DopeNet: Range–Doppler Radar-based UAV Detection Using Convolutional Neural Network

Ali Aouto, Taesoo Jun, Jae-Min Lee and Dong-Seong Kim,
Networked Systems Lab., Dept. of IT Convergence Engineering,
Kumoh National Institute of Technology, Gumi, South Korea.
Email: {ali.aouto,taesoo.jun,ljmpaul,dskim}@kumoh.ac.kr

***Abstract*—The proliferation of commercial unmanned aerial vehicles (UAVs) of various sizes and shapes, equipped with cameras and even signal sabotage devices, has raised concerns regarding privacy and safety. Some websites even offer weapons that can be attached to drones, adding to the security threats. As a result, researchers have been motivated to develop an intelligent system that can be integrated into surveillance systems to classify unauthorized UAVs that are flying in restricted areas. In this paper, we propose a convolutional neural network (CNN) for UAV detection based on Radar image data. The Radar system, called Real Doppler RAD-DAR (Radar with Digital Array Receiver), is a range-doppler system developed by the Microwave and Radar Group. We construct and analyze the CNN by adjusting its hyper-parameters using the RAD-DAR dataset. Our simulation results show that setting the number of filters to 32 results in the best time-wise accuracy. The network achieved an accuracy of 97.63%, which is higher than other benchmark image classifiers. Additionally, we conducted an ablation study to investigate and validate the contribution of each part of the neural network.**

***Index Terms*—Convolutional neural network, Image classification, Range-Doppler radar, UAV.**

I. Introduction

The widespread use of Unmanned Air Vehicles (UAVs), commonly known as drones, has revolutionized the way we think about aerial surveillance, photography, and even package delivery. These autonomous flying machines have gained immense popularity in recent years, with many manufacturers creating UAVs for various applications.

While UAVs were initially developed for military use, their use has now expanded to a broad range of applications. In the past, the Radioplane Company's design for drones was purchased by the US Army during World War II, with nearly 15,000 drones produced. However, they were the size of regular fighter planes, making their production a challenge. Advancements in technology have now made it possible to create drones as small as the size of one's pocket.

The diversity of applications for UAVs is vast. They are widely used in photography, allowing for stunning aerial shots that were once impossible. Search and rescue operations benefit significantly from UAVs, as they can reach remote areas where humans cannot. UAVs are also useful in object detection, making it possible to detect changes in a vast area quickly. They are also widely used in surveillance, providing security and safety in areas where human surveillance is challenging.

One of the most exciting areas where UAVs are making a significant impact is package delivery. Many companies have started using UAVs to deliver packages, which is faster, safer, and more cost-effective than traditional delivery methods. With advancements in AI technology and autonomous flying capabilities, UAVs have the potential to transform the logistics industry as we know it.

The huge number of expected UAVs to be flying all around the cities and traveling at high-speed on highways requires an intelligent system that can distinguish between the authorized UAVs and the unauthorized ones. Therefore, for the previous past couple of years many researchers have been trying to propose such a system. Mainly, most of UAVs detection research has been done on RGB image using convolutional neural networks (CNN), where the proposed detection systems are described as collecting RGB images from a camera and then use those images as an input to a CNN that extracts the features which are used to distinguish those UAVs. Another way of detecting UAVs is to use machine learning algorithms to detect UAVs through the sound produced by fans blades. The third common way of detection is to detect the radio frequency between a UAV and its controller. Lastly, is the use of radar to detect UAVs [1].

Unfortunately, all the previously mentioned methods face some serious issues. Vision-based approaches are challenged by the distance between the object and the camera, when the object is far from the camera it is almost impossible to catch the unique features of the object to be detected, while considering that the reliability of this approach is highly affected by the background clarity. Acoustic-based approaches face more major issues such as each UAV from the same type has a slightly different sound, also the sound produced by the blades is detected only if there is no noise in the surrounding area. Such an approach will be unusable if we consider the number of UAVs that will be flying around us in the future. Detecting the RF signal is one of the most reliable ways [2]. But we cannot neglect the reality that there are many UAVs that are able to reach its destination without a connection with the controller.

On the other hand, Radar-based approaches solve most of the mentioned disadvantages. Radar can cover a long area and can provide an early detection of the threat. Also, radar approaches are not affected by the absence of a controller. Such an approach has not been heavily researched since the absence of radar UAV datasets. In 2020, Roldan *et al.* [3]

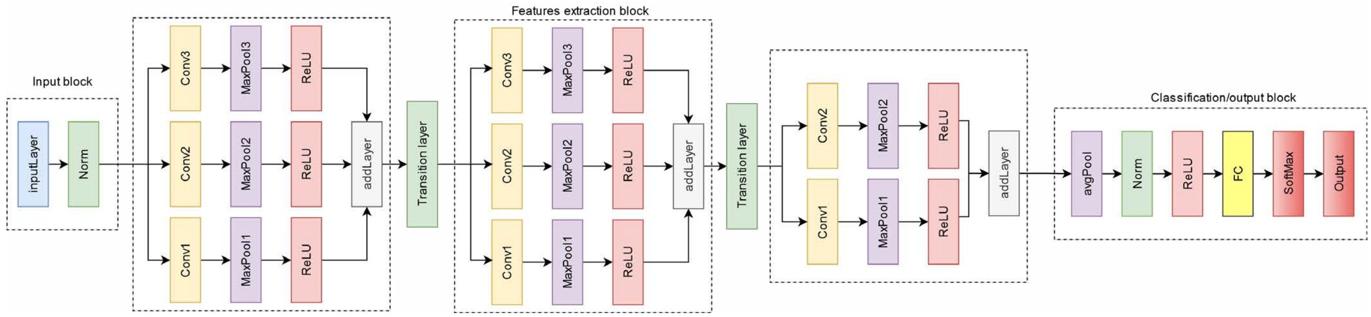

Fig. 1. Proposed network architecture DopeNet.

published a persistent range–Doppler radar dataset that will help the research on this field. It contains more than 17,000 data samples of drones, humans and cars. The mentioned dataset specifications and data collection environment are to be discussed later on.

This paper proposes detection of UAVs using a CNN that behaves as image classifier named DopeNet. where the radar data will be fed to the network after transforming it to a basic RGB image. Then, the data will go through the feature extraction section of the network. After features extraction, the data is to go through the classification section of the network to distinguish between the three different classes provided into the dataset.

The rest of the paper is organized as follows; section II will show some of the related works. Section III will explain the proposed network in this paper. The discussion of the used dataset and its specifications is to be discussed in section IV. Then, section V shows the simulation results that has been acquired from this research and a brief discussion of it. Finally, section VI will conclude the paper and will discuss some future works.

II. Related Works

The research field of UAV detection has been slowing down for the past couple of years. The reason for such decrease is that most of UAVs datasets either have some assumptions that makes it unrealistic, or the dataset does not contain enough samples to obtain a high quality training. It is well known that to obtain a high training quality, a well collected and labeled dataset should be used to train the network and the more data that exists the better reliability that system is considered to have. Thus, we could say that the lack of open to public UAV datasets limits the research in this field. But there is still some state-of-arts that can not be neglected where we will try to cover some in this section.

Mendis *et al.* utilized a deep belief network (DBN) to differentiate between three different micro-drone types. The data was obtained via S-band CW radar and pre-processed to extract the spectral correlation function (SFC), which is the Fourier transform of the correlation function. The DBN employed the resulting SFC as its input feature, and although the accuracy levels were impacted by the signal-to-noise ratio, the system achieved an accuracy rate of approximately 97% for all three classes. In the field of autonomous driving, deep learning techniques have been employed for the classification of ground targets to avert accidents in autonomous driving systems [4].

Zhao *et al.* [5] were able to achieve a comprehensive analysis of the radar point clouds and extract meaningful information about the vehicles present in the scene. This approach allowed for the efficient and accurate identification of vehicle targets in radar data. By combining clustering, feature extraction, and classification techniques. While Passafiume *et al.* [6] has explored the dependability of FMCW radar imagery in categorizing a UAS based on the number of motors it possesses, while also eliminating the influence of rotation speed on the classification process.

Overall, the related works in UAV detection using radar Doppler demonstrate the potential of this approach in providing accurate and reliable UAV detection, and the application of advanced techniques such as machine learning and multi-radar systems can further enhance the detection capabilities.

III. Proposed Model

Deep learning is one of the best methods that has been used for object detection until the day of writing this paper [7]. And inspired by some works that process raw data and transform it into images to use into CNN image classifiers. This paper proposes a convolutional neural network named DopeNet which is an image classifier for UAV detection using the output of range-Doppler radar.

At first, the raw data collected by the radar is shaped into an array, each sample of the dataset we are using in this paper has an array size of 11×61. The raw data is transformed into simple RGB images which will make them look like a heat map, then those images will go through the first stage of the CNN that is the input block which consists of normalization layer, convolution layer and a ReLU rectifier. Secondly, the output of the input block goes to features extraction block which consists of multiple parallel convolutions, each one of those convolutions has a different filter size which are 3×3, 5×5, 7×7. The features extractions block consists of two similar blocks and it squeezes at the third block. Lastly, the extracted features will be sent to the classification block which classifies using Fully Connected layer and a softmax layer. Fig. 1 shows the proposed network architecture. Table I shows the detailed hyperparametes that has been chosen after varying and testing diverse number of filters and examine the model through ablation study.

TABLE I
DETAILED PROPOSED NETWORK ARCHITECTURE OF DOPENET

Layer	Output	Parameters
Conv1	11x61	3x3 conv, stride1
Conv2		5x5 conv, stride 1
Conv3		7x7 conv, stride 1
maxPool1	6x31	3x3 max Pooling, stride 2
maxPool2		5x5 max Pooling, stride 2
maxPool3		7x7 max Pooling, stride 2
avePool	3x16	5x5 average Pooling, stride 2
FC	3	Fully Connected Layer

In neural networks, one of the approaches to increase the velocity of learning process is batch normalization, assuming that the input data is proposed as x and the mini-batch size is given as B, then the normalization output is calculated by

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (1)$$

where the mini-batch mean μ_B and the mini-batch variance σ_B^2 formulas are defined as

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i \quad (2)$$

and

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2 \quad (3)$$

Although, batch normalization is considered to increase the velocity of training networks, but in real-time systems, it results into increasing the prediction process time-consumption. The proposed network is much smaller than the benchmark networks and such a sacrifice to accelerate the training can be neglected since the proposed system has lower time-consumption when it is compared with existing models.

Following the batch normalization layer is a convolutional layer that prepares the data in the input stage to be sent to features extractions. This convolutional layer is set to 1×1. For subsequent blocks a K channels are created, the output matrix of y is calculated while considering that x is the convoluted input with kernel matrix h, the 2-dimensional convolution can be described by the equation:

$$y(i,j) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} h(m,n)\cdot x(i-m, j-n) \quad (4)$$

One of the most crucial components of the neural network (NN) is activation functions. The main purpose of using activation functions is to decide which data is to be activated and which to neglect. Since the start of NN there have been many activation functions that has been proposed and many has been used for a while and disappeared. However, Rectified Linear Unit (ReLU) is one of the most popular and mostly used in image classification [**?**]. RelU is presented by the formula $f(x) = x^+ = max(0, x)$ where x is the input data of a single neuron, Therefore, if x is positive the data will be activated, and the negative data is neglected. The major reason for ReLU popularity is that it has an efficient fast convergence.

IV. DATASET DESCRIPTION AND SIMULATION ENVIRONMENT

The Real Doppler RAD-DAR Dataset is a highly accurate and thoroughly annotated database consisting of over 17,000 examples of cars, drones, and people captured in real-world outdoor scenarios. This innovative dataset was developed by Roldan et al [3]. using a fixed radar system, the Real Doppler RAD-DAR (Radar-Digital Array Receiver), operating at a frequency band centered at 8.75 GHz with a maximum bandwidth of 500 MHz FMCW. After digital signal processing, a 4092x512 matrix is obtained for each scene, with distance cells in rows and Doppler frequencies in columns, expressed in dBm. These matrices are reduced to 11x61 matrices. The detector utilizes CFAR techniques. The dataset contains 5720 car samples, 5065 drone samples, and 6700 human samples.

The neural network employs the stochastic gradient descent with momentum (SDGM) optimizer, which is set to 20 epochs with a starting learning rate of 0.001 that drops by 90% every 5 epochs. The mini-batch size is set at 64. The analysis is performed on a low-cost benchmark system featuring an Intel(R) Core(TM) i7-8700K CPU @3.70GHz, NVIDIA GeForce RTX 2070 GPU, and 32GB RAM. The dataset is divided into 80% for training and 20% for validation.

V. SIMULATION RESULTS

This section will discuss three subsections, the first part will illustrate varying the amount of filters to see which one is going to result in the highest accuracy. The second part discusses an ablation study that has been done to validate the efficiency of each part of the network. The second part will discuss and compare the results with existing benchmark classifiers. We evaluate our proposed system in terms of accuracy and time consumption.

A. Different Number of Filters

Varying the amount of filters is not common in research papers, since it is well known that increasing the number of filters will result in higher accuracy. But, that is the case when there is big number of classes. Since this dataset provides only three different classes, finding the lowest appropriate number of filters will result in reducing computational complexity, which also reduces time-consumption. In this study we vary the filters amount between {8,16,24,32,64,128}. The proposed model results in the highest accuracy at 32. Table II lists the results for each case. The best case was highlighted by bold font and the second best was highlighted by an underline.

TABLE II
PROPOSED DOPENET WHILE VARYING THE NUMBER OF FILTERS

No. of filters	Accuracy (%)
8	96.13
16	96.87
24	97.58
32	**98.63**
<u>64</u>	<u>98.65</u>
128	98.66

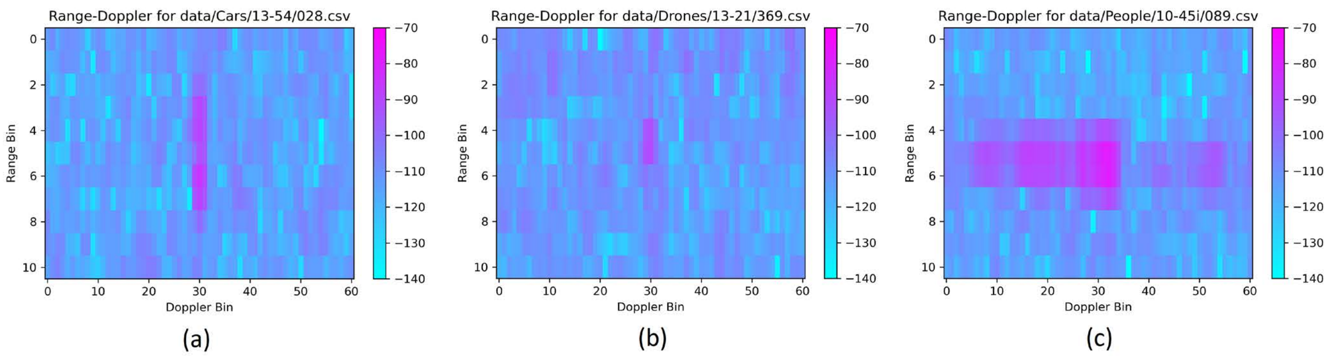

Fig. 2. Visualized samples of the RDRD dataset, the original samples are matrices of 11×61 of radar raw data, for transforming it to RGB images it was assigned a color according to the value of each cell, (a) sample of cars. (b) sample of drones. (c) sample of people.

As Table II shows, when the amount of filters were low the system did not show its best performance. Also, after the number of filters 32, the system performance started to converge around the same values. Although, a high number of filters might perform better on high-quality simulation environment. But, in this study we are considering a low budget environment performance.

B. Ablation Study

To prove that each part of any CNN has to be there for it to show its high performance, many papers consider ablation study where the network is trained and tested after removing some parts of it. If a part is removed and the network still shows similar accuracy, then removing that part would be better to decease the computational complexity and time consumption. Also, performing ablation study makes it easier to understand the effectiveness of each part of the network. DopeNet features extraction stage contains three major blocks, first, we remove the second block. In scenario (b), we remove the third block. Lastly we manipulate the parallel layers to see how it will perform with removing 3×3 [conv1] . Fig. 3 illustrates a visualization of each scenario. Table III shows the accuracy of the different scenarios.

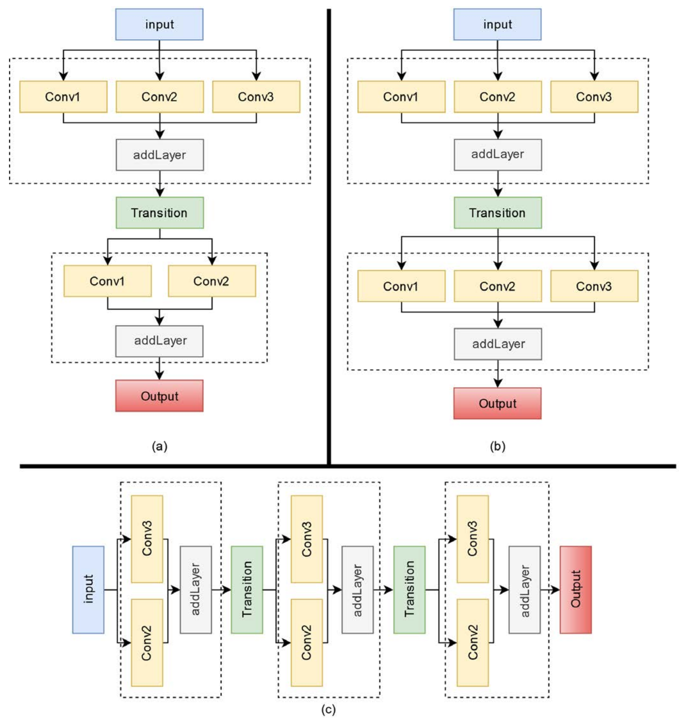

Fig. 3. Three different cases of ablation study application where in (a) the depth decreased, (b) the squeeze block removed, and in (c) one of the parallel convolutions has been removed [conv1].

TABLE III
ABLATION STUDY SCENARIO'S ACCURACY

scenario	Accuracy (%)
(a)	92.61
(b)	91.7
(c)	84.35
DopeNet	98.63

Looking at the results of Table III we could see, in scenario (a), the disappearance of the second convolutional block has reduced the accuracy by 5.02%. On the other hand, the accuracy significantly dropped in the (c) scenario. Such results show that the third parallel convolution with filter size 3×3 is contributing highly to the network accuracy. Therefore, ablation study results show that each part of the network is highly contributing to the final accuracy of DopeNet.

C. Comparison with Existing Models

In this section we show the results of our DopeNet in comparison with existing models. Although, these models have not been built to be used for the purpose of detecting UAVs using range doppler radar, but these are a well-known image classifiers. In our approach, we convert the doppler data to images. Then feed it to an image classifier. Also, to our knowledge, there is no similar approach to our proposal. That is the main reason we had to resolve to use existing image classifiers.

While choosing the models to compare with our model, we considered only small CNNs which will have lower computational complexity just to be fair. We chose ResNet-18 [8], VGG-16 [9], SqueezeNet [10] and DarkNet-19 [11]. In this study we will not be considering the learnable parameters since our network has a low amount of 411,472 which is lower than half of SqueezNet's parameters. Instead, this study will depend on time consumption. Table IV Shows accuracy results from RAD-DAR dataset on the mentioned models. Also, we include our previous work CNN-32DC [12] to compare with our DopeNet.

TABLE IV
COMPARISON ON DIFFERENT MODELS

Network	Accuracy (%)	Time Consumption (ms)
VGG-16	93.87	6.9
SqueezeNet	95.34	6.3
ResNet-18	96.74	5.8
CNN-32DC	96.86	**3.2**
DarkNet-19	97.12	8.65
DopeNet	**98.63**	4.7

DopeNet has shown the highest accuracy in comparison with existing models. Although, our previous work has shown lower time consumption which is expected since CNN-32DC has less learnable parameters, but with considering the sacrifice of 1.5ms we gained 2.23% extra accuracy. DarkNet-19 has shown the second-best results regarding accuracy. On the other hand, it has the highest computational time because of its high number of learnable parameters. Fig 4 shows the results of comparison with existing models in a combo graph of bar for accuracy and line graph for time consumption. VGG-16 has performed poorly on DAR-RAD dataset where both its accuracy and time consumption were unsatisfactory.

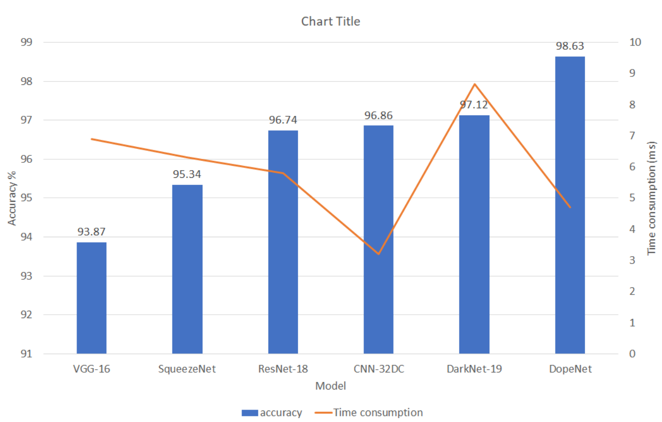

Fig. 4. Comparison of DopeNet with other existing models, DopeNet shows higher performance with second lowest time consumption.

VI. Conclusion and Future Works

In this study, we have introduced a novel neural network model and evaluated its performance on the RDRD dataset. Our experiments demonstrate that the proposed model achieves high classification accuracy while consuming minimal time. We compared the proposed model with existing models and found that our model with 32 convolution filters achieves the highest accuracy with respect to computational complexity. We conducted an ablation study and found that each component of the proposed model contributes to achieving the maximum accuracy. Our results show that despite having fewer learnable parameters, our proposed model outperforms state-of-the-art models in terms of accuracy and prediction time. Moving forward, we plan to extend our model to multi-modal networks for enhanced drone detection and classification in surveillance systems, which will include the other mentioned approaches such as images and RF signals to improve the overall system efficiency. Also, we would like to investigate the probability of acquiring higher accuracy if transformers were to be applied in it.

Acknowledgment

This research work was supported by Priority Research Centers Program through NRF funded by MEST(2018R1A6A1A03024003), the Grand Information Technology Research Center support program (IITP-2023-2020-0-01612) supervised by the IITP by MSIT, Korea. and, by project for Industry-University-Research Institute platform cooperation R&D funded Korea Ministry of SMEs and Startups in 2022.(S3311338).

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," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 68–74, 2018.

[2] T. Huynh-The, Q.-V. Pham, T.-V. Nguyen, D. B. Da Costa, and D.-S. Kim, "Rf-uavnet: High-performance convolutional network for rf-based drone surveillance systems," *IEEE Access*, vol. 10, pp. 49 696–49 707, 2022.

[3] I. Roldan, C. R. del Blanco, A. Duque de Quevedo, F. Ibanez Urzaiz, J. Gismero Menoyo, A. Asensio Lopez, D. Berjon, F. Jaureguizar, and N. Garcia, "DopplerNet: a convolutional neural network for recognising targets in real scenarios using a persistent range–doppler radar," *IET Radar, Sonar & Navigation*, vol. 14, no. 4, pp. 593–600, 2020.

[4] G. J. Mendis, T. Randeny, J. Wei, and A. Madanayake, "Deep learning based doppler radar for micro uas detection and classification," in *MILCOM 2016-2016 IEEE Military Communications Conference*. IEEE, 2016, pp. 924–929.

[5] Z. Zhao, Y. Song, F. Cui, J. Zhu, C. Song, Z. Xu, and K. Ding, "Point cloud features-based kernel svm for human-vehicle classification in millimeter wave radar," *IEEE Access*, vol. 8, pp. 26 012–26 021, 2020.

[6] M. Passafiume, N. Rojhani, G. Collodi, and A. Cidronali, "Modeling small uav micro-doppler signature using millimeter-wave fmcw radar," *Electronics*, vol. 10, no. 6, p. 747, 2021.

[7] B. Major, D. Fontijne, A. Ansari, R. T. Sukhavasi, R. Gowaikar, M. Hamilton, S. Lee, S. Grzechnik, and S. Subramanian, "Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 924–932.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[10] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[11] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[12] A. J. Garcia, A. Aouto, J.-M. Lee, and D.-S. Kim, "CNN-32DC: An improved radar-based drone recognition system based on Convolutional Neural Network," *ICT Express*, vol. 8, no. 4, pp. 606–610, 2022.