Abstract

Currently, music is most commonly listened to as audio recordings; however, MIDI and other file formats detailing in essence only the notes in music were previously popular due to a lack of storage and scalability with processing power. Since musicians organically use the notes and a limited set of commands to generate music, it should be possible to translate from file formats like MIDI to raw audio in an organic way. To start, single timbrally accurate notes need to be generated. This can be achieved by converting audio into an image format, i.e. spectrograms, then generating images using machine learning. A deep convolutional generative adversarial network is used to generate spectrograms to be turned back into audio. Currently, generated audio is unrealistic, but with further model training and optimization, audio of single notes on multiple different instruments could be realistically generated as a stepping stone to converting MIDI files more accurately into raw audio.