1 Abstract

Classifying audio signals by source or by content with machine learning has become a topic of much research in the past few years. Models often involve the production of a spectrogram or feature vector and passing either array into a network of a single type such as an Convolutional Neural Network (CNN) or Multilayer Perceptron (MLP). In this work, we explore a new hybrid neural-network architecture that combines the MLP and CNN models to produce a signal classifier with superior performance over models that rely solely one or the other. This hybrid network uses two branches, one being a CNN to process an image-like 2-dimensional spectrogram, and one being an MLP to process a 1-dimensional feature vector. The output(*) of each branch is then concatenated into a single 1-dimensional dense layer, allowing for any predictions to be a product of both branches. We describe in detail the production and usage of the spectrogram and predictors, as well as how they influence the chosen network architecture. We finish with a practical demonstration in using this classifier model to match chaotically generated wave forms to real-world musical instruments.

^ waveforms from a chaotic music synthesizer to real-world musical instruments

-- We should include mention of the data set we are using -- it was U. Iowa, wasn't it?

(*) I don't think you want to say " output" here, since wouldn't the output be the actual classification result? SInce you did not like my previous expanded suggestion, can you just say that the concatenation is formed at one of the hidden layers to produce a final hybrid architecture.