Detection of nearby UAVs using a multi-microphone array on board a UAV

Aldrich A Cabrera-Ponce, J Martinez-Carranza and Caleb Rascon

Abstract
In this work, we address the problem of UAV detection flying nearby another UAV. Usually, computer vision could be used to face this problem by placing cameras onboard 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 onboard 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 UAV itself through its propellers and motor’s noise. Our proposed CNN is based on Google’s Inception v.3 network. The Inception model is trained with a dataset created by us, which includes examples of when an intruder UAV flies nearby and when it does not. We conducted experiments for off-line and on-line detection. For the latter, we manage to generate spectrograms from the audio stream and process it with the Nvidia Jetson TX2 mounted onboard the patrolling UAV.

Keywords
Detection UAVs, audio classification, spectrograms, microphones

Date received: 15 March 2020; accepted: 7 April 2020

Introduction
Recently, the autonomous UAVs have grown in popularity in aerial robotics since they are vehicles with multiple 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 detect another UAV 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 UAV. One such modality can be audio.

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 UAV) as well as to detect 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 detecting features inside the
spectrogram, sound source detection and localisation may be possible over a UAV.

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. In this work, we propose to use a CNN to classify a spectrogram aiming at detecting if an intruder UAV flies nearby a patrolling UAV (see Figure 1).

We base our CNN-based classification model on Google’s Inception v.3 architecture. The information is separated into two different classes: with and without a UAV. 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. To assess our approach, we have carried out off-line and on-line experiments. The latter means that we have managed to carry out the detection in real-time onboard the patrolling UAV during flight operation.

This paper is organised as follows: the next section provides related works which detect sources in the environment with aerial vehicles; then describes the hardware used; the subsequent section provides a detailed description of the proposed approach; the analysis of the spectrograms for each class is then described; the penultimate section presents the classification results using the proposed approach; and conclusions and future work are outlined in the last section.

Related work

As mentioned earlier, UAVs that solely employ vision may be limited when detecting aerial vehicles in an environment near a flight zone. Thus, works with radars have used the micro-Doppler effect to detect a target or different targets. This is used as classification due to changes in the velocity of the propellers, as well as other features. 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.

As for audio processing techniques, they have been used in aerial robotics for classification, detection, and

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

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

Several strategies employ deep learning for sound classification. For example, the direction of a sound source was estimated using spherical bodies around a UAV and microphones on-board. Furthermore, multiple targets were detected, localised and tracked using audio information and convolutional networks. Deep learning strategies have also been used to detect the presence of different UAVs in outdoor environments, by analysing and classifying their spectrogram-based sound signatures or by merging them with wave radar signals in a convolutional network. However, these strategies are performed from ground stations. There is not much developed when it comes to detecting a UAV from the audio data captured from microphones on-board another UAV.

System overview

The objective of this work is the classification of spectrograms from audio signals to detect an intruder UAV. Hence, we performed an audio classification using spectrograms created from raw files recorded on-board of the Matrice 100. The audio capture system is the 8SoundUSB system of the ManyEars project composed of eight microphones designed for mobile robot audition in dynamic environments, implementing real-time processing to perform sound source localisation, tracking, and separation.

The microphones were mounted over a 3D-printed structure used by Ruiz-Espitia et al. to record the audio during the flight and driven by the on-board Intel Compute Stick using “arecord” (ALSA command-line for sound file recorder) to generate raw files audio. These recordings were carried out in the Centre of Information of the Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), where we took audio information in the presence of an intruder UAV and without intruder. The intruder UAV is a Drone Bebop 2.0 considered like an intruder for flying near the area where the first UAV is localised. The specifications with which audios were recorded are 48 kHz of sampling rate, 240 s of recording time while the intruder UAV flies in the environment, and 198 s of recording time without it. Each recording was performed with different actions such as the activate motors, hovering, and manual flight without the intruder; and flights on the side and over the top of the first UAV using the intruder UAV.

For a clearer understanding of the whole approach, we present the recorded audio files in Figure 2, where we show the microphones mounted over the UAV, the scenarios, the audio recorded by each action and the spectrograms generated.

On the other hand, considering avoid recorder, the audios in raw files then transferred and transformed to the time–frequency domain in a computer in the ground. We decide to use a computer with more

![Figure 2](image_url). General overview to record the audio in two different environments and generate a dataset.
speed and power-efficiency than Intel Compute Stick to transform the audio to spectrograms at the moment of the flight. The NVIDIA Jetson TX2 Module is an embedded device with an NVIDIA Pascal GPU architecture with 256 Cuda cores, a hex-core ARMv8 64-bit CPU complex, and 8GB of LPDDR4 memory with a 128-bit interface. The audio streaming was taken using PortAudio API, which allows the use of two buffers to recollect the audio information and create the spectrogram without loss of information in the audio signal. Besides, the audio streaming was performed in the same scenarios with different actions as shown in Figure 3 without the need to record an audio file.

**Training dataset**

For the off-line experiments, two datasets were created from the spectrograms generated by the recorded data, one for training and the other for validation. For each microphone, a raw audio signal was recorded in a wav format. In total, 200 s corresponding to the “intruder UAV” and 198 s corresponding to the “no-intruder UAV” were used to generate the spectrograms for each microphone. Additionally, an audio file of 460 s was used to generate the validation dataset with 920 spectrograms. Each audio file was segmented in 2-s segments and converted to wav format by applying the Short Time Fourier Transform for each segment with 1024-sample Hann window and 75% overlap. Tables 1 and 2 show the amount of spectrograms generated for each class.

| Action                  | Time  | Spectrograms (by mic) |
|-------------------------|-------|-----------------------|
| Motor activation        | 198.0 s | 98                    |
| Hovering                | 198.0 s | 98                    |
| Flight manual           | 198.0 s | 98                    |
| Flight manual 2         | 198.0 s | 98                    |
|                        | 792.0 s | 3168                  |
| (total time recorded)   |       | (by all mics)         |

| Action                  | Spectrograms (by mic) |
|-------------------------|-----------------------|
| Flight over             | 200.0 s | 100 |
| Flight to the side      | 200.0 s | 100 |
| Flight over 2           | 200.0 s | 100 |
| Flight to the side 2    | 200.0 s | 100 |
|                        | 800.0 s | 3200 |
| (total time recorded)   |       | (by all mics)         |

**Table 3.** Spectrograms of the class “no-intruder UAV”.

| Action                  | Spectrograms |
|-------------------------|--------------|
| Nothing                 | 1000         |
| Motor activation        | 1000         |
| Hovering                | 1000         |
| Flight manual           | 1000         |
|                        | 4000         |

**Table 4.** Spectrograms of the class “intruder UAV”.

| Action                  | Spectrograms |
|-------------------------|--------------|
| Flight over with activation | 1000     |
| Flight around with hovering | 1000     |
| Flight to the side with hovering | 1000     |
| Flight over with hovering     | 1000     |
|                          | 4000        |

For the on-line experiments, where the detection has to be carried out in real-time, that is, onboard the UAV during flight, one dataset was generated with the same specifications of the Hann window and overlap, using the audio streaming as input. In this case, the eight audio channels are coupled together and used to generate a single spectrogram.

For this dataset, we generated 1000 spectrograms for each action presented in Figure 3, obtaining 4000 spectrograms for each class and a total of 8000 spectrograms to the training dataset.

Tables 3 and 4 show the amount of spectrograms generated for each action in real-time.
CNN architecture

The convolutional neural network (CNN) proposed in the classification is based on the architecture of the Google Net Inception v.3 (Figure 4) using Keras 2.1.4 and Tensorflow 1.4.0. We employed a transfer learning strategy, using a model that was already trained on the ImageNet corpus\textsuperscript{20} and augmented it with a new top layer to be trained with our dataset. The resulting model will be focused on recognising the spectrogram images to our application: detecting an intruder UAV flying near another.

The training data set was arranged in folders, each representing one class with approximately 3000 images to generate a model from recorded audios, and 4000 images obtained from audio streaming in real-time. The models inherited the input requirement of the Inception V.3 architecture, receiving as input an RGB image of size $224 \times 224$ pixels. The network was trained with 4500 iterations for the model with recorded audios and 15,000 iterations with the dataset of the audio streaming. 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.

Spectrograms

The spectrograms generated from the audio files recorded and of audio streaming were processed for preliminary analysis to observe if there is a distinguishable change in both the audio and the spectrograms, when an intruder UAV is present or not. The analysis was made in the Audacity software\textsuperscript{21} to visualise the spectrograms while we also listen to the audios. For those spectrograms generated from the audio streaming in real-time, we compared the

![Schematic diagram of inception V3.](image)

**Figure 4.** Schematic diagram of inception V3.

![Comparison between spectrograms of manual control (top) and intruder UAV (bottom).](image)

**Figure 5.** Comparison between spectrograms of manual control (top) and intruder UAV (bottom).

![Spectrograms generated of activate motors (left) and intruder UAV flight (right).](image)

**Figure 6.** Spectrograms generated of activate motors (left) and intruder UAV flight (right).
image with the exact time of the flight to verify if there is an intruder nearby to the main UAV. In Figure 5, we show the possible positions corresponding to these moments marked in a circle belonging to the audio files recorded, and Figure 6 shows the spectrograms generated (as described in ‘Spectrograms’ section) for the two classes. As can be seen, there is a significant amount of high-frequency energy present in the “intruder UAV” class that is not present in the class “no-intruder UAV.”

In Figure 7, we show some segment of spectrograms captured in real-time with the Nvidia Jetson TX2 for both classes. The spectrograms were generated using PortAudio and visualised with OpenCV to then convert it to a message of ROS and publish in a ROS topic. The publication is suitable to perform the image classification, avoiding the delays in the spectrograms.

**Results for the no-intruder/intruder classification**

The experiments focused on image classification of two classes of input spectrograms to detect if there is an intruder UAV or not. In addition, we tested the network for real-time detecting during flights.

**Spectrograms validation**

We performed two experiments to evaluate the classification of model generated from audios recorded. 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 s. Table 5 presents the results of this test where the class “intruder UAV” obtains a percentage of 97.93%, classifying 19 spectrograms incorrectly. Nevertheless, we obtain a lower with the class “no-intruder UAV” than the first class with an 82.28% of accuracy, classifying 163 spectrograms incorrectly.

Evaluating the performance of the classifier, we considered a binary classification which a nearby UAV is considered as a positive sample, where: the true positives (Tp) = 901, true negative (Tn) = 757, false positive (Fp) = 19, and false negative (Fn) = 163. In Table 6, we show the result of accuracy, precision and recall to provide a better understanding of the performance of the classifier.

In the second experiment, we measured the output accuracy of the model for each class, using the spectrograms of the validation dataset as input. We choose 20 spectrograms (10 for each class) to test the model output and recorded the results in Table 7. Although some outputs are below 0.70 (which implies some uncertainty of the model), the final classification is correct in all cases. Besides, the results obtained in this experiment give us a representative view of the expected performance of the model, classifying two types of classes to detect an intruder UAV flying nearby another UAV.

**Real-time detection**

In our initial experiments, we validated the classifier created from audio recorded files, using a validation dataset with 920 images and measuring the accuracy of the model with 10 spectrograms for each class. Nevertheless, we have also conducted additional experiments to carry out intruder UAV detection in real-time. For the latter, we used the trained model from the audio streaming to check out the effectiveness of the classifier by detecting a UAV nearby.

**Table 5.** Validation of classification network.

| Class          | Images | Incorrect | Accuracy (%) |
|----------------|--------|-----------|--------------|
| No-intruder UAV | 920    | 163       | 82.28        |
| Intruder UAV   | 920    | 19        | 97.93        |

Note: In bold, the accuracy percentage for each class.

**Table 6.** Accuracy, precision and recall result.

|          | Precision | Recall |
|----------|-----------|--------|
| 0.90108  | 0.97934   | 0.84680|

**Figure 7.** Examples of spectrogram segments where 1 is the ambient audio; 2 motor activation; 3 to hovering flight, and 4 the manual flight.
The experiments have carried out onboard of the Nvidia Jetson TX2 whose GPU architecture improved the classifier’s performance in terms of processing time, thus enabling real-time detection. In Figure 8, we present the architecture related to the acquisition of the spectrograms and their classification in real-time using ROS.

Table 7. Example of classification with CNN using some test images.

| Class           | Classification |
|-----------------|----------------|
| No-intruder UAV | 0.9782 0.9328 0.7882 0.6380 0.7659 |
| Intruder UAV    | 0.0218 0.0672 0.2118 0.3620 0.2341 |

| Class           | Classification |
|-----------------|----------------|
| No-intruder UAV | 0.7754 0.8576 0.8466 0.8957 0.8231 |
| Intruder UAV    | 0.2246 0.1424 0.1334 0.1043 0.1769 |

| Class           | Classification |
|-----------------|----------------|
| Intruder UAV    | 0.8823 0.9046 0.9470 0.7050 0.9157 |
| No-intruder UAV | 0.1177 0.0954 0.0530 0.2950 0.0843 |

| Class           | Classification |
|-----------------|----------------|
| Intruder UAV    | 0.9314 0.9110 0.7593 0.6527 0.9889 |
| No-intruder UAV | 0.0686 0.0890 0.2407 0.3473 0.0111 |

Note: In bold, the highest classification value returned by the CNN.

Figure 8. Set-up for the real-time UAV classification: spectrograms acquisition and classification synchronized by the robot operating system (ROS).
We performed two flights, one with the intruder UAV flying nearby the Matrice 100 during 335 s and the other without a UAV nearby during 300 s. The first flight consisted of the intruder having to fly around, over and to the side of the Matrice 100, while the CNN ran in real-time classifying the spectrograms to determine whether an intruder UAV was detected or not. For this flight, the Matrice 100 flew in hovering, while the intruder UAV kept a distance between 30 cm until 2 m w.r.t the Matrice 100. The second flight performed the same for the Matrice 100, that is, in hovering, but without the intruder UAV flying around. Table 8 shows the results obtained for these flights with the real-time detection in terms of accuracy and average time of the classification using the Nvidia Jetson TX2.

For these real-time experiments, we report an accuracy of 0.8409 for the intruder UAV detection scenario. This result is suitable to detect an intruder UAV located at 2 m of distance. For the second flight, the CNN obtained an accuracy of 0.7578 for the no-intruder detection scenario. In comparison with the model trained off-line, on-line detection is slightly lower. We argue that this can be improved by adding more

Table 8. Real-time results onboard of the Matrice 100

| Class            | Flight 1 | Flight 2 |
|------------------|----------|----------|
| Intruder UAV     | 0.8409   | 0.2421   |
| No-intruder UAV  | 0.1590   | 0.7578   |
| Average time (s) | 1.2526   | 1.2525   |

Note: The class with highest accuracy is highlighted in bold for each flight test.

Table 9. Real-time intruder UAV detection for different distances.

| Class                | Flight 1 | Flight 2 | Flight 3 |
|----------------------|----------|----------|----------|
| Intruder UAV         | 0.8356   | 0.7205   | 0.6677   |
| No-intruder UAV      | 0.1643   | 0.2794   | 0.3322   |
| Average time (s)     | 1.2432   | 1.2520   | 1.2986   |
| No. of incorrect     | 31       | 53       | 64       |
| detections           |          |          |          |

Note: The class with highest accuracy is highlighted in bold for each flight test.

Figure 9. Image sequence that shows the process of real-time classification to detect an intruder UAV flying nearby the Matrice 100 vehicle, which carried the microphones and computer hardware for the processing, including the Nvidia Jetson TX2 computer.
training data. We could also change Hann’s window with more samples to generate a spectrogram with major quality.

In addition to the above, we performed other three flights with the intruder UAV near the Matrice 100. The motivation for these experiments aimed at assessing the performance of the classifier to detect the intruder UAV flying from different points near the Matrice 100. The first and second flights were performed over and to the side of the Matrice 100. For the third flight, the intruder UAV was located at a distance of 5 m. The results are presented in Table 9, showing the accuracy, average time and the number of spectrograms that were classified incorrectly during the 180 s of the flight, which is indicated as No. of incorrect detections.

As noted in Table 9, the nearer the intruder flew to the Matrice 100, the better the classification accuracy. However, flight 3 shows that the CNN struggles when the intruder is farther (5 m) from the Matrice 100. Yet, a detection accuracy of 0.667 was obtained, which could be still exploited to detect suspicious activity. In Figure 9, we show some images of the experiment performed in real-time using the classifier to detect the intruder UAV.

Conclusion

In this work, we have presented a CNN-based classifier for the detection of an intruder UAV flying nearby another patrolling UAV. The main goal was to carry out the detection by using only audio singles recorded with an array of microphones mounted on the patrolling UAV. A time–frequency spectrogram was used as 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.

We conducted experiments in outdoors to assess the performance of our classifier in off-line and on-line mode. For the former, a database of spectrograms was produced from recorded raw audio signals. For the latter, the audio streaming was processed directly to produce spectrograms in real-time, which were used for training and later on for classification in real-time during a flight. In sum, for off-line detection, our CNN obtained an accuracy of 0.97 for the intruder detection and 0.82 for the no-intruder detection. For the real-time experiments, we achieved an accuracy of 0.84 and 0.75, respectively, with an average time of 1.2 s to generate the spectrogram from the audio stream and classify it with the CNN.

We found that the real-time detection decays when the intruder UAV flies farther than 5 m. However, we believe this can be improved by using higher quality microphones and more training data. Our future work includes detecting the direction of the intruder UAV and its distance w.r.t the patrolling UAV.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Aldrich A Cabrera-Ponce https://orcid.org/0000-0002-9998-7444
Caleb Rascon https://orcid.org/0000-0002-1176-6365

References

1. Ritchie M, Fioranelli F, Borroni H, et al. Multistatic micro-Doppler radar feature extraction for classification of unloaded/loaded micro-drones. IET Radar Sonar Navig 2016; 11: 116–124.
2. Harmanny R, De Wit J and Cabic GP. Radar micro-Doppler feature extraction using the spectrogram and the cepstrogram. In: 11th European radar conference, Rome, Italy, 8-10 October 2014, pp. 165–168. Piscataway, NJ: IEEE.
3. Fioranelli F, Ritchie M, Griffiths H, et al. Classification of loaded/unloaded micro-drones using multistatic radar. Electron Lett 2015; 51: 1813–1815.
4. De Wit J, Harmanny R and Molchanov P. Radar micro-Doppler feature extraction using the singular value decomposition. In: International radar conference, Lille, France, 13-17 October 2014, pp. 1–6. Piscataway, NJ: IEEE.
5. Molchanov P, Harmanny RI, de Wit JJ, et al. Classification of small uavs and birds by micro-Doppler signatures. Int J Microwave Wireless Technol 2014; 6: 435–444.
6. Zhang W and Li G. Detection of multiple micro-drones via cadence velocity diagram analysis. Electron Lett 2018; 54: 441–443.
7. Christian AW and Cabell R. Initial investigation into the psychoacoustic properties of small unmanned aerial system noise. In: 23rd AIAA/CEAS aeroacoustics conference, Denver, Colorado, USA , 5-9 June 2017, p. 4051.
8. Furukawa K, Okutani K, Nagira K, et al. Noise correlation matrix estimation for improving sound source localization by multirotor uav. In: IEEE/RSJ international conference on intelligent robots and systems, Tokyo, Japan, 3–7 November 2013, pp. 3943–3948. Piscataway, NJ: IEEE.
9. Ohata T, Nakamura K, Mizumoto T, et al. Improvement in outdoor sound source detection using a quadrotor-embedded microphone array. In: *IEEE/RSJ international conference on intelligent robots and systems*, Chicago, IL, USA, 14-18 September 2014, pp. 1902–1907. Piscataway, NJ: IEEE.

10. Park S, Shin S, Kim Y, et al. Combination of radar and audio sensors for identification of rotor-type unmanned aerial vehicles (UAVS). In: *IEEE sensors*, Busan, South Korea, 1-4 November 2015, pp. 1–4. Piscataway, NJ: IEEE.

11. Misra P, Kumar AA, Mohapatra P, et al. Aerial drones with location-sensitive ears. *IEEE Commun Mag* 2018; 56: 154–160.

12. Harvey B and O’Young S. Acoustic detection of a fixed-wing UAV. *Drones* 2018; 2: 4.

13. Harvey B and O’Young S. A harmonic spectral beamformer for the enhanced localization of propeller-driven aircraft. *J Unmanned Vehicle Syst* 2019; 7: 156–174.

14. Ruiz-Espiria O, Martinez-Carranza J and Rascon C. Aira-uas: an evaluation corpus for audio processing in unmanned aerial system. In: *International conference on unmanned aircraft systems (ICUAS)*, Dallas, TX, USA, 12-15 June 2018, pp. 836-845. Piscataway, NJ: IEEE.

15. Hoshiba K, Washizaki K, Wakabayashi M, et al. Design of UAV-embedded microphone array system for sound source localization in outdoor environments. *Sensors* 2017; 17: 2535.

16. Kaleem Z and Rehmani MH. Amateur drone monitoring: state-of-the-art architectures, key enabling technologies, and future research directions. *IEEE Wireless Commun* 2018; 25: 150–159.

17. Jeon S, Shin JW, Lee YJ, et al. Empirical study of drone sound detection in real-life environment with deep neural networks. In: *25th European signal processing conference (EUSIPCO)*, Kos, Greece, 28 August - 2 September 2017, pp. 1858–1862. Piscataway, NJ: IEEE.

18. Kim BK, Kang HS and Park SO. Drone classification using convolutional neural networks with merged Doppler images. *IEEE Geosci Remote Sens Lett* 2017; 14: 38–42.

19. Grondin F, Létourneau D, Ferland F, et al. The many-ears open framework. *Autonem Robots* 2013; 34: 217–232.

20. Deng J, Dong W, Socher R, et al. Imagenet: a large-scale hierarchical image database. In: *IEEE conference on computer vision and pattern recognition*, Miami, FL, USA, 20-25 June 2009, pp. 248–255. Piscataway, NJ: IEEE.

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