Sensor signal processing

Adaptive Learning With Surrogate Assisted Training Models for Acoustic Source Classification

G. Zucatelli*, R. Coelho**, and L. Zão†

Laboratory of Acoustic Signal Processing, Military Institute of Engineering, Rio de Janeiro 22290-270, Brazil
**Student Member, IEEE*
***Senior Member, IEEE*
†Member, IEEE

Manuscript received February 26, 2019; revised April 23, 2019; accepted May 13, 2019. Date of publication May 20, 2019; date of current version June 3, 2019.

Abstract—This article presents an adaptive learning solution for the selection of surrogate assisted training models. The main issue is to improve classification of acoustic sources considering that few labeled data are available. In this proposal, acoustic models are initially obtained from real signals. Surrogate models are then applied to assist the original training procedure and achieve improved classification accuracy. Learned models are defined according to the discrimination power among audio classes. Results show that the learning procedure leads to substantial accuracy gain in classification experiments. The proposed solution is also evaluated as a pre-learning step for a dictionary learning algorithm. In this scenario, the average accuracy is improved for a highly nonstationary and a stationary acoustic source.

Index Terms—Sensor signal processing, acoustic source classification, adaptive learning, index of nonstationarity, k-means singular value decomposition (K-SVD), surrogates.

I. INTRODUCTION

The classification of acoustic sources has gained significant attraction in the signal processing research area [1], [2]. Applications of environmental sound recognition include hearing aid, smart home, robot navigation, and surveillance systems. Dictionary learning [1], [3] and deep neural networks [2], [4], [5] are examples of techniques that are commonly adopted for source classification. These solutions generally require large amount of labeled training corpora. This scenario leads to a key challenge for the classification system, i.e., how is the sequence of relevant observations that represents the natural phenomena under analysis selected? Generally, this sequence must enable dimension reduction, representation analysis, and discrimination power to achieve accurate classification results. The learning process becomes even more challenging for signal analysis and classification when training and test datasets are nonstationary.

Recently, active learning (AL) and semi-supervised learning (SSL) techniques have been applied to deal with large amount of unlabeled instances that generally limits the classification accuracy [6], [7]. Both AL and SSL aim to achieve higher classification rates by selecting part of the unlabeled data to train the acoustic models. AL actively detects the most informative part of the unlabeled data, which is then manually labeled through a human annotator. On the other hand, SSL enables the automatic annotation of unlabeled data by using models trained on smaller sets of the labeled part of the database. The combination of AL and SSL has also been investigated to improve classification results with reduced human annotation efforts [6], [8].

This letter introduces an adaptive learning solution for sound classification considering that few labeled samples are available for training. Different from AL, the proposed technique requires no human effort to manually annotate unlabeled data. Instead, surrogate models replace the original acoustic models to improve the classification accuracy. The adoption of surrogate models is motivated by the nonstationary nature of real acoustic signals. In this article, surrogates consider the Kurtosis ratio (K), the power spectral density (PSD), and the index of nonstationarity (INS) [9] of labeled signals. The most informative models are automatically selected to better represent and distinguish the acoustic sources. Furthermore, models may be updated whenever a new set of signals is available for tests.

Classification experiments are conducted considering eight acoustic sources with different nonstationarity degrees for two different scenarios. In the first one, an adaptive learning is applied to the classical classification procedure based on mel-frequency cepstral coefficients (MFCC) and Gaussian mixture models (GMM). The proposed solution detects the most discriminative surrogates to adapt the acoustic models and lead to the best classification rates. Results show an improvement of more than 11 percentage points (p.p.) in the average classification accuracy. Thus, time-varying properties adopted in the surrogate generation (K, PSD, and INS) prove to be essential to guarantee improved sources discrimination. In a second scenario, the k-means singular value decomposition (K-SVD) [10] is applied over the original MFCC matrix to obtain a set of sparse vectors, which are then used to train the GMM of the K-SVD. The proposed approach is defined as a pre-learning strategy for the K-SVD dictionary learning to increase classification rates. This pre-learning step leads to almost 7 p.p. gain in the average accuracy.

II. ADAPTIVE LEARNING APPROACH

The block diagram in Fig. 1 illustrates the proposed adaptive learning solution. The main goal is to select an acoustic model λ_c that better distinguishes a target class c from the others. For that purpose, let $\left\{\Phi_c^0 \,|\, c = 1, \ldots, C\right\}$ denote the set of training acoustic signals available for C different classes. For each $c \in \{1, 2, \ldots, C\}$, a feature matrix $\mathbf{Y}_c^0$ is extracted from Φ_c^0 and is then used to obtain an initial acoustic model λ_c^0. The proposed solution begins with the genera-

Corresponding author: R. Coelho (e-mail: coelho@ime.eb.br).
Associate Editor: F. Falcone.
Digital Object Identifier 10.1109/LSENS.2019.2917661

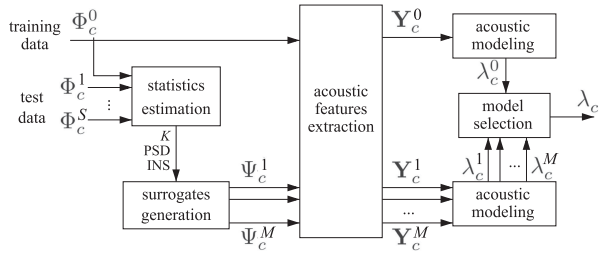


Fig. 1. Block diagram of the proposed adaptive learning.

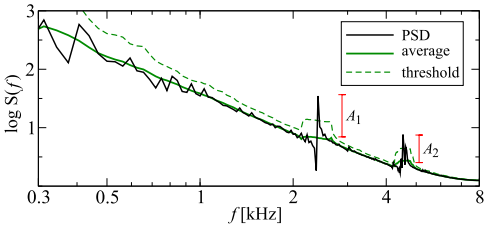


Fig. 2. Illustration of the PSD peak detection. Peaks correspond to frequency bins where the PSD value exceeds the selection threshold.

tion of M nonstationary surrogate signals $\{\Psi_c^m \mid m = 1, \ldots, M\}$ using the statistics of training samples. Initially, these synthetic sequences are generated considering the training samples statistics. Then, a set of matrices $\{\mathbf{Y}_c^m \mid m = 1, \ldots, M\}$ is obtained from these surrogates, leading to a new set of acoustic models $\{\lambda_c^m \mid m = 1, \ldots, M\}$. They are finally compared to the original model λ_c^0 to select the most informative acoustic model λ_c. In a posterior step, surrogates may also be generated applying the statistics of acoustic signals $\Phi_c^1, \Phi_c^2, \ldots, \Phi_c^S$ available for tests. Thus, final models are derived from this new augmented dataset.

A. Surrogate Generation for Learning

Consider an input reference signal $\{x(t)\}$ divided into Q short-time frames $\{x_q(t)\}$, $q = 0, 1, \ldots, Q-1$, with fixed time duration and 50% overlapping. The generation is performed in a frame-by-frame basis, and the algorithm for each frame q is described in three steps. First, a random sequence of uncorrelated samples $\{y_q(t)\}$ is generated with amplitude distribution defined by the Kurtosis ratio of $\{x_q(t)\}$ according to the method introduced in [11]. Next, coefficients of a finite impulse response (FIR) filter are computed on the basis of the target PSD [12]–[14]. In the third step, these uncorrelated samples are filtered in the time domain to obtain artificial samples $\{\bar{y}_q(t)\}$. Finally, samples of each short-time segment are concatenated to form a single surrogate data sequence $\{y(t)\}$.

In this article, filter coefficients computed according to [13] are modified to reproduce peaks that may appear in the PSD of the target real signal. PSD peaks are detected on the basis of a moving average criterion as shown in Fig. 2. In this example, the PSD is obtained from a 32 ms segment of a Train[1] real acoustic signal. Identified peaks (refer to A_1, A_2) correspond to frequency bins whose PSD value exceeds the threshold, which is here defined as the sum of the moving average with the standard deviation of L neighboring points multiplied by a factor F. In Fig. 2, the threshold corresponds to $L = 10$ and $F = 1$. If a total of P peaks are found at frequency bins $f_1, f_2, \ldots, f_P$, new filter coefficients $\{h'(t)\}$ are obtained by the summation of N initial filter coefficients, $\{h(t)\}$ with a set of sine waves that correspond to

[1]Available at http://www.freesfx.co.uk

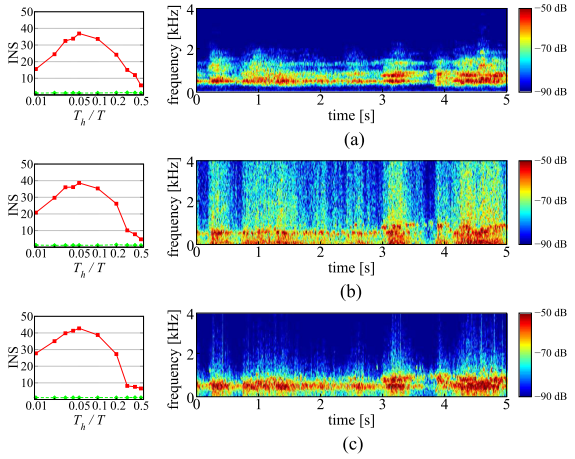


Fig. 3. INS and spectrograms of (a) Dogs acoustic signal and (b)–(c) two surrogates generated with different values of L and F.

PSD peaks, i.e.,

$$h'(t) = h(t) + \sum_{p=1}^{P} A_p \sin(2\pi f_p t)\,,\ 1 \le t \le N \tag{1}$$

where A_p amplitudes are given by the difference between PSD peaks and the moving average.

In summary, computation of filter coefficients $\{h'(t)\}$ is the starting point to obtain a sample sequence $\{\bar{y}_q(t)\}$ with Kurtosis and PSD defined by $\{x_q(t)\}$. Then, sequence $\{y_q(t)\}$ is generated with Kurtosis ratio $K_{y,q}$ according to [11]. Filtered samples $\bar{y}_q(t)$ are obtained by the convolution $\bar{y}_q(t) = y_q(t) * h'(t)$. In this case, synthetic samples reproduce the decaying rate β_q and the most prominent peaks present at the PSD of $\{x_q(t)\}$. Finally, the Q short-time segments are overlapped and added to form the surrogate signal $y(t)$.

If the INS estimated from the surrogate sample sequence $(\overline{\text{INS}})$ differs from the target value (INS_{tar}), amplitudes A_p are adjusted by $A_p \leftarrow r^2 A_p$, where $r = \text{INS}_{\text{tar}}/\overline{\text{INS}}$. The generation method is then applied to obtain a new sample sequence. This procedure is repeated until the INS estimated from $y(t)$ is considerably close to INS_{tar}, which corresponds to $r \approx 1$. In this article, the generation stops when r lies in the range [0.85, 1.15].

Surrogate signals obtained with the proposed method are able to reproduce the nonstationarity behavior and the time-frequency characteristics of the target signal. As an example, surrogates are generated to represent a Dogs[1] real nonstationary acoustic signal. Fig. 3 depicts INS values and spectrograms from the real signal and two surrogates. These surrogates were obtained with different parameters: $L = 16$ and $F = 1.6$ for Fig. 3(b), and $L = 64$ and $F = 2.0$ for Fig. 3(c). INS values are calculated considering different observation scales T_h/T, where T_h is the length adopted in the short-time spectral analysis, and $T = 5$ s is the total duration. Green dashed lines refer to the stationarity threshold $\gamma \approx 1$. Note that the INS of the real signal and surrogates are considerably similar. Moreover, the spectrogram energy of both surrogates are mainly concentrated at the same regions of the real signal. The choice of parameter values clearly leads to different surrogate spectrograms. In this example, the surrogate signal of Fig. 3(c) is the best candidate to achieve improved accuracy in the classification task.

B. Adaptive Learning for Sound Classification

After the generation of nonstationary surrogates, the adaptive learning solution is applied to determine which acoustic model λ_c^m,

$m = 1, \ldots, M$, better represents the target class $c \in \{1, \ldots, C\}$. Given a set of test signals, the classification is initially conducted with original models $\lambda_c = \lambda_c^0$ obtained from training signals. Let R represent the percentage of correctly recognized trials obtained in this classification procedure. In the proposed solution, a feature matrix $\mathbf{Y}_c^m$ is extracted from each surrogate Φ_c^m (see Fig. 1). These matrices are then used to obtain a set of acoustic models $\left\{\lambda_c^m \mid c = 1, \ldots, C;\ m = 1 \ldots, M\right\}$. Then, for each class $c \in \{1, 2, \ldots, C\}$, consider R_c^m the classification accuracy obtained by individually replacing λ_c by a new model λ_c^m $(1 \leq m \leq M)$. If any of these models lead to an increase in the recognition rate R, it is selected as the most informative acoustic model λ_c according to the maximum classification accuracy, i.e.,

$$\lambda_c \leftarrow \lambda_c^{\hat{m}}, \quad \text{where } \hat{m} = \arg \max_{1 \leq m \leq M} R_c^m. \tag{2}$$

After the selection of each model λ_c, the sound classification procedure may be repeated considering all adapted models. The proposed learning solution is also adaptive in the sense that a new set of surrogate signals can be generated whenever a new dataset is available. Then, acoustic models may be selected from this augmented dataset according to (2) in order to achieve a higher classification accuracy.

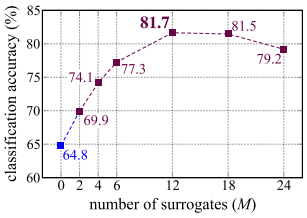
III. PRE-LEARNING SOLUTION FOR K-SVD

The adaptive learning with nonstationary surrogates is also proposed as a pre-learning strategy for the K-SVD. To this end, acoustic source classification is considered with sparse coding based feature [1]. This approach combines the K-SVD dictionary learning algorithm [10] and the orthogonal matching pursuit (OMP). Given a matrix of interest $\mathbf{Y}$, the sparse coding procedure aims to better express each column of $\mathbf{Y}$ as a linear combination of T_0 atoms of a dictionary $\mathbf{D}$. Therefore, the K-SVD objective function can be written as $\min_{\mathbf{D},\mathbf{X}}\{\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2\}$ subjected to $\|\mathbf{x}_i\|_0 \leq T_0\ \forall i$, where $\mathbf{x}_i$ is the ith column of $\mathbf{X}$. The K-SVD iteratively solves this minimization problem by updating each column of $\mathbf{D}$ and its corresponding relevant coefficients on $\mathbf{X}$ through a generalization of the k-means clustering method. In this letter, a dictionary $\mathbf{D}_c$ is generated from each training feature matrix. The sparse coefficients are obtained through the OMP for each training and test feature matrix. All OMP reconstruction coefficients are concatenated to form the sparse feature. For the pre-learning strategy, the sparse features extraction is applied over the matrices $\left\{\mathbf{Y}_c^m \mid c = 1, \ldots, C,\ m = 1, \ldots, M\right\}$ obtained after surrogate generation. The idea is to select and add the most relevant information to the sparse feature vectors and, thus, increase class discrimination.

IV. EXPERIMENTS AND RESULTS

Multiclass classification experiments are conducted with acoustic signals of eight sources obtained from the freeSFX[1] database: Chainsaw, Dogs, Fan, Rain, Shower, Siren, Subway, and Waterfall.[2] Three segments sampled at 22 050 Hz with time duration of 5 s are selected for each source, two for training (one for acoustic models generation and one for surrogate sequences selection) and one for tests. According to the maximum INS value ($\text{INS}_{\max}$) estimated from training signals, these acoustic sources are defined as Chainsaw ($\text{INS}_{\max} = 130$) and Siren ($\text{INS}_{\max} = 149$) are highly nonstationary; Dogs ($\text{INS}_{\max} = 37$) and Subway ($\text{INS}_{\max} = 40$) are nonstationary, whereas Fan, Rain, Shower, and Waterfall are stationary, i.e., INS values are close to the stationarity threshold.

In the first scenario of the classification of acoustic sources, feature matrices are composed of 12-dimensional MFCC vectors extracted from frames of 20 ms with 50% overlapping. During the training phase, each 12×500 matrix is used to obtain a GMM with 5 components. Each feature vector extracted from the remaining signals is then classified according to the maximum likelihood criterion. The adaptive learning is implemented with 2, 4, 6, 12, 18, and 24 surrogates considering short-time frames of 512 samples. In this scenario, acoustic models of four sources are improved with the surrogates: Chainsaw, Dogs, Fan, and Shower. As an example, the surrogate depicted in Fig. 3(c) leads to the most discriminative acoustic model for the Dogs source.



Parameter values:

M	L	F
2	32	1.2, 1.6
4	16, 32	1.2, 1.6
6	16, 32, 64	1.2, 1.6
12	16, 32, 64	0.8, 1.2, 1.6, 2.0
18	16, 32, 64	0.8, 1.0, 1.2, 1.4, 1.6, 2.0
24	16, 32, 64	0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0

Fig. 4. Average classification accuracies obtained with different number of surrogates.

Table 1. Classification Accuracies (%) Obtained in the MFCC + GMM Scenario Without Adaptive Learning.

actual source	classified source							
	Chainsaw	Dogs	Fan	Rain	Shower	Siren	Subway	Waterfall
Chainsaw	**5.8**	0.0	53.0	0.0	39.6	1.6	0.0	0.0
Dogs	24.0	**42.6**	5.4	0.0	0.6	26.8	0.6	0.0
Fan	1.0	0.0	**32.4**	0.2	2.0	3.2	61.2	0.0
Rain	0.0	0.0	1.4	**92.2**	0.0	0.0	0.4	6.0
Shower	1.8	0.0	0.2	0.2	**97.6**	0.0	0.0	0.2
Siren	1.4	0.0	0.8	0.0	15.6	**82.2**	0.0	0.0
Subway	2.4	0.0	10.2	0.4	1.0	0.0	**86.0**	0.0
Waterfall	0.0	0.0	12	12.4	5.6	0.2	0.0	**69.8**

Table 2. Classification Accuracies (%) Obtained in the MFCC + GMM Scenario With Adaptive Learning.

actual source	classified source							
	Chainsaw	Dogs	Fan	Rain	Shower	Siren	Subway	Waterfall
Chainsaw	**95.2**	0.0	0.4	0.0	0.0	4.4	0.0	0.0
Dogs	22.8	**57.6**	5.2	0.0	0.0	13.8	0.6	0.0
Fan	0.0	0.0	**74.2**	0.2	1.0	0.0	24.6	0.0
Rain	0.2	0.0	1.4	**91.8**	0.0	0.2	0.4	6.0
Shower	5.2	0.0	2.8	0.2	**91.4**	0.4	0.0	0.0
Siren	72.6	0.0	1.8	0.0	0.0	**25.6**	0.0	0.0
Subway	0.0	0.2	10.8	0.6	0.0	0.2	**88.2**	0.0
Waterfall	1.4	0.0	3.6	15.8	0.6	0.2	0.0	**78.4**

Classification accuracies obtained from different values of M are illustrated in Fig. 4. These results consider a total of 4000 trials. It also includes the result achieved without the adaptive learning, which corresponds to $M = 0$. The total number of surrogates M is achieved by setting the parameters L and F according to the table in the right of Fig. 4. Note that the classification accuracy is improved in 12.5 p.p. with only six surrogates, from 64.8% to 77.3%. The adaptive learning with $M = 12$ attains an average accuracy of 81.7%, i.e., 16.9 p.p. higher than the classification without learning. Furthermore, the adoption of more than 12 surrogates does not lead to higher classification rates. Thus, subsequent experiments are performed with test signals considering $M = 12$ surrogates.

Tables 1 and 2 present confusion matrices obtained without and with the adaptive learning technique, respectively. Boldface values refer to the accuracy of each source. It can be noticed that the proposed approach increases the classification rates for five acoustic sources. For the highly nonstationary Chainsaw, the classification accuracy is improved from 5.8% to 95.2%. An interesting gain of 41.8 p.p. is also found for the stationary Fan source. The assisted surrogate generation

[2] Acoustic signals are available at lasp.ime.eb.br.

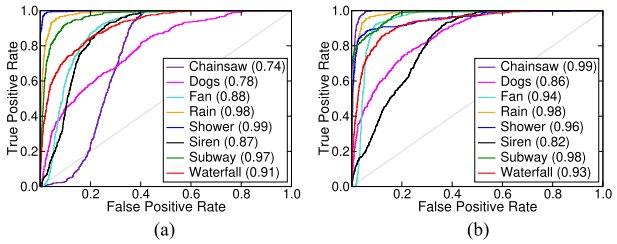


Fig. 5. ROC curves and corresponding AUC in the MFCC + GMM scenario (a) without and (b) with the adaptive learning.

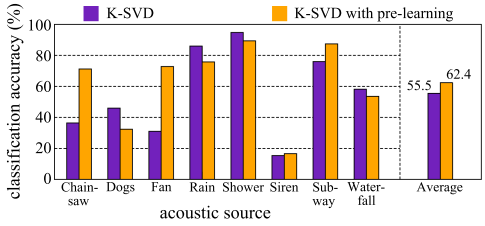


Fig. 6. Classification accuracy obtained with and without pre-learning for different acoustic sources.

is able to represent any index of nonstationarity. Thus, the adaptive learning may be applied to any acoustic source, including those with INS ≈ 1, i.e., stationary signals. The average accuracy is improved in 11.7 p.p., from 63.6% to 75.3%.

Receiver operator characteristic (ROC) curves and the area under the curve (AUC) are adopted as complementary evaluation measures for the confusion matrices (see Tables 1 and 2). Fig. 5 illustrates these plots computed for each acoustic source with and without the adaptive learning technique. Note that the adaptive learning solution improves the AUC values for five noise sources. Particularly for the Chainsaw source, AUC is improved from 0.74 to 0.99. This reinforces the substantial classification accuracy improvement presented in the confusion matrices for the Chainsaw source.

In the second scenario, the adaptive learning is adopted as a pre-learning step for K-SVD. It means that the K-SVD algorithm is used to generate sparse feature vectors from MFCC matrices extracted from the surrogates during the training phase. For each acoustic source, the K-SVD is set to 80 iterations in order to learn a 12×12 dictionary $\mathbf{D}_c$. The OMP was used for the sparse coding solution with $T_0 = 3$ nonzero elements. Experiments consider a total of 200 random initializations for the K-SVD algorithm. The dictionary $\mathbf{D}_c$ adopted for the acoustic source classification is the one that achieves the lowest reconstruction error. Each training and test MFCC matrix is then represented by a linear combination of each dictionary $\mathbf{D}_c$ using the OMP with $T_0 = 6$. The sparse feature vector is composed by concatenating 96 (8×12) reconstruction coefficients obtained with OMP. In terms of computational costs, the generation of $M = 12$ surrogates shows similar values to the K-SVD and OMP algorithms. Thus, the combination of the preprocessing with the K-SVD and OMP consumes about twice the processing time of the K-SVD and OMP algorithms without the preprocessing step.

The classification results obtained without and with the pre-learning strategy are shown in Fig. 6. The pre-learning is implemented considering the same MFCC matrices that lead to the classification results in Table 2. Note that the use of nonstationary surrogates leads to interesting accuracy gain for the Chainsaw (highly nonstationary) and Fan (stationary) sources, i.e., 34.8 and 41.8 p.p., respectively. Furthermore, a 11.4 p.p. improvement is achieved for the classification of the Subway source. The proposed pre-learning procedure also attains an average accuracy of 62.4%, which is 6.9 p.p. higher than that of the K-SVD without pre-learning. This result validates the use of nonstationary surrogates for dictionary learning based classification.

V. CONCLUSION

This article introduced an adaptive learning approach based on surrogate assisted training models. The adaptive learning procedure determines the acoustic models that better discriminate audio classes. Experimental results show that the proposed solution improves discrimination power, i.e., substantial gain is achieved in the average classification accuracy considering MFCC and GMM. It means that Kurtosis, PSD and INS are essential properties of the real signals to increase classification rates. The proposed method also contributes as a pre-learning strategy for the K-SVD dictionary learning. Results demonstrate that the surrogate generation enables good representation of different nonstationary acoustic signals. Moreover, the selection of acoustic models from surrogates proves that the representation analysis and discrimination power are achieved together with dimension reduction.

ACKNOWLEDGMENT

The work of R. Coelho was supported in part by the National Council for Scientific and Technological Development (CNPq) under Grant 307866/2015, in part by the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro under Grant 203075/2016, and in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil under Grant Code 001.

REFERENCES

[1] S. Zubair, F. Yan, and W. Wang, "Dictionary learning based sparse coefficients for audio classification with max and average pooling," *Digit. Signal Process.*, vol. 23, no. 3, pp. 960–970, May 2013.

[2] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 279–283, Mar. 2017.

[3] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, Mar. 2011.

[4] X. Cui, V. Goel, and B. Kingsbury, "Data augmentation for deep convolutional neural network acoustic modeling," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2015, pp. 4545–4549.

[5] B. Mandal, N. B. Puhan, and A. Verma, "Deep convolutional generative adversarial network-based food recognition using partially labeled data," *IEEE Sensors Lett.*, vol. 3, no. 2, pp. 1–4, Feb. 2019.

[6] W. Han *et al.*, "Semi-supervised active learning for sound classification in hybrid learning environments," *PLoS One*, vol. 11, no. 9, pp. 1–23, Sep. 2016.

[7] Z. Shuyang, T. Heittola, and T. Virtanen, "Active learning for sound event classification by clustering unlabeled data," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2017, pp. 751–755.

[8] Z. Zhang, E. Coutinho, J. Deng, and B. Schuller, "Cooperative learning and its application to emotion recognition from speech," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 1, pp. 115–126, Jan. 2015.

[9] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: A time-frequency approach," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3459–3470, Jul. 2010.

[10] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[11] R. Webster, "A random number generator for ocean noise statistics," *IEEE J. Ocean. Eng.*, vol. 19, no. 1, pp. 134–137, Jan. 1994.

[12] M. Al-Alaoui, "Novel digital integrator and differentiator," *Electron. Lett.*, vol. 29, no. 4, pp. 376–378, Feb. 1993.

[13] L. Zão and R. Coelho, "Generation of coloured acoustic noise samples with non-Gaussian distribution," *IET Signal Process.*, vol. 6, no. 7, pp. 684–688, Sep. 2012.

[14] R. Santana and R. Coelho, "Low-frequency ambient noise generator with application to automatic speaker classification," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 175, pp. 1–7, Aug. 2012.