Landon Buell

Kevin Short

Abstract v5

15 September 2020

Classifying audio signals by source 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 a Convolutional Neural Network (CNN) or Multilayer Perceptron (MLP). In this study, we explore a new hybrid neural-network architecture that combines the CNN and MLP 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 2D spectrogram, and the other being an MLP to process a 1D feature vector. Within the model, a hidden layer combines activations from the two branches by concatenating them into a single 1D 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 waveforms from a chaotic music synthesizer to real-world musical instruments. Training data was provided by Philharmonia Symphony Orchestra and University of Iowa's Electronic Music Studios.