IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 25, 2020, accepted December 8, 2020, date of publication December 11, 2020, date of current version January 6, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3044153

A Weighted Partial Domain Adaptation for Acoustic Scene Classification and Its Application in Fiber Optic Security System

NINGYU HE AND JIE ZHU, (Member, IEEE)
Shanghai Jiao Tong University, Shanghai 200240, China
Corresponding author: Ningyu He (ningyuhe_ruby@163.com)

ABSTRACT Domain adaptation (DA) is a technology that transfers knowledge from the source domain to the target domain. General domain adaptation assume that the source and the target domain have the same label space. However, in practical application tasks, the label of the target domain is often only a subset of the source label. For this situation, partial domain adaptation is usually used as an effective solution to transfer knowledge from a large number of labeled data sets to unlabeled micro data sets. In this article, a weighted partial domain adaptation method is proposed to solve the Acoustic Scene Classification (ASC) problem. Our method establish a connection between source and target domains to do the partial domain adaptation. Experiments are carried out on TUT and ESC-50 datasets which show that our method achieves state-of-the-art results. What is more, we apply the algorithms to an optical fiber perimeter security system to complete early warning by identifying intrusion signals.

INDEX TERMS Partial domain adaptation, generative adversarial training, multi-weighting scheme, acoustic scenes classification, optical fiber perimeter security system.

I. INTRODUCTION

Acoustic scenes classification (ASC) is the task for assigning the labels to the audio signals to determine the environment in which the signals are collected [1], [2]. In recent years, domain adaptation method (DA) based on deep learning (DL) has been proved to be an effective trick to solve classification tasks [3]. Existing domain adaptation methods generally assume that source and target domain share the same label space. However, in real ASC task, signals to be classified are usually only the part of the training data set. Standard DA is difficult to obtain satisfactory classification results in this situation. Therefore, partial domain adaptation (PDA) [4] is proposed to transfer knowledge from source dataset with sufficient labels to target dataset with fewer labels.

Domain adaptation (DA) is now widely used in computer vision [5], image recognition [6], natural language processing [7] and other fields. However, up to date, only a few studies applied domain adaptation (DA) methods to acoustic scenes classification (ASC) task. In 2018, IEEE Audio and Acoustic Signal Processing (AASP) proposed an ASC task with mismatched recording devices [8]. In this task, the data collected by each recording device can be regarded as a separate data domain. Therefore, DA is naturally used to solve the problem. S.Gharib et al used generative adversarial networks as feature extractors to extract domain-invariant features for domain adaptation [9]. K.Drossos et al replaced the adversarial adaptation process with Wasserstein Generative Adversarial Networks (WGAN) to improve the transfer effect [10]. However, the above methods are still based on the assumption that two domains have the same label spaces. In this article, we are addressing the transfer problem where target labels are the part of source labels.

Therefore, we propose a weighted partial domain adaptation method based on generative adversarial learning. We establish a connection between two generators to preserve the class-level structure during domain adaptation. A multi-weighting scheme is proposed not only to complete the selection of shared categories in the source domain, but also to help the network to distinguish whether the sample belongs to shared categories. Experiments are conducted on widely used acoustic classification datasets TUT dataset [9] and ESC-50 dataset [11], [12]. Results show that our method improves more than 20% classification accuracy in both dataset after

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

domain adaptation. What is more, we apply our method to a perimeter security system [13] and achieve good alarm function.

The rest of the article is organized as follows. Section II gives a brief description on related work. In Section III, we introduce our method in details. Section IV provides the experiments on TUT and ESC-50 dataset. Our algorithm is also applied to a fiber optic security system which is also introduced in Section IV. Finally, we conclude the article in Section V.

II. RELATED WORK

A. DOMAIN ADAPTATION

Domain adaptation (DA) is a representative method in transfer learning, which uses sufficient source domain samples to improve the performance of the target domain model. Here we introduce the basic ideas of domain adaptation. Source and target domain are noted as $D_S = \langle X_S, Y_S \rangle$ and $D_T = \langle X_T, Y_T \rangle$, where X represents the data distribution and Y represents the label processing. The key to the domain adaptation algorithm is to design a classifier H over x. The expected error of the H over its input x can be expressed as follow.

$$\epsilon(H, Y) = \mathcal{L}(H(x), Y(x))$$

where $\mathcal{L}$ is the loss function, $\epsilon(H, Y)$ indicates the differences between the output of the classifier H and the label Y. The goal of domain adaptation is to adjust H to get the small error $\epsilon_S(H, Y)$ in the source domain and adapt it to the target domain with low vaule of $\epsilon_T(H, Y)$.

The classifier H with the low $\epsilon_S(H, Y)$ can be obtained from classical training on the source domain D_S. However, H cannot be optimized by retraining on the target domain D_T due to the lack of labels. Therefore, the focus of domain adaptation changes to reduce the discrepancy between X_S and X_T.

B. MAXIMUM MEAN DISCREPANCY

As mentioned above, the goal of domain migration is to reduce the discrepancy between the source domain and the target domain. Maximum mean discrepancy (MMD) [14], [15] can reflect the similarity between two distributions, so it is often used as an important indicator in the domain adaptation algorithm to measure the discrepancy between the two domains.

Specifically, the statistical test method based on MMD refers to the following method. For samples of two distributions, we calculate the mean value of the samples on the continuous function f and take the mean difference as the mean discrepancy of the two distributions. Then the MMD is obtained by looking for the continuous function f in the sample space to make this mean discrepancy have the maximum value. Assume that m and n are two data sets sampled from two distributions p and q respectively and $\mathcal{F}$ is used to represent a continuous function set in the sample space. Then MMD can be represented by the following formula:

$$MMD[\mathcal{F}, p, q] = \sup_{f \in \mathcal{F}} \left(\mathbb{E}_{m\sim p}[f(m)] - \mathbb{E}_{n\sim q}[f(n)]\right)$$

C. ADVERSARIAL-TRAINING BASED DOMAIN ADAPTATION METHODS

Many domain adaptation algorithms have achieved good adaptation effects based on adversarial training. Such as Adversarial Discriminative Domain Adaptation (ADDA) [16], Multi-Adversarial Domain Adaptation (MADA) [17] and Conditional Adversarial Domain Adaptation (CDAN) [18]. In these methods, a domain discriminator is trained to minimize the domain discrepancy. Aadversarial losses can make sure that the learned function can transfer an individual source sample to the desired domain. However, previous methods only focus on the global transform. Since the discriminator can reduce the domain discrepancy, it destroys the class semantic feature in each category. Therefore, a partial domain adaptation algorithm is proposed based on adversarial training to overcome this shortcoming.

III. PROPOSED METHOD

In this section, our weighted partial domain adaptation method is introduced in details. Some mathematical notation is set to interpret our algorithm. Soucrce and target domain are defined as $D_S = \langle X_S, Y_S \rangle$ and $D_T = \langle X_T, Y_T \rangle$, where X represents the data and Y represents the label. And in the acoustic scenes classification (ASC) task of this article, the data of the same category in the source domain and the target domain have the same feature distribution ($M_S = M_T$), while the target label is a subset of the source label ($Y_T \in Y_S$).

A. NETWORK FRAMEWORK

The weighted partial domain adaptation we proposed is based on the theory of generative adversarial neural networks (GAN) [19]. The overall framework of our network is shown in Figure.1.

B. GENERATIVE TRAINING

As shown in Figure.1, two generators G_s and G_t are built in source and target domain respectively. G_s aims to generate simliar data F_t based on the source data X_s. G_t does the same job to generate the fake data F_s. The training loss of the generator in the source domain is consistent with the GAN network.

$$\begin{aligned}\mathcal{L}^s_{GAN}(X_s) &= \mathcal{L}_{dis_s} + \mathcal{L}_{cls_s}\\ \mathcal{L}_{dis_s} &= \mathbb{E}\left[\log C_s(X_s)\right] + \mathbb{E}\left[\log(1 - C_s(G_s(X_s)))\right]\\ \mathcal{L}_{cls_s} &= \mathbb{E}\left[\log C_s(X_s, Y_s)\right] + \mathbb{E}\left[\log C_s(G_s(X_s), Y_s)\right]\end{aligned}$$

where $\mathcal{L}_{dis_s}$ is the discrimination loss and $\mathcal{L}_{cls_s}$ is the classification loss. After that, classifier C_s uses both the real source data X_s and the generated fake data F_t as input for training. A similar training process takes place in the

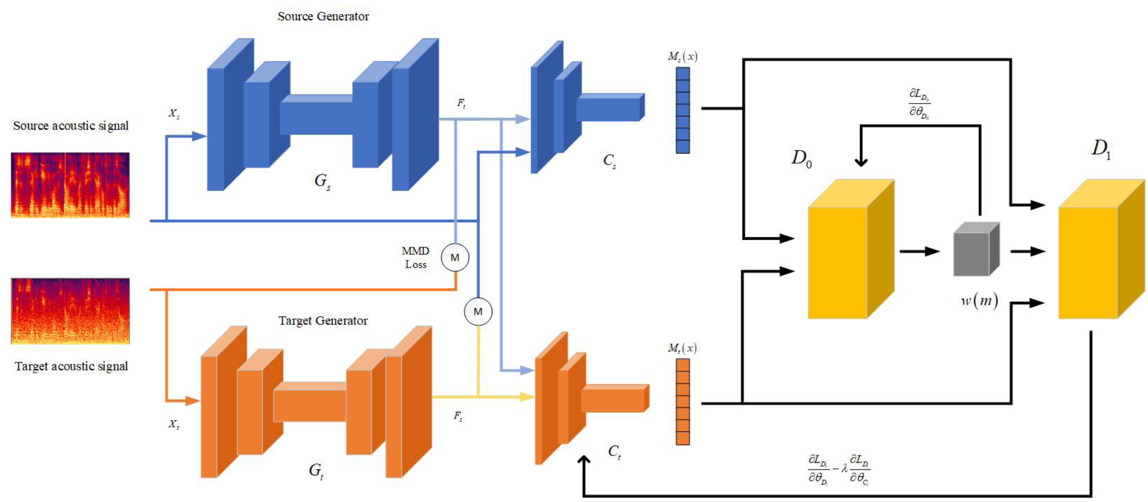


FIGURE 1. **The overall framework of our network.**

target domain.

$$\mathcal{L}_{GAN}^{t}(X_s, X_t) = \mathcal{L}_{dis_t} + \mathcal{L}_{cls_t}$$
$$\mathcal{L}_{dis_s} = \mathbb{E}\left[\log C_t(X_s)\right] + \mathbb{E}\left[\log\left(1 - C_t(G_t(X_t))\right)\right]$$
$$\mathcal{L}_{cls_s} = \mathbb{E}\left[\log C_t(F_t, Y_s)\right] + \mathbb{E}\left[\log C_t\left(F_s, \widehat{Y}_t\right)\right]$$

where $\widehat{Y}_t$ is the pesudo label from the $C_0(X_t)$. Here C_0 is a pre-trained classifier only based on the source data X_s.

Similar to the training of GAN, we put the classifier and generator together for joint generative training. Maximum mean discrepancy (MMD) is selected as an indicator for generator training. MMD can reflect the distribution discrepancy between two domains. In our network, two types of MMD loss are applied to describe the data distribution. One is the global MMD that shows the distance between source and target domain center. The other is the class MMD that calculates the distance between each class data. The whole MMD lose in the network is defined as follows.

$$\mathcal{L}_{MMD}^{s/t} = \mathcal{L}_{gMMD}^{s/t} + \frac{1}{N}\mathcal{L}_{cMMD}^{s/t}$$

where N is the number of data classes. We integrate the generator training loss and MMD loss to complete the generative training of our network. To sum up, generator and classifier are respectively established in the source and target domain. The classifier takes the output of the generator as input and aims to get the best classification result, while the generator aims to minimize the joint loss which is shown as follows. λ in the equation controls the relative weight of two losses.

$$\mathcal{L} = \mathcal{L}_{GAN}^{s} + \mathcal{L}_{GAN}^{t} + \lambda\left(\mathcal{L}_{MMD}^{s} + \mathcal{L}_{MMD}^{t}\right)$$

C. MULTI-WEIGHTING SCHEME

After training the generators and classifiers, a multi-weighting scheme is proposed to complete the subsequent partial domain adaptation. In our method, we propose two weights to do the partial domain adaptation. One is the shared-class weight [20], which is used to select the shared category between two domains. The other is the shared-sample weight. Our method consider the weight of the category and the sample at the same time to achieve better transfer effect.

We first introduce the shared-class weight. We treat the output of the last convolutional layer in the classifier trained in the previous chapter as a feature extractor to get the data feature $M(x)$. The general idea is to learn both discriminator and domain invariant features. Therefore, the discriminator loss in our network is similar to the GAN minimax loss:

$$\min_{M_s, M_t}\max_{D}\mathcal{L}(D, M_s, M_t) = \mathbb{E}_{x\sim Z_s(x)}\left[\log D(M_s(x))\right] + \mathbb{E}_{x\sim Z_t(x)}\left[\log\left(1 - D(M_t(x))\right)\right]$$

where M_s and M_t are the features from source and target data. D is the domain discriminator to identify whether the features come from the source or target domain. The loss minimizes the data distribution divergence on the feature space M while produces a stricter bound for the discriminator D to achieve the more accurate identification results.

In the process of training the discriminator D, for any $M_s(x)$ and $M_t(x)$, training the discriminator D is to maximize the loss:

$$\begin{aligned}\max_{D}&\,\mathcal{L}(D, M_s, M_t)\\ &= \int_x Z_s(x)\log D(M_s(x)) + Z_t(x)\log\left(1 - D(M_t(x))\right)dx\\ &= \int_m Z_s(m)\log D(m) + Z_t(m)\log(1 - D(m))\,dm\end{aligned}$$

where $m = M(x)$ is the feature sample after extraction. And the theoretical optimal solution D of this optimization problem can be obtained by Leibniz's rule.

$$D^{*}(m)=\frac{Z_s(m)}{Z_s(m)+Z_t(m)}$$

As mentioned above, the shared-class weight is proposed to determine whether the features extracted are from independent or shared classes in two domains. Fortunately, we find that the optimum D^* is a good indicator. It can be found that when the value of $D^*(m)$ is closes to 1, features are more likely to come from the particular classes in the source domain. Conversely, when the value of $D^*(m)$ is closes to 0, features are more likely to come from the shared classes. Therefore, the shared-class weight we propose is related to $D^*(m)$, and the relationship between the two is as follows.

$$\tilde{w}(m)=1-D^{*}(m)=\frac{Z_t(m)}{Z_s(m)+Z_t(m)}$$

It is clear from the equation that the shared-class can reflect the distribution of data and give larger weight to similarly distributed data to achieve the purpose of selecting shared classes between two domains. The shared-class weight w_{sc} is normalized for network training as follows.

$$w_{sc}(m)=\frac{\tilde{w}(m)}{\mathbb{E}_{m\sim Z_s(m)}\tilde{w}(m)}$$

The training process of w_{sc} is done through the discriminator D_0 (shown in Figure.1). Another discriminator D_1 with the weighted data is trained to reduce the shift on the shared classes. After adding the weight w_{sc}, the traing goal our network turns to:

$$\begin{aligned}&\min_{M_s,M_t}\max_{D}\mathcal{L}(D,M_s,M_t)\\&=\mathbb{E}_{x\sim Z_s(x)}\left[w_{sc}(m)\log D(M_s(x))\right]\\&\quad+\mathbb{E}_{x\sim Z_t(x)}\left[\log(1-D(M_t(x)))\right]\end{aligned}$$

For the discriminator D_1, w_{sc} can be seen as a constant. So the optimum D_1 can also be obtained by Leibniz's rule.

$$D_1^{*}(m)=\frac{w_{sc}(m)Z_s(m)}{w_{sc}(m)Z_s(m)+Z_t(m)}$$

After the shared classes between the two domains are determined, shared-sample weights w_{ss} are proposed to improve our weighting scheme. A novel shared-sample classifier C_{ss} (not shown in Figure.1) is established to distinguish whether a sample belongs to shared classes. We normalize the output of the C_{ss} by the sigmod function to obtain the shared-sample weight w_{ss}.

$$w_{ss}(x)=sigmoid(C_{ss}(M(x)))$$

The larger $w_{ss}(x)$ is, the more likely sample x is relevant to shared classes. Therefore, the multi-weighting we proposed can not only select shared classes for partial domain migration, but also can correct misclassified samples and greatly improve the fault tolerance of the model. In summary, we design a multi-weighting algorithm based on the generation of the generative adversarial training to complete the partial domain adaptation.

IV. EXPERIMENTS AND APPLICATIONS

Several experiments are carried out on TUT and ESC-50 Acoustic Scenes dataset. We compared our algorithm with SOTA domain adaptation algorithms. Experimental results show that our method achieves better domain adaptation effect. The migrated network can achieve higher classification accuracy.

A. EXPERIMENTS ON TUT AND ESC-50 DATASET

TUT dataset is widely used in Acoustic Scenes Classification task. The dataset includes audio recordings collected under ten different acoustic scenes, such as 'park', 'metro station', 'street traffic' and so on. Each class of audio data is recorded by three different acquisition devices and marked with A, B and C to distinguish. Device A is a professional recording device that can capture high quantity audio data, while device B and C are common recording devices. In our experiments, we regard the data from device A as the source domain data and data from B, C as the target domain data to do the domain adaptation.

The ESC-50 dataset is a labeled collection of 2000 environmental audio recordings suitable for Acoustic Scenes Classification. The dataset consists of 5-second-long recordings organized into 50 semantical classes (with 40 examples per class). In our domain adaptation experiments, we divide the data under each type of label into two parts, one as the source domain data and the other as the target domain data.

In both TUT and ESC-50 dataset, we take the energy spectrum of the audio signal as the input to the network and use the deep network as the feature extractor to extract abstract features instead of traditional speech features. Feature extractor is designed as a residual convolution neural network with long short term memory (LSTM) [21]. The network establishes five layers of convolution layers and three layers of LSTM to extract the time-frequency features of the input. Moreover, residual concepts are introduced in to address the possible degradation problem. Domain discriminators are simply designed as four layers of fully connected layers.

In order to verify the effect of partial domain adaptation algorithm, the data in the source domain selects all classes of data, while the data in the target domain only selects part classes of data. We select different types of target data for multiple experiments.

Figure.2 records the average classification accuracy without domain adaptation in TUT dataset. On the contrary, average accuracy after domain adaptation is shown in Figure.3. It is clear that target data without domain adaptation is difficult to obtain good recognition results on the classifier trained in the source domain. The average recognition accuracy is only 20.4%. This is because the differences in the acquisition equipment will greatly affect the characteristics of the audio signals. In addition, the shift on label domain also increases the difficulty of classification task. In contrast, results in Figure.3 demonstrate the effectiveness of our algorithm.

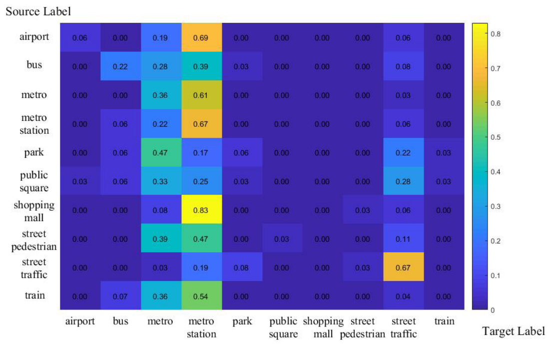


FIGURE 2. TUT classification results before domain adaptation.

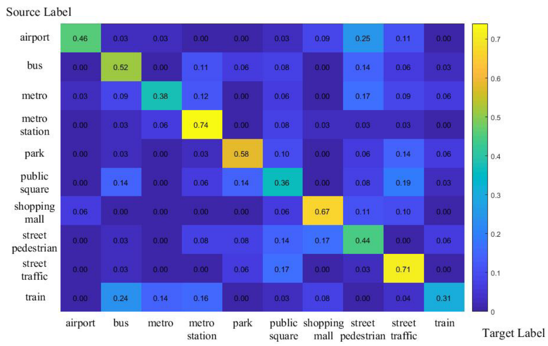


FIGURE 3. TUT classification results after domain adaptation.

TABLE 1. Ablation Experiments on Multi-Weighting Scheme on TUT Dataset.

Ablation experiments	$S \rightarrow T$	$T \rightarrow S$	Avg(%)
Without w_{sc}	59.16 ± 0.18	46.01 ± 0.26	52.59
Without w_{ss}	46.51 ± 0.23	42.38 ± 0.12	44.45
With multi-weighting	62.05 ± 0.12	50.57 ± 0.23	56.31

TABLE 2. Ablation Experiments on Multi-Weighting Scheme on ESC-50 Dataset.

Ablation experiments	$S \rightarrow T$	$T \rightarrow S$	Avg(%)
Without w_{sc}	79.09 ± 0.35	76.59 ± 0.57	77.84
Without w_{ss}	72.42 ± 0.29	70.37 ± 0.39	71.40
With multi-weighting	82.40 ± 0.02	80.92 ± 0.26	81.66

It is obvious that the recognition accuracy on the diagonal of Figure.3 is significantly improved. The recognition accuracy of any class has been improved to more than 30% and the average classification accuracy rate has increased to 56%.

An ablation experiment is designed to reflect the effectiveness of our multi-weighting scheme. We cancel the shared-class weight w_{sc} and the shared-sample weight w_{ss} separately, and conduct ablation experiments on the TUT and ESC-50 dataset. The results of the ablation experiments are recorded in Table 1 and Table 2.

It can be seen from the Table 1 and the Table 2 that the multi-weighting scheme we designed effectively improve the domain adaptation results of the network. Furthermore, the improvement of the shared-class weight w_{sc} to the network is significantly greater than the shared-sample weight w_{ss}. It is reasonable that w_{sc} can help judge whether the extracted features are from independent or shared classes, which plays a decisive role in our partial domain adaptation methods.

More experiments are proposed to prove the superiority of our method. Comparisons are carried out between our method and SOTA domain adaptation methods. Deep Adaptation Network (DAN) [22], Joint Adaptation Network (JAN) [23], Adversarial Discriminative Domain Adaptation (ADDA) [16], Multi-Adversarial Domain Adaptation (MADA) [17], Conditional Adversarial Domain Adaptation (CDAN) [18], Selective Adversarial Networks (SAN) [24], Asymmetric Tri-training for unsupervised Domain Adaptationare (ATDA) [18] and Multiple Instance Detection Network (MIDN) [25] are selected for comparison experiments.

For the reliability of the experiment, we not only transfer the knowledge from the source domain to the target domain ($S \rightarrow T$), but also convert the target domain and the source domain ($T \rightarrow S$). The results of the comparative experiments on TUT dataset are recorded in Table 3 and results on ESC-50 dataset are shown in Table 4.

TABLE 3. Accuracy(%) on TUT Dataset.

Domain adaptation methods	$S \rightarrow T$	$T \rightarrow S$	Avg
DAN	48.73 ± 0.29	46.70 ± 0.24	47.72
JAN	59.16 ± 0.18	46.01 ± 0.26	52.59
ADDA	49.01 ± 0.27	47.96 ± 0.30	48.49
MADA	56.77 ± 0.20	46.89 ± 0.29	51.83
CDAN	57.37 ± 0.21	49.56 ± 0.17	53.47
SAN	56.75 ± 0.36	49.29 ± 0.25	53.02
ATDA	47.35 ± 0.34	46.91 ± 0.27	47.13
MIDN	58.67 ± 0.25	49.74 ± 0.26	54.21
Our method	62.05 ± 0.12	50.57 ± 0.23	56.31

TABLE 4. Accuracy(%) on ESC-50 Dataset.

Domain adaptation methods	$S \rightarrow T$	$T \rightarrow S$	Avg
DAN	73.95 ± 0.43	69.33 ± 0.35	71.64
JAN	77.42 ± 0.29	70.37 ± 0.39	73.90
ADDA	75.12 ± 0.31	70.65 ± 0.17	72.89
MADA	78.67 ± 0.23	74.28 ± 0.30	76.48
CDAN	78.18 ± 0.16	75.42 ± 0.25	76.80
SAN	76.38 ± 0.33	78.86 ± 0.25	77.62
ATDA	73.22 ± 0.33	72.42 ± 0.45	72.82
MIDN	77.37 ± 0.24	77.53 ± 0.19	77.45
Our method	82.40 ± 0.02	80.92 ± 0.26	81.66

As shown in Table 3 and Table 4, our method performs best across both transfer tasks. It outperforms the second best method by more than 3% on both two dataset. We raise average accuracy from the baseline DAN of 71.64% to 81.66% on ESC-50 dataset which indicates that the multi-weighting scheme we proposed is reasonable and effective.

B. APPLICATION IN PERIMETER SECURITY SYSTEM

Since our method has achieved good recognition accuracy on acoustic classification task, we apply it to a perimeter security system and achieve satisfactory results.

With the development of society, the intelligent perimeter security system has been applied in various occasions. So we design a perimeter security system [13], [21] based on the optical fiber sensors. System collects external intrusion signals through optical fiber sensors and analyzes the types of intrusion signals to achieve the purpose of early warning. The overall framework of the perimeter security system we designed is shown in Figure.4.

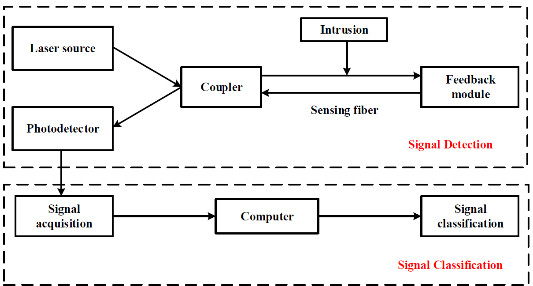


FIGURE 4. **The overall framework of perimeter security system.**

In our designed perimeter security system, the optical fiber sensor collects signals from the vibration on the optical path. The collected signal can be regarded as an audio signal. Thus the perimeter security system can be seen as an alternative audio scene classification task. What is more, although the security system has retrained various types of intrusion signals, when the security system works, each intrusion signal is individually detected and identified, which also coincides with the thought of partial adaptation. The training data contains a variety of intrusion signals, including vehicles passing by, man-made mining, etc. However, the intrusion signal is finally divided into two classes according to their labels, which is only divided into harmful intrusion signals and harmless intrusion signals. Our algorithm is only used in the previous intrusion signal classification task.

TABLE 5. **Alarm Accuracy in Different Environment.**

Classification methods	Sunny days	Rainy days
DNN	76.3%	74.8%
BPNN	86.9%	82.7%
SVM	87.6%	84.5%
Our method	90.2%	86.7%

Traditional recognition algorithms like Back Propagation Neural Networks (BPNN), Support Vector Machines (SVM) and Deep Neural Networks (DNN) are used to conduct comparative experiments with our method. And the results are shown in Table 5. It is clear from the Table 5 that our method achieves the best recognition accuracy in any environment. This proves that our algorithm can achieve the good domain adaptation effect on different data sets, which is effective and universal.

V. CONCLUSION

In this article, a weighted partial domain adaptation method is proposed for acoustic scenes classification task. We expand on the ideas of Generative Adversarial Networks and establish a connection between the two generators in the source and target domain. Therefore, generators can preserve the class-level structure while generating the data samples. What is more, a multi-weighting scheme is proposed to complete the partial domain adaptation. The shared-class weights obtained through discriminator training can help us find shared categories between domains. Moreover, the shared-sample weights serve as a good supplement to describe the association between samples and shared classes. Experiments are taken on TUT and ESC-50 dataset among our method and the SOTA domain adaptation algorithms (DAN, JAN, ADDA, MADA, CDAN, SAN, ATDA, MIDN). Results show that our method outperforms the second best method by more than 3% on both two dataset. What is more, our method is applied to the optical fiber security system and achieves good results, which proves that our algorithm has a strong universality.

REFERENCES

[1] D. Giannoulis, E. Benetos, D. Stowell, M. Rossignol, M. Lagrange, and M. D. Plumbley, "Detection and classification of acoustic scenes and events: An IEEE AASP challenge," in *Proc. IEEE Workshop Appl. Signal Process. to Audio Acoust.*, Oct. 2013, pp. 1–4.

[2] A. Mesaros, T. Heittola, and T. Virtanen, "Acoustic scene classification: An overview of dcase 2017 challenge entries," in *Proc. 16th Int. Workshop Acoustic Signal Enhancement (IWAENC)*, Sep. 2018, pp. 411–415.

[3] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, nos. 1–2, pp. 151–175, May 2010.

[4] G. Yang, H. Xia, M. Ding, and Z. Ding, "Bi-directional generation for unsupervised domain adaptation," in *Proc. AAAI*, 2020, pp. 6615–6622.

[5] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Transferring visual category models to new domains," in *Proc. ECCV*, 2010, pp. 1–5.

[6] Q. Wang, J. Gao, and X. Li, "Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4376–4386, Sep. 2019.

[7] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2006, pp. 22–23.

[8] A. Mesaros, T. Heittola, and T. Virtanen, "A multi-device dataset for urban acoustic scene classification," Tech. Rep., 2018.

[9] S. Gharib, K. Drossos, E. akir, D. Serdyuk, and T. Virtanen, "Unsupervised adversarial domain adaptation for acoustic scene classification," Tech. Rep., 2018.

[10] K. Drossos, P. Magron, and T. Virtanen, "Unsupervised adversarial domain adaptation based on the wasserstein distance for acoustic scene classification," Tech. Rep., 2019.

[11] V. Boddapati, A. Petef, J. Rasmusson, and L. Lundberg, "Classifying environmental sounds using image recognition networks," *Procedia Comput. Sci.*, vol. 112, pp. 2048–2056, Dec. 2017.

[12] Y. Tokozume, Y. Ushiku, and T. Harada, "Learning from between-class examples for deep sound recognition," Tech. Rep., 2017.

[13] N. He, J. Zhu, and L. Li, "[ieee 2018 international joint conference on neural networks (IJCNN) - rio de janeiro, Brazil (2018.7.8-2018.7.13)] 2018 international joint conference on neural networks (IJCNN)—An optic-fiber fence intrusion recognition system using the optimized curve," Tech. Rep., 2018.

[14] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, pp. 723–773, Mar. 2012.

[15] F. Liu, W. Xu, J. Lu, G. Zhang, and D. J. Sutherland, "Learning deep kernels for non-parametric two-sample tests," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 1–29.

[16] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7167–7176.

[17] Z. Pei, Z. Cao, M. Long, and J. Wang, "Multi-adversarial domain adaptation," 2018, *arXiv:1809.02176*. [Online]. Available: http://arxiv.org/abs/1809.02176

[18] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1640–1650.

[19] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *Comput. Sci.*, May 2015.

[20] J. Zhang, Z. Ding, W. Li, and P. Ogunbona, "Importance weighted adversarial nets for partial domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8156–8164.

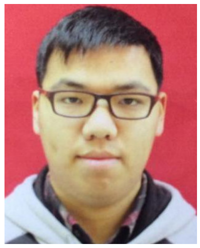
[21] N. He and J. Zhu, "Deep learning approach for audio signal classification and its application in fiber optic sensor security system," in *Proc. 9th Int. Conf. Inf. Sci. Technol. (ICIST)*, Aug. 2019, pp. 263–267.

[22] M. Long and J. Wang, "Learning transferable features with deep adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 97–105.

[23] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2208–2217.

[24] Z. Cao, M. Long, J. Wang, and M. I. Jordan, "Partial transfer learning with selective adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2724–2732.

[25] P. Tang, X. Wang, X. Bai, and W. Liu, "Multiple instance detection network with online instance classifier refinement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2843–2851.



NINGYU HE received the B.S. degree from the School of Electronic Engineering, Xidian University, China, in 2017. He is currently pursuing the Ph.D. degree with Shanghai Jiao Tong University, China. He is also with the Department of Electronic Engineering, Shanghai Jiao Tong University. His research interests include audio signal processing, image processing, and deep learning.

JIE ZHU (Member, IEEE) received the Ph.D. degree in communications and information systems from Shanghai Jiaotong University. In 1997, he went to Bell Labs, NJ, USA, for cooperative scientific research. In 2000, he was a Senior Visiting Scholar with the Dresden University of Technology, Germany, for visiting research. He is currently a Professor with the Department of Electronic Engineering and also a Doctoral Supervisor in electronic science and technology. And, he went to USA, Europe, Japan, South Korea, and other countries to participate in international conferences and academic exchanges for many times.