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. Methods often involve the production of a spectrogram or feature vector and passing these arrays 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 network uses two branches, one that uses a CNN to process an image-like spectrogram, and one that uses an MLP to process a feature vector. The output of each branch is concatenated and the two models are merged to produce a single prediction. We detail the production and usage of the spectrogram and predictors, and how they influence the chosen network architecture. We finish with practical example in using this classifier model to match chaotically generated wave forms to real-world musical instruments.