Practically Classifying Unmanned Aerial Vehicles Sound using Convolutional Neural Networks

DongHyun Lim, HeonGyeom Kim, SangGi Hong, SangHee Lee
Department of Computer Science and Engineering
Chungnam National University
Email: {chiyaky, heongyeom.kim, sanggi.hong11, sanghee.lee1992}@gmail.com

GaYoung Kim
Department of Naval Architecture and Ocean Engineering
Chungnam National University
Email: gykim8798@gmail.com

Austin Snair, Lucy Gotwals
Department of Computer Information and Technology
Purdue University
Email: asnair, lgotwals@purdue.edu

John C. Gallagher
Department of Computer Science and Engineering
Wright State University
Email: john.gallagher@wright.edu

***Abstract*—This work aims to analyze the effectiveness of simple neural networks which determine whether Unmanned Aerial Vehicles (UAVs) are harmless or not depending on their payload via sound events. The goal of this work is to contribute to a real-time UAV detection system which requires acceptable accuracy, minimal cost, and high speed, while keeping the system minimally complex for general usage in law enforcement applications. Specifically, our work is based on classifying the sound events of UAVs and their spectrum data generated by Fast Fourier Transform (FFT) [1]using a single layered Convolutional Neural Networks (CNNs) and a simple voting system to determine accurate results. We recorded the sound produced by Phantom 2 which contains a payload or not and general background noise. And we built a machine learning model to classify different types of recordings. In conclusion, our neural networks showed an accuracy of 99.92% in short processing time.**

Keywords—Audio Categorization, Audio Classification, UAV Categorization, Machine Learning, Convolutional Neural Network, Payload Detection

I. INTRODUCTION

Motivation. New technological advances in UAV systems have developed strong and fast aerial objects with minimal cost. However, due to these improvements, UAVs are increasingly used by criminals and terrorists to attack targets. For example, in October 2016, IS terrorists used UAV to carry explosives and killed two Kurdish soldiers. It could cause loss of life on a large scale if employed. This includes highly urbanized areas, where striking large numbers of people is easy. In conclusion, detecting and categorizing UAVs has become and will continue to become extremely important as terrorists discover and invent new technology.

Existing works. The detection and analysis of UAVs' sound [2][3]and defining its harmful activity has already been studied in depth. Previous works approached this problem using deep neural networks and requiring expensive infrastructure.[4][5]. However, common targets of terrorists equipped with UAVs usually center around crowded places such as big cities, where one could do damage to many people without difficulty. In other words, UAV detection systems will mostly be employed in urban areas by local police departments. In short, a detection system and its sensors should have a low price and high performance in a practical purpose. Thus we aimed to build a simple neural network with inexpensive sensors, whereas other studies used high performance machine and deep neural networks which would be hard to employ practically.

We also took the Doppler effect of UAVs [6]into consideration. We figured out that most commercial UAVs move fast enough to cause the Doppler shift in the frequency of the sound they produce. Our solution was to add a convolution layer to detect the same shape of audio sound that shifts over. Thus, we implemented one convolution layer [7]into our model and compared two models in which one is a single fully connected neural network and the other has an extra convolution layer. As a result, accuracy of a single fully connected neural network showed 91.972% accuracy and the CNN model showed 99.92% accuracy.

Contributions. The following statements summarize our contributions:

- Develop a simple, practical and efficient neural network that does not require a high performance machine.
- Show the effectiveness of using CNN to counter the Doppler effect.

II. METHOD

A. UAV

We considered that 1 kilogram is a minimum weight of baggage used for terrorism. So we chose the Phantom 2 from DJI corporation, as the UAV should be able to carry on above 1 kilogram.

B. Data Collection

We collected the data by using single condenser microphones equipped with FET amplifiers installed into tin cases. The cost of the microphone is about $10.00 U.S. each. Each microphones was mounted at the approximate focal point of plastic bowls installed on the ground. Each of the assembly was connected to a laptop computer built-in consumer grade stereo inputs (44.1kHz sampling). So, the laptop collect the audio samples. These sound samples were taken in varying weather and times of year in Western Ohio.

Sound events collected with microphones were processed by Fast Fourier Transform to make spectrogram data. The window size of FFT is 0.18 seconds. Each of dataset has normalized power at frequencies from 100 Hz to 9000 Hz. The bandwidth of each frequency bin is 100 Hz. So each data set has 90 columns, and each column has all the normalized audio power from -50 Hz through +50 Hz. All powers in each frequency bin sum up to 1. The number of data which we collected is 100,000. The dataset is labeled as three classes which are loaded Phantom 2, unloaded Phantom 2, and noise.

C. Convolutional Neural Network

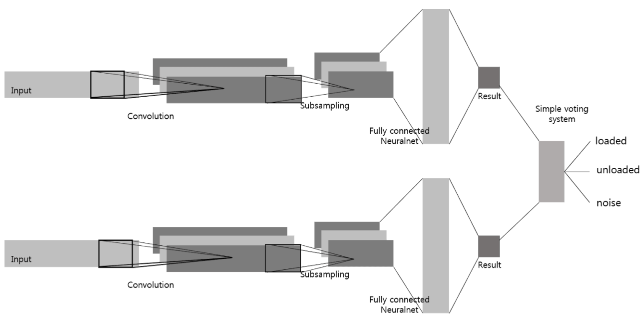

Fig. 1: Whole structure of the classifier

Although the final goal is to build a 3-way classifier, we did not build one directly. We built two binary classifiers and merged them into one. As can be seen in Figure 1, our model consists of two separate CNNs that are identical, and a voting system which puts together the output of the two CNNs. One CNN predicts how likely it is that we are listening to a loaded Phantom 2, and the other CNN predicts how likely it is that we are listening to an unloaded Phantom 2.

We trained the two CNNs separately with 50,000 data, respectively. The voting system makes a final prediction as to whether we are listening to a Phantom 2 with load, a Phantom 2 without load, or environmental noise, and the unloaded dataset.

This model design makes the system economical, as the only necessary work in attaching a new UAV is training the model, then attaching to what is already implemented. That work costs less than training a complex neural network for multi-classification.

Table I: Model description about each layer

Convolutional Layer	Filter size: 10 Step Size: 1
Activation Function	ReLU
Subsampling Layer	Window Size: 5 Step Size: 2
Fully-connected Layer	Sigmoid function

The purpose of using the CNN, rather than other types of neural networks, is to counter the Doppler effect. Our target model, Phantom 2, moves at the speed of 15 m/s^2 at its fastest and 12 m/s on average. Under the assumption that the air doesn't move relative to the microphone and the Phantom 2 moves toward the microphone at the speed of 12 m/s, the Doppler effect expects that the observed frequency be 1.03655 times the emitted frequency. The Phantom 2 emits a sound higher-pitched than 3000 Hz, which is interpreted into more than 100 Hz of shift to its observed frequency.

This amount of change is enough to confuse the neural networks trained with frequency based learning. Filters of CNNs are trained to recognize patterns which could appear anywhere in the frequency range, to counter the Doppler shift.

D. Voting System

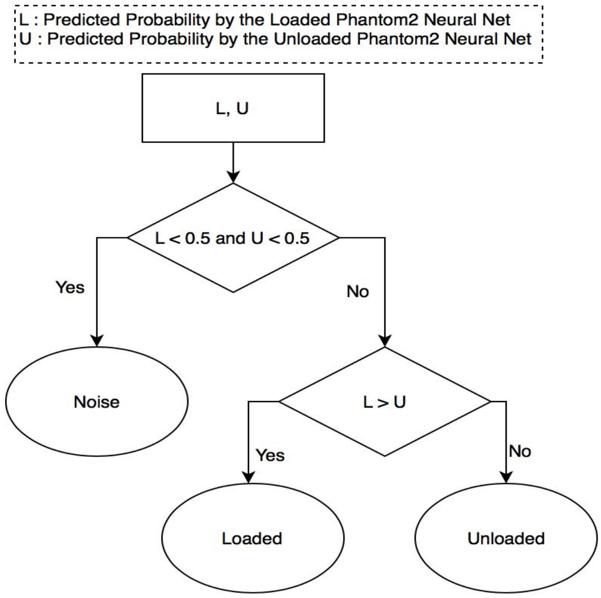

Fig. 2: Voting System in Testing

Given the test dataset, our voting system uses the output of two neural nets to predict the result. One is the loaded Phantom 2 neural net trained with loaded data. The other is the unloaded Phantom 2 neural net trained with unloaded data. It therefore classifies the input as loaded Phantom 2, unloaded Phantom 2, or noise. If both the predicted probability by loaded Phantom 2 neural net and unloaded Phantom 2 neural net are below 50%, the prediction is noise. If the predicted probability by loaded Phantom 2 neural net is greater than the predicted probability by unloaded Phantom 2 neural net, the prediction is loaded Phantom 2, otherwise, it is the unloaded Phantom 2.

III. Experiment

A. Test Setup

We gave 50,000 dataset to our model for a test. The dataset was recorded at the same environment that training dataset made but never used for training.

B. Result

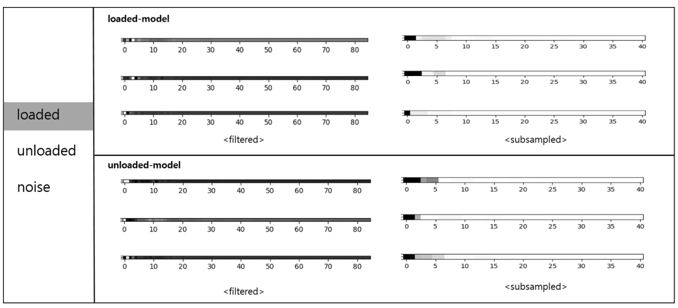

Fig. 5: In each model, activated parts (black area) by filter and subsampling of loaded data.

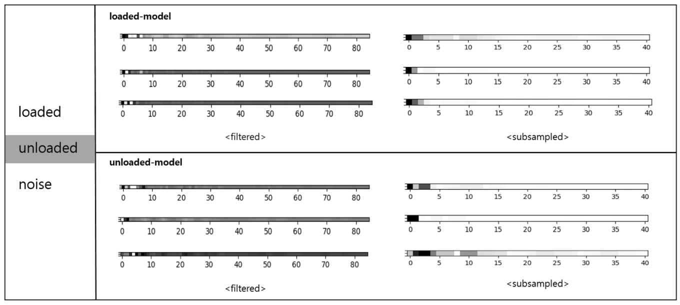

Fig. 6: In each model, activated parts (black area) by filter and subsampling of unloaded data.

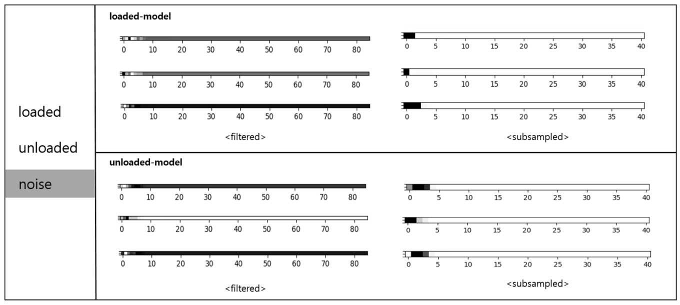

Fig. 7: In each model, activated parts (black area) by filter and subsampling of noise data.

In these figures, we can find which parts are main frequencies used for figuring out what the sounds of the drone are. Three sets of data at the left side in upper right window are data after convolutional layer of loaded Phantom 2 neural net, and three sets of data at the right side of same window is data after subsampling (max-pooling) layer of loaded Phantom 2 neural net. and data in the lower right window are the same as the unloaded Phantom 2 neural net.

In subsampled data, most activated parts (portrayed in the white area above) have disappeared, due to the ReLU layer which is between convolutional layer and subsampling layer. Activated parts are only left on the range of 200 Hz ~ 2200 Hz, As this result, we can consider to make narrow frequency range of data used for classifying Phantom 2 with baggage and Phantom 2 without baggage.

TABLE II: The accuracy for each used bandwidth

Bandwidth	150~2650	2150~4650	4150~6650	6150-8650
Accuracy	99.518	87.114	60.858	53.97

With this result, we tested model with various bandwidth portion of frequency. The frequency bandwidth used for each test is 2500 Hz. As expected, we can see that the accuracy is high in the 150 to 2750 Hz range. This can be seen as an important feature in distinguishing loaded Phantom 2, unloaded Phantom 2, and noise, and the accuracy is lowered as the frequency band shifts to larger Hz.

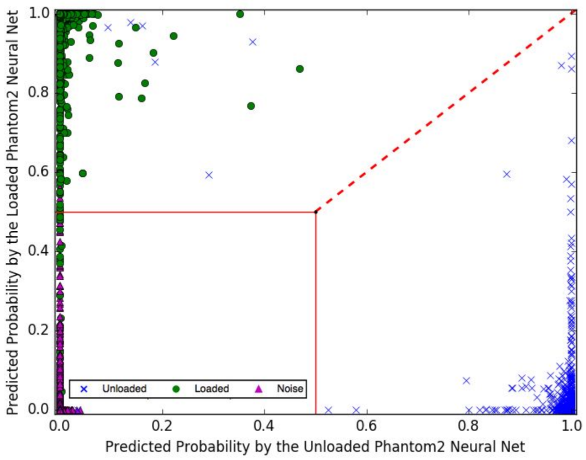

Fig. 8: Actual/Prediction test data distribution

(Figure 8) The figure shows the test data distribution by using our voting system. X-axis shows the predicted probability by unloaded Phantom 2 neural net. Y-axis shows the predicted probability by loaded Phantom 2 neural net. Actual label is shown as colors and shapes. 'X' shape points represent actual unloaded data. Triangle points represent actual loaded data. And 'O' shape points represent noise data. The red boundary lines are drawn to show the boundaries of each prediction categories.

TABLE III: Actual/Prediction

Predict / Actual	Loaded	Unloaded	Noise
Loaded	15363	0	21
Unloaded	12	15371	1
Nose	0	5	19227

In TABLE II, each row represented actual labeled data. Each column represented prediction by our model. The total number of test data is 50,000. In actual loaded labeled data, 21 data are predicted as noise. So, it can detect loaded Phantom 2 about 99.86%. In actual unloaded labeled data, 12 data are predicted as loaded. So, it misclassify unloaded Phantom 2 as loaded Phantom 2 about 0.07%. So, it misclassify noise as unloaded Phantom 2 about 0.02%. Total accuracy is 99.92%.

IV. Conclusion

This paper presents that simple CNN model can be used for classification of harmful UAV. The model shown that classification accuracy is above 99.92 with data collected by common-level microphone.

The concern of work is that the model can be vulnerable by the noise makes pattern of UAV faint because the model doesn't have system to prevent noise.

In the future work we will implement noise canceling system on the model to make it reliable and attach the classifier of UAV model to make model determine the model of UAV now hearing and whether UAV is harmful or not automatically.

Acknowledgment

This research was supported by the Ministry of Science, ICT & Future Planning (MISP), Korea, under the National Program for Excellence in SW)(과제번호) supervised by the Institute for Information & communications Technology Promotion (IITP)

References

[1] Heckbert, Paul "Fourier Transform and the Fast Fourier Transform(FFT) Algorithms," Computer Graphics 2, 15-456 (Feb. 1995), revised Jan. 27, 1998.

[2] S. Jeon, J. Shin, Y. Lee, W. Kim, Y. Kwon and H. Yang, "Empirical Study of Drone Sound Detection in Environment with Deep Neural Networks" in

[3] B. Kim, H. Kang and S. Park, "Drone Classification Using Convolutional Neural Networks With Merged Doppler Images" in IEEE Geoscience and Remote Sensing Letters, Vol. 14. No. 1, January 2017

[4] Lin Shi, Yujing He, and KyungHi Chang, "Acoustic-based Classification for Drone Identification using Hidden Markov Model with MFCC with MFCC Technique" In The Korean Institute of Communications and Information Science, June 2016

[5] F. Fioranelli, M. Ritchie, H. Griffiths, H.Borrion, "Classification of loaded/unloaded micro-drones using multistatic radar" In Electronics Letters (Volume: 51, Issue: 22, 10 22 2015)

[6] P. Molchanov, K. Egiazarian, J. Astola, R. I. A. Harmanny, J. J. M. de Wit, "Classification of small UAVs and birds by micro-Doppler signatures" In Processing of the 10th European Rader Conference, 2013

[7] Y. Lecun, L.Bottou, Y. Bengjo, P. Haffner, "Gradient-based learning applied to document recognition" In Proceedings of the IEEE (Volume: 86, Issue: 11, Nov 1998)