Convolutional Neural Networks for Analyzing Unmanned Aerial Vehicles Sound

HyunJong Kim[2], Daeun Kim[1], SungWook Park[1], Sukhoon Lee[1], Shulin Li[3], Kyle Moffitt[3], Eric T. Matson[3], John C. Gallagher[4]

[1]Department of Computer Science and Engineering, Chungnam National University

[2]Department of Mechatronics Engineering ,Chungnam National University

[3]Department of Computer information and Technology, Purdue University

[4]Department of Computer Science and Engineering, Wright State University

Abstract—**Convolutional Neural Network is one of several models of deep learning. It is applied in various fields such as image recognition and natural language processing. In this study, we design a system to detect the presence of dangerous substances on the basis of sound generated by drones using convolution neural network. In the paper, the sound of recorded drones is pre-processed to spectral data by Fast Fourier Transform (FFT) and Mel-Frequency Cepstrum (MFCC) and given as the input value to the CNN model.**

I. Introduction

With the development of technology, the application of drones has become more active, like delivering stuffs between areas. According to the previous research [1], it is possible to realize a courier service with the drones, increasing the convenience is also possible to expect. However, the spread and development of these drones also increased the security risk. Terrorists, criminals and malicious users possibly can attack by manipulating drones carrying dangerous contraband such as explosives. Therefore, it is necessary to develop approaches to protect people from these threats.

According to those previous studies [2][3][4], the research, determine the danger by the drone sound, has accomplished. Thoroughly, the sound of a drone is converted to a spectrum through Fast Fourier Transform (FFT) and is given to the Conventional Neutron Network (CNN) model as an input value to judge whether it is threatened [5]. However, when the sound of the drones is transformed into a spectrum and is given as an input value, the input range is usually small which leads to the results which are greatly affected by the noise. In order to decrease the noise influence maximally, this study sampled the drones sound data by a larger size of about 5 to 10 seconds compared to those previous researches, so that the result were not affected by the temporary noise.

The 5 to 10 seconds data sample fragments were numerous for the machine running and the CNN learning which could take a long time. A Mel-Frequency Cepstrum (MFCC) was used to avoid consuming a relatively large amount of time. Since the feature of the data can be compacted by the MFCC procedure, at the end, the size of the data can be reduced to 1/90. Therefore, the training time can also be reduced accordingly.

The result data columns based on CNN models learning was classified into "loaded drones", "unloaded drones" and "noise". The false negatives were defined as "Loaded but not loaded" or "Loaded but noise". Weighting is used to minimize such false negatives due to the false negatives cause severe problems in reality.

II. Study Methods

A. Data Collection

The DJI Phantom 2 was used for collecting the sound datas. Its specification shown as below.

Weight : 1000g

Maximum Take-off Weight : 1300g

Maximum Flight Speed : 15m/s

The sound data of a Phantom 2 movement was collected at an airfield. The original collection included various conditions of Phantom 2, such as not loaded, loaded, and the background noises. The collection did not involve any other drones sound.

B. Data Preprocessing

Four preprocessing methods were performed on the collected data to apply the large number of data to the model efficiently.

The first procedure was normalization. Operating the normalized input data had the advantage of finding an optimization point quickly for the gradient descent method. Furthermore, normalization allows performing adequate learning instantaneously by eliminating the small learning rate set disadvantage.

The second technique was the pre-emphasis filter. During the second half of the Fourier transform, the frequency response at the low-frequency component was much larger than the frequency response at the high-frequency component, so that a pre-emphasis filter was used to produce a substantially relatively constant frequency response in all frequency bands.

The third approach was Short Time Fourier Transform (STFT). When STFT applied, the window type was Hamming, the length of the window was 863, which corresponded to 0.0116 seconds at 44.1kHz sampling rate. It converted the voice data into a two-dimensional spectrogram. Scilicet, the STFT transformed the input data format to fit the CNN model and enabled sound analysis, which was impossible in the time domain, while adding the frequency domain.

The fourth is MFCC. The data obtained from STFT was numerous due to both the necessary sound data and the counterproductive noise data existed in the 1024 frequency sample. MFCC was used to extract the important feature as well as reduce the data quantity to about 1/90.

C. Convolutional Neural Network

The model used CNN to classify the drone sound into three categories: loaded drones, unloaded drones and noise.

The CNN model consist of the following layers:

TABLE I. CNN Layers

Layer Name	**Size**
1_Convolutional	window size: 2x4x1 number of filter : 32
1_Relu	activation function
1_MaxPool	kernel size : [1,2,4,1] stride : [1,2,4,1]
2_Convolutional	window size: 2x9x32 number of filter : 64
2_Relu	activation function
2_MaxPool	kernel size : [1,1,2,1] stride : [1,1,2,1]
3_Fully Connected Neural Network	input size: 24192 output size : 100
3_Relu	activation function
3_DropOut	keeping rate : 70%
4_Fully Connected Neural Network	input size : 100 output size : 3

In reality, the cases, detecting being loaded but not being loaded or being classified as noise, were called the false negative. False negatives should be minimized because they would be more likely to cause risks and damages once occur. Therefore, during the CNN model learning process, the weight of the false negative in the result was given a large number, so that the cost was adjusted effectively, therefore, the probability of the false negative was reduced.

III. Experiment Result

The test was contributed by 56,000 data samples which was processed in the exact same manner as the learning data. The high accuracy of 99.5% was obtained and 0.005% of the result were false negatives.

TABLE II. Result Table

Reality Test Result	**Noise**	**Unloaded Drone**	**Loaded Drone**
Noise	46766	60	174
Unloaded Drone	0	5000	0
Loaded Drone	3	0	3997

IV. Conclusion

Based on the experiment, it was conspicuous that the loading condition of drones can be distinguished through data preprocessing with a functional CNN model. Furthermore, the derivation of false negatives, which could be a problem remaining in the past studies, can be solved by weighting adjustment.

Finally and hopefully, this research will be developed on a mobile phone or a portable device so that the drone voice information can be easily recorded. The mobile application should also include functions judging whether or not the drone is loaded in real time.

V. Acknowledgement

This research was supported by the MIST(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW(2015-0-00930) supervised by the IITP(Institute for Information & communications Technology Promotion)

References

[1] Seung Jin Shin, 2016, “Possibility of logistics delivery through drone (drone) and overseas case," Monthly KOTI Magazine on Transport, pp. 45~53.

[2] Dong-Hyun Lee, 2017, "Convolutional Neural Network-based Real-Time Drone Detection Algorithm," The Journal of Korea Robotics Society, Vol. 12, No. 4, pp. 425~431.
[3] Kee-Woong Lee, Kyoung-Min Song, Jung-Hwan Song, Chul-Ho Jung, Woo-kyung Lee, Myeong-Jin Lee and Yong-Kyu Song, 2017, "Implementation of Radar Drone Detection Based on ISAR Technique," The Journal of Korean Institute of Electromagnetic Engineering and Science, Vol. 28, No. 2, pp. 159~162.
[4] Lin Shi, Yujing He and KyungHi Chang, 2016, "Acoustic-based Classification for Drone Identification using Hidden Markov Model with MFCC Technique," Proceedings of Symposium of the Korean Institute of communications and Information Sciences, , pp. 204~205.
[5] SangGi Hong, GaYoung Kim, HeonGyeom Kim, SangHee Lee, DongHyun Lim, Austin Snair, Lucy Gotwals and Jone C. Gallagher, 2017, "Practically Classifying Unmanned Aerial Vehicles Sound using Convolutional Neural Networks," Proceedings of the Korean Information Science Society, pp. 1856~1858.