inter-noise 2021
1–5 August | Washington, DC

Abnormal drone noise detection system based on the microphone array and self-supervised learning

Hao Wu[1]
Tsinghua University
Haidian District,Beijing, China 100084

Huitian Jiang[2]
University of Electronic Science and Technology of China
2006 Xiyuan Avenue, West Hi-Tech Zone, Chengdu, China 611731

Haifeng Wen[3]
University of Electronic Science and Technology of China
2006 Xiyuan Avenue, West Hi-Tech Zone, Chengdu, China 611731

Chuang Shi[4]
University of Electronic Science and Technology of China
2006 Xiyuan Avenue, West Hi-Tech Zone, Chengdu, China 611731

ABSTRACT

Drone noise is primarily generated from its rotating blades, providing plentiful information on the condition of the drone. In the production line, the abnormal sound detection (ASD) system has shown advantages, such as non contact, ease of deployment, and capability to locate the faulty products at a relatively low cost. Therefore, this study aims to develop an abnormal drone noise detection system (ADNDS) based on the microphone array and self-supervised learning. The microphone array is a part of the data acquisition module that picks up the drone noise remotely. There are eight microphones in the array that is constantly pin pointing to the direction of the drone in air by beamforming. The acquired drone noise samples are extremely unbalanced, as abnormal samples are difficult to collect. Hence, a self-supervised learning strategy is adopted by creating auxiliary classification tasks to determine feature representations of the drone noise samples. With intentional consideration to reduce the complexity, the trained deep learning model can be successfully deployed on an embedded system with no graphics processing units (GPUs).

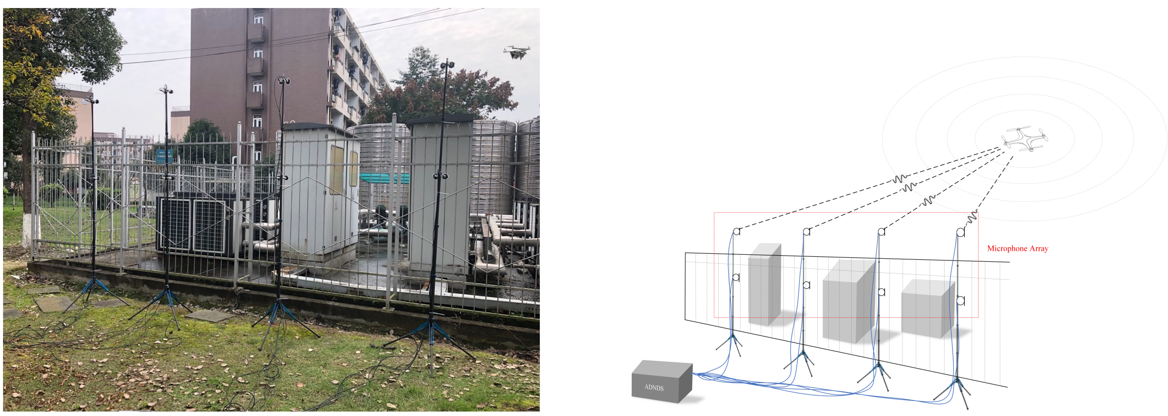
[1]hao-w19@mails.tsinghua.edu.cn
[2]huitianjiang@std.uestc.edu.cn
[3]wenhaifeng@std.uestc.edu.cn
[4]shichuang@uestc.edu.cn (corresponding author)

1. INTRODUCTION

The drone, being small, portable and exquisite, can be easily manipulated. It has been widely used in military, commercial and civilian fields, such as search and rescue, ground surveillance, collecting data, package delivery, and agricultural applications [1–4]. With the increasing number of drones, the incidences of drone crashes have shown a steady increase. When a drone is in the middle of a mission, a malfunction crash not only disrupts the mission. It can also cause great economic loss for the precision equipment attached to it or placed on the ground below it. Serious defects and incidences may also cause major accidents. Resultantly, drone usage safety issues have necessarily received extensive attention and research [5–7]. At present, the main reasons for the failure of drone functions are structural defects, software failure, sensor failure, bad weather, and improper manipulation. The current solution to prevent drone malfunction is to add sensing circuits and detection programs to the drone, to forewarn abnormal situations. However, these practices will undoubtedly lead to an increase in the weight and power consumption of the drone. This brings about new challenges to the design of drones, as well as creating new problems when manipulating them.

When the drone is in operation, the rotating blades will produce a loud sound, which provides plentiful information regarding the condition of the drone. Therefore, this paper proposes a non-contact method to detect the anomaly states of the drone by analyzing the noise it emits. Figure 1 shows the photo of the ADNDS in a prototyping experiment. The proposed ADNDS adopts a microphone array on the ground to capture the noise of the drone, avoiding any changes to the structure of the drone. Moreover, the analysis of the drone noise is carried out by an external processing unit that does not consume resources of the drone. The ADNDS is an application of the abnormal sound detection (ASD) technique, where the mainstream choices of models are based on deep learning [8]. As compared to conventional models, the deep learning model optimizes feature representations by a data-driven training process.

A sufficiently large number of both normal and abnormal samples of the drone noise are necessary in order to carry out the supervised training. However, the drone noise samples are distributed in a very unbalanced manner. It is often much easier to collect normal samples and extremely costly to obtain abnormal ones. Therefore, the ADNDS must adopt the self-supervised learning strategy by creating auxiliary classification tasks [9, 10]. External data sets are also included to assist the training process, since the other types of sound samples are naturally counted as abnormal samples of the drone noise. Lastly, with the consideration to reduce complexity, the trained model can be deployed on an embedded system with no GPUs.



(a) (b)

Figure 1: (a) Photo and (b) illustration of a prototype of the ADNDS.

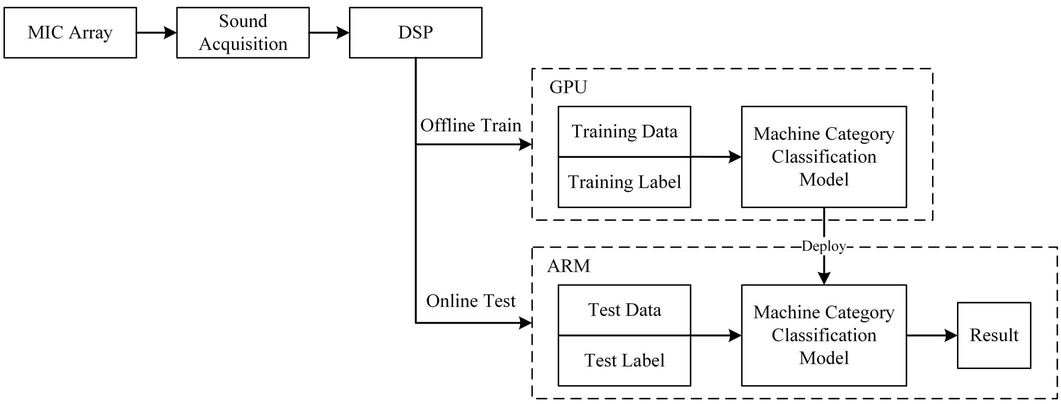


Figure 2: Block diagram of the ADNDS.

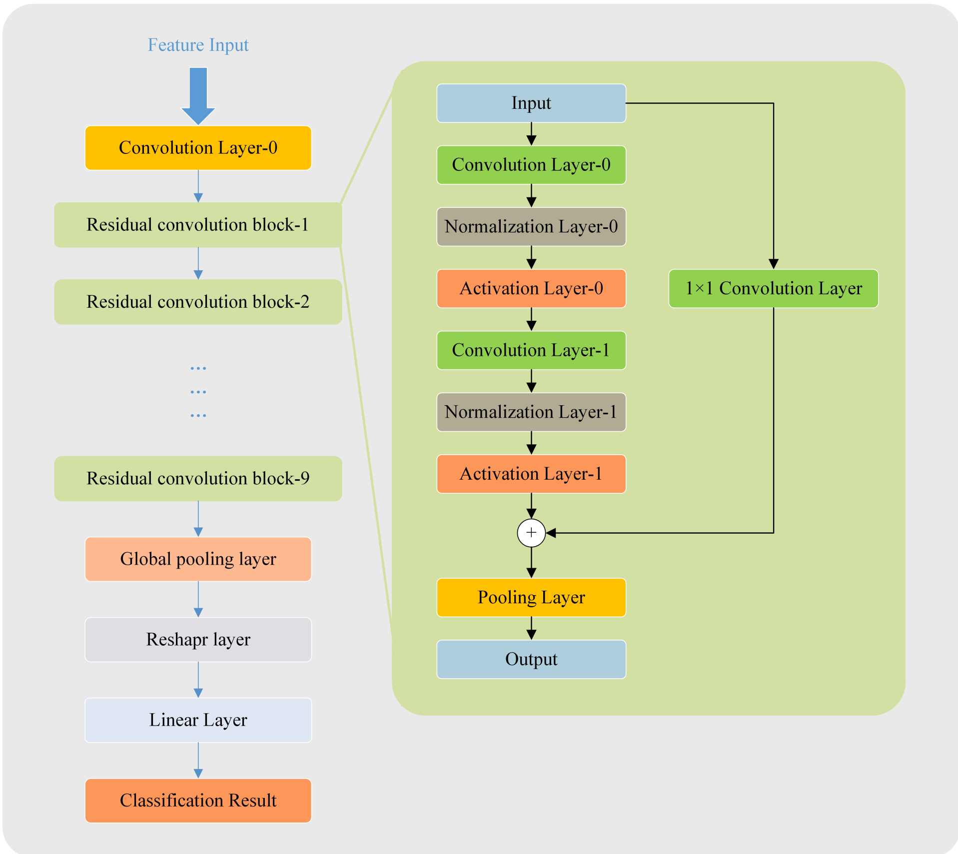


Figure 3: Deep learning model of the ADNDS.

2. ABNORMAL DRONE NOISE DETECTION SYSTEM

The block diagram of the ADNDS is shown in Figure 2. When the ADNDS is set up to detect the status of a drone flying over a designated area, its microphone array is placed outside the designated area and constantly follows the direction of the drone by beamforming [11]. The drone noise is acquired and analyzed in real time. The acquisition is controlled by a digital signal processor (DSP), where the beamforming algorithm and the short-time Fourier transform (STFT) are implemented. In the development phase of the ADNDS, drone noise samples stored in the memory of the DSP are uploaded to the cloud sever equipped with GPUs, in order for the deep learning model to be trained by the self-supervised learning strategy.

The deep learning model of the ADNDS is a residual neural network (ResNet), as shown in Figure 3. Previous studies have demonstrated that neural networks with residual connections are

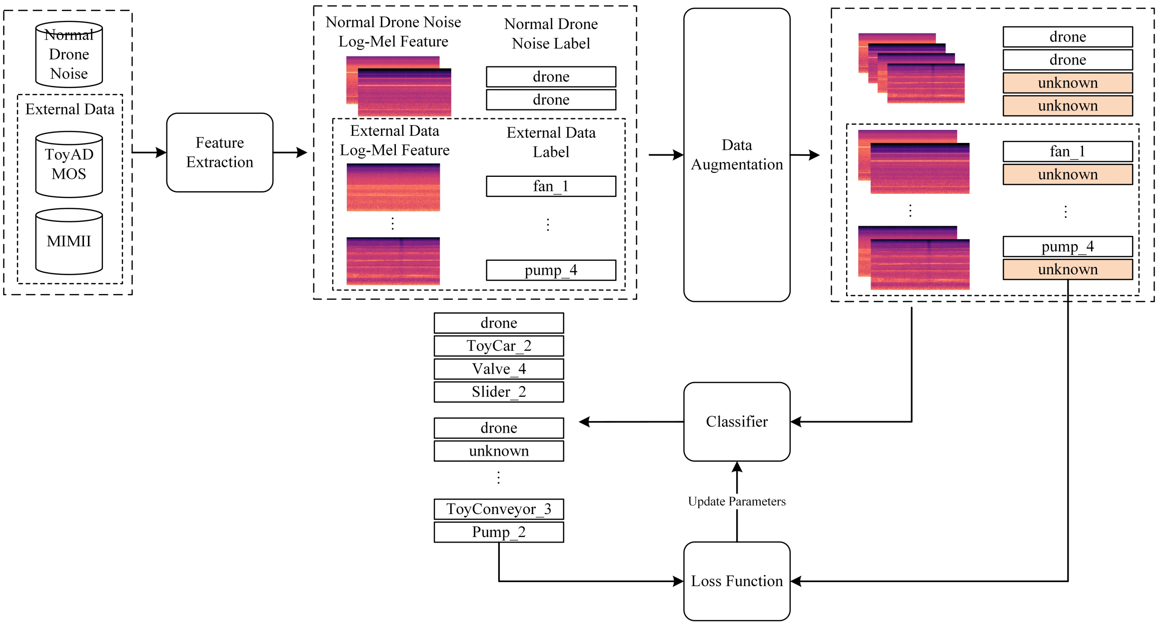


Figure 4: Self-supervised learning strategy for the ADNDS.

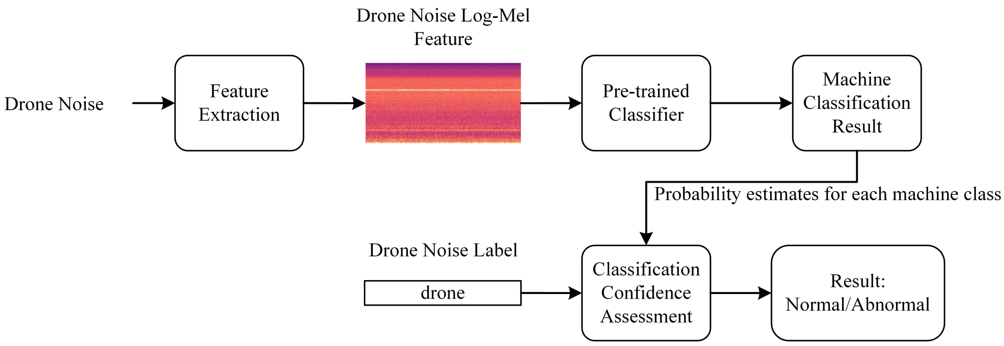


Figure 5: Lab test procedure of the ADNDS.

able to learn more efficient feature representations, making them easier to capture complex feature representations [12]. The residual connection also allows the gradient to be transmitted back to the previous convolutional layer during the back propagation process by jumping and connecting different convolutional layers, which effectively solves the gradient disappearance problem. More importantly, the ResNet requires only a small number of parameters, resulting in a low computational complexity. The structure of the convolution block is also shown in Figure 3. In each convolution block, there are two convolution layers, two normalization layers, two activation function layers and one pooling layer. Once the deep learning model of the ADNDS is trained, an embedded system is connected to the DSP to estimate the status of the drone with no involvement of GPUs. In this case, the ADNDS can be powered by a Lithium battery pack for over 8 hours.

2.1. Self-Supervised Learning Strategy

The drone noise samples are extremely unbalanced, because abnormal samples are difficult to collect. In the situations when only normal samples are available in the data set, conventional supervised learning strategies cannot function. Alternatively, the self-supervised learning strategy introduces auxiliary tasks to accurately model the distribution of the normal samples in the feature space specified by the auxiliary tasks [13]. The abnormal sample will exhibit a distinguishable difference in the same feature space. In the ADNDS, the machine category classification task is selected as an auxiliary task to achieve self-supervised learning, as shown in Figure 4.

Table 1: ResNet settings for the machine category classification.

Network layer	Convolution kernel	Pooling layer	Number of channels
Convolutional layer-0	[3, 7]	[1, 1]	16
Residual convolution block-1	[3, 3]	[4, 4]	16
Residual convolution block-2	[3, 3]	[2, 2]	32
Residual convolution block-3	[3, 3]	[2, 2]	64
Residual convolution block-4	[3, 3]	[2, 2]	64
Residual convolution block-5	[3, 3]	[2, 2]	64
Residual convolution block-6	[3, 3]	[1, 1]	64
Residual convolution block-7	[3, 3]	[1, 1]	128
Residual convolution block-8	[3, 3]	[1, 1]	128
Residual convolution block-9	[3, 3]	[1, 1]	128

Table 2: Accuracy of the abnormal drone noise detection.

Test sample time length	Test result (accuracy rate)
10s	95.7% (±2.9%)
60s	96.5% (±1.3%)

The ToyADMOS and MIMII data sets are adopted in addition to the on-site drone noise samples collected by the ADNDS [14, 15]. Six types of machine sound samples are selected from the ToyADMOS and MIMII data sets. They are the fan, valve, slide rail, pump, toy car and toy conveyor. The drone noise samples provide the seventh type, i.e. the drone. The sampling rate of all the samples is set to 48 kHz by re-sampling. The STFT spectrogram is calculated in the DSP with a 1024-point hamming window and the hop size of 512, and the log-mel spectrum feature is further extracted by a 128-order mel filter bank that covers the frequency range from 0 to 22.05 kHz. To alleviate the sparsity of the machine category distribution, another "unknown" type of machine sound samples is constructed by spectral distortion, which is realized on the log-mel spectrum by image distortion. Furthermore, 64 consecutive frames are randomly selected from the original log-mel spectrum to train a more robust model. The mixup method with a hyper-parameter of 1.0 is adopted for data augmentation. The training loss is the binary cross-entropy loss function. The ResNet settings for the machine category classification is presented in Table 1. The parameters of the convolution kernel and the number of channels is applied to all the convolutions in the residual block.

2.2. Lab Test Result

After the training of the deep learning model with the self-supervised learning strategy, the performance of the ADNDS is evaluated. A decision fusion method is used in the evaluation. When the time length of the input audio feature is greater than 64 frames, the decision fusion method is used to obtain the overall result. Specifically, the input audio feature is divided into a group of feature vectors. Each of the feature vector consists of 64 frames, and are sent to the deep learning model for evaluation. Finally, the evaluation results of all the feature vectors are averaged out to obtain the overall anomaly detection result. The aforementioned lab test procedure is shown in Figure 5.

Table 2 shows the lab test results. The deep learning model can effectively detect the abnormality in the drone noise. The accuracy of the abnormal drone noise detection result is higher than 95%. The trained model shows adaptability to the drone noise samples of different time lengths. When there is a longer drone noise sample, a higher accuracy is achieved. The same results can be obtained on the cloud sever and a low-cost embedded system such as a Raspberry Pi.

3. CONCLUSIONS

This paper presents a system-level solution to the safety issues of drones deployed over a designated area. The drone noise is used to identify abnormalities of the drone when it is in flight. The proposed ADNDS uses a microphone array to collect the noise emitted by the drone remotely. The beamforming algorithm is implemented on the DSP to track the drone in real time and calculate the spectrogram that serves as the input to the deep learning model. The self-supervised learning strategy enables the deep learning model to be trained with only normal drone noise samples. Lab test results demonstrate a high accuracy of detecting the abnormal drone noise samples. Last but not the least, the trained model can run on the embedded system with no GPUs. The cost and power consumption of the ADNDS is greatly reduced.

4. ACKNOWLEDGEMENTS

This manuscript is prepared based on the research work supported by the National Natural Science Foundation of China and the Civil Aviation Administration of China (Joint Grant No. U1933127).

REFERENCES

[1] C. Xu, X. Liao, H. Yue, X. Deng and X. Chen, “3-D path-searching for UAVs using geographical spatial information,” *Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, 2019.

[2] N. Kumar, M. Ghosh and C. Singhal, “UAV network for surveillance of inaccessible regions with zero blind spots,” *Proceedings of the 2020 IEEE Conference on Computer Communications Workshops*, Toronto, Canada, 2020.

[3] D. Wang, P. Hu, J. Du, P. Zhou, T. Deng and M. Hu, “Routing and scheduling for hybrid truck-drone collaborative parcel delivery with independent and truck-carried drones,” *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10483-10495, 2019.

[4] Y. Guo et al., “A drone-based sensing system to support satellite image analysis for rice farm mapping,” *Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, 2019.

[5] Z. Liguo, W. Lu, J. Shude, Z. Wangwang and W. Feng, “UAV flight reappearance and failure analysis,” *Proceedings of the 2018 Chinese Automation Congress*, Xian, China, 2018.

[6] B. Wang, P. H. Ng, B. M. N. A. K. Elhadidi, H. S. Ang and S. K. Moon, “Failure analysis and finite element simulation for structural systems in an unmanned aerial vehicle,” *Proceedings of the 16th International Conference on Ubiquitous Robots*, Jeju, Korea, 2019.

[7] T. Okutake, N. Uchida and N. Yamamoto, “A collaborative safety flight control system for multiple drones,” *Proceedings of the 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, Fukuoka, Japan, 2016.

[8] H. Wen, C. Shi, D. Yao, S. Zhang, and J. Nie, “Automatic anomalous sound detection systems for unaligned multi-device machine sound samples,” *Proceedings of the 27th International Congress on Sound and Vibration*, Prague, Czech, 2021.

[9] Y. Koizumi, S. Saito, H. Uematsu, Y. Kawachi and N. Harada, “Unsupervised detection of anomalous sound based on deep learning and the Neyman–Pearson lemma,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 1, pp. 212-224, 2019.

[10] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo and Y. Kawaguchi, "Anomalous sound detection based on interpolation deep neural network," *Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, Barcelona, Spain, 2020.

[11] C. Shi, W. S. Gan, Y. K. Chong, A. Apoorv, and K. S. Song, "A vehicular noise surveillance system integrated with vehicle type classification," *Proceedings of the 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Kaohsiung, China, 2013.

[12] T. Yao, C. Shi, and H. Li, "A CRNN network with multi-time-scale pooling and densely connected GRU layers for sound event detection," *Proceedings of the 27th International Congress on Sound and Vibration*, Prague, Czech, 2021.

[13] D. Yao, S. Zhang, T. Yao, C. Shi, and J. Nie, "Anomalous electrical transformer sound detection using self-supervised learning with auxiliary classification tasks and data augmentations," *Proceedings of the 27th International Congress on Sound and Vibration*, Prague, Czech, 2021.

[14] Y. Koizumi, S. Saito, H. Uematsu, N. Harada and K. Imoto, "ToyADMOS: A dataset of miniature-machine operating sounds for anomalous sound detection," *Proceedings of the 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, United States, 2019.

[15] P. Harsh, T. Ryo, I. Takeshi, E. Takashi, N. Yuki, S. Kaori and K. Yohei, "MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection," *Proceedings of the 2019 Detection and Classification of Acoustic Scenes and Events Workshop*, New York, United States, 2019.