IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 27, 2018, accepted December 13, 2018, date of publication January 8, 2019, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2888882

Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network

ADITYA KHAMPARIA[1], DEEPAK GUPTA[2], NHU GIA NGUYEN[3], ASHISH KHANNA[2], BABITA PANDEY[4], AND PRAYAG TIWARI[5]

[1]School of Computer Science and Engineering, Lovely Professional University, Phagwara 144401, India
[2]Maharaja Agrasen Institute of Technology, New Delhi 110086, India
[3]Graduate School, Computer Science, Duy Tan University, Da Nang 550000, Vietnam
[4]Department of Computer and Information Technology, Babasaheb Bhimrao Ambedkar University, Lucknow 226025, India
[5]Department of Information Engineering, University of Padova, I-35131 Padua, Italy

Corresponding author: Nhu Gia Nguyen (nguyengianhu@duytan.edu.vn)

This work was supported in part by the Duy Tan University. The authors would like to thank the reviewers in advance for their comments and suggestions.

ABSTRACT In every aspect of human life, sound plays an important role. From personal security to critical surveillance, sound is a key element to develop the automated systems for these fields. Few systems are already in the market, but their efficiency is a point of concern for their implementation in real-life scenarios. The learning capabilities of the deep learning architectures can be used to develop the sound classification systems to overcome efficiency issues of the traditional systems. Our aim, in this paper, is to use the deep learning networks for classifying the environmental sounds based on the generated spectrograms of these sounds. We used the spectrogram images of environmental sounds to train the convolutional neural network (CNN) and the tensor deep stacking network (TDSN). We used two datasets for our experiment: ESC-10 and ESC-50. Both systems were trained on these datasets, and the achieved accuracy was 77% and 49% in CNN and 56% in TDSN trained on the ESC-10. From this experiment, it is concluded that the proposed approach for sound classification using the spectrogram images of sounds can be efficiently used to develop the sound classification and recognition systems.

INDEX TERMS Deep learning, convolutional neural network, tensor deep stacking networks, spectrograms.

I. INTRODUCTION

In recent years, research on automatic sound recognition has gained momentum and has been used in multidisciplinary fields like multimedia [1], bioacoustics monitoring [2], intruder detection in wildlife areas [3], audio surveillance [4] and environmental sounds [5]. Sound recognition problem consists of three different stages as pre-processing of signals, extraction of specific features and their classification. Signal pre-processing divides the input signal to different segments which used for extracting related features. Feature extraction reduces the size of data and represent the complex data as feature vectors. Crossing rate, pitch and frame features used in speech recognition applications were classified using various classifiers like decision trees, random forest and k nearest neighbor. Spectrogram image features (SIF), Stabilized auditory image (SAI) and Linear prediction coefficients (LPC) are used widely in recent years. Moreover, usage of different machine learning and soft computing techniques like Hidden and Gaussian mixture model, random forest, multi-layer perceptron and emerging deep learning networks in sound recognition system resulted in performance enhancement of sound recognition and classification systems.

In recent years SIF generates sound waves which provides more accurate results in noisy conditions. These sound waves are made up of high pressure and low-pressure regions moving through a medium. Such high- and low-pressure regions forms a specific type of pattern to every distinguish sound. These waves have few characteristics like wavelength, frequency, wave speed and time periods [6]. These characteristics are used to classify the sounds into different categories like humans do. As shown in Fig. 1, a spectrogram is a way to visualize the frequency spectrum of the sound wave. In simple words, it is a photograph of the frequency spectrum present in the sound wave [7].

The generated spectrogram of the sound signal is infrequent so that noise intensity is found in lower region and strong components are found in higher region of the generated spectrogram. The generated spectrogram images can be used together with various machine learning classifiers. In the study of Sun *et al.* [8] they proposed an integrated