Improving label efficiency through multi-task learning on auditory data

Anonymous Author(s)
Affiliation
Address
email

Abstract
Collecting high-quality, large scale datasets typically requires significant resources. The aim of the present work is to improve the label efficiency of large neural networks operating on audio data through multitask learning with self-supervised tasks on unlabeled data. To this end, we trained an end-to-end audio feature extractor based on WaveNet that feeds into simple, yet versatile task-specific neural networks. We describe three self-supervised learning tasks that can operate on any large, unlabeled audio corpus. We demonstrate that, in a scenario with limited labeled training data, one can significantly improve the performance of a supervised classification task by simultaneously training it with these additional self-supervised tasks. We show that one can improve performance on a diverse sound events classification task by nearly 6% when jointly trained with up to three distinct self-supervised tasks. This improvement scales with the number of additional auxiliary tasks as well as the amount of unsupervised data. We also show that incorporating data augmentation into our multitask setting leads to even further gains in performance.