Anti-Drone System: A Visual-based Drone Detection using Neural Networks

Ann Janeth Garcia, Jae Min Lee, Dong Seong Kim
Networked Systems Laboratory, Department of IT Convergence Engineering,
Kumoh National Institute of Technology, Gumi, South Korea.
{ajg.garcia, ljmpaul, dskim}@kumoh.ac.kr

***Abstract*—A system that secures an area from trespassing drones that might bring threat is in demand these days since drones became easily-available to the public and it became easier to operate. This paper proposes an anti-drone system that uses visual sensing to detect drones. A Faster R-CNN (Region-based Convolutional Neural Network) with ResNet-101 (Residual Neural Network-101) networks are used in this paper using a dataset from the SafeShore project. The network's accuracy is 93.40% and it has successfully detected drones in the simulation that has been done.**

***Index Terms*—Drone detection, faster r-cnn, resnet-101, video surveillance.**

I. Introduction

Nowadays, drones became easily-available, inexpensive, and easy to operate making it as a tool in parcel deliveries, film shootings, disaster surveillance and sometimes, a hobby for some people. But aside these recreational activities and advantages that drones bring, others also use it for trespassing and spying private properties for drones with built-in cameras. Some are also used to carry explosives called *kamikaze* drones or smuggle illegal drugs across border lines. With these, many had created anti-drone systems that will secure a specific area against these dangerous situations brought by drones or trespassing drones or UAVs (Unmanned Aerial Vehicles) that puts great threats. These anti-drone systems can either detect a trespassing drone, classify the type of drone and some systems can also jam the signal that is used by the controller to communicate or control the drone.

Several surveillance technologies have been studied and proposed for drone detection or recognition, tracking, and/or jamming of signal using radar, audio, video, and radio frequency (RF) data [1]. In paper [2], the authors proposed an anti-drone system that uses radar technology for the detection of drones. This paper proposes a system that will track the drone and jam the signal that connects the user and the drone. The authors' aim is to provide long-term surveillance to the specific area that is being guarded and temporary protection for mass events.

On paper [3], the authors' used and Recurrent Neural Network (RNN), Convolutional Neural Network (CNN) and Convolutional Recurrent Neural Network (CRNN) in the proposed audio-based drone detection system. The authors' used a personal dataset that contains audio recorded samples of drones where background noises are present and for RF, Nguyen *et al.* in [5], they proposed two technical approaches in detecting drones. The first approach is waiting for a reflected signal which bounces off the drone after sending a continuous radio signal. The second approach is passive listening, the system will observe, extract, and analyze present signals using only a receiver. For vision-based drone detection, Aker *et al.* of [4] formed an artificial dataset by combining real images subtracted from its background and used a detection method based on CNN.

As can be seen CNN, R-CNN and C-RNN are readily used due to the success recorded in the use of deep learning in classification projects where accuracy is critical and of high importance such as in security [6], health [7] and automatic modulation classification [8], [9].

A. Surveillance Technologies

1) Radar: In paper [2], it is an anti-drone system that uses radar technology for the detection of drones. This paper proposes a system that will secure a predefined area against approaching unwanted drones by tracking and jamming the signal that is used by the controller to communicate with the drone. This is to prevent potential hazards that the unwanted drone may bring. The authors' goal is to provide long-term surveillance and provide temporary protection for mass events.
2) Audio: There are also papers that uses acoustic signals coming from the drones for detection. On paper [3], the authors' used different Deep Learning techniques namely Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Convolutional Recurrent Neural Network (CRNN). They based the performance of their system using their personal dataset that contains audio recorded samples of drones. Their way of drone-detecting has been employed where background noises are present.
3) Video: For video technology, Aker *et al.* of [4] created a large artificial dataset by merging real images, subtracted from background and used an end-to-end subject detection method that is based on CNN as their standalone approach in drone detection.
4) Radio Frequency (RF): Nguyen *et al.* in [5] proposed two technical approaches in detecting drones using radio frequency signals. The first approach is sending a radio signal continuously with a transmitter antenna and then

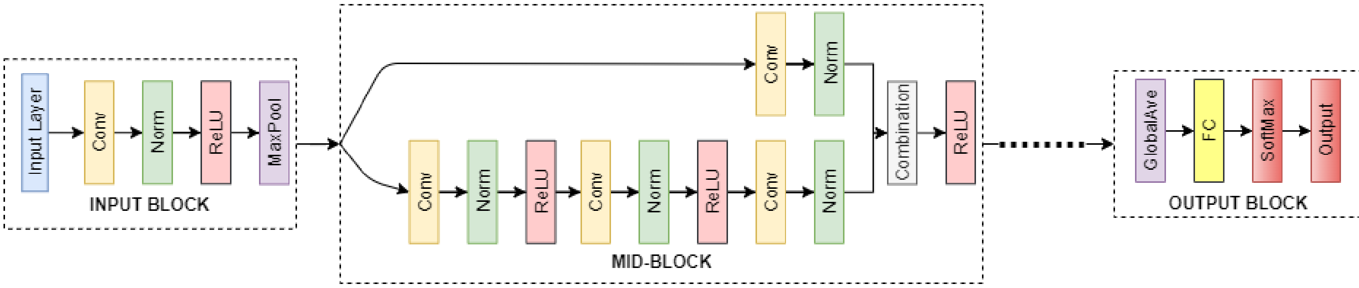

Fig. 1. Proposed network structure with 358 layers implemented in MATLAB using Deep Network designer for vision-based drone detection system.

wait for a reflected signal which bounces off the drone. The second approach is passive listening, the system will observe present signals and then extract and analyze it using only a receiver antenna. The challenges for this paper are the range became a problem over longer distances and the radio frequency (RF) signal is greatly used by Wi-Fi devices in urban environments making is difficult to scan and capture the signal instantly.

In this paper, the authors will focus on video surveillance technology that proposed a Faster R-CNN (Region-based Convolutional Neural Network) with ResNet-101 (Residual Neural Network-101). ResNet-101 is a pre-trained network available in MATLAB and is selected because of the different scales of drones present in the dataset used. Faster R-CNN is an improved Fast R-CNN that is a lot faster and has less processes involved.

This paper is composed with the following contents. For section II, it will discuss the background of the different surveillance technologies available with its limitations, advantages, and disadvantages. Section III is the datasets' description and Section IV is the systems' network structure. Section V is the results and the last section would be the conclusion and acknowledgment of this paper.

II. Datasets

SafeShore dataset that is available in http://safeshore.eu/dataset/ and is used for this vision-based drone-detection system. To obtain this dataset, the authors have to agree with the data usage provided in the link, and the password will be sent via email after it has been approved by the creators of the dataset.

This dataset is created to address the gap that has been seen by its sponsor, European Commission for the H2020 - SafeShore project. This dataset has been used for coastal border surveillance and to increase internal security by preventing cross-border crime such trafficking in human beings and the smuggling of drugs. Drones that are too small, has small radar cross sections which is difficult to be detected by radar systems. Some sensing techniques such as LIDAR, RADAR and the other surveillance technologies that has been discussed from the previous section are effective to use but could not give the satisfying level of accuracy that the European Commission aims for.

To create this dataset, a 3D LIDAR has been used with the aid of a passive acoustic sensor, passive radio detector and video analyzation to create a virtual drone shield across a large area in the coastal border for about a distance of 1500-1800 meters.

III. Network Structure

TABLE I
Comparison Table

Network	Description	Processing time
R-CNN	Uses selective search with specific features in extracting regions, includes multiple steps	40-50 s
Fast R-CNN	Whole image is processed in ConvNet, has issues in selective search	2 s
Faster R-CNN	Fast, reduced steps	0.02 s

```
(1) net = resnet101;
    lgraph = layerGraph(net);
```

```
(2) layersToRemove = {
        'fc1000'
        'prob'
        'ClassificationLayer_predictions'
        };

    lgraph = removeLayers(lgraph, layersToRemove);
```

```
(3) newLayers = [
        fullyConnectedLayer(numClassesPlusBackground, 'Name', 'rcnnFC')
        softmaxLayer('Name', 'rcnnSoftmax')
        classificationLayer('Name', 'rcnnClassification')
        ];

    lgraph = addLayers(lgraph, newLayers);

    lgraph = connectLayers(lgraph, 'pool5', 'rcnnFC');
```

Fig. 2. MATLAB codes used in the creation of the proposed network with its specific functions.

For visual sensing, the authors made use of Faster R-CNN (Region-based Convolutional Neural Network) with ResNet-101 (Residual Neural Network-101) as the base network for the detection of drones. ResNet-101 is selected because the dataset that is used contains different scales of drones and Faster R-CNN is used because it is the improved version of R-CNN and Fast R-CNN. Table 1 shows the differences of these three networks.

Fig. 1 shows the network structure designed using deep network designer in MATLAB that is 358 layers deep. ResNet-101 is a pre-trained network that is already available in the said program and to create Faster R-CNN, the authors then first created an R-CNN network, then modified the last three layers of that network to improve it to a Fast R-CNN, and then modified again the last three layers and added a feature extraction layer *res4b22_relu* which is a type of ReLU (Rectified Linear Unit) that is the feature extraction layer used in ResNet-101 and contains 1024 filters. After this comes the additional layer, RPN (Region Proposal Network) and the ROI (Regions of Interest) pooling layer then the classifier layer will give the output.

Fig. 2 shows the MATLAB codes used in creating the proposed network. *Code 1* that is boxed in the figure will reload the ResNet-101. It is a pre-trained in MATLAB which is the base network of the proposed network. With the loaded R-CNN, the network will be improved to Fast R-CNN and then Faster R-CNN by removing the last three layers using *Code 2* and then replacing it with new layers as shown in *Code 3*.

IV. Results

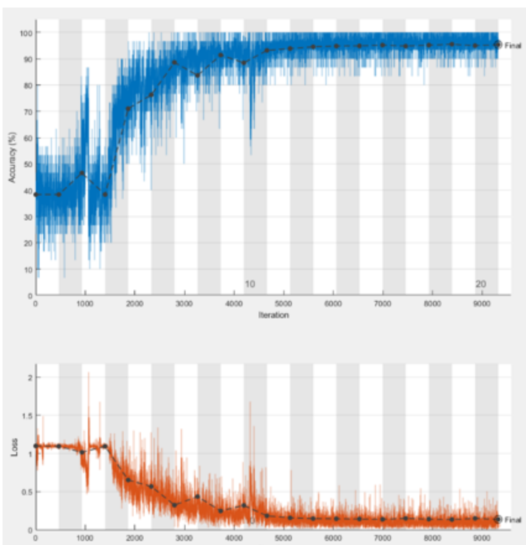

Fig. 3. 93.40% accuracy for the proposed network, Faster R-CNN with ResNet-101.

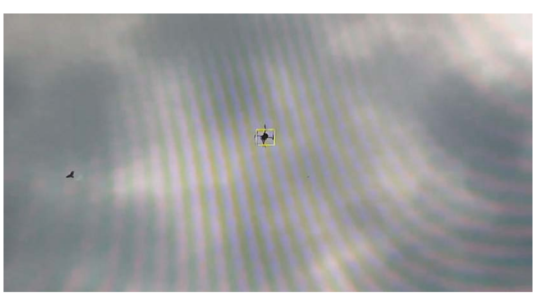

Fig. 4. Simulation sample using the SafeShore dataset with the drone detected in the yellow square.

Fig. 3 shows the training accuracy of Faster R-CNN with a base network of ResNet-101 is 93.40% and is simulated as shown in Fig. 4. As can be seen in Fig. 4, the drone is detected in the yellow square, not confusing it with the bird that is also present in the frame shown.

V. Conclusion

The proposed network for visual sensing shows a high accuracy of 93.40% and has successfully detected the drone in the test video simulation from the SafeShore dataset without confusing the drone with a bird that is also present in the said data. As shown in the frame, the drone and bird are shown in the same frame, but the drone is the one that is only detected by the proposed drone detection system.

For future works, the authors would like to recommend a combination of other neural networks together with Faster R-CNN or another neural network and observe which network combination will provide a much better accuracy and also combining vision-based drone detection system with drone acoustic signal detection using a multimodal neural network that can also classify drones using own dataset.

Acknowledgement

This work was supported by Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2018R1A6A1A03024003).

References

[1] X. Shi, C. Yang, W. Xie, C. Liang, Z. Shi and J. Chen, "Anti-Drone System with Multiple Surveillance Technologies: Architecture, Implementation, and Challenges," in IEEE Communications Magazine, vol. 56, no. 4, pp. 68-74, April 2018, doi: 10.1109/MCOM.2018.1700430.

[2] T. Multerer et al., "Low-cost jamming system against small drones using a 3D MIMO radar based tracking," 2017 European Radar Conference (EURAD), Nuremberg, 2017, pp. 299-302, doi: 10.23919/EURAD.2017.8249206.

[3] S. Al-Emadi, A. Al-Ali, A. Mohammad and A. Al-Ali, "Audio Based Drone Detection and Identification using Deep Learning," 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), Tangier, Morocco, 2019, pp. 459-464, doi: 10.1109/IWCMC.2019.8766732.

[4] C. Aker and S. Kalkan, "Using deep networks for drone detection," 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, 2017, pp. 1-6, doi: 10.1109/AVSS.2017.8078539.

[5] P. Nguyen, M. Ravindranatha, A. Nguyen, R. Han, T. Vu. (2016). Investigating Cost-effective RF-based Detection of Drones. In Proceedings of the 2nd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use (pp. 17-22). ACM. 17-22. 10.1145/2935620.2935632.

[6] A. Aouto, T. Huynh-The, J. Lee and D. Kim, "Pose-Based Identification Using Deep Learning for Military Surveillance Systems," 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea (South), 2019, pp. 626-629, doi: 10.1109/ICTC46691.2019.8939983.

[7] T. Huynh-The, C. Hua, N. A. Tu and D. Kim, "Physical Activity Recognition with Statistical-Deep Fusion Model using Multiple Sensory Data for Smart Health," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2020.3013272.

[8] A. P. Hermawan, R. R. Ginanjar, D.S. Kim and J. M. Lee "CNN-Based automatic modulation classification for beyond 5G communication", IEER Communication letters vol 24, no 5, pp 2020

[9] T. Huynh-The, C. Hua, Q. Pham and D. Kim, "MCNet: An Efficient CNN Architecture for Robust Automatic Modulation Classification," in IEEE Communications Letters, vol. 24, no. 4, pp. 811-815, April 2020, doi: 10.1109/LCOMM.2020.2968030.