INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

# Automatic Exploration of Optimal Data Processing Operations for Sound Data Augmentation Using Improved Differentiable Automatic Data Augmentation

*Toki Sugiura*[1], *Hiromitsu Nishizaki*[1]

[1]Graduate School of Medicine, Engineering, and Agricultural Sciences, University of Yamanashi
t_sugiura@alps-lab.org, hnishi@yamanashi.ac.jp

## Abstract

Data augmentation is one of the methods used to robustly train machine learning models with a small dataset. This method randomly applies pre-defined data processing operations to input data, regardless of the characteristics of the input data. However, some data processing operations may be inappropriate for certain data. In this study, we propose a new method to automatically search for the best data processing operations for each sound file to be input into a sound classification neural network. The proposed method is an improvement on the previously proposed differentiable automatic data augmentation (DADA), which uses a differentiable neural network to select the optimal data processing operations. We evaluated our proposed method on an acoustic scene classification task on the ESC-50 dataset and demonstrated that the proposed method can train a more robust model compared to the original DADA-based data augmentation.

**Index Terms**: acoustic scene classification, data augmentation, differentiable automatic data augmentation

## 1. Introduction

When developing deep learning models, it is generally easier to achieve good models by training on a large dataset. However, in some cases, it may not be feasible to prepare a large dataset due to the cost of labeling or the scarcity of sample data. Therefore, a widely used technique is to pseudo-increase the number of data by applying some data processing to the data in the currently available dataset and making minor changes to the original data. This process is called data augmentation (DA). For instance, in image recognition tasks, images can easily be rotated, moved, and zoomed without reducing image quality and used as augmented data [1]. Acoustic data processing methods, such as Gaussian noise addition [2], time stretching [3], time masking [4], and mix-up [5, 6], have also been proposed for the acoustic scene classification task [7, 8] addressed in this paper.

Naturally, the choice of data processing operation for DA depends on the data. For example, in the field of image recognition, affine-transformations can be applied to data with minimal degradation, but such processing is unsuitable for signal data such as sound. Rotation and flipping are useful for object recognition and detection tasks in the image processing area, but they are not suitable for character recognition tasks. In speech recognition and acoustic scene classification tasks, data processing operations such as time stretching and time masking are frequently used but are not applicable to anything other than signals. Therefore, selecting and applying the appropriate data processing operation for the specific task is a crucial factor in improving model accuracy. However, training the model each time the DA design strategy is altered and seeking the appropriate combination of data processing operations require significant computational resources and human effort.

Therefore, an automatic optimization method for data augmentation called automatic data augmentation (ADA) has been proposed. Automatic data augmentation automatically searches for the optimal combination of data processing operations from a set of pre-designed data processing operations. This optimal combination of data processing operations is generally called a policy. A processing operation for a given data, or a combination of several of them, is called a "*sub-policy*," and a DA strategy that performs the optimal combination of sub-policies is called a "*policy*" in this paper. Research on ADA began with AutoAugment [9], and several methods have been proposed, including [10, 11, 12, 13].

However, although these studies can search for effective policies for most of the data in the training dataset, they have not been able to perform a policy search that focuses on individual data. For example, in an object recognition task that involves recognizing an image of a "car" and an image of an "apple," applying a data processing operation that performs color transformation to the "car" image can result in the transformed image being identified as a "car." However, in the case of an "apple" image, color is also an essential recognition factor, and the same data processing operation cannot be applied as in the case of a "car."

Therefore, in this paper, we propose an ADA that is adaptive to the nature of individual data. We apply the proposed ADA to the ESC-50 dataset [14], which is commonly used in acoustic scene classification tasks, to verify its effectiveness. If the proposed method's effectiveness is demonstrated on the dataset, it can be expected to be applied to other tasks, such as environmental sound analysis. Our proposed ADA is based on, and improves upon, differentiable automatic data augmentation (DADA) [13], which has been successfully applied to image classification tasks. In the original DADA, the selection of sub-policies (a combination of multiple data processing operations) is based on categorical distributions, but the proposed method uses a neural network to extract features of the input data and reflect them in the sub-policy search of the DADA. In addition, we introduce the concept of Faster AutoAugment (FAA) [12] to constrain the distribution of the embedding vectors of the data after data processing to be close to the distribution of the embedding vectors of the original data before processing. This can suppress inappropriate data processing.

In the experiment, we evaluated the performance of the proposed method for automatically exploring the optimal DA policy in the acoustic scene classification task using the ESC-50 dataset. The acoustic scene classification model was trained using the DA policy explored with the proposed method, improving accuracy by 2.1% compared to those using the DA policy explored with the original DADA. These results demonstrated the effectiveness of the proposed method in the acoustic scene classification task.

The contributions of this work can be summarized in the following two points:


10.21437/Interspeech.2023-202

- The automatic search for a DA policy that takes into account the characteristics of individual sound data was shown to be effective in the acoustic scene classification task.
- In addition, we demonstrated that performing the DA policy search using a method that minimizes the distance between the distributions of the embedding vectors in the dataset before and after data augmentation is effective.

## 2. Automatic Data Augmentation

Automatic data augmentation (ADA) is a method of searching for a DA policy that minimizes the validation loss by applying a data processing operation during model training and generalizes the dataset so that the model avoids over-training. Assuming $D_{train}$ and $D_{valid}$ are the training and validation data, respectively, and $\mathcal{T}$ is the data augmentation policy, ADA can be formalized as finding an algorithm that solves the following two-layer optimization problem:

$$\min_{\mathcal{T}} \mathcal{L}\left(\theta^* \mid D_{valid}\right) \text{ s.t. } \theta^* \in \operatorname*{argmin}_{\theta} \mathcal{L}\left(\theta \mid \mathcal{T}\left(D_{train}\right)\right) \quad (1)$$

where $\theta$ represents the parameters of a classification model, and $\mathcal{L}\left(\theta \mid D\right)$ is the loss for a set $D$.

In ADA, when training classification models, it is common to optimize the DA policy using a reduced dataset from which a portion of the full-size dataset is taken, and then apply the optimized DA policy to the full-size dataset to train the classification model from scratch.

### 2.1. AutoAugment

AutoAugment [9] was one of the first studies to focus on ADA. It solved the problem formulated in Equation (1) by fully training a small model multiple times with different DA policies on a subset of the training dataset and using the validation loss as the reward function for reinforcement learning.

The DA policy optimized by AutoAugment consists of several sub-policies. A sub-policy is a concatenation of a few data processing operations. A data processing operation has an intensity parameter $\mu$, which indicates how strongly the input data should be processed, and a probability parameter $p$, which is the probability that the operation will be applied.

Recurrent neural networks are used to optimize the DA policy in AutoAugment by optimizing which data processing operations to choose, its intensity $\mu$, and its probability of application $p$. The disadvantage of this method is its computational cost, and the policy search is extremely time-consuming.

### 2.2. Faster AutoAugment

Faster AutoAugment (FAA) is an improvement on the AutoAugment described in Section 2.1, which requires a huge computational cost to find a DA policy. FAA allows for a DA policy to be obtained with a realistic computational cost. Specifically, we approximated the parameters ($\mu$, $p$, and sub-policy selection) used for the DA policy in AutoAugment to be differentiable and designed an objective function to optimize them using back-propagation. This allowed for a fast optimization of the DA policy.

If the parameters of the DA policy can be replaced by differentiable ones, optimization can be achieved by back-propagation with an appropriate objective function. A candidate for this objective function could be verification loss minimization, as in differentiable architecture search (DARTS) [15]. However, this approach can be very time-consuming and memory-intensive. To avoid this problem, FAA assigns an objective function to a different approach.

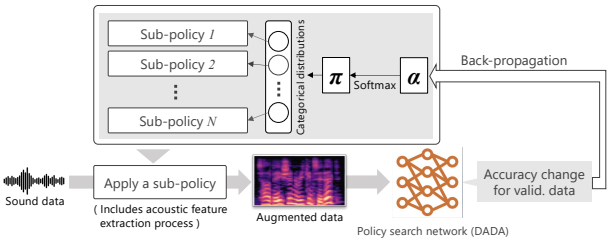


Figure 1: *Policy search with DADA.*

Data augmentation can be considered as a process of filling in missing data in the training dataset. Therefore, we consider it as a problem of minimizing the distance between the distribution of embedding vectors in the original data and the distribution of embedding vectors after data augmentation. The Wasserstein GAN (W-GAN) [16] with gradient penalty [17] can be used as this objective function, which minimizes the distance between these distributions. Unlike GANs for image generation [18], this model does not have a typical image generator using a conventional neural network. Instead, the DA policy is optimized and processes the data with several predefined sub-policies. In other words, the DA policy is optimized by considering the DA policy as the generator and the DA evaluation model as the discriminator.

### 2.3. Differentiable Automatic Data Augmentation

In this study, we propose an extended method for optimizing a DA policy using DADA, which has already been proposed in image recognition research [13]. To do this, we first give a brief description of DADA. DADA for acoustic scene classification model training has already been studied [19], and this study is an extension of this work. Basically, DADA is a derivative of AutoAugment; the contributions of DADA are the following two points:

1. Efficient DA policy optimization, and
2. Introduction of an accurate gradient estimator.

First, regarding the first point, DADA is a simple sampling-based optimization method in which two sub-optimizations, "DA policy optimization" and "classification model training," are repeated until convergence, as in AutoAugment. However, since this sequential optimization is computationally expensive, the parameters of the DA policy and the classification model are simultaneously optimized by the stochastic gradient descent method with reference to DARTS [15]. This one-pass strategy can significantly reduce the computational cost.

The classical gradient estimation method for the second point is the Gumbel-Softmax estimation method [20]. However, the gradient estimated by Gumbel-Softmax is biased. To overcome this, RELAX [21], an unbiased gradient estimator, was introduced. Figure 1 shows an overview of the policy search with DADA. Unlike Faster AA, the DA search model in DADA uses a classification model; the policy can be updated by using how the accuracy of the model changes after the DA search due to the DA policy as a criterion for determining the appropriateness of the DA.

The parameters of DADA are the same as those of AutoAugment, namely, the choice of sub-policy, its intensity parameter $\mu$, and the probability parameter $p$. These are differentiable in the same way as in Faster AA. However, the choice of the data processing method is optimized differently than in Faster AA.

In the case of DADA, all possible sub-policies are generated in advance from the set of data processing operations to be explored. Among these sub-policies, policy optimization is performed to select the appropriate sub-policies for the dataset.

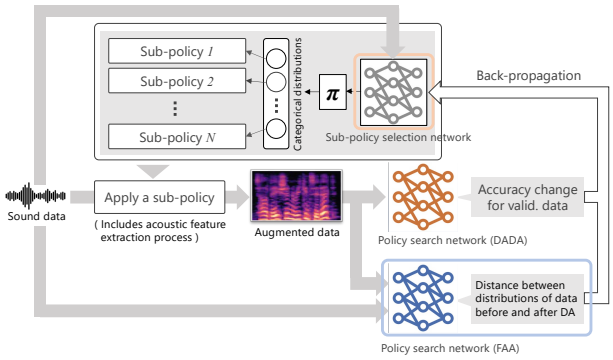


Figure 2: *Policy search with the improved DADA.*

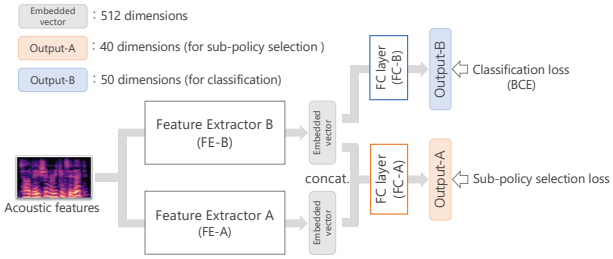


Figure 3: *The sub-policy selection network.*

A categorical distribution is used for sub-policy selection. When selecting the sub-policy, we sample from the categorical distribution $\mathrm{Cat}\,(\mathbf{s} \mid \boldsymbol{\pi})$ with probability $\boldsymbol{\pi}$. $\mathbf{s}$ is the set of sub-policies. That is, the priority of which sub-policy should be selected in the Categorical distribution is determined by $\boldsymbol{\pi}$. The probability $\boldsymbol{\pi}$ is calculated as a softmax for parameter $\boldsymbol{\alpha} = \alpha_{1:N}$, which defines the priority of the $N$ sub-policies. However, since Categorical and Bernoulli distributions are not differentiable, they are relaxed to a differentiable state using the Gumbel Softmax gradient estimation method [20]. For more accurate gradient estimation, we introduced an unbiased gradient estimator called RERAX [21].

The above represents the preliminary preparation for DADA. The search for the optimal policy is then performed by optimizing the parameter $\boldsymbol{\alpha}$ through a procedure that includes parameter update, loss computation, and gradient computation.

# 3. Improved DADA

The ADA techniques briefly introduced in Section 2 solve the problem of finding the optimal DA policy for a dataset. However, as mentioned in Section 1, it is possible that there are inappropriate sub-policies for some data. Therefore, in this study, we improve DADA to develop a method for finding the optimal DA policy that matches the characteristics of the sound data, so that an appropriate data processing operation is selected for each individual data.

## 3.1. Sub-policy selection network

Figure 2 represents a schematic diagram of our proposed ADA. In DADA, sub-policy selection was based on the parameter $\boldsymbol{\alpha} = \alpha_{1:N}$, which defines the priority of sub-policy selection, and a categorical distribution was used to assign a selection probability to each sub-policy. While the DADA method estimated this parameter $\boldsymbol{\alpha}$ directly, the proposed method estimates $\boldsymbol{\alpha}$ using a sub-policy selection network (ResNet34 [22] is used in this study). This is expected to allow the optimal sub-policy to be selected using the characteristics of the sound data.

As shown in Figure 3, this sub-policy selection network

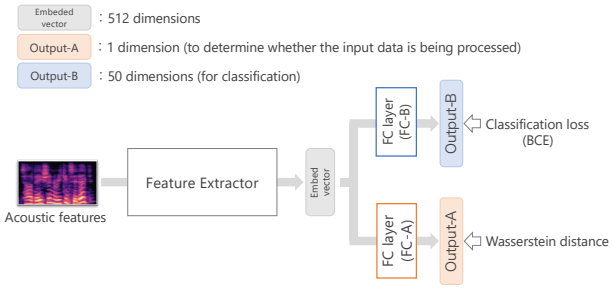


Figure 4: *The policy search network based on FAA.*

consists of two feature extractors (FEs) and two fully connected (FC) layers. The input sound data (acoustic features) are output by the two FEs. "Output-A" is a sub-policy selection vector and corresponds to the alpha parameter of DADA. The vector dimension is 40, which corresponds to the number of sub-policies to be selected in this work. "FE-A" has weights that are pre-trained without data augmentation using the dataset to be searched for the optimal policy. These weights are not updated during policy optimization. "FE-B" is trained during the policy search using the training data after DA. In the proposed method, the two FEs can be considered as trained with two different types of data, one before and one after DA. We believe that this would allow for an effective DA policy search with Output-A.

Next, we describe the training method of the sub-policy selection network. Output-A is used to optimize the DA policy selection with DADA, and Output-B is used for error backpropagation with class classification losses to train FE-B. Since these losses cannot be obtained simultaneously from a single sound data, the sub-policy selection network is optimized twice per training step. First, the data prior to DA are fed into the sub-policy selection network, and the appropriate sub-policy for the data is selected based on Output-A. The data are processed using the selected sub-policy and the gradient of the sub-policy is calculated using DADA. Based on this information, only "FC-A" is trained to obtain Output-A of the sub-policy selection network. At this point, the weights of the two FEs are kept fixed. The data after DA obtained in the above procedure are then fed into the sub-policy selection network. Now, FC-B and FE-B are trained to obtain Output-B based on the class classification loss of Output-B. The binary cross entropy (BCE) is used for the class classification loss of Output-B.

## 3.2. Minimization of distance between distributions

Another improvement is the application of the FAA framework to DADA. As shown in the lower part of Figure 2, a new model for "Policy search network (FAA)" is added. This model serves as a discriminator for the Wasserstein GAN [16] in the FAA framework.

The specific model structure is shown in Figure 4. The model takes sound data (acoustic features) and produces two outputs, Output-A and Output-B. Output-A is a one-dimensional output to determine whether the input data are being processed. The Wasserstein GAN used in FAA calculates the Wasserstein distance, which is the distance between data distributions, based on this true/false decision. The Wasserstein distance can be used as a loss function, and the generator (DA policy) is optimized so that the Wasserstein distances between data distributions are closer before and after DA. Output-B is the output of the 50 dimensions[1] for classification. FAA uses class classification loss in addition to Wasserstein distance to find a DA policy that is close to the nature of the original data

[1]This is due to the use of the ESC-50 dataset.

Table 1: *List of DA operations.*

(a) For waveform domain

| DA operation | Explanation |
|---|---|
| Non | Nothing to do |
| RandomFlip | Reverse in time direction |
| RandomScale | Extension/shortening in time direction |
| Gaussian | Adding Gaussian noise |

(b) For mel-spectrogram domain

| DA operation | Explanation |
|---|---|
| Non | Nothing to do |
| FrequencyMasking | Mask some data on the frequency axis |
| TimeMasking | Mask some data on the time axis |
| TimeStretch | Extension/shortening in time direction |

and less prone to misclassification.

# 4. Experiment

## 4.1. Experimental setup

### *4.1.1. Dataset*

The proposed method is evaluated on an acoustic scene classification task using the ESC-50 [14] dataset, which contains a total of 2,000 sound files of 50 classes of environmental sounds. On the other hand, the normal dataset contains 400 evaluation data, and the remaining 1,600 data are divided into 1,360 and 240 files, which are used as training and validation data, respectively. On the other hand, a reduced dataset is used for the DA policy search. This is a 1:1 split of the 1,600 data, excluding the evaluation data from the normal dataset, and is used for training and validation of the policy search network.

### *4.1.2. Classification model*

The ResNet34 model [22] was used as the classification model. A 64-dimensional log mel-spectrogram was used for acoustic features for the classification model training and policy search. We used BCE as the loss function during model training, the optimization function was Adam with 0.00025 of learning rate, and the mini-batch size was set to 32.

Although the accuracy could be improved by using a more powerful model, such as a transformer [23], a simple model was adopted because the purpose of this paper is to compare ADA methods and to demonstrate the effectiveness of the proposed ADA method.

### *4.1.3. DA sub-policy*

The data processing operations are classified into two categories, one for the waveform domain (Table 1 (a)) and the other for the mel-spectrogram domain (Table 1 (b)), each with four different data types. In this study, two of these data processing procedures were combined to form sub-policies. A total of 40 sub-policies were created. The combinations of data processing operations in the waveform and mel-spectrogram domains are shown in Figure 5.

### *4.1.4. Training conditions for policy search*

We used Adam as the optimization function for the neural networks for policy search, including DADA and the proposed method. The learning rate was set to 0.00025, and the mini-batch size was commonly set to 32. Classification loss and Wasserstein loss were used in the optimization of the FAA model, combined in a ratio of 10:1.

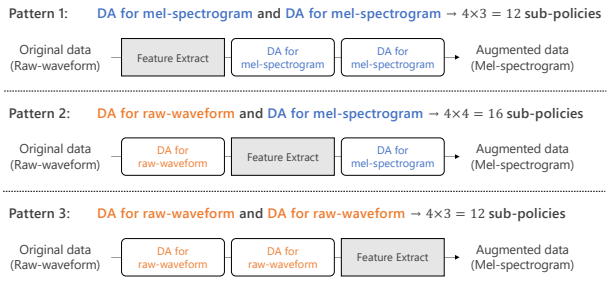


Figure 5: *Combinations of DA operations for raw-waveform and mel-spectrogram data.*

Table 2: *F1 scores [%] by ADA methods.*

| ADA methods | F1 score |
|---|---|
| w/o DA | 70.5 |
| DADA (baseline) | 73.9 |
| DADA+SPS | 75.3 |
| DADA+FAA | 75.6 |
| DADA+SPS+FAA (proposed) | **76.0** |

### *4.1.5. Evaluation metric*

The evaluation of the investigated DA policies is based on the performance of the acoustic scene classification model learned during the DA processing using the policies. The macro average of the F1 scores for each class was used as the evaluation metric for the acoustic scene classification model. The model was trained five times under the same conditions and the average of these training sessions was used as the final evaluation metric.

## 4.2. Result and discussion

In the experiment, we evaluated acoustic scene classification models trained under the following five ADA conditions: (**w/o DA**) no DA applied, (**DADA, baseline**) original DADA applied, (**DADA+SPS**) DADA plus a subpolicy selection model, (**DADA+FAA**) DADA plus FAA, and (**DADA+SPS+FAA**) proposed method. Table 2 presents the classification performance for the models trained with each ADA method.

The classification performance of the model trained without any DADA was 70.5%, and the application of ADA improved the classification performance. Among the ADA methods, the baseline DADA exhibited the lowest classification performance. The introduction of sub-policy selection models and FAA were individually effective and exhibited higher performance compared to the baseline DADA. The proposed method had the highest classification performance, namely 76.0%, which was an improvement of 2.1% on the baseline. This confirmed the effectiveness of the proposed method.

# 5. Conclusions

In this paper, we proposed an improved version of DADA for sound data that "reflected the characteristics of individual data" and "closed the data distribution before and after DA." The results indicate that the proposed method could train a classification model that performs better than the original DADA.

In the future, we intend to confirm the effectiveness of the proposed method on other model structures besides ResNet34 and on tasks dealing with other sound data.

# 6. Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 21H00901.

# 7. References


[1] S. Connor and K. M. Taghi, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 20, pp. 14–48, 2019.

[2] Y. Shi, L. Chao, Z. Zhang, L. Yiye, D. Wang, T. Javier, T. Zheng, and Y. Li, "Noisy training for deep neural networks in speech recognition," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 1, no. 2, pp. 1–14, 2015.

[3] J. Driedger and M. Müller, "A review of time-scale modification of music signals," *Applied Sciences*, vol. 6, no. 2, p. 57, 2016.

[4] D. S. Park, W. Chan, Y. Zhang, C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition," in *Proceedings of Interspeech 2019*, 2019, pp. 2613–2617. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2019-2680

[5] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proceedings of the International Conference on Learning Representations (ICLR2018)*, 2018, pp. 1–13.

[6] V. Verma, A. Lamb, C. Beckham, A. Najafi, I. Mitliagkas, D. Lopez-Paz, and Y. Bengio, "Manifold Mixup: Better Representations by Interpolating Hidden States," in *Proceedings of the 36th International Conference on Machine Learning (ICML2019)*, 2019, pp. 6438–6447.

[7] J. Salamon and J. P. Bello, "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.

[8] A. Mesaros, T. Heittola, A. Diment, B. Elizalde, A. Shah, E. Vincent, B. Raj, and T. T. Virtanen, "DCASE 2017 Challenge setup: Tasks, datasets and baseline system," in *Proceedings of the DCASE 2017 - Workshop on Detection and Classification of Acoustic Scenes and Events*, Munich, Germany, Nov 2017. [Online]. Available: https://hal.inria.fr/hal-01627981

[9] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR2019)*, 2019, pp. 113–123.

[10] S. Lim, I. Kim, T. Kim, C. Kim, and S. Kim, "Fast autoaugment," in *Proceedings of the Neural Information Processing Systems (NIPS2019)*, 2019, pp. 1–11.

[11] D. Ho, E. Liang, X. Chen, I. Stoica, and P. Abbeel, "Population based augmentation: Efficient learning of augmentation policy schedules," in *Proceedings of the International Conference on Machine Learning (ICML2019)*, 2019, pp. 2731–2741.

[12] R. Hataya, J. Zdenek, K. Yoshizoe, and H. Nakayama, "Faster autoaugment: Learning augmentation strategies using backpropagation," in *Proceedings of the European Conference on Computer Vision (ECCV2020)*, 2020, pp. 1–16.

[13] Y. Li, G. Hu, Y. Wang, T. Hospedales, N. M. Robertson, and Y. Yang, "DADA: Differentiable automatic data augmentation," in *Proceedings of the European Conference on Computer Vision (ECCV2020)*, 2020, pp. 580–595.

[14] K. J. Piczak, "ESC: Dataset for Environmental Sound Classification," in *Proceedings of the 23rd ACM International Conference on Multimedia*, 2015, pp. 1015–1018. [Online]. Available: https://doi.org/10.1145/2733373.2806390

[15] H. Liu, K. Simonyan, and Y. Yang, "DARTS: Differentiable architecture search," in *Proceedings of the International Conference on Learning Representations (ICLR2019)*, 2019, pp. 1–13.

[16] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of International conference on machine learning (ICML2017)*, 2017, pp. 214–223.

[17] G. Ishaan, A. Faruk, A. Martin, D. Vincent, and C. Aaron, "Improved Training of Wasserstein GANs," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS2017)*, 2017, pp. 5769–5779.

[18] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proceedings of the 4th International Conference on Learning Representations, ICLR 2016*, 2016, pp. 1–16.

[19] T. Sugiura, A. Kobayashi, T. Utsuro, and H. Nishizaki, "Automatic selection of appropriate data augmentation operation for acoustic scene classification model training," in *Proceedings of the 2022 IEEE 11th Global Conference on Consumer Electronics (GCCE)*, 2022, pp. 348–351.

[20] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," in *Proceedings of the International Conference on Learning Representations (ICLR2017)*, 2017, pp. 1–13.

[21] W. Grathwohl, D. Choi, Y. Wu, G. Roeder, and D. Duvenaud, "Backpropagation through the void: Optimizing control variates for black-box gradient estimation," in *Proceedings of the International Conference on Learning Representations (ICLR2017)*, 2017.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[23] A. F. R. Nogueira, H. S. Oliveira, J. J. M. Machado, and J. M. R. S. Tavares, "Transformers for urban sound classification – a comprehensive performance evaluation," *Sensors*, vol. 22, no. 22, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/22/8874