A HIERARCHIC MULTI-SCALED APPROACH FOR SOUND EVENT DETECTION

Fabio Vesperini[1]*, Diego Droghini*[1]*, Daniele Ferretti*[1]
Emanuele Principi[1]*, Stefano Squartini*[1]*, Leonardo Gabrielli*[1]*, Francesco Piazza*[1]

[1] Politecnic University of Marche, Information Engineering Dept., Ancona, Italy,
{d.droghini, v.vesperini, d.ferretti}@pm.univpm.it
{e.principi, s.squartini, l.gabrielli, f.piazza}@univpm.it

ABSTRACT

Index Terms— One, two, three, four, five

1. INTRODUCTION

The field of computational auditory scene analysis (CASA) cover many topics. Nowadays, one of the most importat topic is the automatic sound event detection (SED). SED is defined as the task of analysing a continuous audio signal in order to extract a description of the sound events occurring in the audio stream. This description is commonly expressed as a label that marks the start, the ending, and the nature of the occurred sound (e.g., children crying, cutlery, glass jingling). Task 2 of DCASE challange 2017 [1] consists in determining the precise onset of three types of sounds: babycry, glassbreak and gunshoot.

2. PROPOSED METHOD

algo composto da 4 stadi principali: eat extr - event detection - multiscaled detection refinement - event onset annotation

2.1. Feature Extraction

descrivere logmel

2.2. Multilayer Perceptron Neural Network

descriveree a cosa adibita la prima rete descriere come sono organizzati inpute e label per il training descrivere cosa la predictionrow

2.2.1. Post Processig

come viene fatto (conv + media + th) per poi ottenere delle regioni contigue che proponiamo come eventi

2.3. Convolutional Neurl Network

lavora su base chunk 20 frame: organizzazione input (chunk da 20 non overlappati) e label descrizione di come lavorano i kernel per efatizzare la multiresolution approach cleassificazione chunk ovellappati (chunk size -1)

2.3.1. Post Processig

per ogni seq analizzo tutti gli eventi. Scarto quelli classificati con bck . Prendo il primo evento classificato come non bck perche lo scopo beccare l onset

3. EXPERIMENTAL SET-UP

In the following section .....

3.1. Onset Detection Stage

random search per la ricerca di parametri di layout della rete NN: validation split del dcase Per goni rete stata effettuata un gridsearch sui parametri di post processing per l ottimizzazione del ER E' stata selezionata la rete che ha ottenuto l' ER pi basso La rete selezionata stata trainata nuovamente con aggiungendo al trainset delle sequenze contenente gunshot per bilanciare i secondi di materiale degli eventi

3.2. Multiscaled Refinement Decision Stage

Per trainare la cnn abbiamo data in ingresso alla prima rete delle sequenze audio di solo bck. Gli eventi selezionati da questa rete rappresentano il materiale di training della classe bck per la cnn. Per le altre classi di eventi abbbiamo preso le porzioni di soli eventi mixed to bck relativi alle mixture fornite dal dcase e gli isolated events.

Abbiamo generato una stratified validation split del dataset appena descritto. Abbiamo effetuato una valutazione della cnn subase evento con la fmeasure. Finally, sul validati

3.3. Evaluation Phase

Descrizione con img della fase di valutazione Scelta th 0.20 invece che 0.25: per favorile meno deletion a discapito delle insertion. La cnn pensa a eliminare le insertion classificandole come bck

4. RESULTS

Error rate su base evento prima rete: Fmeasure cnn:

Risultato finale 0.18 pi report per classi (evetuale discussio sui babycry che nn vengono classificati bene)

4.1. Real Scenario application

Descrizione scenario reale: trainig cnn su 4 classi risultato finale 0.23

5. CONCLUSION

Fig. 1.

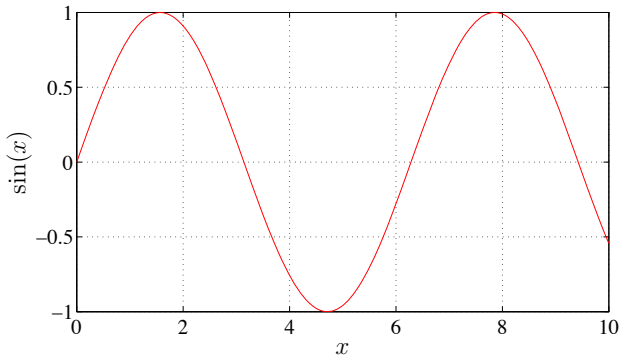

Figure 1: Example of a figure with experimental results.

$$\Delta^2 p(x, y, z, t) - \frac{1}{c^2} \frac{\partial^2 p(x, y, z, t)}{\partial t^2} = 0, \qquad (1)$$

6. REFERENCES

[1] http://www.cs.tut.fi/sgn/arg/dcase2017/.