Intro and Background

Solving classification problems by way of machine learning has been evolving at an astonishing pace. Currently, there are many different architectures that can be used to do this. Deep learning models are often able to solve classification problems with very high levels of accuracy, especially with image classification. Audio classification, however is often considered a more challenging problem to address. By applying the techniques and methods used for image classification, results in audio classification have been able to achieve the same solid results. Real world applications using these methods are quite prevalent. From Alexa and Siri to a wide variety of voice dictation products, audio classification is used. These all rely on some sort of machine learning, which are built off of large datasets.

Anytime a machine learning process is to be employed, having initial quality data to start with is key. Properly classified data is imperative for being able to create a model that is possible of delivering consistent and reliable results. A large dataset enables better training, and usually will result in improved results. When working with smaller datasets however, this can make quality training less fruitful which will result in decreased performance from the model. One method to overcome this is through data augmentation; altering an existing set of data in some way to produce new instances which are properly classified. By using augmentation, one can effectively increase the size of their dataset for training a model.