UAV payload detection using deep learning algorithms

*Note: Sub-titles are not captured in Xplore and should not be used

Ilmun Ku
Artificial Intelligence Convergence
Hankuk University of Foreign Studies
Seoul, South Korea
mun90505@hufs.ac.kr

Seungyeon Roh
Computer Science and Engineering
Konkuk Univeristy
Seoul, South Korea
shtmddus99@konkuk.ac.kr

Gyoungyoung Kim
Computer Science and Engineering
Sunmoon University
Asan, South Korea
kky57389@sunmoon.ac.kr

Charles Taylor
Computer and Information Technology
Purdue University
West Lafayette, United States
taylo869@purdue.edu

Yaqin Wang
Computer and Information Technology
Purdue University
West Lafayette, United States
wang4070@purdue.edu

Eric Matson
Computer and Information Technology
Purdue University
West Lafayette, United States
ematson@purdue.edu

***Abstract*—In recent years, the technology behind Unmanned Aerial Vehicles (UAVs) has continually advanced. However, with these developments, malicious activities employing the use of UAVs have also been on the rise. Within this study, Deep Learning(DL) algorithms are utilized to detect and classify UAVs transporting payloads based on the sound they release. In order to exercise DL algorithms on a set of data, a sufficient amount of audio data is necessary to obtain a more reliable result. So UAV sound recordings have been collected alongside the use of data augmentation to secure a satisfactory sample size for testing purposes. Afterward, a feature-based classification was applied to the audio identifying each UAV's payload (or lack thereof). Lastly, both Convolutional Neural Network(CNN) and Recurrent Neural Network(RNN) are supplied with the final dataset and evaluated for their abilities to correctly categorize the absence or presence of a UAV's payload solely through audio.**

***Index Terms*—UAVs, payload detection, deep learning**

I. Introduction

Over the years as Unmanned Aerial Vehicle (UAV) technology has continued to advance drones have become much more accessible to the public. While UAV accessibility has been steadily growing, malicious activities have also become increasingly common. Especially UAVs with payloads that can easily be employed to endanger innocent civilians or Government Dignitaries with their airborne contents. These unknown packages could be potentially transporting harmful materials, explosives, etc. through public airspace.

In 2019, a Houthi drone was observed targeting senior Yemeni military officers and even exploded over a military base, killing six and wounding twelve [1]. In another instance, a UAV was seen carrying a bottle with unidentified contents and even landed on the Japanese Prime Minister's living quarters [2]. Additionally, in 2018 there was an infamous incident involving Venezuela's President Maduro who was attacked by two small drones carrying explosives [3].

In [4], Machine Learning (ML) algorithms were applied to detect UAVs carrying payloads on the basis of the sound that they emit. Feature extraction in addition to a combination of other features is proposed for preparing data for further analysis. In this study, however, these methods will be utilized in order to train and evaluate Deep Learning (DL) models as opposed to ML Algorithms. Although in the past the study had been limited to using only the ML algorithms for detecting UAVs due to the limited processing power available when using a MacBook Air with only 8GB of RAM.

In order to deploy DL algorithms to the classification model, extra resources and additional data are required. Furthermore, data collection has been conducted with a Raspberry Pi in previous research. Although there are more effective methods, with this study there is an aim to develop a method that is convenient and easy to use by anyone. In accordance with [5], data augmentation methods are pivotal for the smooth and continuous improvement of audio classification performance utilizing DL algorithms.

The rest of this paper is organized as follows. Section 2 inspects the current acoustic identification methods for UAV classification and the payload classification. Section III presents the methodology suggested for the feature extraction methods and two different DL models. Furthermore, data augmentation methods would be proposed. Section IV describes our experiments and results. Finally, Section V carries the conclusion and future works.

II. Literature Review

2.1 Audio UAVs classification

Identify applicable funding agency here. If none, delete this.

As the threat of malicious drones has continued to rise, UAVs have gradually become an important issue. Thus, various UAVs and their associated fields of research have evolved to be a field teeming with interesting and new research. Numerous detection and classification methods have been developed, including image classification, audio classification, utilization of radar, and so on. Since audio data is less affected by weather conditions and light when compared to image data, audio UAV classification can realistically be considered as a reliable methods to detect UAVs.

In the study [6], research about autonomous detection systems for Unmanned Aerial Vehicles (UAVs) based on acoustic signatures has been deeply examined. In order to train DL models, a large amount of drone audio data is essential in order to detect the differences between them. The researchers recorded a UAV audio clip with an iPhone, along with a clip consisting solely of background noise were both recorded and properly formatted. Splitting the separate clips into single-second segments for processing. This method of data processing is dependent upon transforming audio clips into spectrograms for further analysis.

Moreover, the real-life scenario can be realistically simulated by combining drone audio and real-life background noise. In the study, researchers made use of two different brands of UAVs for the purpose of data collection. Then samples of background noise, alongside samples of loaded and unloaded drones were recorded for testing purposes [6]. The collected data was then merged, and CNN, RNN, and Convolutional Recurrent Neural Network (CRNN) were utilized to evaluate neural network performance. As a result, CNN outperformed RNN and CRNN in accuracy and F1 score. Nonetheless, the performance of CRNN had a near insignificant difference in proportion to the performance of CNN.

In another study, research utilizing a combination of audio data and visual data has been applied [7]. Audio and image datasets were obtained through several sensors placed within the zone that the researchers had marked off for audio collection at various distances. Feature extraction was performed by applying the Mel-frequency Cepstral Coefficient (MFCC) descriptor and Image feature extraction had been implemented through the usage of AlexNet [8]. Afterward, a Support Vector Machine (SVM) algorithm was deployed for the analysis of the image and auditory datasets. Several different kernel methods were tested on the dataset. From the methods tested RBF/Gaussian kernel's analysis of the combined audio and imagery dataset produced the highest rate of accuracy. This accuracy rate was noticeably higher than the other methods: linear, and polynomial kernels.

According to auxiliary research, UAVs can be identified through ML algorithms based on UAV sound signals. Five features, including MFCC, chroma, tonnetz, contrast, and mel were analyzed for their effects on sound data analysis. To better train the ML model, a combination of those features is then exploited to produce more consistent results [8]. The ML model can produce an acceptable result with a relatively small dataset. In this study, features extracted from raw data are sent to the ML model as input, with that raw data the model is then trained, and subsequent model evaluation is then performed upon the data. Certain combinations of features were found to produce better performance than others, so among the combinations analyzed, MFCC and chroma were found to bring about good performance. Although there were certain limitations to this experiment as there were only two UAVs being utilized, and the dataset the researchers made use of was considerably small.

In agreement with [9], feature extraction from UAV audio datasets was carried out by applying MFCC. The researcher had even employed classification analysis by utilizing Hidden Markov Model (HMM). A collection of Twenty-four MFCCs and a collection of thirty-six MFCCs are procured in order to extract feature vectors. These specific amounts are chosen to demonstrate a correlation between the sample size and the reliability of the results. A classifier based on HMM identifies their two features. As the larger the sample size of sounds the model includes in training, the higher the average recognition rate will rise. Therefore, they have concluded that the method in this study is more effective in distinguishing UAV sounds regardless of whether or not they are located within a noisy environment.

In consonance with [10], UAV detection and direction-finding were conducted utilizing Bartlett and Capeon's methods of cross-correlation function. The datasets that were classified were acoustic emission that was collected via the microphone array. In spectral power density (SPD), the first 10 spectral peaks included 80% of the acoustic signal energy. For this reason, first spectral peaks are used when UAV detecting and recognizing. The dataset consists of numerous kinds of UAV acoustic signatures in different conditions. Their research resulted in achieving the best accuracy with analysis performed upon the MFCC of UAV acoustic emissions with up to 100m distance.

2.2 UAV payload detection

Since drones have become rather widely used around the globe, there are hazards that come with the rise in the popularity of UAVs, namely harmful gas or explosives. There are quite a few papers detailing practical and potential methods of detecting a UAV's payload. In one paper, UAV payload identification had been effectuated through the usage of sound signals and a classification system making use of a Machine learning (ML) algorithm[10]. Features were extracted from the audio dataset by using librosa since it is a feature-based classification. 5 features: mfcc, chroma, tonnetz, contrast, mel, and their different combinations were trained by the ML model, several different algorithms were thus utilized including SVM, Neural Network(NN), Gaussian Naive Bayes(GNB), and K-Nearest Neighbors (KNN). The ensuing comparison between the differing combinations and their rates of performance were then put into focus[10]. The dataset used for testing consisted of sound samples involving both loaded drones and unloaded drones. Their payload method involved hanging a

500ml bottle of water from the base of the drone. The result of the training was that the combination shows higher accuracy than individuals. Among individuals, MFCC and chroma show high performance. Albeit there are caveats, for instance, the small dataset, or the usage of only a single payload during audio collection.

In a separate study loaded and unloaded UAV detection utilized a solely image-based dataset and was conducted by YOLOv2 [11]. For data collection purposes, a DJI Phantom 2 drone with an attached 100g object for payload was used. To overcome the data shortage, drone images were collected from open source. The performance of loaded and unloaded UAV detection turned out to be 74.97% of mean-average precision. The paper proposed future work involving studying different types of object detectors such as Fast R-CNN, and Mask R-CNN that are believed to improve performance [11]. One study differentiated loaded and unloaded drone images by classifying them through the use of a residual convolutional neural network [12].

Similarly, another team of researchers took images of both the loaded UAV and unloaded UAV using the same DJI Phantom 2 drone, this time with a weighted object attached weighing 1000g. After data collection it was determined that more data was required, so data augmentation was performed on the set of images with different transformations for example rotation, resizing, the addition of Gaussian noise, cropping, and so on. A training of 96% was achieved through the application of ResNet-34. Throughout the research written upon this topic in the past there have been many different methods of detection tested, one study sought to detect hovering micro drones with differing sizes of loaded objects loaded via multistatic radar [13]. Detected Micro-Doppler signatures on radar were noticeably different when spotted without payload, this study tested both a 200g payload and a 500g payload while hovering. Even utilizing classification methods such as Naïve Bayes and discriminant analysis, both of them performed with above 90% accuracy. UAV sound classification was also used in conjunction with the LSTM-CNN architecture in order to classify drones [14]. The multiple labels of collected data were as follows: 'loaded', 'unloaded', and 'no drone'. The feature extraction is a data processing process involving the time domain, frequency domain, and MFCC. At last, training on the dataset was proceeded using a stacked bidirectional LSTM-CNN structure. 'Stacked BiLSTM-CNN structure' is a combined structure with BiLSTM layers and CNN layers. The result of the combined structure was an accuracy rating of 94.28%, so it is considered to be an efficient model.

References

[1] M. al Kibsi, "Houthi drone targets senior yemeni officers, kills five soldiers," Available at https://www.aljazeera.com/news/2019/1/10/houthi-drone-targets-senior-yemeni-officers-kills-five-soldiers (2022/05/18).

[2] W. Ripley, "Drone with radioactive material found on japanese prime minister's roof," Available at https://www.cnn.com/2015/04/22/asia/japan-prime-minister-rooftop-drone/index.html (2022/05/18).

[3] E. Kelly, "Venezuela drone attack: Here's what happened with nicolas maduro," Available at https://www.usatoday.com/story/news/politics/2018/08/06/venezuela-drone-attack-nicolas-maduro-assassination-attempt-what-happened/913096002 (2022/05/18).

[4] Y. Wang, F. E. Fagiani, K. E. Ho, and E. T. Matson, "A feature engineering focused system for acoustic uav payload detection," *the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022)*, pp. 470–475, 2022.

[5] S. Wei, S. Zou, F. Liao *et al.*, "A comparison on data augmentation methods based on deep learning for audio classification," *2019 2nd International Conference on Computer Information Science and Artificial Intelligence(CISAI 2019)*, pp. 1–8, 2020.

[6] S. Al-Emadi, A. Al-Ali, A. Mohammad, and A. Al-Ali, "Audio based drone detection and identification using deep learning," pp. 459–464, 2019.

[7] S. Jamil, M. Rahman, A. Ullah, S. Badnava, M. Forsat, S. S. Mirjavadi *et al.*, "Malicious uav detection using integrated audio and visual features for public safety applications," *Sensors*, vol. 20, no. 14, p. 3923, 2020.

[8] Y. Wang, F. E. Fagian, K. E. Ho, and E. T. Matson, "A feature engineering focused system for acoustic uav detection," pp. 125–130, 2021.

[9] L. Shi, I. Ahmad, Y. He, and K. Chang, "Hidden markov model based drone sound recognition using mfcc technique in practical noisy environments," *Journal of Communications and Networks*, vol. 20, no. 5, pp. 509–518, 2018.

[10] V. Kartashov, V. Oleynikov, I. Koryttsev, S. Sheiko, O. Zubkov, S. Babkin, and I. Selieznov, "Use of acoustic signature for detection, recognition and direction finding of small unmanned aerial vehicles," pp. 1–4, 2020.

[11] U. Seidaliyeva, M. Alduraibi, L. Ilipbayeva, and A. Almagambetov, "Detection of loaded and unloaded uav using deep neural network," pp. 490–494, 2020.

[12] U. Seidaliyeva, M. Alduraibi, L. Ilipbayeva, and N. Smailov, "Deep residual neural network-based classification of loaded and unloaded uav images," pp. 465–469, 2020.

[13] F. Fioranelli, M. Ritchie, H. Griffiths, and H. Borrion, "Classification of loaded/unloaded micro-drones using multistatic radar," *Electronics Letters*, vol. 51, no. 22, pp. 1813–1815, 2015.

[14] D. Utebayeva, M. Alduraibi, L. Ilipbayeva, and Y. Temirgaliyev, "Stacked bilstm-cnn for multiple label uav sound classification," pp. 470–474, 2020.