IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2021.DOI

RF-UAVNet: High-Performance Convolutional Network for RF-based Drone Surveillance Systems

THIEN HUYNH-THE[1], (Member, IEEE), QUOC-VIET PHAM[2], (Member, IEEE), TOAN-VAN NGUYEN[3], (Member, IEEE), DANIEL BENEVIDES DA COSTA[4], (Senior Member, IEEE), and DONG-SEONG KIM[1], (Senior Member, IEEE)

[1]Department of IT Convergence, Kumoh National Institute of Technology, Gumi, Gyeongsangbuk-do 39177, Republic of Korea (email: thienht@kumoh.ac.kr, dskim@kumoh.ac.kr).
[1]Korean Southeast Center for the 4th Industrial Revolution Leader Education, Pusan National University, Busan 46241, Republic of Korea (email: vietpq@pusan.ac.kr).
[3]Department of Electrical and Computer Engineering, Utah State University, UT 84321, USA (email: van.nguyen@usu.edu).
[4]AI & Telecom Research Center, Technology Innovation Institute, 9639 Masdar City, Abu Dhabi, United Arab Emirates and also with the Future Technology Research Center, National Yunlin University of Science and Technology, Douliou, Yunlin 64002, Taiwan, R.O.C. (e-mail: danielbcosta@ieee.org).

Corresponding author: Thien Huynh-The (e-mail: thienht@kumoh.ac.kr).

This research was financially supported by National Research Foundation of Korea (NRF) through Creativity Challenge Research-based Project (2019R1I1A1A01063781), and in part by National Research Foundation of Korea (NRF), Institute for Information & Communications Technology Planning & Evaluation (IITP) funded by the Ministry of Education, Science and Technology under Grant 2018R1A6A1A03024003 and Grant IITP-2020-2020-0-01612.

ABSTRACT In recent years, the increasing popularity of unmanned aerial vehicles (UAVs) has arisen from the emergence of cutting-edge technologies deployed in small and low-cost devices. With the great capability of friendly uses and wide applications for multiple purposes, amateur drones can be piloted to effortlessly access any geographical area. This poses some difficulties in monitoring and managing drones that may invade private or limited-access areas. In this paper, we propose a radio-frequency (RF)-based surveillance solution to effectively detect and classify drones, and recognize operations by leveraging a high-performance convolutional neural network. The proposed network, namely RF-UAVNet, is specified with grouped one-dimensional convolution to significantly reduce the network size and computational cost. Besides, a novel structure of multi-level skip-connection, for the preservation of gradient flow, incorporating multi-level pooling, for the collection of informative deep features, is proposed to achieve high accuracy via learning efficiency improvement. In the experiments, RF-UAVNet yields the accuracy of 99.85% for drone detection, 98.53% for drone classification, and 95.33% for operation mode recognition, numbers which outperform the current state-of-the-art deep learning-based methods on DroneRF, a publicly available dataset for RF-based drone surveillance systems.

INDEX TERMS Convolutional neural network, deep learning, drone detection, drone classification, drone surveillance.

I. INTRODUCTION

Recently, small unmanned aerial vehicles (UAVs), also known as drones, have received explosive interest unprecedentedly for numerous applications in diverse domains, which range from aerial photography to disaster management, agriculture, and communications [1], [2]. However, this rapid development of amateur drones poses many critical threats to public security and personal privacy [3], where drones are controlled to enter restricted zones intentionally without authentication. Drone surveillance has become a potential solution to effectively cope with the above-mentioned problem [4], in which drone detection and classification play an important role in many advanced anti-drone systems. Several drone detection approaches based on radar, audio, video, and radio-frequency (RF) [5] have been developed in surveillance systems, however, they have suffered from different limitations: ineffectiveness of radar-based approaches for small drones detection, high sensitivity to noise and limited working range of audio-based approaches, and weather constraints with object occlusion of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3172787, IEEE Access

video-based approaches [6]. Compared with the others, the RF-based approach is more favorable to drone detection and classification thanks to substantial reliability and outstanding performance [7] besides easy implementation and long operation range (approximately 1400 ft) [8].

Several RF-based drone detection methods have been introduced with feature extraction and machine learning (ML) algorithms. In [9], a reliable detection method was introduced to warn the presence of drones by analyzing the fast Fourier transform (FFT) of drone-controller RF signals. Shoufan *et al.* [10] learned the random forest classifier with time-domain features to improve the accuracy of drone pilot identification. In [11], a regular ML framework with frequency-domain features and neural networks was studied to detect drones via WiFi interferences. Deep learning (DL) [12] with recurrent neural network (RNN) and convolutional neural network (CNN) [13]–[15] has been exploited to process RF signals in wireless communications and improve the performance of RF-based drone surveillance systems. The superiority of DL with great capacity for learning relevant features at multi-scale RF signal resolutions was articulated in [16] when compared with various traditional ML algorithms for drone detection. However, the primitive architecturein [16] suffered from high implementation complexity and low accuracy. Recently, some RNN and CNN architectures have been designed with cascade structures to process time-series data in different classification tasks, such as wearable-based physical activity recognition [17] and motor fault diagnosis [18].

In this work, we present an efficient drone surveillance method using supervised learning with deep architectures to monitor and manage known drones. Especially, we propose RF-UAVNet, a novel deep CNN architecture for three tasks: drone detection, drone classification, and operation mode recognition. We design the deep network architecture with multiple convolutional layers, where each layer is specified by several one-dimensional (1D) asymmetric filters of sizes 1×3 and 1×5 to extract local features of raw 1D signal at multi-scale resolutions (from coarse to fine). We leverage grouped convolution instead of standard convolution to significantly reduce the number of network parameters and the processing cost. Moreover, a multi-level skip-connection is cleverly designed in association with a multi-level pooling to regulate the gradient flow and collect more global underlying features, which in turn increase the overall accuracy of surveillance systems. We evaluate the performance of RF-UAVNet on a practical DroneRF dataset [19] and investigate with various hyper-parameters and architecture configurations to corroborate the efficiency of RF-UAVNet for different drone surveillance tasks. The proposed method is compared with state-of-the-art approaches with the same condition, showing that RF-UAVNet is superior to other deep networks. The primary contributions of this paper can be summarized as follows:

- We propose RF-UAVNet to separately learn three common tasks in drone surveillance systems, in which the network architecture is specified by asymmetric filters to calculate the local correlations of RF signals. We deploy grouped convolution to reduce the number of trainable parameters and computational cost. We further design an advanced structure with multi-level skip-connection and multi-level pooling to improve the accuracy of three tasks.
- The proposed network achieves the overall accuracy of 99.85% for drone detection, 98.53% for drone classification, and 95.33% for operation recognition on the DroneRF dataset while reducing the number of parameters by around 78% and the computational cost by around 82%. In the method comparison, RF-UAVNet outperforms several existing DL-based models in terms of accuracy for different drone surveillance tasks.

The remainder of this paper is organized as follows: Section II briefly surveys the available literature on drone detection methods. In Section III, we present the proposed RF-based method for drone detection, drone classification, and operation recognition, wherein a high-performance CNN is introduced to learn RF signals. The diversified experiments with numerical results are provided in Section IV. Finally, Section V concludes this paper.

II. STATE-OF-THE-ART TECHNIQUES

A. RADAR-BASED TECHNIQUE

The principle of radar-based methods is electromagnetic backscattering to identify an aerial object by measuring the radar cross-section (RCS) signature. Compared with aircraft, drones are more challenging for RCS-based detection because of their small size and low-conductivity materials that induce low RCS. In [20], the micro-Doppler signature with time-domain analysis performed better than the Doppler-shift signature to increase the accuracy of clutter/target discrimination. The micro-Doppler signature was also adopted for drone classification in a non-cooperative drone surveillance system [21] with support vector machine (SVM) and decision tree (DT) classifiers. Recently, a passive radar with long-term evolution (LTE) downlink signals [22] was exploited to detect small UAVs. Geng *et al.* [23] utilized frequency-modulated continuous waveform to specify an effective LTE downlink signal feature, which in turn increases the accuracy of drone detection.

B. AUDIO-BASED TECHNIQUE

This technique can detect, classify, and localize drones based on the sounds of the engine and high-speed rotating propellers using acoustic sensors (e.g., microphones). In [24], a data-driven method was proposed to detect drones in a heavy rain environment, in which a regular ML workflow was developed to extract FFT-based frequency features and learn the classification model with SVM. Anwar *et al.* [25] improved the performance of audio-based drone detection systems with linear predictive cepstral coefficients and Mel-frequency cepstral coefficients. In [26], Uddin *et al.* calculated the power spectral density of acoustic signals with independent component analysis and learned a classification

model with a k-nearest neighbor (k-NN) to discriminate drones and other objects, such as birds, airplanes, thunderstorms, rain, and wind. Audio-based approaches are usually sensitive to ambient noise in crowded urban areas and require some advanced designs of microphone array [8].

C. VIDEO-BASED TECHNIQUE

Video-based drone detection is typically the moving object detection in computer vision, in which an object can be detected and localized by analyzing visual features (e.g., color, texture, and shape) in images and videos. With the great capability of learning representational features automatically, some recent methods have exploited DL with CNN architectures to detect and identify drones using color camera [27] and depth camera [28]. To overcome the limited field of view, Rozantsev *et al.* [27] deployed a surveillance system with moving cameras to track small fast-flying drones. In [29], a high-performance CNN architecture was designed to boost the accuracy of drone detection while satisfying computational complexity for practical implementations. The network architecture has multiple spatial attention modules to highlight small and ambiguous drones in an image for better localization and detection. Besides high sensitivity to illumination, these approaches face many challenging issues, such as occlusion and multi-object interference.

D. RF-BASED TECHNIQUE

By exploiting the intercepted RF signals between drones and ground controllers, several RF-based drone detection methods have been proposed with the advantages of day-and-night working under all weather conditions for various scenarios. In [30], a passive cost-efficiency RF sensing method was proposed with discrete wavelet transform and short-time fast Fourier transform to extract drone body shifting and vibration, which in turn improves the detection accuracy of surveillance systems. In [31], Ezuma *et al.* analyzed intercepted RF signals at multi-resolution in the wavelet domain and then learned a binary classifier with Naïve Bayes and Markov models. Several methods have classified drones by learning RF fingerprints of WiFi and Bluetooth interference signals with ML algorithms. For instance, Bisio *et al.* [32] estimated the number of packets and the packet inter-arrival time of WiFi fingerprints to discriminate between different drones. In [33], Alipour-Fanid *et al.* optimally selected the L1-norm regularization-based descriptive statistics of traffic fingerprints to achieve comparable detection accuracy with low computation.

Lately, various DL models have been deployed to improve the detection and classification accuracy of drone surveillance systems. In [34], three deep neural networks (DNNs) were built with the same architecture for three separated tasks: drone detection, drone classification, and operation recognition, where each network was designed with three hidden layers to learn the discrimination model from FFT features of RF signals. Allahham *et al.* [35] designed a multi-channel convolutional network (MC-CNN) by alternately constructing 1D convolutional layers and max-pooling layers. In [36], a 5-layer DNN was deployed to estimate the direction of arrival (DoA) of single-channel RF signals for aerial drone localization, in which the number of neurons in the input layer was identical to the number of directional antenna elements. Most of the existing RF-based methods present low accuracy in drone detection and classification tasks because of the weak discrimination of ML algorithms [31]; meanwhile, DL with primitive architectures cannot optimize learning efficiency.

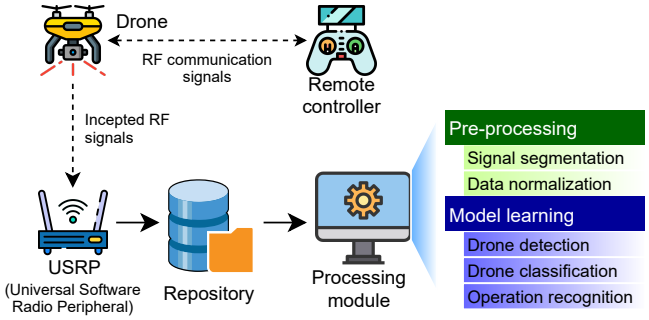


FIGURE 1. System setup for RF-based UAV detection, classification, and operation mode recognition.

In general, the drone signatures, detection range, advantages, and disadvantages of above-mentioned drone detection approaches are summarized in Table 1. It is worth noting that the detection ranges are typically derived from the literature, that means, they might vary in the practice up on the type of drones with different hardware specifications, the algorithms of deployment, and the surveillance environments.

III. METHODOLOGY

A. SYSTEM MODEL

The proposed RF-based UAV surveillance system is presented in Fig. 1. The system comprises several primary modules: drones, a remote controller (a.k.a., flight control unit), an RF sensing module, and a processing module with a database repository. Some drones for multi-purpose civil applications are specified by different technical specifications, such as maximum operation range and connectivity. The remote controller, included in the drone package, sends flight commands to the target drones and also receives the response of operating status. To intercept the drone-controller communication, an RF sensing module is configured with the assumption that the WiFi frequency is known (the operation frequency can be determined by some passive frequency scanning techniques). The intercepted RF signals can be captured by software-defined radio configurable devices and then stored in a local database repository for processing hereafter.

B. RF DATABASE DESCRIPTION

Regarding the aforementioned system, DroneRF [19], a large-scale RF dataset, is introduced for different tasks: drone detection, drone classification, and operation mode recogni-

TABLE 1. Summary of existing drone detection approaches.

Approach	Drone signature	Detection range	Advantages	Disadvantages
Radar	RCS Micro Doppler	$\leq$ 3000 ft	Easy installation	Require large mono- or multi-static RF nodes Small RCS caused by low flying attitude Expensive device
Audio	Time-frequency feature	$\leq$ 30 ft	Cheap sensors Easy implementation Accessible equipment	Sensitive to ambient noise Limited range Complex microphone array arrangement
Video	Visual features Motion features	$\leq$ 300 ft	Easy installation Accessible equipment	Line of sight necessary High resolution camera requirement Weather constraint
RF	WiFi fingerprint Time-frequency features	$\leq$ 1400 ft	Cheap sensors Easy installation	Multipath and non-line of sight High signal-to-noise ratio necessary Vulnerable to interference

TABLE 2. Overall specifications of thee three drones under consideration.

Specifications	Parrot Bebop	Parrot AR Drone	DJI Phantom 3
Dimensions (mm)	$330 \times 380 \times 36$	$517 \times 517 \times 127$	$520 \times 490 \times 290$
Weight (g)	400	420	1216
Max. flight speed (m/s)	13	11	16
Max. wireless range (m)	300	50	1000
Connectivity	WiFi	WiFi	WiFi
Operation Frequency	2.4 GHz and 5 GHz	2.4 GHz	2.4−2.485 GHz

TABLE 3. Specifications of the USRP-2943 RF receivers.

Specifications	Values
Number of channels	2
Range of operation frequency	1.2−6.0 GHz
Frequency step	< 1 KHz
Gain range	0−37.5 dB
Gain step	0.5 dBm
Maximum instantaneous real-time bandwidth	40 MHz
Maximum I/Q sample rate	200 MS/s
Digital-to-analog converter (DAC) resolution	14 bit

tion. Four operation modes, denoted as modes 01 to 04, of three different drones (Parrot Bebop, Parrot AR Drone, and DJI Phantom 3, where some main specifications are listed in Table 2) are considered as follows:

- Powering on and connecting to the controller.
- Hovering automatically without physical intervention and user commands.
- Flying without video recording.
- Flying with video recording.

Two USRP-2943 RF receivers with the specifications summarized in Table 3 are synchronized to collect the lower (L) and upper (H) half of the frequency band signals with the sampling rate of 40 MHz and the gain of 30 dB. The dataset has 227 RF signal segments, where each segment has two amplitude records (L and H bands), as shown in Fig. 2(a). It is noted that the Phantom drone is represented with one mode of drone connection (the data of other modes are corrupted in the collection procedure). Besides, the data of no drone scenario (or background activity) is recorded to serve for the detection task, which identifies the presence of a drone. The detailed experiments for the RF signal acquisition are presented in the original work [19].

RF data in DroneRF should be partitioned into multiple signal frames with the length of 10^4 samples (long enough to represent radio characteristic) using the non-overlap windowing mechanism [34]. Accordingly, we segment 227K signal frames, where each frame is formed in a high-dimensional array $\mathcal{S}$ with the size of $1 \times 10000 \times 2$. The information of dataset distribution is provided in Table 4. Due to the critical bias caused by two RF receivers with different RF band configurations and three drones without identical technical specifications, the amplitude data needs to be normalized by scaling values to the range $[0, 1]$ as

$$\hat{x}_i^{(L)} = \frac{x_i^{(L)} - \mathrm{x}_{\min}^{(L)}}{\mathrm{x}_{\max}^{(L)} - \mathrm{x}_{\min}^{(L)}}, \quad \hat{x}_i^{(H)} = \frac{x_i^{(H)} - \mathrm{x}_{\min}^{(H)}}{\mathrm{x}_{\max}^{(H)} - \mathrm{x}_{\min}^{(H)}}, \tag{1}$$

where $\mathrm{x}_{\min}^{(L)}, \mathrm{x}_{\max}^{(L)}, \mathrm{x}_{\min}^{(H)}, \mathrm{x}_{\max}^{(H)}$ denote the minimum and maximum amplitude values of the lower and upper half bands of all drones. An example of partitioning RF signal into fixed-sized frames is illustrated in Fig. 2(b).

C. RF-UAVNET: RF-BASED UAV DETECTION-CLASSIFICATION CONVOLUTIONAL NETWORK

This part presents an efficient CNN, denoted RF-UAVNet, for low-cost and high-accuracy drone detection, drone classification, and operation recognition based on RF signals, where the overall architecture is shown in Fig. 3(a). Regarding the detailed architecture, an input layer is first specified with the size of $1 \times 10000 \times 2$ to facilitate partitioned frames. To reduce the spatial size of feature maps to significantly save computation at deeper layers, a 1D regular convolutional (conv) layer is specified by the filters (so-called kernels) of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3172787, IEEE Access

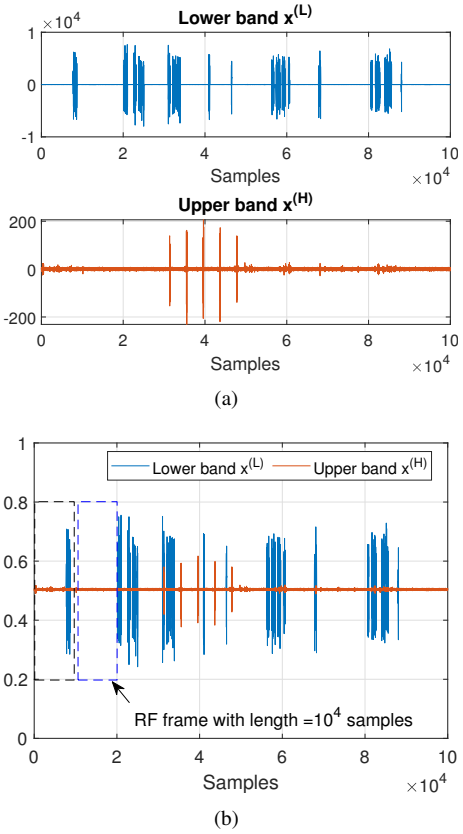


FIGURE 2. Example of Bebop drone's signal: (a) the raw amplitude samples of the lower and upper bands and (b) the normalized samples with fixed-sized partition using non-overlap windowing mechanism.

TABLE 4. Dataset Distribution of DroneRF.

Task	Class	No. segments	No. frames
Drone detection	Drone	186	186,000
	No drone	41	41,000
Drone Classification	AR	81	81,000
	Bebop	84	84,000
	Phantom	21	21,000
	No drone	41	41,000
Operation recognition	AR mode 01	21	21,000
	AR mode 02	21	21,000
	AR mode 03	21	21,000
	AR mode 04	18	18,000
	Bebop mode 01	21	21,000
	Bebop mode 02	21	21,000
	Bebop mode 03	21	21,000
	Bebop mode 04	21	21,000
	Phantom mode 01	21	21,000
	No drone	41	41,000

size 1×5 with the stride of $(1, 5)$. Notably, the number of channels for each filter in a regular conv layer is always equal to the number of channels (or the third dimension) of the input to the layer. For example, because the input having two channels connects with the conv layer directly, the number of channels for each filter is two. The conv layer is followed by a batch normalization (norm) layer and an exponential linear unit (eLU) activation layer, where this layer group is denoted r-unit in Fig. 3(b). Fundamentally, the convolution between a kernel with weights w_i and an input map u_i at any specific coordinate (x, y) in the spatial domain is formulated as

$$z_{x,y} = \sum_i w_i u_i + b, \tag{2}$$

where b is the scalar bias. The volume convolution of K channels (a.k.a., the depth size of the input) is the sum of K resulting convolution scalar z. The norm layer performs value normalization to its input with the mean μ_B and variance σ_B^2 calculated for each mini-batch and input channel as

$$\hat{z}_{x,y} = \frac{z_{x,y} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \tag{3}$$

where $\epsilon = 10^{-5}$ is the constant scalar to prevent the numerical uncertainty caused by a very small variance. To possibly process the inputs with zero mean and non-optimal variance passing to the layer that follows the norm layer, the norm layer then shifts and scales the normalized values by

$$v_{x,y} = \kappa \hat{z}_{x,y} + \xi, \tag{4}$$

where κ and ξ denote, respectively, the scale factor and offset parameters that are updated during network training. Once the training process is finished, the norm layer calculates the mean and variance over the whole training set. In the testing process to predict the class of new input, the norm layer uses the trained mean and variance instead of the mean and variance of a mini-batch to normalize the activation. In a CNN for classification, batch normalization is typically used between conv and eLU layers to accelerate the training of networks and reduce the sensitivity to network initialization [37]. The eLU layer then performs nonlinearity by outputting the identical value on positive inputs and the exponential operation result on negative inputs, i.e.,

$$o_{x,y} = \mathrm{eLU}\left(v_{x,y}\right) = \begin{cases} v_{x,y}, & \text{if } v_{x,y} \geq 0, \\ \alpha\left(e^{v_{x,y}} - 1\right), & \text{if } v_{x,y} < 0, \end{cases} \tag{5}$$

where $\alpha = 1$ is the nonlinear parameter. Compared with the rectified linear unit commonly developed in several convolutional networks, eLU can achieve the learning convergence faster with a higher accuracy in some classification tasks [38]. With 64 filters of size 1×5 and the stride of $(1, 5)$ specified for the conv layer, the width dimension of output maps reduces five times and the depth dimension is identical to the number of filters. At this point, with an input signal frame $\mathcal{S} \in \mathbb{R}^{1\times 10000\times 2}$, we determine the output of the r-unit $\mathbf{F}_{\mathrm{OUT}}^{\mathrm{r-unit}} \in \mathbb{R}^{1\times 2000\times 64}$.

As the primary modules of RF-UAVNet to extract the local correlations of amplitude samples at multi-scale feature representations (as illustrated in Fig. 4), four processing units (denoted g-unit) are arranged in cascade throughout the network architecture in Fig. 3(a). In each g-unit, a grouped convolutional (g-conv) layer with the filters of size 1×3 and the stride of $(1, 2)$ is followed by an eLU layer (see Fig. 3(c)). Regarding the grouped convolution, the filters are separated into different groups for processing, where each group will

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3172787, IEEE Access

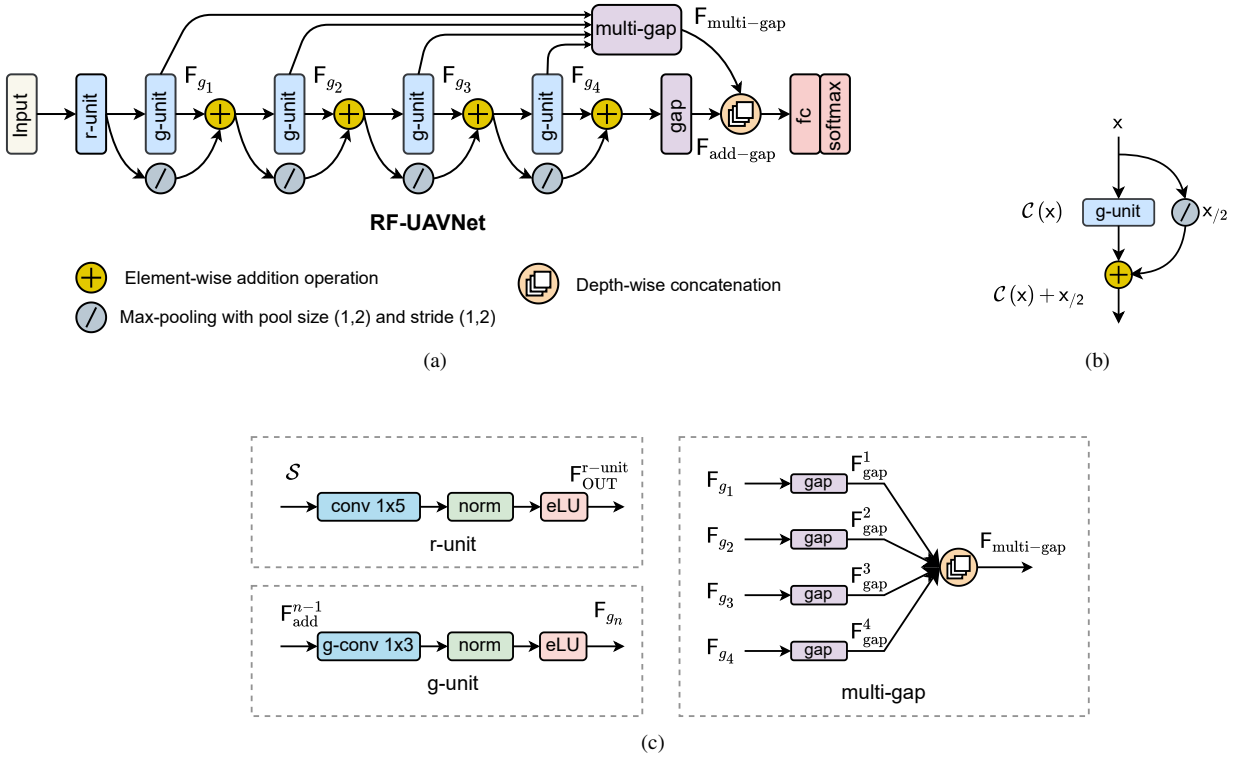


FIGURE 3. RF-UAVNet: (a) overall architecture, (b) skip-connection, and (c) network components: r-unit, g-unit, and multi-gap.

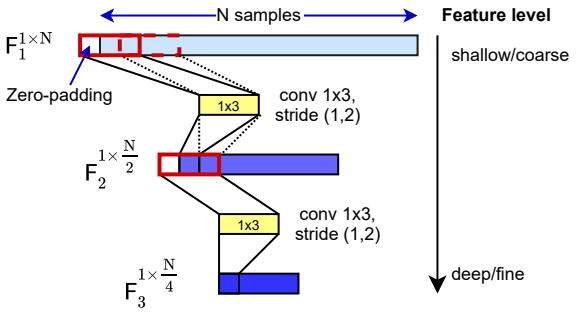


FIGURE 4. Example of the 1D convolution operation to extract local features at multi-scale resolutions from coarse to fine.

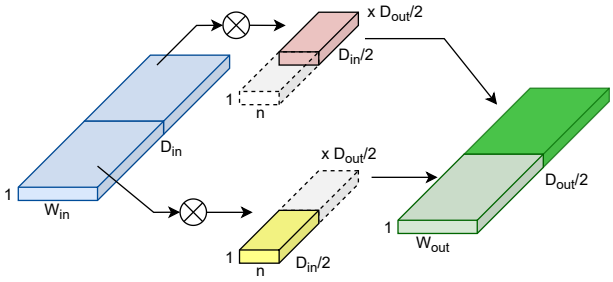


FIGURE 5. Example of the grouped 1D convolution with two filter groups.

serve as standard convolution for a certain number of input maps. Compared with the standard convolution, the grouped one reduces the number of parameters and computational cost significantly further as the number of non-overlapping filter groups increases, while the learning efficiency is mostly maintained with the sparse relationship among filters [39]. With G groups, the complexity of g-conv is $\frac{D_{\text{in}} \times D_{\text{out}}}{G}$, which is lower than that of conv by G times [40], where D_{in} and D_{out}, respectively, denote the depth size of input and output. An example of grouped convolution with two groups is illustrated in Fig. 5, where the first group (yellow) convolves with the first half of the input and the second group (red) convolves with the second half. The number of filter groups must evenly divide the number of channels of the input layer. The number of channels in the output of g-conv layer is $\text{numGroups} \times \text{numFiltersPerGroup}$, where numGroups and numFiltersPerGroup, respectively, denote the number of groups and the number of filters for each group. In this work, we specify the g-conv layers by eight groups with each group having eight filters to always produce 64 feature maps at the output of g-conv layers. We denote by $\mathbf{F}_{g_{i=1,\ldots,4}} \in \mathbb{R}^{1 \times W_{g_i} \times 64}$ with $W_{g_i} = \{1000, 500, 250, 125\}$ the outputs of g-units, where W_{g_i} is the horizontal size of the output maps resulted in the g-conv layer with the stride of $(1, 2)$ in the i-th g-unit.

To increase the overall system accuracy via well-handling gradient flow in RF-UAVNet, we study a multi-level skip-connection. Inspired by [41], skip-connection associates the outputs of different g-units by adopting element-wise addition operation. Given the output of r-unit $\mathbf{F}_{\text{OUT}}^{\text{r-unit}}$, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3172787, IEEE Access

procedure of the first skip-connection can be formulated with an element-wise addition (add) layer as

$$\mathbf{F}^1_{\text{add}} = \mathcal{F}\left(\mathbf{F}^{\text{r-unit}}_{\text{OUT}}\right) = \mathbf{F}_{p_1} + \mathbf{F}_{g_1}, \tag{6}$$

where $\mathbf{F}_{p_1} = \mathcal{P}_{\max}\left(\mathbf{F}^{\text{r-unit}}_{\text{OUT}}\right)$ and $\mathbf{F}_{g_1} = \mathcal{C}\left(\mathbf{F}^{\text{r-unit}}_{\text{OUT}}\right)$; $\mathcal{P}_{\max}$ denotes the max-pooling operation with the pool size of $(1, 2)$ and the stride of $(1, 2)$; $\mathcal{C}$ denotes the united operation of grouped convolution, batch normalization, and eLU activation. The max pooling (maxpool) layer is adopted to pick on the most salient features [42] and to align the output to spatial size of $\mathbf{F}_{g_1}$. The element-wise addition operation in (6) is done by an addition (add) layer, where all inputs to this layer must have the same dimension. For example, by downsampling $\mathbf{F}^{\text{r-unit}}_{\text{OUT}}$ in the horizontal dimension with the stride of $(1, 2)$, the output of maxpool layer $\mathbf{F}_{p_1} \in \mathbb{R}^{1\times 1000\times 64}$ has the same dimension with $\mathbf{F}_{g_1} \in \mathbb{R}^{1\times W_{g_1}\times 64}$. In neural networks, the vanishing gradient problem usually occurs when the gradient elements (the partial derivatives with respect to the network parameters) become exponentially small; hence, the update of the parameters with the gradient is almost negligible. Based on the principle of residual block [41], the skip-connection allows the network to learn the residual $\mathcal{C}(\cdot)$ instead of learning the true output $\mathcal{F}(\cdot)$ to nearly maintain the gradient flow. To put it simply, the attenuated gradient caused by activation functions is consolidated with the identity information by skipping layers. The structure of skip-connection deployed in RF-UAVNet is generally described in Fig. 3(b). For a multi-level mechanism, the n-th skip-connection with $n \geq 2$, in general, can be expressed as follows:

$$\begin{aligned}\mathbf{F}^n_{\text{add}} &= \mathcal{F}\left(\mathbf{F}^{n-1}_{\text{add}}\right) = \mathbf{F}_{p_{n-1}} + \mathbf{F}_{g_{n-1}} \\ &= \mathcal{P}_{\max}\left(\mathbf{F}^{n-1}_{\text{add}}\right) + \mathcal{C}\left(\mathbf{F}^{n-1}_{\text{add}}\right).\end{aligned} \tag{7}$$

For the purpose of tracking the dimension of feature maps along with the backbone of RF-UAVnet in the corporation with multi-level skip-connection, the output halves the width dimension of feature maps for each time of passing the input through a skip-connection. In particular, we obtain $\mathcal{F}: \mathbf{F}^1_{\text{add}} \in \mathbb{R}^{1\times W_{g_1}\times 64} \leftarrow \mathbf{F}^{\text{r-unit}}_{\text{OUT}}$ for the first skip-connection and $\mathcal{F}: \mathbf{F}^n_{\text{add}} \in \mathbb{R}^{1\times W_{g_n}\times 64} \leftarrow \mathbf{F}^{n-1}_{\text{add}} \in \mathbb{R}^{1\times W_{g_{n-1}}\times 64}$ for the n-th skip-connection with $n \geq 2$, where W_{g_n} is the width of output maps. It is worth noting that the element-wise addition layer does not change the feature dimension. The proposed multi-level skip-connection allows the gradient information to be mostly maintained at different feature levels. Consequently, the vanishing/exploding gradient problem can be prevented effectively when the network goes deeper.

A global average pooling (gap) layer is configured to calculate the mean of each feature map, which is usually located before fully connected (fc) layers for network size reduction and overfitting alleviation with no learnable parameters. For example, the gap layer transforms the dimensions from $1 \times W \times D$ to $1 \times 1 \times D$ by averaging across with the pool size of $(1, W)$ for every maps along the channel dimension. To collect the meaningful knowledge from the preceding layers, we propose a multi-level global average pooling (multi-gap) module as described in Fig. 3(c). The outputs of different g-unit $\mathbf{F}_{g_i}$ are first forwarded to the gap layers individually to obtain the global features

$$\mathbf{F}^i_{\text{gap}} = \mathcal{P}_{\text{avg}}\left(\mathbf{F}_{g_i}\right), \text{ with } i = 1, \ldots, 4, \tag{8}$$

where $\mathbf{F}^i_{\text{gap}} \in \mathbb{R}^{1\times 1\times 64}$ and $\mathcal{P}_{\text{avg}}$ denotes the average pooling operation with different pool sizes. The resulting global features are then stacked by a depth-wise concatenation (concat) layer as

$$\mathbf{F}_{\text{multi-gap}} = \mathcal{D}\left(\mathbf{F}^1_{\text{gap}}, \mathbf{F}^2_{\text{gap}}, \mathbf{F}^3_{\text{gap}}, \mathbf{F}^4_{\text{gap}}\right), \tag{9}$$

where $\mathcal{D}$ denotes the concatenation operation. It is noted that a concat layer takes the inputs having the same height and width and aggregates them along the depth (or channel) dimension. The output $\mathbf{F}_{\text{multi-gap}} \in \mathbb{R}^{1\times 1\times 256}$ has the number of channels (or depth size) that is the sum of those of all inputs.

At the end of network, the output of multi-gap $\mathbf{F}_{\text{multi-gap}}$ combines with the output of the last skip-connection $\mathbf{F}_{\text{add-gap}} = \mathcal{P}_{\text{avg}}\left(\mathbf{F}^4_{\text{add}}\right)$ by a concat layer, see Fig. 3(a) as

$$\mathbf{F}_{\text{final}} = \mathcal{D}\left(\mathbf{F}_{\text{multi-gap}}, \mathbf{F}_{\text{add-gap}}\right), \tag{10}$$

where $\mathbf{F}_{\text{final}} \in \mathbb{R}^{1\times 1\times 320}$, which contains the synthetic information of signal characteristics at multiple scales, is flattened into a single vector before feeding to the fc layer defined by k neurons and followed by a softmax layer. The number of neuron in fc layer is identical to the number of classes in a given dataset for a particular task. Regarding different drone surveillance tasks studied in this work, we use the same CNN architecture with a minor change in the fc layer. Particularly, RF-UAVNet is with $k = 2$ for drone detection, $k = 4$ for drone classification, and $k = 10$ for operation recognition (see Table 4). The detailed network configurations are further provided in Table 5. The network is finalized with a softmax layer that performs a normalized exponential function to transform the scores $\mathbf{s} = \{s_j\}_{j=1}^k$ deduced by the fc layer to the class probability distribution

$$p_j = \frac{e^{s_j}}{\sum_{i=1}^k e^{s_i}}, \tag{11}$$

where p_j is the j-th class probability of an input data with $0 \leq p_j \leq 1$ and $\sum_{i=1}^k p_i = 1$. The softmax function can be considered the multi-class generalization of the logistic sigmoid function. The network assigns each input to one of k mutually exclusive classes based on the output of softmax function and computes the cross entropy loss for multi-classes classification as follows:

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{j=1}^{k} \nu_{ij} \ln\left(\upsilon_{ij}\right), \tag{12}$$

where N is the number of training signal frames, ν_{ij} denotes the ground-truth of the i-th input signal associated with the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3172787, IEEE Access

TABLE 5. RF-UAVNet Configuration (Task 01: Drone Detection, Task 02: Drone Classification, and Task 03: Operation Recognition).

Layer	Network description			Output Size
	Task 01	Task 02	Task 03	
input	RF signal frame			$1 \times 10000 \times 2$
r-unit	64 filters 1×5, stride $(1, 5)$			$1 \times 2000 \times 64$
g-unit	8 groups of 8 filters 1×3, stride $(1, 2)$			$1 \times 1000 \times 64$
g-unit	8 groups of 8 filters 1×3, stride $(1, 2)$			$1 \times 500 \times 64$
g-unit	8 groups of 8 filters 1×3, stride $(1, 2)$			$1 \times 250 \times 64$
g-unit	8 groups of 8 filters 1×3, stride $(1, 2)$			$1 \times 125 \times 64$
gap	pool size $(1, 125)$			$1 \times 1 \times 64$
multi-gap	4 gap layers + depth-wise concat			$1 \times 1 \times 256$
concat	depth-wise concat			$1 \times 1 \times 320$
fc	$k = 2$	$k = 4$	$k = 10$	$1 \times 1 \times k$

j-th class, and v_{ij} denotes the output resulted by the network for the class j of the input signal i.

The configurations of the network training process are given as follows: the optimizer is the stochastic gradient descent with momentum, the momentum factor is 0.95, the L_2 regularization factor is 0.0001, the maximum number of epochs for training is 90, the initial learning rate is 0.01 (dropping to 0.001 after 45 epochs for a better training convergence), and the mini-batch size is 128. The network is evaluated on a platform using 3.70-GHz CPUs, 32GB RAM, and a single NVIDIA GTX 1080Ti GPU.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section evaluates the performance of RF-UAVNet for drone surveillance using the 10-fold cross-validation protocol [34]. Concretely, the DroneRF dataset is randomly partitioned into ten non-overlapping folds, where one fold is used as the validation set for testing, and the remainder is used to train the network. This process is then repeated ten times to evaluate the whole dataset, and the overall performance is reported by averaging the performance of all folds.

Specifically, four principal experiments along with insightful discussions are delivered as follows:

- In the first experiment, we benchmark the performance of RF-UAVNet for three drone surveillance tasks (i.e., drone detection, drone classification, and operation recognition denoted by tasks 01-03) on the DroneRF dataset.
- The second experiment evaluates the parameter sensitivity of RF-UAVNet to the overall system performance with different numbers of filter groups in g-conv layers.
- The third experiment is an ablation study, in which RF-UAVNet is compared with its variants to demonstrate the advantages of grouped convolution and skip-connection incorporated with multi-gap.
- The last one compares RF-UAVNet with other state-of-the-art DL-based approaches, which also exploit DNN and CNN models for drone surveillance.

The performance of drone detection and classification is measured using the accuracy and F1-score metrics. The MATLAB codes of our work can be freely accessed on the GitHub repository[1].

A. MODEL PERFORMANCE

In the first experiment, we evaluate the performance of RF-UAVNet on the DroneRF dataset for different tasks, where the numerical results of confusion matrices are shown in Fig. 6. For drone detection with the result given in Fig. 6(a), RF-UAVNet reaches the overall correct identification rate of 99.85%. For drone classification, RF-UAVNet yields the overall accuracy of 98.55%, whereas Phantom presents the worst classification rate of 96.41%. Phantom and Bebop are confused with AR by around 2.80% and 2.04%, respectively. For operation recognition which is more challenging than drone detection and classification with more classes for discrimination, the network achieves the overall accuracy of 95.33%. From the confusion matrix presented in Fig. 6(c), Bebop mode 04 (lying with video recording) presents the worst recognition rate of 88.37%, which is mostly confused with Bebop mode 03 (flying without video recording) by approximately 8.63%. Interestingly, high confusion is found between different modes of the Bebop drone, for instance, mode 01 (powering on and connecting) and mode 02 (hovering) are misclassified as mode 03 with the error rate of 4.31% and 5.74%, respectively.

B. PARAMETER SENSITIVITY

This experiment investigates the relationship between the overall performance and the network complexity with various numbers of filter groups ($\mathrm{numGroups}$) defined in g-conv layers. The complexity is evaluated by the network size, which is measured by the number of trainable parameters, and the computational cost, which is estimated by the number of floating point operations (FLOPs). In Fig. 7, we report the results of three tasks with $\mathrm{numGroups} = \{2, 4, 6, 8, 16, 32\}$, where the number of filters in each group is automatically determined to produce the same number of feature maps. In general, the cost efficiency of RF-UAVNet gains along with the increment of $\mathrm{numGroups}$, whereas the performance of three drone surveillance tasks gets degradation. For instance, from Figs. 7(a) and 7(b), the F1-score of drone detection and classification reduces by around 0.01% and 0.14%, respectively, when we increase $\mathrm{numGroups}$ from 4 to 8 groups. However, when we increase $\mathrm{numGroups}$ from 8 to 16 groups, the F1-score decreases severely by 0.62% for drone detection and 3.08% for drone classification. For operation recognition, the performance deterioration is nearly identical in each time of doubling $\mathrm{numGroups}$ (approximately the accuracy of 2.41% and the F1-score of 2.57%). Concerning network complexity, the number of parameters and FLOPs reduce significantly for small numbers of filter groups. Because of the same architecture configuration of r-unit and g-units to extract features, the size difference of RF-UAVNets for different tasks derives from the numbers of weights and biases to densely associate all neurons in the concat layer

[1]https://github.com/ThienHuynhThe/RF-based-Drone-Surveillance-with-DL.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3172787, IEEE Access

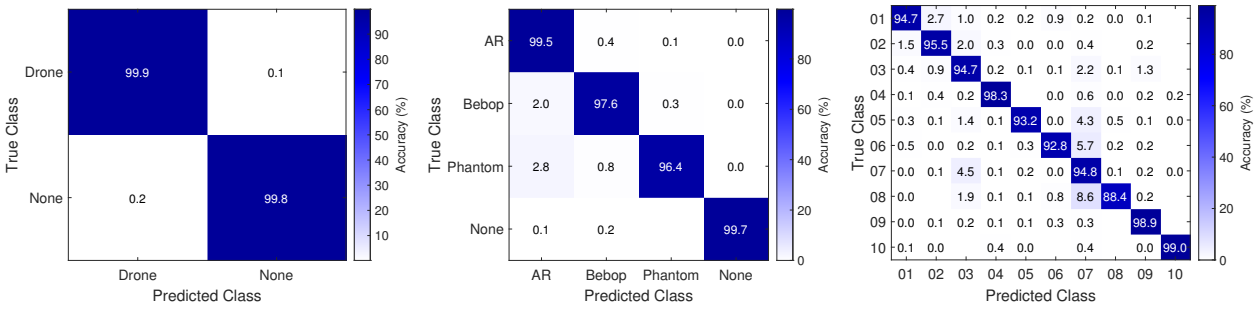


FIGURE 6. Confusion matrices: (a) drone detection, (b) drone classification, and (c) operation mode recognition, where classes 01-04 are of the AR drone, classes 05-08 are of the Bebop drone, class 09 denotes the Phantom mode 01, and class 10 denotes RF background activity without drone detection.

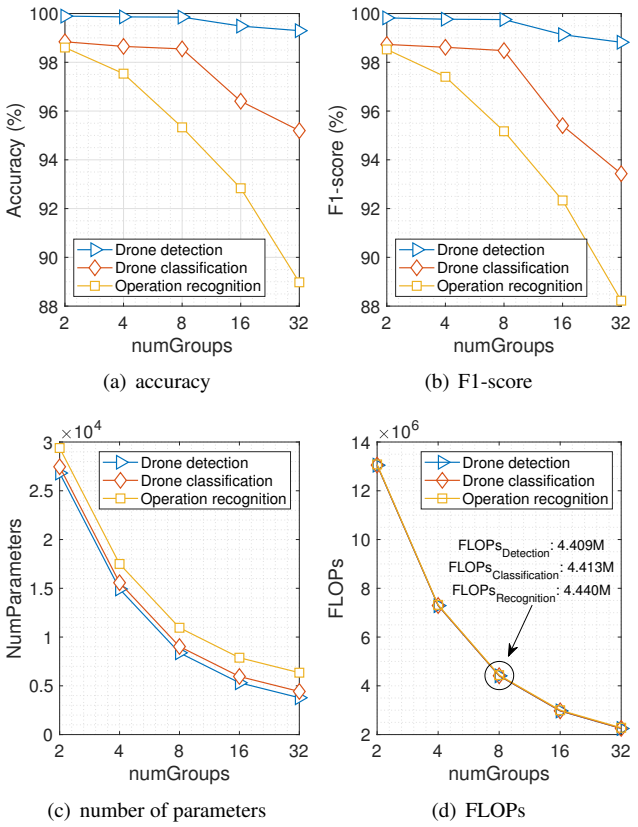


FIGURE 7. Performance and complexity of RF-UAVNet with various number of filter groups (numGroups).

with k neurons in the fc layers. With $k = 10$, the size of RF-UAVNet for operation recognition is bigger than those of the networks for drone detection and drone classification as shown in Fig. 7(c). Notably, the computation of RF-UAVNet mostly is for feature calculation in convolutional layers, while the computational cost consumed in the fc layer is insignificant. This explains a small gap in FLOPs for three different tasks in Fig. 7(d). As a result, we specify $\mathrm{numGroups} = 8$ to achieve a comfortable balance between performance and complexity.

We further investigate the performance of RF-UAVNet with different common activation functions: rectified linear unit (ReLU), leaky ReLU, hyperbolic tangent (Tanh), and eLU. Based on the accuracy and F1-score results in Table 6, eLU shows the best performance for drone detection and drone classification, whereas being worse than LeakyReLU for operation recognition with insignificant gaps. Despite being better than Tanh for drone classification, ReLU presents worse results in drone detection and operation mode recognition experiments. It is noted that the activation layers do not have parameters to learn and take up a small fraction of the overall network computation.

TABLE 6. Performance of RF-UAVNet With Different Activation Functions.

Activation	Accuracy (%)			F1-score (%)		
	Task 01	Task 02	Task 03	Task 01	Task 02	Task 03
ReLU	98.69	97.91	94.14	97.89	97.10	94.14
Tanh	98.97	96.82	94.52	98.32	95.41	94.44
LeakyReLU	99.77	97.80	95.36	99.61	97.82	95.01
eLU	99.85	98.55	95.33	99.75	98.48	95.06

C. ABLATION STUDY

The third experiment evaluates the effectiveness of grouped convolution incorporated with multi-layer skip-connection and multi-gap mechanisms, which are exploited in RF-UAVNet. Specifically, we compare RF-UAVNet with three cut-off variants, denoted by Net-A, Net-B, and Net-C, in terms of accuracy and complexity. The detailed architectures of these CNNs are given as follows and summarized in Table 7:

- Net-A: an alternative architecture of RF-UAVNet using standard conv layers without the incorporated structure of skip-connection and multi-gap.
- Net-B: an alternative architecture of RF-UAVNet using g-conv layers without skip-connection and multi-gap (or Net-A with g-conv).
- Net-C: an alternative architecture of RF-UAVNet using g-conv layers and skip-connection without multi-gap (or Net-B with skip-connection).

It is noted that all networks are evaluated with the same training configuration. Compared with Net-A which uses

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3172787, IEEE Access

TABLE 7. Performance Comparison Between RF-UAVNet and Its Cut-off Variants. Noted: complexity is reported with RF-UAVNet for Operation Recognition.

Network	Strategy				Accuracy (%)			F1-score (%)			Approx.Complexity	
	conv	g-conv	skip-conn	multi-gap	Task 01	Task 02	Task 03	Task 01	Task 02	Task 03	No.params	FLOPs
Net-A	✓				99.70	92.72	83.98	99.49	90.98	82.55	50K	24.5K
Net-B		✓			99.50	91.14	83.21	99.15	90.43	81.86	8.4K	4.4M
Net-C		✓	✓		99.71	95.77	92.62	99.51	93.75	92.06	8.4K	4.4M
RF-UAVNet		✓	✓	✓	99.85	98.55	95.33	99.75	98.48	95.06	11K	4.4M

TABLE 8. Performance Comparison Between RF-UAVNet and Other Deep Networks.

Activation	Accuracy (%)			F1-score (%)			Approx. Complexity (%)		
	Task 01	Task 02	Task 03	Task 01	Task 02	Task 03	No. params	FLOPs	Speed (ms)
DNN [34]	99.71	84.52	46.83	99.57	78.81	43.02	5.1M	5.2M	1.24
1D-CNN [16]	99.81	85.45	59.19	99.61	84.68	55.11	63K	23M	1.98
MC-CNN [35]	99.95	94.55	87.37	99.46	91.01	77.02	47K	47M	2.64
RF-UAVNet	99.85	98.55	95.33	99.75	98.48	95.06	11K	4.4M	1.31

standard conv layers, Net-B with g-conv layers remarkably reduces the network size (from 50K to 8.4K parameters) and computational cost (from 24.5 MFLOPs to 4.4 MFLOPs, where MFLOPs denotes megaFLOPs), while it suffers a trivial reduction of performance. Net-C, which is upgraded from Net-B with the multi-layer skip-connection, achieves a performance improvement in terms of accuracy and F1-score without the increment of the number of parameters and FLOPs. By leveraging a sophisticated-designed architecture, wherein g-conv layers associate with multi-layer skip-connection and multi-gap mechanisms, RF-UAVNet successfully obtains a twofold objective: low complexity and high accuracy. Concretely, RF-UAVNet is approximately 80% lower than Net-A in terms of complexity and effectively performs drone classification and operation recognition better than Net-B by around $7.41 - 12.12\%$. Notably, RF-UAVNet is bigger than Net-B and Net-C in terms of the network size because the multi-gap structure for multi-level feature maps aggregation increases the number of parameters in the fc layer.

D. METHOD COMPARISON

In the last experiment, the RF-UAVNet is compared with some recent DL models deployed for drone surveillance, including DNN [34], 1D-CNN [16], and MC-CNN [35], in terms of performance and complexity, where the numerical results, including accuracy, F1-score, number of trainable parameters, FLOPs, and processing speed (a.k.a., the average prediction time of a signal frame) are reported in Table 8. In general, the performance gap between the comparison methods in drone detection is trivial (around $0.1 - 0.5\%$), whereas the gaps in drone classification and operation recognition are more significant. DNN, which is simply designed with three hidden layers, reports the worst performance, especially with the operation recognition task with the overall accuracy of 46.8% and the F1-score of 43.0%. By leveraging convolutional architectures, 1D-CNN with multiple 1D conv layers associated with average pooling layers in a cascade structure performs better than DNN in drone classification and operation recognition. Compared with DNN, 1D-CNN improves accuracy and F1-score by up to 12.4% and 12.1%, respectively. With a depth-wise concatenation layer to selectively aggregate features in multiple channels, MC-CNN reaches the drone classification accuracy of 94.6% and the operation recognition accuracy of 87.4%, which significantly outperforms DNN by $9.1 - 28.2\%$. Despite being worse than MC-CNN by a tiny gap of drone detection (around $0.1 - 02\%$), RF-UAVNet, which effectively incorporates the multi-level skip-connection and multi-gap mechanisms, performs drone classification and operation recognition more precisely with the higher accuracy of 4.0% and 7.9% and with the greater F1-score of 7.5% and 18.1%, respectively. For complexity, DNN has the biggest size with around 5.1M parameters because of specifying large numbers of neurons in hidden layers for dense connection, but its speed is fastest with a very simple network structure. Compared with DNN, 1D-CNN and MC-CNN are much more lightweight but have higher computation costs and classify more slowly. RF-UAVNet presents the smallest network size with 11K parameters and the cheapest computation with 4.4 MFLOPs thanks to the deployment of grouped convolution (instead of regular convolution in 1D-CNN and MC-CNN). With the sophisticated structure of multi-level skip-connection and multi-level pooling, our network performs a little slower than DNN. Accordingly, the proposed CNN has demonstrated superiority in terms of accuracy and complexity against other existing deep models for drone surveillance.

V. CONCLUSION AND DISCUSSIONS

In this paper, we presented an efficient RF-based surveillance approach to manage the flying activities of registered drones, in which we proposed a cost-efficient and high-accuracy CNN, namely RF-UAVNet, to effectively detect and classify drones and recognize their operation modes. RF-UAVNet was characterized by grouped convolutional layers to reduce the network size and computational cost significantly, where each g-conv layer was specified by asymmetric 1D kernels

to capture the temporal correlations as local features of RF signals. Remarkably, to increase the overall accuracy of systems, we designed the multi-level skip-connection and multi-gap mechanisms to, respectively, prevent the vanishing gradient effectively and collect the useful global features at different signal resolutions. In the experiments, we analyzed RF-UAVNet exhaustively with different architecture configurations on the DroneRF dataset to demonstrate the effectiveness of grouped convolution to reduce complexity and multi-level skip-connection incorporated with multi-gap to improve accuracy. RF-UAVNet achieved very high accuracy for different drone surveillance tasks (approximately 99.9% for drone detection, 98.6% for drone classification, and 95.3% for operation recognition) with low complexity (11K parameters and 4.4 MFLOPs), which could be a favorable approach for onboard deployment in portable anti-drone systems. Furthermore, the proposed deep network considerably outperformed state-of-the-art DL models for drone surveillance in terms of accuracy and F1-score.

In the future, we intend to create our own practical dataset of RF signals for different drone surveillance tasks, in which each signal should be represented by a complex envelop form with in-phase and quadrature components. Besides, we will extend the dataset into more conditions and scenarios, including increasing the number of drones, recording RF signals under different channel impairments (e.g., fading and additive noise) in both indoor and outdoor environments, and varying speed, altitude, and distance between drones and the RF sensing module.

REFERENCES

[1] S. Hussain, S. A. Chaudhry, O. A. Alomari, M. H. Alsharif, M. K. Khan, and N. Kumar, "Amassing the security: An ECC-based authentication scheme for internet of drones," *IEEE Syst. J.*, vol. 15, no. 3, pp. 4431–4438, Sep. 2021.

[2] Q.-V. Pham, T. Huynh-The, M. Alazab, J. Zhao, and W.-J. Hwang, "Sum-rate maximization for UAV-assisted visible light communications using NOMA: Swarm intelligence meets machine learning," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10 375–10 387, Oct. 2020.

[3] I. Bisio, C. Garibotto, H. Haleem, F. Lavagetto, and A. Sciarrone, "On the localization of wireless targets: A drone surveillance perspective," *IEEE Netw.*, vol. 35, no. 5, pp. 249–255, Sep./Oct. 2021.

[4] H. Fu, S. Abeywickrama, L. Zhang, and C. Yuen, "Low-complexity portable passive drone surveillance via SDR-based signal processing," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 112–118, Apr. 2018.

[5] B. Taha and A. Shoufan, "Machine learning-based drone detection and classification: State-of-the-art in research," *IEEE Access*, vol. 7, pp. 138 669–138 682, 2019.

[6] M. M. Azari, H. Sallouha, A. Chiumento, S. Rajendran, E. Vinogradov, and S. Pollin, "Key technologies and system trade-offs for detection and localization of amateur drones," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 51–57, Jan. 2018.

[7] G. Ding, Q. Wu, L. Zhang, Y. Lin, T. A. Tsiftsis, and Y.-D. Yao, "An amateur drone surveillance system based on the cognitive internet of things," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 29–35, 2018.

[8] I. Bisio, C. Garibotto, F. Lavagetto, A. Sciarrone, and S. Zappatore, "Unauthorized amateur UAV detection based on WiFi statistical fingerprint analysis," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 106–111, Apr. 2018.

[9] P. Nguyen, H. Truong, M. Ravindranathan, A. Nguyen, R. Han, and T. Vu, "Matthan: Drone presence detection by identifying physical signatures in the drone's RF communication," in *Proc. 15th Annual Int. Conf. Mob. Syst. Appl. Serv. (MobiSys)*, Niagara Falls, New York, USA, Jun. 2017, pp. 211–224.

[10] A. Shoufan, H. M. Al-Angari, M. F. A. Sheikh, and E. Damiani, "Drone pilot identification by classifying radio-control signals," *IEEE Trans. Inf. Forensic Secur.*, vol. 13, no. 10, pp. 2439–2447, Oct. 2018.

[11] H. Zhang, C. Cao, L. Xu, and T. A. Gulliver, "A UAV detection algorithm based on an artificial neural network," *IEEE Access*, vol. 6, pp. 24 720–24 728, 2018.

[12] T. Huynh-The, Q.-V. Pham, T.-V. Nguyen, T. T. Nguyen, R. Ruby, M. Zeng, and D.-S. Kim, "Automatic modulation classification: A deep architecture survey," *IEEE Access*, vol. 9, pp. 142 950–142 971, 2021.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436–444, May 2015.

[14] T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "MCNet: An efficient CNN architecture for robust automatic modulation classification," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 811–815, Apr. 2020.

[15] G. B. Tunze, T. Huynh-The, J.-M. Lee, and D.-S. Kim, "Sparsely connected cnn for efficient automatic modulation recognition," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 557–15 568, Dec. 2020.

[16] S. Al-Emadi and F. Al-Senaid, "Drone detection approach based on radio-frequency using convolutional neural network," in *Proc. IEEE Int. Conf. Inform. IoT Enabling Technol. (ICIoT)*, Doha, Qatar, Feb. 2020, pp. 29–34.

[17] C. Yang, W. Jiang, and Z. Guo, "Time series data classification based on dual path cnn-rnn cascade network," *IEEE Access*, vol. 7, pp. 155 304–155 312, 2019.

[18] F. Wang, R. Liu, Q. Hu, and X. Chen, "Cascade convolutional neural network with progressive optimization for motor fault diagnosis under nonstationary conditions," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2511–2521, Apr. 2021.

[19] M. S. Allahham, M. F. Al-Sa'd, A. Al-Ali, A. Mohamed, T. Khattab, and A. Erbad, "DroneRF dataset: A dataset of drones for RF-based detection, classification and identification," *Data in Brief*, vol. 26, p. 104313, Oct. 2019.

[20] F. Hoffmann, M. Ritchie, F. Fioranelli, A. Charlish, and H. Griffiths, "Micro-doppler based detection and tracking of UAVs with multistatic radar," in *Proc. IEEE Radar Conf. (RadarConf)*, Philadelphia, USA, May 2016, pp. 1–6.

[21] M. Jahangir, B. I. Ahmad, and C. J. Baker, "Robust drone classification using two-stage decision trees and results from SESAR SAFIR trials," in *Proc. IEEE Int. Radar Conf. (RADAR)*, Washington, DC, USA, Apr. 2020, pp. 636–641.

[22] Y. Dan, J. Yi, X. Wan, Y. Rao, and B. Wang, "LTE-based passive radar for drone detection and its experimental results," *The Journal of Engineering*, vol. 2019, no. 20, pp. 6910–6913, Oct. 2019.

[23] Z. Geng, R. Xu, and H. Deng, "LTE-based multistatic passive radar system for UAV detection," *IET Radar Sonar Navig.*, vol. 14, no. 7, pp. 1088–1097, Jul. 2020.

[24] X. Yue, Y. Liu, J. Wang, H. Song, and H. Cao, "Software defined radio and wireless acoustic networking for amateur drone surveillance," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 90–97, Apr. 2018.

[25] M. Z. Anwar, Z. Kaleem, and A. Jamalipour, "Machine learning inspired sound-based amateur drone detection for public safety applications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2526–2534, Mar. 2019.

[26] Z. Uddin, M. Altaf, M. Bilal, L. Nkenyereye, and A. K. Bashir, "Amateur drones detection: A machine learning approach utilizing the acoustic signals in the presence of strong interference," *Computer Communications*, vol. 154, pp. 236–245, Mar. 2020.

[27] A. Rozantsev, V. Lepetit, and P. Fua, "Detecting flying objects using a single moving camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 879–892, May 2017.

[28] A. Carrio, J. Tordesillas, S. Vemprala, S. Saripalli, P. Campoy, and J. P. How, "Onboard detection and localization of drones using depth maps," *IEEE Access*, vol. 8, pp. 30 480–30 490, 2020.

[29] H. Sun, J. Yang, J. Shen, D. Liang, L. Ning-Zhong, and H. Zhou, "TIB-net: Drone detection network with tiny iterative backbone," *IEEE Access*, vol. 8, pp. 130 697–130 707, Jul. 2020.

[30] P. Nguyen, H. Truong, M. Ravindranathan, A. Nguyen, R. Han, and T. Vu, "Cost-effective and passive RF-based drone presence detection and characterization," *GetMobile: Mobile Comp. and Comm.*, vol. 21, no. 4, pp. 30–34, Feb. 2018.

[31] M. Ezuma, F. Erden, C. K. Anjinappa, O. Ozdemir, and I. Guvenc, "Detection and classification of UAVs using RF fingerprints in the presence of Wi-Fi and bluetooth interference," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 60–76, Nov. 2020.

[32] I. Bisio, C. Garibotto, F. Lavagetto, A. Sciarrone, and S. Zappatore, "Blind detection: Advanced techniques for WiFi-based drone surveillance," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 938–946, Jan. 2019.

[33] A. Alipour-Fanid, M. Dabaghchian, N. Wang, P. Wang, L. Zhao, and K. Zeng, "Machine learning-based delay-aware UAV detection and operation mode identification over encrypted Wi-Fi traffic," *IEEE Trans. Inf. Forensic Secur.*, vol. 15, pp. 2346–2360, 2020.

[34] M. F. Al-Sa'd, A. Al-Ali, A. Mohamed, T. Khattab, and A. Erbad, "RF-based drone detection and identification using deep learning approaches: An initiative towards a large open source drone database," *Futur. Gener. Comp. Syst.*, vol. 100, pp. 86–97, Nov. 2019.

[35] M. S. Allahham, T. Khattab, and A. Mohamed, "Deep learning for RF-based drone detection and identification: A multi-channel 1-D convolutional neural networks approach," in *Proc. IEEE Int. Conf. Inform. IoT Enabling Technol. (ICIoT)*, Doha, Qatar, Feb. 2020, pp. 112–117.

[36] S. Abeywickrama, L. Jayasinghe, H. Fu, S. Nissanka, and C. Yuen, "RF-based direction finding of UAVs using DNN," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, Chengdu, China, Dec. 2018, pp. 157–161.

[37] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. - Volume 37*, Lille, France, Jul. 2015, pp. 448–456.

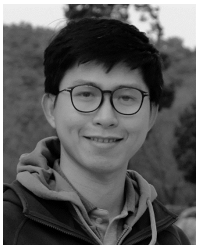
[38] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," *arXiv preprint arXiv: 1511.07289*, 2015.

[39] Y. Ioannou, D. Robertson, R. Cipolla, and A. Criminisi, "Deep roots: Improving CNN efficiency with hierarchical filter groups," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 5977–5986.

[40] Z. Zhang, J. Li, W. Shao, Z. Peng, R. Zhang, X. Wang, and P. Luo, "Differentiable learning-to-group channels via groupable convolutional neural networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct.-Nov. 2019, pp. 3541–3550.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[42] J. Nagi, F. Ducatelle, G. A. Di Caro, D. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Kuala Lumpur, Malaysia, Nov. 2011, pp. 342–347.



THIEN HUYNH-THE (Member, IEEE) received the B.S. degree in electronics and telecommunication engineering from Ho Chi Minh City University of Technology and Education, Vietnam, in 2011, and the Ph.D. degree in computer science and engineering from Kyung Hee University (KHU), South Korea, in 2018. From March to August 2018, he was a Postdoctoral Researcher with KHU.

He is currently a Postdoctoral Research Fellow with ICT Convergence Research Center at Kumoh National Institute of Technology, South Korea. He is awarded with the Superior Thesis Prize by KHU in 2018 and the Golden globe award 2022 for Vietnamese young scientists by Ministry of Science and Technology of Vietnam in 2020. His current research interest includes radio signal processing, digital image processing, computer vision, wireless communications, and deep learning.

QUOC-VIET PHAM (Member, IEEE) received the B.S. degree in electronics and telecommunications engineering from Hanoi University of Science and Technology, Vietnam, in 2013, and the Ph.D. degree in telecommunications engineering from Inje University, South Korea, in 2017. From September 2017 to December 2019, he was with Kyung Hee University, Changwon National University, and Inje University, on various academic positions.

He is currently a Research Professor with Pusan National University, South Korea. He has been granted the Korea NRF Funding for outstanding young researchers for the term 2019–2023. His research interests include convex optimization, game theory, and machine learning to analyze and optimize edge/cloud computing systems and future wireless systems. He received the Best Ph.D. Dissertation Award in Engineering from Inje University, in 2017. He received the top reviewer award from the IEEE Transactions on Vehicular Technology, in 2020. He is an Editor Journal of Network and Computer Applications (Elsevier) and a lead guest editor of the IEEE Internet of Things Journal.

TOAN-VAN NGUYEN (Member, IEEE) received the B.S. degree in Electronics and Telecommunications Engineering and the M.S. degree in Electronics Engineering from HCMC University of Technology and Education, Vietnam, in 2011 and 2014, respectively, and the Ph.D. degree in Electronics and Computer Engineering from Hongik University, South Korea, in 2021. He is currently a Postdoctoral Researcher with the Electrical and Computer Engineering Department at Utah State University, Logan, Utah, USA. His current research activity is focused on the mathematical modeling of 5G networks and machine learning for wireless communications.

DANIEL BENEVIDES DA COSTA (Senior Member, IEEE) was born in Fortaleza, Ceará, Brazil, in 1981. He received the B.Sc. degree in Telecommunications from the Military Institute of Engineering (IME), Rio de Janeiro, Brazil, in 2003, and the M.Sc. and Ph.D. degrees in Electrical Engineering, Area: Telecommunications, from the University of Campinas, SP, Brazil, in 2006 and 2008, respectively. His Ph.D thesis was awarded the Best Ph.D. Thesis in Electrical Engineering by the Brazilian Ministry of Education (CAPES) at the 2009 CAPES Thesis Contest. From 2008 to 2009, he was a Postdoctoral Research Fellow with INRS-EMT, University of Quebec, Montreal, QC, Canada. From 2010 to 2022, he was with the Federal University of Ceará, Brazil. From January 2019 to April 2019, he was Visiting Professor at Lappeenranta University of Technology (LUT), Finland, with financial support from Nokia Foundation. He was awarded with the prestigious Nokia Visiting Professor Grant. From May 2019 to August 2019, he was with King Abdullah University of Science and Technology (KAUST), Saudi Arabia, as a Visiting Faculty, and from September 2019 to November 2019, he was a Visiting Researcher at Istanbul Medipol University, Turkey. From 2021 to 2022, he was Full Professor at the National Yunlin University of Science and Technology (YunTech), Taiwan. Since 2022, he is with the AI & Telecom Research Center at the Technology Innovation Institute (TII), in Abu Dhabi, UAE. He is Editor of several IEEE journals and has acted as Symposium/Track Co-Chair in numerous IEEE flagship conferences.

DONG-SEONG KIM (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, Republic of Korea, in 2003. From 1994 to 2003, he worked as a full-time Researcher at the ERC-ACI, Seoul National University, Seoul. From March 2003 to February 2005, he worked as a Postdoctoral Researcher with the Wireless Network Laboratory, School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA. From 2007 to 2009, he was a Visiting Professor with the Department of Computer Science, University of California at Davis, Davis, CA, USA. He is currently the Director of the ICT Convergence Research Center (Grand ICT Program), Kit Convergence Research Institute supported by Korean Government with the Kumoh National Institute of Technology. His current main research interests include real-time IoT, industrial wireless control networks, networked embedded systems, and fieldbus. He is a Senior Member of the ACM.