國 立 中 央 大 學

資訊工程學系

碩士論文

基於深度學習之聲音辨識及偵測

Sound Classification and Detection using Deep Learning

研 究 生：Dang Thi Thuy An

指導教授：王家慶 教授

中 華 民 國 106 年 06 月

NATIONAL CENTRAL UNIVERSITY

Department of Computer Science

Master Thesis

Sound Classification and Detection using Deep Learning

研 究 生 ： Dang Thi Thuy An

指導教授：Jia-Ching Wang

中 華 民 國 106 年 06 月

National Central University Library
Letter of Authorization for Electronic Theses and Dissertations

(The latest version since Feb. 2017)

This license authorizes my complete electronic thesis **(not in paper format, see footnote 1)** be archived and read in the "National Central University Library Electronic Theses & Dissertations System ".

( ) Agree (available immediately)
( ) Agree (available in ____/____/____ (yyyy/mm/dd))
(✓) Disagree, because: It will be submitted to the journal

and be archived and read in NDLTD(National Digital Library of Theses and Dissertations in Taiwan)

( ) Agree (available immediately)
( ) Agree (available in ____/____/____ (yyyy/mm/dd))
(✓) Disagree, because: It will be submitted to the journal

I undertake to submit my thesis, to National Central University Library and NDLTD(National Digital Library of Theses and Dissertations in Taiwan), non-exclusively and voluntarily. Based on the concept of sharing sources and mutually benefitted cooperation and in aims to repay the society and for the academic usage, I hereby grant a non-exclusive, for the full term of copy right protection, license to National Central University Library and NDLTD(National Digital Library of Theses and Dissertations in Taiwan), (a) to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internetf and distribute my thesis, in microform, paper, electronic and/or any other formats, regardless of place, time, and frequency;
(b) to authorize, sub-license, sub-contract or procure any of the acts mentioned in paragraph (a). And grant the reader the right to read, download and print out non-profitably.

Signature of the Author: Dang Thi Thuy An
Student Number: 104522608
Thesis Title: Sound Classification and Detection using Deep Learning
Supervisor: Jia-Ching Wang
Faculty, Department, School:
Computer Science and Information Engineering ☐ PH. D ☑ Master
Date: 107/07/25 (year, month, date).

Footnote :

1. The license is limited to only electronic files. In terms of Regulation of Copyright No 15, article 3, theses in paper can be published by the library. If you have patent or submitting concerns and do not want to publish your theses in paper, please fill out the Proclamation form.
2. This form requires the author's own signature, and it should be attached to a page after the paper theses' cover. (type-up signature is allowed for full electronic theses)
3. Readers should observe the Regulation of copyright while searching, reading, downloading and printing theses online non-profitably.

National Central University

Advisor's Recommendation for Graduate Students

This thesis is by Dang Thi Thuy An (Author) of the graduate program in Master degree, CSIE, entitled: Sound Classification and Detection using Deep Learning, which is written under my supervision, and I agree to propose it for examination.

Advisor Jia-Ching Wang

2017/06/30 (YYYY/MM/DD)

National Central University

Verification Letter from the Oral Examination Committee

for Graduate Students

This thesis is by Dang Thi Thuy An (Author) of the graduate program in Master Degree, CSIE, entitled: Sound Classification and Detection using Deep learning, who is qualified for master degree through the verification of the committee.

Convener of the degree examination committee

Jia-Yu Lin

Members

Min-Liang Wang

Jia-Ching Wang

Date: 2017/07/11

(YYYY/MM/DD)

中文摘要

本研究開發了各種深度學習模型，以在現實環境中進行聲學場景分類(ASC)和聲音事件檢測(SED)。我們利用卷積神經網絡(CNN) 及時間遞歸神經網絡(RNN) 用於音頻信號處理的優點來建立模型。CNN 對於提取多維數據的空間信息提供了一個有效率的方法，而 RNN 在學習具有時間順序的數據是強大的。我們的實驗在 DCASE 2017 challenge 的三個開發數據集中進行，包括聲學場景數據集，稀有聲音事件數據集和復音聲音事件數據集。為了避免過度擬合問題，我們採用一些數據增加技術，例如以給定的概率中斷輸入值到零，增加高斯噪聲或改變聲音的響度。

提出的方法的性能對於三個 DCASE 2017 challenge 的數據集優於基礎方法。聲學場景分類的準確度相對於基礎方法提高了 7.2%。對於罕見的聲音事件檢測，我們的方法平均誤差率為 0.26，F 評分為 85.9%，而基礎方法為 0.53 和 72.7%。對於復音聲音事件檢測，我們的方法的誤差率改進為 0.59，而基礎方法為 0.69。

Abstract

In this work, we develop various deep learning models to perform the acoustic scene classification (ASC) and sound event detection (SED) in real life environments. In particular, we take advantages of both convolution neural networks (CNN) and recurrent neural networks (RNN) for audio signal processing, our proposed models are constructed from these two networks. CNNs provide an effective way to capture spatial information of multidimensional data, while RNNs are powerful in learning temporal sequential data. We conduct experiments on three development datasets from the DCASE 2017 challenge including acoustic scene dataset, rare sound event dataset, and polyphonic sound event dataset. In order to reduce overfitting problem as the data is limited, we employ some data augmentation techniques such as interrupting input values to zeros with a given probability, adding Gaussian noise, and changing sound loudness.

The performance of proposed methods outperforms the baselines of DCASE 2017 challenge over all three datasets. The accuracy of acoustic scene classification improves 7.2 % in comparison with the baseline. For rare sound event detection, we report an average error rate of 0.26 and F-score of 85.9% compared to 0.53 and 72.7% of baselines. For polyphonic sound event detection, our method obtains a slight improvement on error rate of 0.59 while the baseline of 0.69.

Acknowledgements

The work presented in this thesis has been carried out at the Department of Computer Science and Information Engineering in National Central University, Taiwan during the years 2015-2017.

First of all, I wish to express my deepest gratitude to my research advisor, Professor Jia-Ching Wang, for guiding and encouraging me in my research. The fact that the thesis is finished at all is in great part of his endless enthusiasm for talking about my work.

I also specially thank to Mr. Toan Vu. He greatly supported me for theoretical and helped me take my initial thesis proposal and develop it into a true body of work, resulting in several conference and workshop papers together.

The financial support provided by National Central University fellowship program and advisor Professor Jia-Ching Wang is gratefully acknowledged.

Table of Contents

List of Figures

List of Tables

List of symbols and abbreviations

ANNs	Artificial neural networks
ASC	Acoustic scene classification
BLSTM	Bidirectional long short term memory
BRNNs	Bidirectional recurrent neural networks
BPTT	Backpropagation through time
CNNs	Convolutional neural networks
CRNN	Convolutional neural network
DNNs	Deep neural networks
FFT	Fast Fourier transform
GMM	Gaussian mixture model
HMM	Hidden Markov model
LSTM	Long short-term memory
NMF	Non-negative matrix factorization
NNs	Neural networks
ReLU	Rectified linear unit
RNNs	Recurrent neural networks
pRCNN	parallel convolutional neural network
SED	Sound event detection
SGD	Stochastic gradient descent
STFT	Short time Fourier transform
SVM	Support vector machine