Analysis of Commercial Drone Sounds and Its Identification

Sinwoo Yoo and Hyukjun Oh

September 10, 2020

Analysis of commercial drone sounds and its identification

ABSTRACT

The usage of quadcopter types of drones is now on a mature and practical stage. Numbered companies are manufacturing them and expanding their application. Considerable characteristics of this type of flying object that has the maneuverability and practicality is now being focused on how we control this among our urban life from its offensive liabilities. They are small enough to avoid many current airborne detection systems and cheap enough to use them as disposable. In this paper, we tried to analyze the sounds of a subset of commercial drones as quadcopter types and also built a trained simple non-linear neural network filter to classify them among the given sound samples. We borrowed Mel-frequency cepstral coefficients as the well-known methodology of sound analysis but including some of the parameter adjustments, and applied LeNet neural network filter structure for its simple classification. In order to maintain the information of adjacent samples among the series of wave samples, 2-D spectrogram planning was applied as the input signal of the filter. Most of the frequencies from drones were observed as gathered around 3 to 5Khz, up to 10Khz, and adjusted LeNet architecture could classify over 10 types of drone categories with over 95% of accuracy.

KEYWORDS

Drone sound, MFCC, LeNet, Convolutional Neural Network, Neural Network Capacity, quadcopter

1 INTRODUCTION

A quadcopter drone propelled by 4 rotors is easy to build with reasonable prices, wide use of cheap flight controllers, and highly efficient BLDC motor - Lithium batteries [1] by recent market. And due to the unique flight pattern calls the most popular in many business verticals such as filming, [2] entertainment, [2] or the new trend of the logistics market. [3] Because of the distributed in 4 ways to obtain required lift the over1all size of its propeller is relatively small compares to the other similar previous types of drones such as a traditional helicopter, this quadcopter drone makes unique sounds. It is the reason that usually this type of quadcopter is being indicated as drone because it sounds similar to the drone in nature.

Based on the specifications mentioned above, it is very maneuverable [4] as it sized so it is quite tricky to detect and hard to hold back from the asset we want to keep from [5-6].

Many companies are investing their assets into research to secure a practical solution to make it possible with regarding to the mentioned issues. Some of them utilize electromagnetic radio waves to neutralize them in the air as detected, or some of them use a physical methodology to hold it down [7-8]. But most of these methods or related solutions are not feasible without preliminary awareness of approaching drones to the countermeasures. Securing practical and effective solutions of neutralization is very hard to be flawless even in these days, and it gets worse as the current trends of rapid improvements of the key technologies of parts that consist of the drone itself. One of the promising methodologies of detecting it is the use of its unique sound itself because it is not possible to eliminate the fundamental mechanism that makes it possible the drone can fly. Using the sound the drone produces while it maneuvers has a limitation that it's only useful when the drone flies into the range of the sensitivity of the sensor, but still, it's very promising because it doesn't have to pinpoint the location of the drone approaching and also it's possible to do that with multiple sensors and the certain algorithm.[9]

In this manner, we tried to perform a brief analysis of it from some popular commercial drones in the market and tested whether we can apply this information into the one of simplest but adjustable LeNet neural network filters.

2 FUNDAMENTAL FEATURES OF THE DRONE SOUND SIGNALS AND ITS COMPONENTS

2.1 Signal Characteristics

The unique sound patterns from quadcopter drones generated by some key factors that are rooted in the specific geometrics of it.

One of the distinctive geometrical characteristics of a quadcopter drone is actually not directional itself rather be indicated by only its flight controller. This is one of the main reasons that quadcopter drone is able to have its considerable maneuverability. Also, the cause of this the sound characteristics from most of the quadcopter drones consists very specific several harmonics that are from each BLDC motor and propeller.

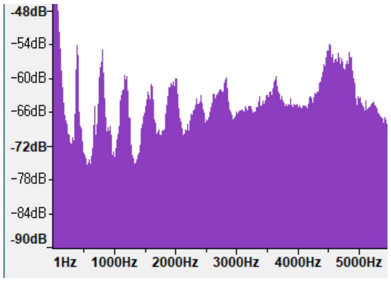

Figure 1: Frequency distribution from one of the collected sound sample with noise – Bebop 2, Parrot

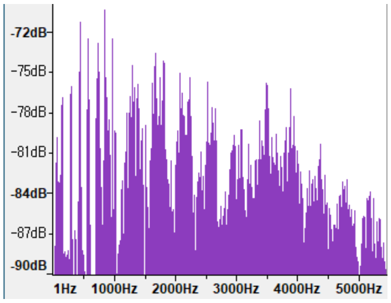

Figure 2: Frequency distribution from one of the collected sound sample – Inspire 2, DJI

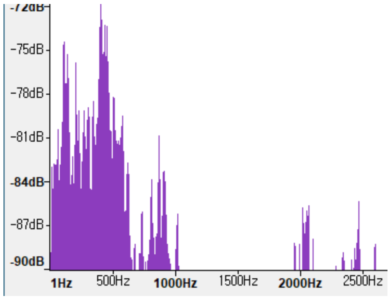

Figure 3: Frequency distribution from one of the collected sound sample – Human voice, mixed words speaking in Korean and English

Each of propellers and BLDC motors has their unique harmonics band from the vibration depending on which RPM runs on while the flight controller gives a continuous signal in order to sustain with the required valance. But in situations without dramatic changes in its altitude or hassle acceleration, it shows a very certain signal power distribution which seems one of the unique characteristics that is considered for further discussion as comes up later. On the contrary, based on the biological structure of a human vocal cord, figure 3 shows clearly that human voices from words and sentences are being grouped by much gathered in narrower bands. While quadcopter drone flying the rotating propellers hitting the air surrounding produces many other harmonics that convoluted with other adjacent frequencies besides the major ones that come directly from the moving parts themselves. Figure 2 shows a good example of the sound characteristics from most of the quadcopter drones consist of very specific several harmonics that are from each BLDC motor and propeller.

Table 1: ACOUSTIC NOISE MEASURED FOR BLDC DRIVE [10]

Speed (RPM)	Background noise level (dB)	Motor noise level (dB)	Difference (dB)
100	48	72. 29	24. 29
300	48. 3	75	26. 7
500	49. 48	73. 54	24. 06
800	47. 56	77	29. 44
1000	50	82	32
1500	48. 5	86. 52	38. 02
2000	49. 23	89. 2	39. 97
2500	52. 45	91. 26	38. 81

2.2 Preprocessing

The raw sound wave signal propagated over the air is a form of a series of samples located on the time domain, and it’s hard to analyze in this dimension, usually, the proper transformation process is followed as preprocessing for the better feature extraction. The sound spectrogram is one effective way to obtain information including a power spectral density distribution among sampled frequencies.

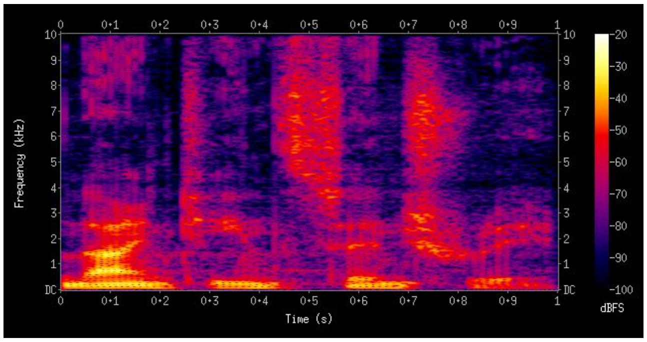

Figure 4: Spectrogram of the spoken words "nineteenth century". Frequencies are shown increasing up the vertical axis, and time on the horizontal axis, figure referred from Wikipedia.org

Calculating Mel-frequency kepstrum coefficients is one of the improved methodologies from a simple spectrogram derived from the short-time Fourier transform, which has been widely utilized in especially speech recognition. Generating Mel-frequency kepstrum coefficients follows certain computational steps –

slicing, FFT, windowing, filter-bank, log-amplitude, Mel-scaling and smoothing, and DCT. [11]

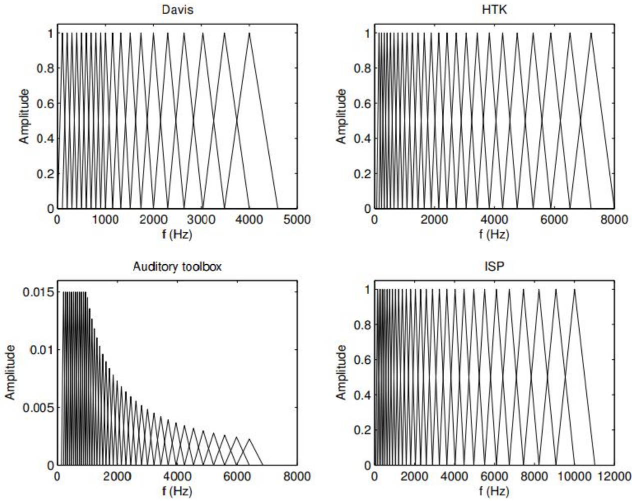

Figure 5: Shows different implementations of the Mel filter bank. Note the different scaling of the frequency axes in the plots. [12]

Filtering with different filter-banks apply the obtained power spectrum from the short-time FFT into different targets or purpose of the goal. Every detail to calculate these filter-banks is available in many other papers [11][12] so will not be covered in this paper. With regarding the perspective of the given sound signal as the purpose, overall frequency band from the sounds of quadcopter drones has unique harmonics but is distributed up to around 10Khz and their bandgaps are not wide as we peeked already, so the algorithm that has been used in this paper referred to HTK filter-bank but only the number of banks are selected as 40 rather than the one originally suggested 24 [13] and trimmed as a fewer bin length later. While the major application that the method of preprocessing in this paper applied for is focused on specifically the human vocal sound or speech itself, the difference by the sound signals from quadcopter drones should apply needful adjustment as parameters in the preprocessing sequence.

2.3 LeNet Neural Network

One of considerable factor of applying any of the information into any of modern neural network filters implies there is a specific part that the signal contains the distinctive or essential features in. There are many suggested or practically proven methodologies already in the field of speech recognition and it is already widely being used for many services effectively yet but due to many specific differences among other types of sound signals including the one as quadcopter drones', there are not many of them like the one mentioned already. There have been previous in-depth studies within the past several years and based on references those suggested models with proportional simulations and tests practically have not suggested with an efficient and robust model yet. [14][15][16]

LeNet neural network model was announced in 1989 with its first original form and shows its performance with the meaningful efficient network model as introducing a convolutional neural network for the first. [17] A two-dimensional plane that the preprocessed information is prepared gets into the input of the suggested convolutional computation sequence and one of its promising foundation that makes LeNet neural network filter working is that it maintains the spatial information by given 2nd order of Euclidian space so the information that is included among the input data does put weights gradually during the following optimization and backpropagation updates. LeNet has several hyperparameters that could be adjusted for better outcomes and training results depending on the given typical sets of possible datasets.

It is very hard to estimate how much of the capacity the neural network filter has and there aren't any practical metrics to measure the upper bound of the learning capacity of any given neural network which is structured indecent levels of nonlinearity. So using LeNet in this analysis could be assumed that has no issues unless with any of possible stochastic deductions of measurable elements in LeNet neural network.[18]

3 EXPERIMENTAL AND COMPUTATIONAL DETAILS

3.1 Preprocessing

The unique sound patterns from quadcopter drones generate some key factors from the specific geometric structure of it as flying and maneuvering. While a human vocal cord is not a static or hardly casted material so it could distort the major tones with slight frequency dispersions or spreads, quadcopter drone has very solid and fixed geometry and distinctive signal sources such as BLDC motors, propellers, arms, and its sculpture. Still, all variations of RPM controlling, and possible maneuvers make frequency slides and dopplers, but it is considerably limited comparing the others that can be regarded. One of the distinctive differences we could peek from the overall spectrum analysis in order to use the Mel-frequency kepstrum coefficient method as the preprocessing for dimension reduction is the sound signal from quadcopter drones is observed as doubled compares to the humans one, and these narrow tone bands allow us to reduce filter-bank gaps down.

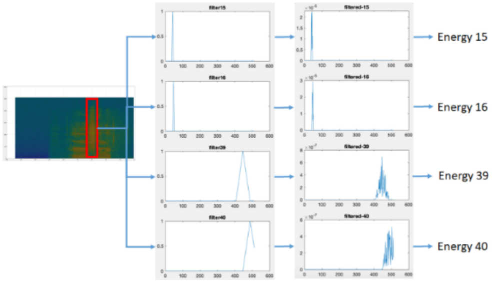

Figure 6: Tried to apply different cases of filter-banks and Mel-frequency bin selection. We tried them up to 40 and extracted bins fewer than 10

Since of many practical limitations to build the evenly distributed dataset as regarding the type of the source signal we need to consider that this empirical selection could be biased or not well arranged, but we designed the overall simulator with the utmost flexibility so was able to adjust maximum hyperparameters as below, thanks for the convolutional neural network model that helps it possible with the essential architecture that keeps the information of time and spatial domain in both.

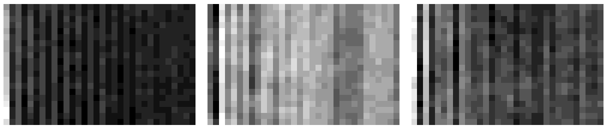

Figure 7: MFCC spectrograms that were used for input vectors. Selected 32 bins from generated total 40 frequency bins, for 20 samples by 1024 SFFT points under 48Khz sample rates. From the left, Background noise, Bebop 2 (Parrot), and F450 (DJI, Custom)

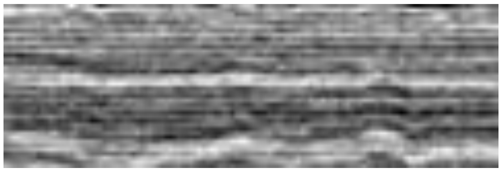

Figure 8: MFCC spectrogram, Mavic Air from DJI, Hovering

Table 2: All tested preprocessing parameters

Test Parameters	Configurations
Sample rate	44.1 ~ 48Khz
Normalization	Ortho-normal DCT
Mel-Frequency bins	8 ~ 40
Stride window	FFT size / 2
SFFT length	512 ~ 2048
MFCC frame window	10 ~ 20
Spectrogram dimension	(8 ~ 40) x (10 ~ 20)

Librosa one of the popular open sourced projects for music and audio analysis was used as complete computation sequences to obtain Mel-frequency Kepstrum coefficients as given input data. Based on the overall signal pattern that was used for the input ground fact (label) as recorded as the noisy and bulky environment in this analysis and experiment, using optimizers without a momentum factor inside didn't help for any of effective training while the other ones have. Due to the completely limited and reduced the total number of the calculation elements consist of the entire this neural network which was intended in order to apply the following evaluation on NVIDIA Jetson Xavier device, we downsized the resolution of the input signal as under 1000 pixels and now concatenated SFFT coefficients keep being influenced by the power of other adjacent frequency bands and while/non-white noises. Pixels within the size of the kernels for the convolutional computation but doesn't have any adjacent pixels in a time-domain within the equivalent period of the FFT size could be considered as noise elements from a bad SNR or some erupted noise such like a human voice, car engine noise, bird chirps, and etcetera. Since most of the waveforms for the input vector of the network have their unique buzzing sound pattern long enough, longer than the FFT window size, as long as we can recognize any of straight or curved "lines" from Mel-frequency kepstrum coefficient spectrograms as 2-d planning input vectors we can expect that this is what the convolutional kernels may find their own way of shaping along with the results from the cost function and following optimizer algorithm.

3.2 Training

Entire training and evaluation were designed and performed by the Google TensorFlow framework described as referred to overall trend timelines as figure 9 and 10, respectively.
Including one eraser category, prepared sample classes were 10 types, over 2.4Gbytes of stereo wave files were used for this supervised learning. Many different variances by combinations with preprocessed input signals, overall learning process that being assessed by updates of the summation of cross-entropy losses, and following accuracies show quite promising results.

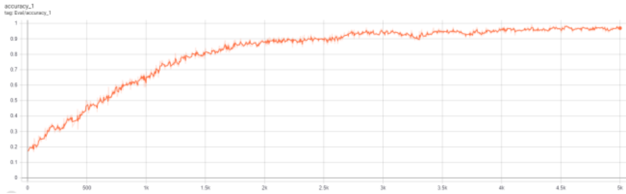

Figure 9: Accuracy trend timeline with ADAM optimizer, learning rate

Table 3: All tested parameters for the performed training

Test Parameters	Configurations
Optimizers	Adam, adadelta, adagradDA
Batch sizes	128 ~ 256
Cost function	Cross entropy sigma with 11 logits
framework	Tensorflow

The learning process uses several types of optimizer models that are supported by google TensorFlow library set - AdamOptimizer, AdadeltaOptimizer, and AdagradDAOptimizer. The size of the batch regularization window was also applied in some different combinations from 128 to 256 max. Only softmax cross-entropy with multiple outputs was used to calculate the cost with the one-hot encoding labels.

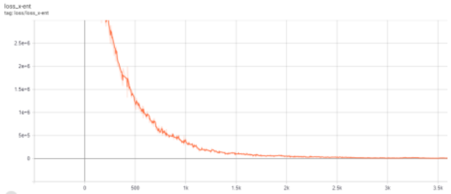

Figure 10: Decreasing trend of the calculated overall loss from the cost function – 11 logits merged by cross-entropy function

By analyzing the graphs and recorded scalar vectors as weights and biases from the trained network as Figure 11, we can assume that this network does not use all its total computational elements yet. The overall distribution we could see there are clearly numbers that have zero counts among the network elements, which means that the applied dropout sequence neutralizes some synapses to secure a certain level of generalizations.

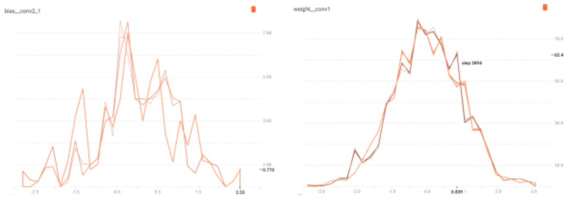

Figure 11: Tensorboard marker can show there are zero-valued elements were recorded as like from the histogram in left, not like the one right (the second convolutional layer)

But one thing that we need to be sure of is that the convolutional computation sequence does not encode any of the information as continuously provided from the input into another degree of information of any distinct set of the group like linguistic words or vector sets. Unlike some other network that has the outputs as totally altered dimensions from the input signals, such as an autoencoder or many variances of recursive networks, only extracted features from the input exist on the 2-d plan kernels which have fixed coordination direction as the time and frequency domain. It is only able to work as long as the absolute Euclidean plane, so if we need to use these outcomes into a practical application with evaluation signals, that signal should be obtained by the same quadcopter drones but also by exactly the same sensors. Once any of the non-linear factors can be a role during the input signal is being secured, we cannot rely on the extracted features anymore. This is the reason that the general speech recognition model generates the corresponding words or some of the vector subsets while the voice recognition does only the voiceprints which can be regarded as a transformation process.

4 RESULT AND DISCUSSION

4.1 Overall performance

After around 5K iteration as repeated with 128 to 256 batch-sized regularizations and up to 40% of dropout ratio from 20% the model shows almost perfect accuracies, around 96% overall. Some of the learning cases show some overfitted situation around 20% of the dropout ratio, but at the point of around 40% of a synapse-toggling achieved the maximum accuracy in a stable with repeated training and testing. Figure 9 shows the overall learning curves during the iteration, and figure 10 is for the sum of cross-entropy error from the 11 logits output in the same way as the previous one.

4.2 Consideration

As expected, weights and biases on kernels of the convolutional filters can be adapted by specific features for the input signal patterns. With regarding spatial information, dealing with these sound signals into 2-d planning is obviously reasonable and has been justified. Therefore, we can consider that the overall outcomes and its performance seem quite promising to utilize it towards other networks or applications.

Further research will cover the better hybrid neural network model over the larger computational elements within in order to overcome the input signal sensor sensitivity and the information dimension alternation

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support from Nano UAV Intelligence Systems Research Laboratory at Kwangwoon University, originally funded by Korean Defense Acquisition Program Administration (DAPA) and Agency for Defense Development (ADD).

REFERENCES

[1] Lee, D. W. (2014). Development of BLDC motor and multi-blade fan for HEV battery cooling system. International Journal of Automotive Technology, 15(7), 1101-1106.

[2] Mademlis, I., Mygdalis, V., Raptopoulou, C., Nikolaidis, N., Heise, N., Koch, T., ... & Metta, S. (2017). Overview of drone cinematography for sports filming. In European Conference on Visual Media Production (CVMP), short.

[3] Murray, C. C., & Chu, A. G. (2015). The flying sidekick traveling salesman problem: Optimization of drone-assisted parcel delivery. Transportation Research Part C: Emerging Technologies, 54, 86-109.

[4] Fotouhi, A., Ding, M., & Hassan, M. (2017, June). Understanding autonomous drone maneuverability for internet of things applications. In 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM) (pp. 1-6). IEEE.

[5] Sosnowski, T., Bieszczad, G., Madura, H., & Kastek, M. (2018, May). Thermovision system for flying objects detection. In 2018 Baltic URSI Symposium (URSI) (pp. 141-144). IEEE.

[6] Ezuma, M., Ozdemir, O., Anjinappa, C. K., Gulzar, W. A., & Guvenc, I. (2019, January). Micro-UAV detection with a low-grazing angle millimeter wave radar. In 2019 IEEE Radio and Wireless Symposium (RWS) (pp. 1-4). IEEE.

[7] Kutinlahti, V. P. (2019). Antenna for Directive Energy Device Against Drones.

[8] Birch, G. C., Griffin, J. C., & Erdman, M. K. (2015). UAS Detection, Classification, and Neutralization: Market Survey 2015. Sandia National Laboratories.

[9] Damarla, T. (2015, November). Detection of Gunshots using Microphone Array mounted on a moving Platform. In 2015 IEEE SENSORS (pp. 1-4). IEEE.

[10] Pindoriya, Rajesh & Mishra, Anshul & Singh, Bharat & Kumar, Rajeev. (2018). An Analysis of Vibration and Acoustic Noise of BLDC Motor Drive. 1-5. 10.1109/PESGM.2018.8585750.

[11] Logan, B. (2000, October). Mel frequency cepstral coefficients for music modeling. In Ismir (Vol. 270, pp. 1-11).

[12] Sigurdsson, S., Petersen, K. B., & Lehn-Schiøler, T. (2006, October). Mel Frequency Cepstral Coefficients: An Evaluation of Robustness of MP3 Encoded Music. In ISMIR (pp. 286-289).

[13] Ganchev, T., Fakotakis, N., & Kokkinakis, G. (2005, October). Comparative evaluation of various MFCC implementations on the speaker verification task. In Proceedings of the SPECOM (Vol. 1, No. 2005, pp. 191-194).

[14] Kim, J., & Kim, D. (2018). Neural network based real-time UAV detection and analysis by sound. Journal of Advanced Information Technology and Convergence, 8(1), 43-52.

[15] Jeon, S., Shin, J. W., Lee, Y. J., Kim, W. H., Kwon, Y., & Yang, H. Y. (2017, August). Empirical study of drone sound detection in real-life environment with deep neural networks. In 2017 25th European Signal Processing Conference (EUSIPCO) (pp. 1858-1862). IEEE.

[16] Carrio, A., Sampedro, C., Rodriguez-Ramos, A., & Campoy, P. (2017). A review of deep learning methods and applications for unmanned aerial vehicles. Journal of Sensors, 2017.

[17] Bengio, Y., & LeCun, Y. (2007). Scaling learning algorithms towards AI. Large-scale kernel machines, 34(5), 1-41.

[18] Achille, A., & Soatto, S. (2018). Information dropout: Learning optimal representations through noisy computation. IEEE transactions on pattern analysis and machine intelligence, 40(12), 2897-2905.