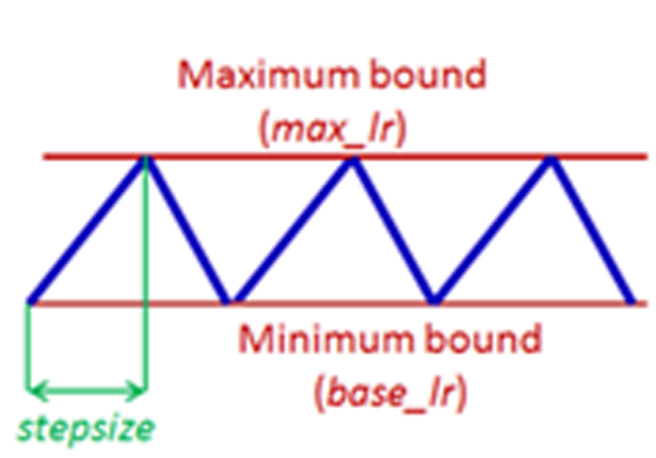
In Montreal, researchers Zaragoza and Pertusa [2] applied a modified CRNN to address the Optical Music Recognition (OMR) task. Inputting a set of monophonic images to their network, they successfully reproduced each musical score with the average classification error 2% on the symbol level. Their network consisted of a series of convolutional filters for feature extraction and recurrent block modules to cover the sequential nature of the music.

Methodologies

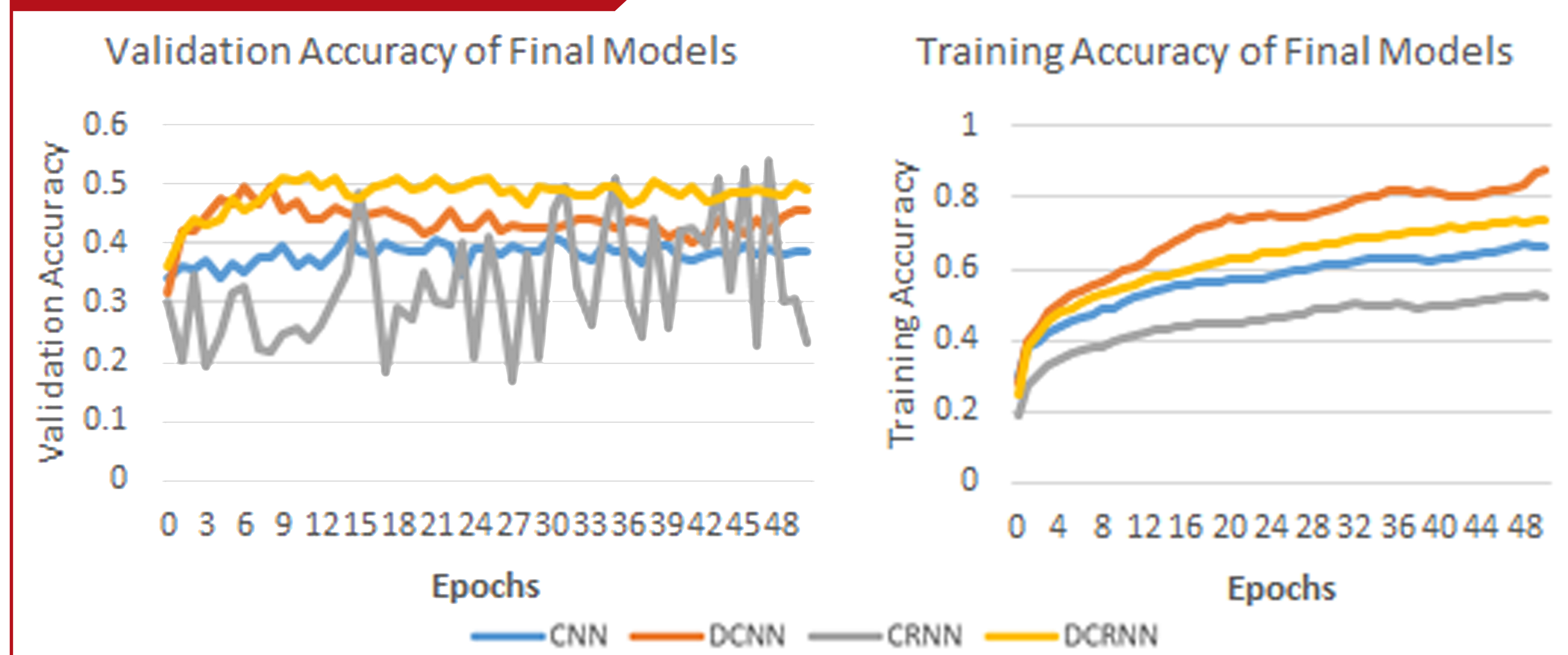
Architecture



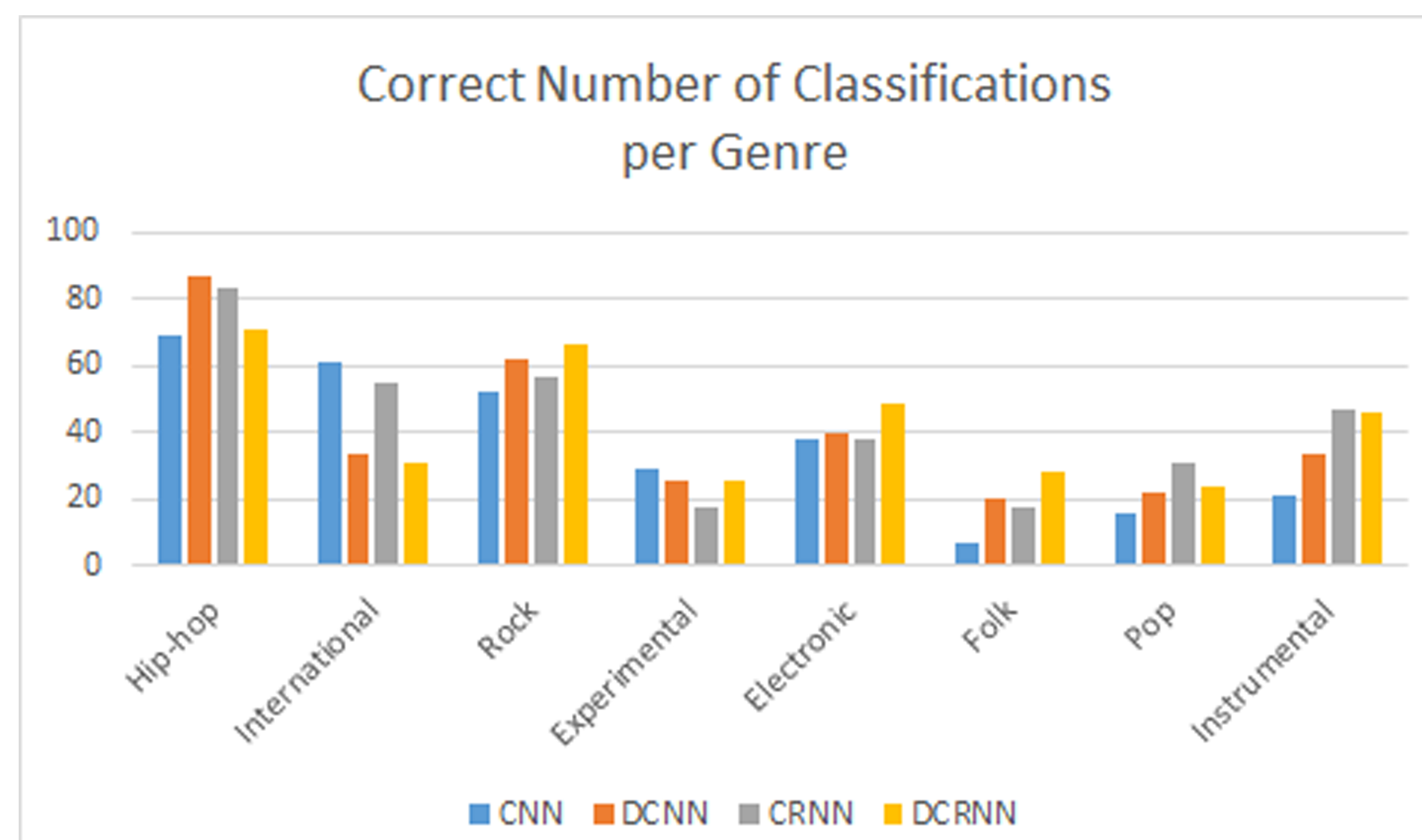
Cyclic Learning Rate



Results



	Standard Learning Rate		Cyclic Learning Rate	
	Epochs Required	Test Accuracy	Epochs Required	Test Accuracy
CNN	29	37.75%	15	36.63%
DCNN	9	36.50%	7	41.63%
CRNN	47	38.13%	48	43.38%
DCRNN	12	42.63%	23	40.00%



Conclusion

In conclusion, our experiments demonstrate overall improved test accuracy with faster learning with the addition of RNN and duplicate convolution layers to the basic CNN model. Improvements in results could also be realized with the removal of genres with poor representation, utilizing raw data in addition to the spectrogram, and by increasing hardware capacity to boost parameter count and image resolution. This would only further prove our studies and the applicability of these model architectures.

References

[1] H.Yang & W.Q. Zhang (2019) Music Genre Classification Using Duplicated Convolutional Layers in Neural Networks, ISCA
[2] J. Calvo-Zaragoza, J.J. Valero-Mas & A. Pertusa (2017) End-to-End Optical Music Recognition Using Neural Networks, ISMIR, 2017
[3] M. Dong (2018) Convolutional Neural Network Archives Human-Level Accuracy in Music Genre Classification, Cornell University ArXiv
[4] H.Yang & W.Q. Zhang (2019) Music Genre Classification Using Duplicated Convolutional Layers in Neural Networks, ISCA
[5] M. Defferrart, K. Benzi, P. Vandergheynst & X. Bresson (2017) FMA: A Dataset For Music Analysis, Cornell University ArXiv
[6] P.M. Radiuk (2017) Impact of Training Set Batch Size on the Performance of Convolutional Neural Networks for Diverse Datasets, Information Technology and Management Science
[7] S.L. Smith, P. Kindermans, C. Ying & Q. Le (2018) Don't Decay The Learning Rate, Increase The Batch Size, OverReview.org
[8] M.D. Zeiler (2012) Adadelta: An Adaptive Learning Rate Method, Cornell University Arxiv,
[9] D.P. Kingma & J.L. Ba (2015) Adam: A Method For Stochastic Optimization, Cornell University ArXiv
[10] G. Hinton, N. Srivastava & Kevin Swesky (2014) Overview of Mini-Batch Gradient Descent, University of Toronto
[11] L.N. Smith (2017) Cyclical Learning Rates for Training Neural Networks, Cornell University ArXiv