Detection of nearby UAVs using CNN and Spectrograms

Aldrich A. Cabrera-Ponce *[1], J. Martinez-Carranza[1,2], and Caleb Rascon[3]

[1]Instituto Nacional de Astrofisica, Optica y Electronica , Puebla, Mexico
[2]University of Bristol, Bristol, UK
[3]Universidad Nacional Autonoma de Mexico, Ciudad de Mexico, Mexico

ABSTRACT

In this work, we address the problem of drone detection flying nearby another UAV. Usually, computer vision could be used to face this problem by placing cameras on board the patrolling UAV. However, visual processing is prone to false positives, sensible to light conditions and potentially slow if the image resolution is high. Thus, we propose to carry out the detection by using an array of microphones mounted with a special array on board the patrolling UAV. To achieve our goal, we convert audio signals into spectrograms and used them in combination with a CNN architecture that has been trained to learn when a UAV is flying nearby and when it is not. Clearly, the first challenge is the presence of ego-noise derived from the patrolling drone itself through its propellers and motor's noise. Our proposed CNN is based on the Google's Inception v.3 network. The Inception model is trained with a dataset created by us, which includes examples of when an intruder drone flies nearby and when it does not. We tested our approach with three different drone platforms, achieving a successful detection of 97.93% for when an intruder drone flies by and 82.28% for when it does not. The dataset used for this work will be available as well as the code.

1 INTRODUCTION

Recently, the autonomous drones have grown in popularity in aerial robotics since they are vehicles with multiples capabilities, with the help of on-board sensors such as Inertial Measurement Unit (IMU), laser, ultrasonics, and cameras (both monocular and stereo). Visual sensors can be used to generate maps, for 3D re-construction, autonomous navigation, search and rescue, and security applications. However, these applications face serious problems when attempting to identify another drone in circumstances where the visual range is lacking, which can cause collisions, putting bystanders at risk in public places. Thus, it is necessary to have strategies that employ other modalities other than vision to ensure the discovery of an intruder drone. One such modality can be audio.

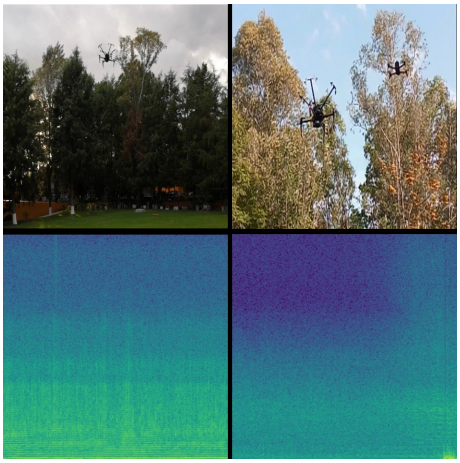


Figure 1: Classification of audio in two different environments. Left: spectrogram of an intruder aerial vehicle nearby. Right: spectrogram without an intruder vehicle nearby. `https://youtu.be/B32_uYbL62Y`

Audio processing has been a topic of research for years, which includes the challenge of recognising the source of an audio signal. In aerial robotics, the signals usually tend to present noise that disturbs the original signal, making the recognition an even more difficult task. However, if this is successful, it can be used to find the relative direction of a sound source (such as another drone) as well as identify other sounds in different distance ranges. A useful manner with which audio is represented in this type of applications is in the time-frequency domain, in which the spectrogram of the signal is manipulated as if it were an image. These images allow a detailed inspection of the noise of the rotors to analyse vibration and prevent future failures in the motors. By identifying features inside the spectrogram, sound source identification and localisation may be possible over a drone.

Recent works employ deep learning strategies (such as Convolutional Neural Networks, CNN) to classify sound sources, and many of these methods aim to learn features from a spectrogram. We propose to use a CNN to identify when there is or may not be an aerial vehicle near our drone from a given input spectrogram (See Fig.1).

We base our CNN-based classification model on the

*Department of Computer Science at INAOE. Email addresses: {aldrichcabrera, carranza}@inaoep.mx

Google's Inception v.3 architecture. The information is separated in two different classes: with and without a drone. Each class has 3000 spectrograms for training. Each spectrogram is manipulated as though it is an image, with each pixel representing a time-frequency bin, and its colour representing its energy magnitude. Moreover, our approach aims to classify with a high level of performance over different aerial platforms.

This paper is organised as follows: Section 2 provides related works which identify sources in the environment with aerial vehicles; Section 3 describes the hardware used; Section 4 provides a detailed description of the proposed approach; Section 5 describes the analysis of the spectrograms for each class; Section 6 presents the classification results using the proposed approach; and conclusions and future work are outlined in Section 7.

2 RELATED WORK

As mentioned earlier, drones that solely employ vision may be limited when identifying aerial vehicles in an environment near a flight zone. Thus, works with radars have used the micro-doppler effect to identify a target [1] or different targets [2]. This use, as the basis for classification, the change of the audible frequency due to changes in the velocity of the propellers [3,4], as well as other features [5]. Additionally, when this effect is represented by its cadence frequency (CFS), it can be used to recognise other descriptors like shape and size, achieving the classification of multiple classes of aerial vehicles [6].

As for audio processing techniques, they have been used in aerial robotics for classification, detection, and analysis of rotors, to analyse annoyances generated by the UAV's noise through psycho-acoustic and metrics of noise [7]. Likewise, they have been used for the localisation of sound sources [8], reducing the effect ego-noise of the UAV's rotors and localise the source in high noise conditions in outdoor environments [9]. These auditory systems have been used in conjunction with radars and acoustic sensors, showing good performances when identifying UAVs in public places [10] and detecting sound sources in spaces of interest [11]. Even though these alternatives have been developed, the audio processing area of research over a drone is a challenging task that still has considerable room to develop.

On the other hand, good acoustic identification using harmonic spectrums can help avoid collisions between two fixed-wing aircraft by increasing the detection range of an intruder UAV to 678 meters [12]. This localisation range can be further improved by 50% (while reducing computational costs) by using harmonic signals [13]. In [14], a design for positioning an 8-microphone array over a drone is presented, aimed to detect distinct nearby UAVs from a given drone. This design is useful for detection, localisation, and tracking intruder drones operating close to undesired areas such as airports or public environments.

There are several strategies that employ deep learning for sound classification. For example, the direction of a sound source was estimated using spherical bodies around a drone and microphones on-board in [15]. Furthermore, multiple targets were detected, localised and tracked using audio information and convolutional networks in [16]. Deep learning strategies have also been used to identify the presence of different drones in outdoor environments, by analysing and classifying their spectrogram-based sound signatures in [17] or by merging them with wave radar signals in a convolutional network [18]. However, these strategies are performed from ground stations. There isn't much developed when it comes to identifying a UAV from the audio data captured from microphones on-board another UAV.

3 SYSTEM OVERVIEW

The primary aerial vehicle from which all test are carried out is a quad-rotor "Matrice 100", manufactured by DJI. This platform is popular because it can carry multiple sensors for outdoor navigation and autonomous flight, as it can bare a load of up to 1000 grams.

The audio capture system is the 8SoundUSB system that is part of the ManyEars project [19]. It is composed of 8 miniature microphones and an USB-powered external audio interface. The microphones were designed for mobile robot audition in a dynamic environment, implementing real-time processing to perform sound source localisation, tracking, and separation.

For audio recording and processing, we mounted the microphones over the same 3D-printed structure used by [14] to record eight-channel audio in raw format. All of the hardware was driven by the on-board Intel Stick Computer, with 32 GB of RAM and Linux Ubuntu 16.04 LTS. The recordings were carried out in two different environments: with an intruder drone and without an intruder drone. The intruder drone was a Drone Bebop 2.0, manufactured by Parrot, which is known for its stability and ease of control.

The place where the recordings were made was in the Centre of Information of the Instituto Nacional de Astrofisica, Optica y Electronica (INAOE), where there is a considerable large area that is appropriate for flying multiple drones at once.

The recording process is shown in Figure 2. First, the microphones are placed in the Matrice 100, with microphones 1, 2, 3 and 4 mounted in the front and microphones 5, 6, 7 and 8 mounted in the back. Then, an expert pilot controls the drone while the audio is recorded on-board the Intel Stick Computer.

The specifications with which audios were recorded are:

- The sampling rate is 48 kHz to allow a considerable amount of original resolution which can later be reduced if need be.
- Recording time: 240 seconds in the environment

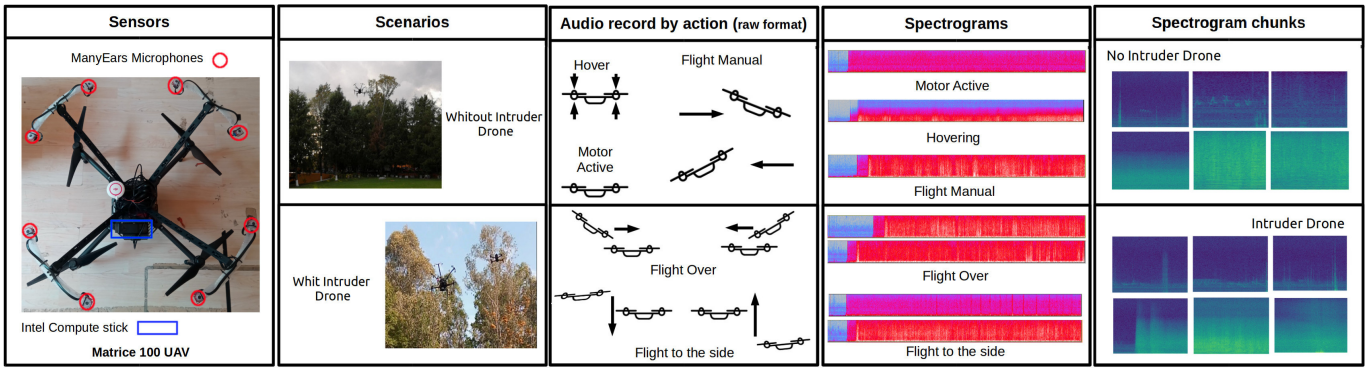


Figure 2: General overview to record the audio in two different environments and generate a dataset.

with an intruder drone, labelled as the class "intruder drone"; and 198 seconds in the environment without it, labelled as the class "no intruder drone."

- The drones performed different actions while recording in both environments. In the environment "intruder drone", these actions were: on the ground with just the motors activated, hovering, and manual flight. In the environment "intruder drone", the actions were: the intruder drone flying on the side of the drone and over the top of the drone.

- The audio files were then manually transferred to a computer in the ground. This was done to avoid latency issues in the wireless transfer. The audio files were then transformed to the time-frequency domain, generating a spectrogram for each microphone (as detailed in the following section).

4 TRAINING DATASET

The training dataset was created from the spectrograms generated by the recorded data, and apart from the training dataset, a testing dataset was created to validate the system. Each audio file was segmented in 2-second segments. The Short Time Fourier Transform was applied to each segment, with a 1024-sample Hann window (to avoid spectral leakage) and 75% overlap. The audio files in the negative class "no intruder drone" include recordings of the air blowing through the trees, voices, cars, people and the noise of the Matrice's motors. The positive class "intruder drone" includes the recording of 200 seconds of the intruder drone flying on the side and over the Matrice 100. The Tables 1 and 2 show the spectrograms generated for each action which make up the whole of the training data set.

4.1 CNN architecture

We propose a convolutional neural network (CNN) as our classification model. This network is based on the architecture of the Google Net Inception v.3 (as shown in Figure 3)

Action	Time	Spectrograms (by mic)
Motor Activation	198.0 sec	98
Hovering	198.0 sec	98
Flight Manual	198.0 sec	98
Flight Manual 2	198.0 sec	98
	792.0 sec	**3168 (by all mics)**

Table 1: Spectrograms of the class "no intruder drone".

Action	Time	Spectrograms (by mic)
Flight over	200.0 sec	100
Flight to the side	200.0 sec	100
Flight over 2	200.0 sec	100
Flight to the side 2	200.0 sec	100
	800.0 sec	**3200 (by all mics)**

Table 2: Spectrograms of the class "intruder drone".

using Keras and Tensorflow. We employ a transfer learning strategy. Meaning, our system uses a model that was already trained on the ImageNet corpus [20], but we then augmented it with a new top layer to be trained with our recorded data. This is done so that the resulting model is focused in recognising the spectrogram-type of images relevant to our application: identifying an intruder drone flying near another.

The training data set was arranged in folders, each representing one class and baring approximately 3000 images. The model inherited the input requirement of the Inception V.3 architecture, receiving as input an image of size 224 x 224 pixels. The network was trained for 4500 epochs. Since the softmax layers can contain N labels, the training process corresponds to learning a new set of weights; that is to say, it is technically a new classification model.

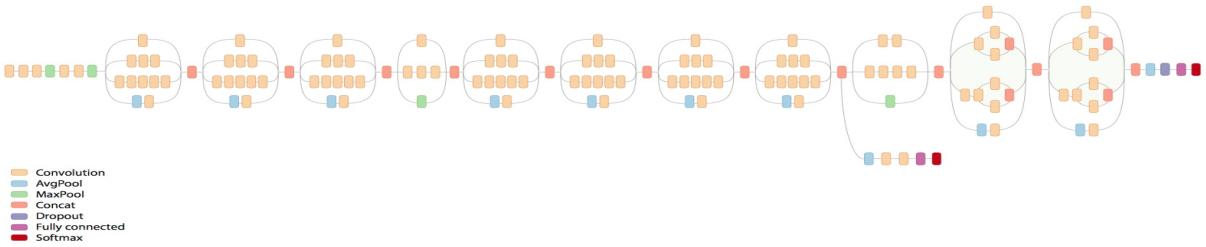


Figure 3: Schematic diagram of Inception V3.

5 SPECTROGRAM ANALYSIS

It is important to manually analyse the resulting spectrograms, to observe (in a preliminary fashion) if both classes are distinguishable to a human listener. The first analysis was made with the Audacity software [21] to visualise the audio data as a spectrogram. Then, the audio files were reproduced to see if a human listener was able to identify the intruder drone flight during the recordings. In Figure 4, the possible positions corresponding to these moments are marked in a circle.

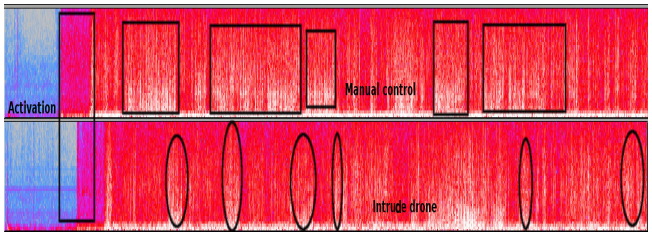


Figure 4: Comparison between spectrograms of manual control (top) and intruder drone (bottom).

Once it was shown that a human listener is able to identify the intruder drone, further analysis was carried out. 2-second time-frequency spectrograms were generated (as described in Section 4) for the two classes, and are shown in Figure 5. As it can be seen, there is an important amount high-frequency energy present in the "intruder drone" class that is not present in the class "no intruder drone."

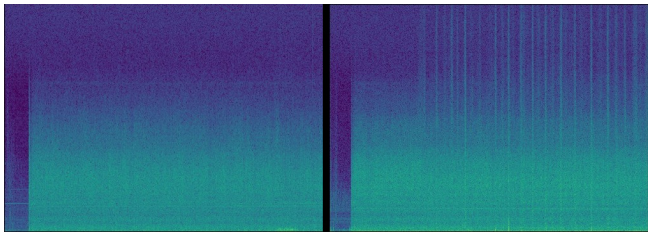


Figure 5: Spectrograms generated of activate motors (left) and intruder drone flight (right).

6 RESULTS OF THE DRONE CLASSIFICATION

The results shown in this section is that of the trained classifier. Its aim is to classify between two classes of input spectrograms. It could be argued that it is actually a verifier. However, we evaluated it as a classifier for future proofing.

6.1 Validation

We performed two experiments to evaluate the classification model. The first experiment shows the overall effectiveness of the model by testing it with 920 images for each class with an average inference time of 0.4503 sec. Table 3 presents the results of this test, and it can be seen that the class "intruder drone" is consistently classified correctly, with only 19 wrong classifications out of 920 tests. However, the class "no intruder Drone" gives a lower accuracy, with 163 images wrong classifications.

Class	Images	Incorrect	Accuracy
No intruder Drone	920	163	**82.28%**
Intruder Drone	920	19	**97.93%**

Table 3: Validation of classification network.

To a better understanding of the performance of the classifier we considered a binary classification where a nearby drone is considered as a positive sample in this way we have the true positive (Tp) = 901, true negative (Tn) = 757, false positive (Fp) = 19 and false negative (Fn) = 163. In Table 4, we show the result of Accuracy, Precision and Recall provide a better understanding of the performance of the classifier.

Accuracy	Precision	Recall
0.90108	0.97934	0.84680

Table 4: Accuracy, precision and recall result.

The second experiment measures the output of the model for each class, given a representative spectrograms to test with. 20 spectrograms were chosen (10 for each class), and the model outputs of each class are shown in Table 5. Although some outputs are below 70% (which implies some uncertainty of the model), the final classification is correct

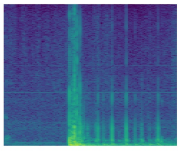
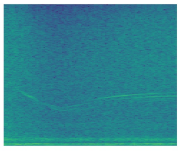
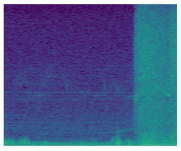
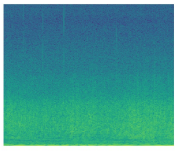
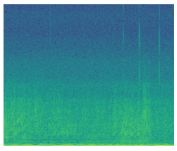
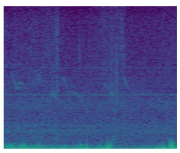
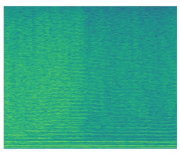
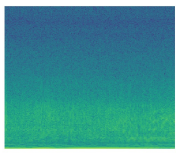
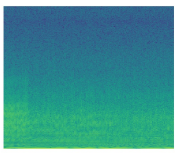
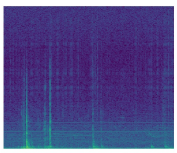
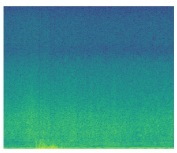
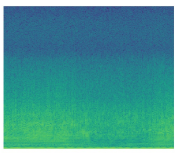
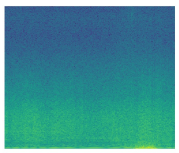
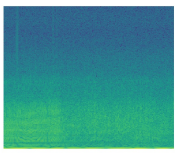
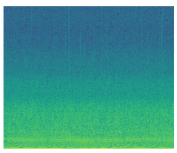
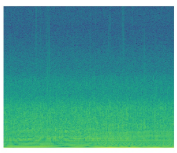
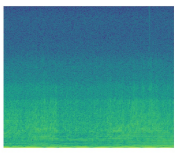
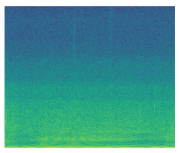
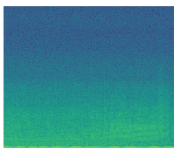
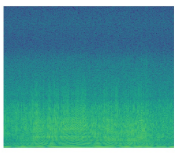
Class	Classification				
					
No intruder Drone	**97.82%**	**93.28%**	**78.82%**	**63.80%**	**76.59%**
Intruder Drone	2.18%	6.72%	21.18%	36.2%	23.41%
					
No intruder Drone	**77.54%**	**85.76%**	**84.66%**	**89.57%**	**82.31%**
Intruder Drone	22.46%	14.24%	15.34%	10.43%	17.69%
					
Intruder Drone	**88.23%**	**90.46%**	**94.70%**	**70.50%**	**91.57%**
No Intruder Drone	11.77%	9.54%	5.30%	29.50%	8.43%
					
Intruder Drone	**93.14%**	**91.10%**	**75.93%**	**65.27%**	**98.89%**
No Intruder Drone	6.86%	8.90%	24.07%	34.73%	1.11%

Table 5: Example of classification with CNN using some test images.

in all cases. These results give us a representative view of the expected performance of the model with the two classes that are relevant to our application: identifying an intruder drone flying near another.

7 Conclusion

In this paper, we proposed a CNN-based classifier of two types of environments: with and without an intruder drone, using only audio captured by a UAV. A time-frequency spectrogram was used as an the signal representation, which is compatible with known CNN-based architectures. We employed a transfer-learning strategy, with which the top layer of a pre-trained Google's Inception V.3 model was modified and trained, which made the training process very efficient. The classifier was evaluated in two experiments, and it achieved a good classification performance in most cases. The work can be further strengthened by using the eight microphones individually to detect the direction of the intruder drone. This would allow a fast enough detection, to give enough time to plan a strategy for collision avoidance.

References

[1] Matthew Ritchie, Francesco Fioranelli, Hervé Borrion, and Hugh Griffiths. Multistatic micro-doppler radar feature extraction for classification of unloaded/loaded micro-drones. *IET Radar, Sonar & Navigation*, 11(1):116–124, 2016.

[2] RIA Harmanny, JJM De Wit, and G Prémel Cabic. Radar micro-doppler feature extraction using the spectrogram and the cepstrogram. In *2014 11th European Radar Conference*, pages 165–168. IEEE, 2014.

[3] F Fioranelli, M Ritchie, H Griffiths, and H Borrion. Classification of loaded/unloaded micro-drones using multistatic radar. *Electronics Letters*, 51(22):1813–1815, 2015.

[4] JJM De Wit, RIA Harmanny, and P Molchanov. Radar micro-doppler feature extraction using the singular value decomposition. In *2014 International Radar Conference*, pages 1–6. IEEE, 2014.

[5] Pavlo Molchanov, Ronny IA Harmanny, Jaco JM de Wit, Karen Egiazarian, and Jaakko Astola. Classification of small uavs and birds by micro-doppler signatures. *International Journal of Microwave and Wireless Technologies*, 6(3-4):435–444, 2014.

[6] Wenyu Zhang and Gang Li. Detection of multiple micro-drones via cadence velocity diagram analysis. *Electronics Letters*, 54(7):441–443, 2018.

[7] Andrew W Christian and Randolph Cabell. Initial investigation into the psychoacoustic properties of small unmanned aerial system noise. In *23rd AIAA/CEAS Aeroacoustics Conference*, page 4051, 2017.

[8] Koutarou Furukawa, Keita Okutani, Kohei Nagira, Takuma Otsuka, Katsutoshi Itoyama, Kazuhiro Nakadai, and Hiroshi G Okuno. Noise correlation matrix estimation for improving sound source localization by multirotor uav. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3943–3948. IEEE, 2013.

[9] Takuma Ohata, Keisuke Nakamura, Takeshi Mizumoto, Tezuka Taiki, and Kazuhiro Nakadai. Improvement in outdoor sound source detection using a quadrotor-embedded microphone array. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1902–1907. IEEE, 2014.

[10] Seongha Park, Sangmi Shin, Yongho Kim, Eric T Matson, Kyuhwan Lee, Paul J Kolodzy, Joseph C Slater, Matthew Scherreik, Monica Sam, John C Gallagher, et al. Combination of radar and audio sensors for identification of rotor-type unmanned aerial vehicles (uavs). In *2015 IEEE SENSORS*, pages 1–4. IEEE, 2015.

[11] Prasant Misra, A Anil Kumar, Pragyan Mohapatra, and P Balamuralidhar. Aerial drones with location-sensitive ears. *IEEE Communications Magazine*, 56(7):154–160, 2018.

[12] Brendan Harvey and Siu OYoung. Acoustic detection of a fixed-wing uav. *Drones*, 2(1):4, 2018.

[13] Brendan Harvey and Siu O'Young. A harmonic spectral beamformer for the enhanced localization of propeller-driven aircraft. *Journal of Unmanned Vehicle Systems*, 7(2), 2019.

[14] Oscar Ruiz-Espitia, Jose Martinez-Carranza, and Caleb Rascon. Aira-uas: An evaluation corpus for audio processing in unmanned aerial system. In *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 836–845. IEEE, 2018.

[15] Kotaro Hoshiba, Kai Washizaki, Mizuho Wakabayashi, Takahiro Ishiki, Makoto Kumon, Yoshiaki Bando, Daniel Gabriel, Kazuhiro Nakadai, and Hiroshi Okuno. Design of uav-embedded microphone array system for sound source localization in outdoor environments. *Sensors*, 17(11):2535, 2017.

[16] Zeeshan Kaleem and Mubashir Husain Rehmani. Amateur drone monitoring: State-of-the-art architectures, key enabling technologies, and future research directions. *IEEE Wireless Communications*, 25(2):150–159, 2018.

[17] Sungho Jeon, Jong-Woo Shin, Young-Jun Lee, Woong-Hee Kim, YoungHyoun Kwon, and Hae-Yong Yang. Empirical study of drone sound detection in real-life environment with deep neural networks. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 1858–1862. IEEE, 2017.

[18] Byung Kwan Kim, Hyun-Seong Kang, and Seong-Ook Park. Drone classification using convolutional neural networks with merged doppler images. *IEEE Geoscience and Remote Sensing Letters*, 14(1):38–42, 2017.

[19] François Grondin, Dominic Létourneau, François Ferland, Vincent Rousseau, and François Michaud. The manyears open framework. *Autonomous Robots*, 34(3):217–232, 2013.

[20] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[21] D Mazzoni and R Dannenberg. Audacity [software]. *The Audacity Team, Pittsburg, PA, USA*, 2000.