World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:15, No:03, 2021

1D Convolutional Networks to Compute Mel-Spectrogram, Chromagram, and Cochleogram for Audio Networks

Authors : Elias Nemer, Greg Vines

Abstract : Time-frequency transformation and spectral representations of audio signals are commonly used in various machine learning applications. Training networks on frequency features such as the Mel-Spectrogram or Cochleogram have been proven more effective and convenient than training on-time samples. In practical realizations, these features are created on a different processor and/or pre-computed and stored on disk, requiring additional efforts and making it difficult to experiment with different features. In this paper, we provide a PyTorch framework for creating various spectral features as well as time-frequency transformation and time-domain filter-banks using the built-in trainable conv1d() layer. This allows computing these features on the fly as part of a larger network and enabling easier experimentation with various combinations and parameters. Our work extends the work in the literature developed for that end: First, by adding more of these features and also by allowing the possibility of either starting from initialized kernels or training them from random values. The code is written as a template of classes and scripts that users may integrate into their own PyTorch classes or simply use as is and add more layers for various applications.

Keywords : neural networks Mel-Spectrogram, chromagram, cochleogram, discrete Fourrier transform, PyTorch conv1d()
Conference Title : ICCMDM 2021 : International Conference on Computer Music and Digital Media
Conference Location : Barcelona, Spain
Conference Dates : March 04-05, 2021