Skysong Innovations

Advancing the Arizona State University
Knowledge Enterprise

1475 N. Scottsdale Road, Suite 200
Scottsdale, AZ 85287-3538
Phone: 480 884 1996 Fax: 480 884 1984

Case ID:M20-093P

Published: 11/2/2020

Inventors

Vivek Sivaraman Narayanaswamy

Sameeksha Katoch

Jayaraman Thiagarajan

Andreas Spanias

Contact

Shen Yan
shen.yan@skysonginnovations.com

Audio Source Separation via Multi-Scale Learning with Dilated Dense U-Nets

Background

Audio source separation refers to the process of extracting constituent sources from a given audio mixture. Despite being a critical component of audio enhancement and retrieval systems, the task of source separation is severely challenged by variabilities in acoustic conditions and the highly ill-posed nature of this inverse problem.

A majority of conventional source separation techniques operate in the spectral domain, specifically the magnitude spectrum. However, by ignoring the crucial phase information, these methods often require extensive tuning of front-end spectral transformations to produce accurate source estimates. Recent approaches have resorted to time-domain processing to bypass the need for front-end transformations. On the other hand, fully time-domain approaches must contend with variable temporal contexts to extract useful features, making network training challenging even with sophisticated sequence models such as long short-term memory (LSTM) and one-dimensional convolutional neural networks (1DCNNs). This motivates the design of architectures that can effectively extract multi-scale features and produce generalizable source estimation models for highly underdetermined scenarios.

Invention Description

Researchers at Arizona State University have developed DDU-Net, a fully convolutional approach for time-domain audio source separation. Designed as a U-Net style architecture, DDU-Net utilizes dilated convolutions to leverage information from exponentially increasing receptive fields and features dense connections to improve the robustness of the training process. The modeling approach can produce multi-scale features which are robust to sampling rate changes and can enable complex temporal modeling. Experiments demonstrate that the improved feature extraction process outperforms state-of-the-art time-domain separation approaches, namely the Wave-U-Net and the WaveNet models.

Potential Applications

- Audio source separation

- Time-domain feature extraction

Benefits and Advantages

- Robust to sampling rate changes

- Efficient model training due to improved gradient flow

- Efficient feature reuse resulting from dense connections between convolutional layers

- Improved local context for source reconstruction from the use of skip connections between layers

Related Publication

Faculty Homepage of Professor Andreas Spanias