# 2020「中技社科技獎學金」

## 2020 CTCI Foundation Science and Technology Scholarship

## 境外生研究獎學金

## Research Scholarship for International Graduate Students

# DCNN based Environmental Sounds Classification by Using Spectrogram Images and Various Data Augmentation Techniques

以光譜圖及不同的資料擴增技術用於深層卷積神經網路的環境聲音分類辨識

**PhD student: Zohaib Mushtaq, Advisor: Prof. Shun-Feng Su**

**Department of Electrical Engineering, National Taiwan University of Science & Technology**

## *Abstract*

The DCNN from scratch and 11 transfer learning models used on the meaningful data augmented spectral images. The concept of fine-tuning layers with optimal learning rates based discriminative learning was also used. This approach achieves the astonishing results on all used datasets. The strategy involved was the selection of the best pre-trained model, discussed in the approach, and tested on the aggregation of various features. This study also generates the two novel acoustic features, based on Mel filter bank, and named as Log (LogMel) L2M, and Log (Log (LogMel)) L3M. The novel data enhancement techniques also proposed which were the mixture of the reinforcement and aggregation of auditory features. This novel data augmentation approach was also tested on the real-time extracted audio recordings, identical to the categories of the used datasets. This proposed approach also has outstanding results on real-time audio data.

## *Research Focus*

### Acoustics Features

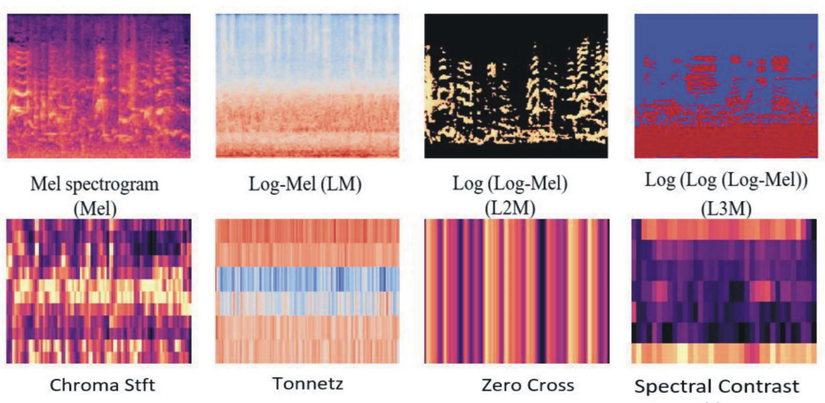


Fig.1.The spectrogram of various acoustic features used in this study.

### Optimal Learning Rates

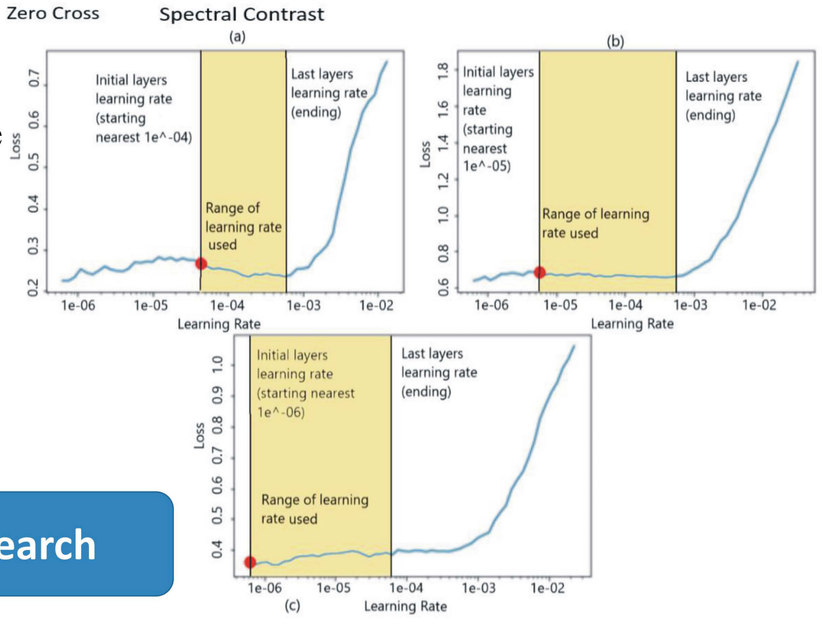


Fig. 3. The optimal learning rates of each dataset used. The red circle indicates the optimal starting learning rate. (a) ESC-10 ($1e^{-4}$,$1e^{-3}$), (b) ESC-50 ($1e^{-5}$,$1e^{-3}$), (c) Us8k ($1e^{-6}$,$1e^{-4}$).

### Implemented Strategies

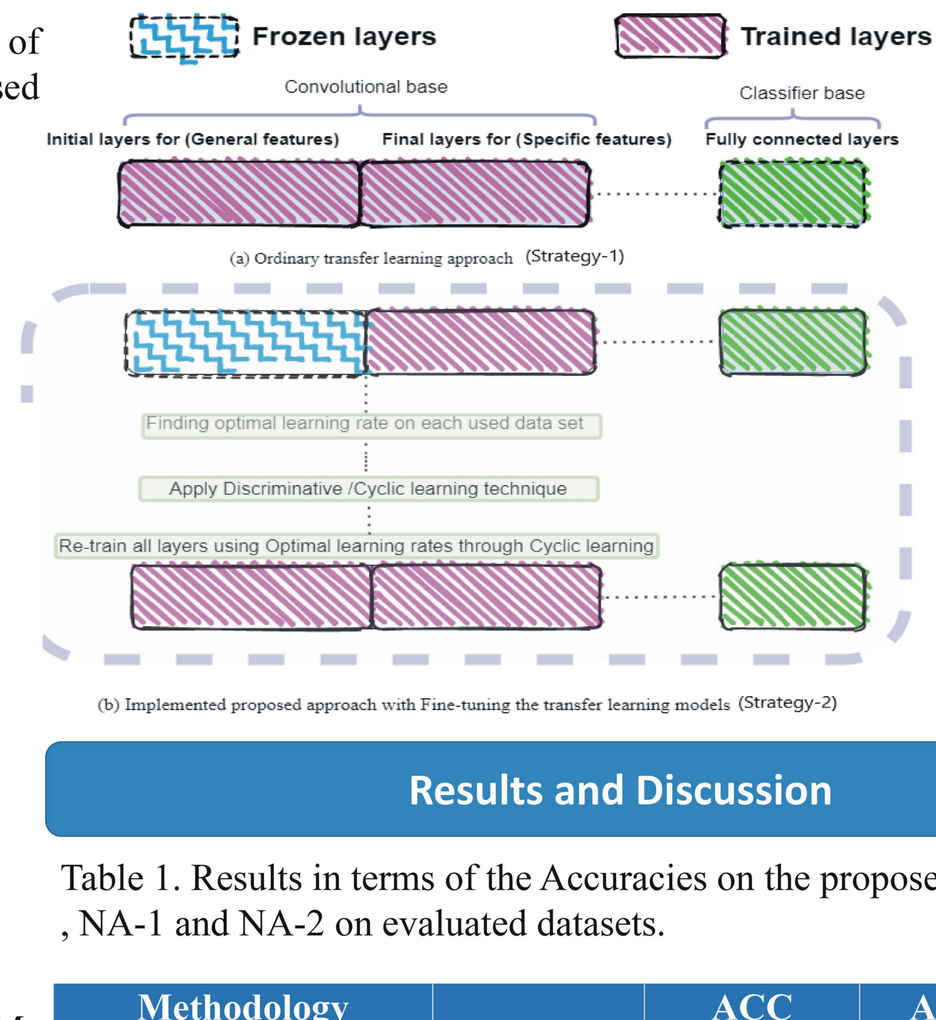


Fig.2.Block diagram of fine-tuned pre-trained weights through Cyclic learning by using optimal learning rates (proposed approach).

### Summary of This Research

The major contributions of this research include two auditory features named as L2M and L3M and the involvement of these SIF with Mel and LM to implement two NA-1 and NA-2. The first NA-1 involves the enhancement of SIF data by combining various spectrogram-based audio features and NA-2 consists of the vertical aggregation of these images in pairs. The baseline datasets used in the experiment were ESC-10, ESC-50, and Us8k. The best accuracy for ESC-10 and ESC-50 datasets has been reported are shown in Table 1.

### Results and Discussion

Table 1. Results in terms of the Accuracies on the proposed approach NAA , NA-1 and NA-2 on evaluated datasets.

| Methodology | | ACC ESC10 | ACC ESC50 | ACC Us8k |
|---|---|---|---|---|
| **Proposed NAA (ResNet-152)** | This study 2020 | 99.04 | 97.30 | 99.49 |
| **Proposed NA-1 (DenseNet-161)** | This study 2020 | 98.71 | 97.05 | 97.98 |
| **Proposed NA-2 (DenseNet-161)** | This study 2020 | 99.22 | 98.52 | 97.18 |

### Selected Journal Publications on this Research

- Z.Mushtaq and S. F.Su, "Environmental sound classification using a regularized deep convolutional neural network with data augmentation," *Appl. Acoust.*, vol. 167, p. 107389, 2020.
- Z. Mushtaq, S.-F.Su, and Q.-V.Tran, "Spectral images based environmental sound classification using CNN with meaningful data augmentation," *Appl. Acoust.*, 2020
- Mushtaq, Z.; Su, S.-F. Efficient Classification of Environmental Sounds through Multiple Features Aggregation and Data Enhancement Techniques for Spectrogram Images. *Symmetry* **2020**, *12*,

財團法人中技社 CTCI FOUNDATION