Abstract

Making sense of the environment by sounds is an important research in machine learning community. In this work, a Deep Convolutional Neural Network (DCNN) model is presented to classify acoustic scenes along with a multiple spectrograms fusion method. Firstly, the generations of raw spectrograms and CQT spectrogram are introduced separately. Corresponding features can then be extracted by feeding these spectrogram data into the proposed DCNN model. To fuse these multiple spectrogram features, two fusing mechanisms, namely the voting and the SVM methods, are designed. By fusing DCNN features of the raw and CQT spectrograms, the accuracy is significantly improved in our experiments, comparing with the single spectrogram schemes. This proves the effectiveness of the proposed multi-spectrograms fusion method.

Introduction

In our DCASE2017 ASC submission, we also use a deep convolutional neural network (DCNN) based method to classify the acoustic scenes. Specifically, we produce multiple spectrograms from audio files which are used to train a DCNN model. We have explored two different productions of spectrogram: raw spectrogram and Constant-Q-Transform (CQT) spectrogram[6]. According to the sliding window width and shift step length, multiple raw spectrograms with different resolutions are general- ed. The classification performances of the DCNN model with multi-resolution raw spectrogram and CQT spectrograms are compared respectively. Next, we use the DCNN model to extract features, instead of classifying directly. A feature fusion method is applied in our submission. We have tried the fusion of features extracted from raw spectrograms with different resolutions, as well as the fusion of CQT spectrograms combined with raw spectrograms. Among our experiments, the CQT plus raw spectrograms fusion has achieved the best performance.

Methods and Materials

As we can see, the four fusion solutions have achieved satisfactory results. All of the four accuracies are greater than 0.89. Actually, the highest one is 0.8986 and the lowest one is 0.8939. It is easy to find that the difference of accuracies among these four are very slight. However, compared to the best results of raw spectrogram and CQT spectrogram solutions (0.8536 and 0.8052 respectively), the improvements in accuracies of these fusion solutions are still significant, which proves the effectiveness of our multiple spectrograms fusion. Similarly, Table 5 shows the accuracy superiority of SVM method over the voting in the fusion scenarios. To better understand the fusion performance, the class-wise accuracies of the best result, namely R706+CQ784(SVM), are further given in Table 6.

Results

Table 5: Accuracies of multiple spectrograms fusion solutions

<table>
<thead>
<tr>
<th>Method</th>
<th>Folder 1</th>
<th>Folder 2</th>
<th>Folder 3</th>
<th>Folder 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>R706+CQ784 (Voting)</td>
<td>0.8641</td>
<td>0.8217</td>
<td>0.8547</td>
<td>0.8504</td>
<td>0.8536</td>
</tr>
<tr>
<td>R706+CQ784 (SVM)</td>
<td>0.8567</td>
<td>0.8681</td>
<td>0.8635</td>
<td>0.8556</td>
<td>0.8620</td>
</tr>
<tr>
<td>R8752+CQ784 (Voting)</td>
<td>0.8521</td>
<td>0.8502</td>
<td>0.8547</td>
<td>0.8504</td>
<td>0.8536</td>
</tr>
<tr>
<td>R8752+CQ784 (SVM)</td>
<td>0.8567</td>
<td>0.8681</td>
<td>0.8635</td>
<td>0.8556</td>
<td>0.8620</td>
</tr>
</tbody>
</table>

SUBMISSION RESULTS

All the development data are utilized to train the model, and the submitted results are tested on this final model. According to the fusion methods, two systems are included in our submission to the DCASE2017 challenge (task 1). The first one is DCNN based voting system, which fuses the raw (R706) and CQ784 spectro-grams by voting method (namely the R706+CQ784 (Voting) solu- tion). The second one is DCNN based SVM system, which fuses the same data by SVM method (namely the R706+CQ784 (SVM) solution).

Conclusions

The main contributions of this work lie in two aspects as follows. First, a deep CNN model is presented, which is originated from [5] and is improved to be more suitable for our problem. Second, a multi-spectrogram fusion method is proposed. Multiple spectrograms are fed into the same DCNN model and the corresponding features are fused to improve the accuracy of classification. In this work, the raw spectrogram and the CQT spectrogram are studied. The best accuracy of using the raw spectrograms is 0.8536, and the one of using CQT spectrograms is 0.8052. Although the accuracy of using CQT spectrograms is unsatisfactory, it can significantly improve the accuracy when fused with the raw spectrogram. The best result of the fusion scheme is 0.8986 and outperforms the best results of the single spectrogram schemes by more than 0.045.

References


Contact

Xing Xiaotao
South China Normal University
1299670261@qq.com