Sponsoring Committee: Professor Juan Pablo Bello, Chairperson
Professor Robert Rowe
Doctor Eric J. Humphrey

An Abstract of

AUTOMATIC MUSIC TRANSCRIPTION IN THE DEEP LEARNING ERA:

PERSPECTIVES ON GENERATIVE NEURAL NETWORKS

Jong Wook Kim

Program in Music Technology
Department of Music and Performing Arts Professions

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
Steinhardt School of Culture, Education, and Human Development
New York University
2020

ABSTRACT

The problem of automatic music transcription (AMT) is considered by many researchers as the holy grail of the field, because of the notorious complexity and difficulty of the problem. Meanwhile, the current decade has seen an unprecedented surge of deep learning where neural network methods have achieved tremendous success in many machine learning tasks including AMT. The success of deep learning is largely enabled by the ever-increasing amount of available data and the innovation of GPU hardware, allowing a deep learning model to enjoy the increased capacity to process such scale of data. While having more data and higher capacity translates better performance in general, there still remains the question of how to design an AMT model that can effectively incorporate the inductive bias for the task and best utilize the increased capacity.

This thesis hypothesizes that an effective way to address this question is through the use of generative neural networks. Starting with a simplified setup of monophonic transcription, we learn the effectiveness of convolutional representation and the roles of dataset choices in data-driven models for music analysis. In the subsequent chapters, we examine the applications of deep generative models in music analysis and synthesis tasks, by introducing a WaveNet-based music synthesis model that learns a multi-dimensional timbre representation and a music language model applied in an adversarial manner to improve a piano transcription model. Finally, we combine the analysis and synthesis methods to develop a multi-instrument polyphonic music transcription system. From these observations, we conclude that deep generative models can be used to improve AMT in many ways, and they will be a crucial component for further advancing AMT.