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

Digital Object Identifier 10.1109/ACCESS.2023.DOI

CNN-Based UAV Detection and Classification Using Sensor Fusion

HUNJE LEE[1], SUJEONG HAN[1], JEONG-IL BYEON[1], SEOULGYU HAN[1], RANGUN MYUNG[1], JINGON JOUNG[2], (Senior Member, IEEE), AND JIHOON CHOI[1], (Senior Member, IEEE)

[1]School of Electronics and Information Engineering, Korea Aerospace University, Gyeonggi-do 10540, South Korea (e-mail: dlgjswp@kau.kr, paransea03@kau.kr, start2ji@kau.kr, gshan96@kau.kr, audfkddns@kau.kr, jihoon@kau.ac.kr)

[2]School of Electrical and Electronics Engineering, Chung-Ang University, Seoul 06974, South Korea (e-mail: jgjoung@cau.ac.kr)

Corresponding author: Jihoon Choi (e-mail: jihoon@kau.ac.kr).

This research was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIT) under Grants 2021R1A4A2001316 & 2022R1F1A1073999 & 2022R1A2C1003750.

ABSTRACT This paper proposes a detection and classification method for unmanned aerial vehicles, commonly called drones, using sensor fusion schemes. Datasets for drone detection and classification are collected by field measurements of actual drones using the optical camera, radar, and audio microphone as well as obtained from open online databases. In the first stage of the proposed method, drone detection and classification are conducted using the convolutional neural network (CNN) models separately trained by the optical images, radar range-Doppler maps, and audio spectrograms. Then, the CNN output probabilities are combined by the multinomial logistic regression to improve the drone surveillance accuracy through the fusion of the optical, radar, and audio sensors. Numerical simulations are performed with the experimental data and the open datasets. From the results, it is verified that the proposed sensor fusion method can improve the drone detection accuracy by up to 24.6% and can enhance the drone classification accuracy by up to 36.0% in terms of the F-score, compared to individual sensing schemes.

INDEX TERMS UAV detection, UAV classification, sensor fusion, convolutional neural network (CNN), multinomial logistic regression

I. INTRODUCTION

Unmanned aerial vehicles (UAVs), commonly called drones, are widely used in our lives with various applications for military missions, agriculture, entertainment, safety diagnosis, disaster relief, shipping, and wireless communications. With the rapid expansion of the drone industry, we are exposed to potential threats by drones, such as security area invasion, privacy infringement, and destructive terrors. Considering recent advances in UAV flight systems, anti-drone (or counter drone) technologies have been actively investigated to protect essential facilities and areas from accidental or intentional intrusion of drones [1]–[8]. Some civilian drone manufacturers have embedded geofencing software to prevent drones from flying over no-fly and flight-restricted zones, such as government buildings and airports. However, it is formidable and challenging to apply geofencing to military-purpose drones and to enforce flight restrictions on all civilian drones. Therefore, deploying anti-drone systems for protecting security-sensitive areas is very important.

The anti-drone system requires real-time detection of drones, estimation of location, and classification of drone types (or models) to determine if the object is a threatening drone. Practically, it is challenging to detect a drone because of its small size, low flying speed, low altitude, low radar cross section (RCS), and low vibration. To overcome these difficulties, several surveillance techniques have been devised based on the video sensor, radar sensor, acoustic sensor, and radio frequency (RF) receiver. Each detection method has complementary advantages and disadvantages. Drone detection using video images is a sort of object detection problem which has been extensively studied in the field of pattern recognition and computer vision, and many research results have been reported based on image features such as colors, line shapes, geometric forms, and edges [9]–[12], as well as based on motion features such as the object velocity, moving direction, and flight pattern [13]–[15]. Whereas optical cameras provide low-cost detection and fine-grained tracking of drones, there are shortcomings like the relatively short detection range, high sensitivity to weather conditions, and invisibility by obstacles. In an attempt to find drones under low light conditions, thermal infrared cameras detect the heat

emitted from motors, batteries, and internal hardware [16], [17]. Thermal detection enables drone surveillance at night, yet the practical detection range is significantly shorter than other surveillance methods.

Meanwhile, radar is widely used for the surveillance of large aircraft, yet it is not easy to detect drones with radar due to the limited RCS, low speed, and low altitude. By virtue of recent advances in radar system technology, it is possible to detect extremely small targets including drones [18]. Radar surveillance is a promising technology due to the long detection range, the high position accuracy, the weather independence, the capability for multi-target detection, and the night operability [19]–[23]. For example, the micro-Doppler signatures caused by the rotation of rotors and propellers can be used to detect and classify small drones with high accuracy [24], [25]. Further research has been conducted to improve the detection granularity via the multi-channel passive radar [26], [27] and to provide more advanced features like high resolution and phase interferometry through the frequency-modulated continuous wave (FMCW) radar [28]–[30]. Despite these advantages, high-power radar is strictly regulated in crowded urban areas, and the drone detection radar necessitates a relatively high cost for installation and operation. Thus, it is difficult to construct an anti-drone system only using radar except in the government and military areas.

Alternatively, we can exploit the sounds emitted from the rotors and propellers including inherent drone features. Acoustic drone detection can be accomplished with a single microphone [31] and multiple microphones [32] by analyzing the acoustic signatures in the time and/or frequency domains. Various techniques are jointly considered to improve the acoustic detection performance: a noise reduction technique is employed in [33]; machine learning and deep learning approaches are utilized in [31], [34]–[36]; and the use of acoustic sensors equipped with drones has been investigated for target localization in [36]–[38]. In [36], to mitigate the influence of noise, acoustic features are extracted by the short-time Fourier transform (STFT) in combination with convolutional neural networks (CNNs) . Moreover, machine learning is applied, followed by feature extraction methods such as mel-frequency cepstral (MFC) coefficients and linear predictive cepstral coefficients in [31], and similarly, the independent vector analysis is employed for feature extraction from sounds in [39]. Low-cost implementation is possible for anti-drone systems using acoustic sensors, whose detection performance is robust to light and weather conditions and relatively less affected by obstacles. However, the detection range is relatively shorter than other methods (up to a few hundred meters), and the detection accuracy can be significantly degraded in the presence of background noise.

The use of an RF scanner is another promising approach for drone detection. RF scanning devices capture wireless signals for controlling a drone that contains various sensing data for navigation, flight commands, and so on [40]. Commercial drones use RF signals typically in the range of 2.4 GHz to 5 GHz reserved for industrial, scientific, and medical radio bands (ISM bands), which can be detected by the RF scanner [41], [42]. As the frequency used by an illegal drone is usually unknown, an RF scanner hops among multiple frequency bands in order to find a control signal in all possible frequency ranges [43]. RF-based drone detection and identification methods can be further enhanced by using machine learning [44], [45] and deep learning [46]. The RF scanner is robust to weather conditions enabling long-range and low-cost surveillance of drones, if the frequency bands and/or the control protocols are known. However, this method has some limitations for drone detection, if the control information is transferred without following a standard communication specification or autonomous flight is used without communicating between a drone and its controller. Also, the performance can be deteriorated by interferences from other RF signals [47].

Drone detection using individual sensors reveals problems in specific scenarios due to the drawbacks of the aforementioned sensing methods. In an attempt to improve detection performance, sensor fusion technologies have been investigated [48]. The first approach for sensor fusion is to use two or more different sensors simultaneously for accurate and reliable detection. Several sensor fusion techniques are devised by combining optical and acoustic sensors in [49]–[51] and by concatenating signatures of acoustic, optical, and radar sensors in [52]. Both audio and video streams are concurrently used for drone detection by extracting features and feeding to a classifier [50]. A deep neural network (DNN) to process the RF sensing data is concatenated with a CNN to process the visual sensing data to form a combined DNN for sensor fusion [53]. The second approach is to use one sensor for acquisition and the other sensor for verification, that is, one sensor with a more extended range detects the presence of a drone, and the other sensor with higher accuracy confirms the initial detection results by adjusting the parameters such as the angle of arrival (AoA) and the zoom level of the camera. A new procedure for drone detection and tracking is developed based on the fusion of daylight camera, thermal camera, and acoustic sensors in [54], and also, visual and radar sensors are combined for drone detection in [55]. Sensor fusion can enable more reliable, robust, and accurate surveillance for drones in various operating scenarios, yet requiring higher system complexity and deployment cost. For example, multiple sensors need to be synchronized in time for detection targets from the combined sensing data, and parameter optimization for joint detection is required to enhance the performance. In other words, sensor fusion methods necessitate an elaborate design and experimental validation for accelerating the development of a practical anti-drone system.

In this paper, we consider the detection of illegal drones which very rarely communicate with the controller or exploit autonomous flight. To this end, we collect experimental data for drone detection through field measurement using the optical camera, the FMCW radar, and the acoustic microphone. Then, we perform the detection and classification of

drones using individual sensing data and combined data. The contribution of this paper is summarized as follows.

- Through field measurements at Korea Aerospace University (KAU) and Chung-Ang University (CAU), experimental sensing data are obtained for the optical image, the radar range-Doppler map, and the acoustic spectrogram. We collect experimental data on drones in flight using three types of drones with different sizes. Sensing data for non-drone objects are obtained by field measurements and also acquired from open datasets for machine learning [56], [57]. The optical images are scaled considering the input image size of CNN models; the FMCW radar echoes are converted to the range-Doppler map by radar signal processing; and the acoustic signals collected by a microphone are used to form the spectrogram through the MFC filtering and the STFT.
- Deep learning models are developed corresponding to the optical image, the range-Doppler map of FMCW radar, and the spectrogram of acoustic signals for detecting and classifying drones. By employing the transfer learning technique, the first-stage CNN models are modified for drone detection with a single output node for binary classification, and the second-stage CNN models are revised for drone classification with multiple output nodes equal to the number of drone types.
- To improve the detection and classification accuracy, a new sensor fusion method is proposed based on a multinomial logistic regression model [58]. In the training period, the coefficients for combining three kinds of sensors are optimized using the probabilities of three CNN output nodes and the ground trueth label. In the test period, concurrently measured data are used for drone detection and classification based on the trained logistic regression models.
- Numerical simulations are performed with the experimental data and the open datasets based on the CNN models corresponding to individual sensors. Moreover, the probability datasets obtained from the CNN output nodes are exploited to train and test the proposed multinomial logistic regression model for sensor fusion. It is verified that the proposed three-sensor fusion method improves the drone detection accuracy by $1.4\% \sim 24.6\%$ and enhances the drone classification accuracy by $9.4\% \sim 36.0\%$ in terms of the F-score, compared to individual sensing schemes.

From the results, it is verified that the proposed sensor fusion method can improve the detection accuracy by up to 24.6% and can enhance the drone classification accuracy by up to 36.0% in terms of the F-score, compared to individual sensing schemes.

The organization of this paper is as follows. Section II presents the measurement setup to obtain experimental data for individual sensors using commercial drones, and Section III introduces the drone detection and classification techniques using the optical camera, the FMCW radar, and the acoustic microphone. In Section IV, we propose a new sensor fusion method based on the multinomial logistic regression model. Section V presents numerical results for evaluating various drone detection and classification schemes, and Section VI provides concluding remarks and future research issues.

Notations: Superscripts T and -1 denote transposition and inversion, respectively, for any scalar x, vector $\boldsymbol{x}$, or matrix $\boldsymbol{X}$. $\mathbf{1}$ denotes the all-ones column vector; $\mathrm{diag}(\boldsymbol{x})$ returns a diagonal matrix whose main diagonal elements are equal to $\boldsymbol{x}$; $\frac{\boldsymbol{y}}{\boldsymbol{x}}$ stands for elementwise division between vectors $\boldsymbol{y}$ and $\boldsymbol{x}$; and $\frac{\partial \boldsymbol{y}^T}{\partial \boldsymbol{x}}$ means a matrix whose (m,n)th element is $\frac{\partial y_n}{\partial x_m}$ where x_m and y_n are the mth and nth elements of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively.

TABLE 1. Specifications of the drones used in experiments.

Parameter	Inspire2	Mavic3	Phantom4
Size	42.5cm(W) × 42.7cm(L) × 31.7cm(H)	34.75cm(W) × 28.3cm(L) × 10.07cm(H)	28.9cm(W) × 28.9cm(L) × 19.6cm(H)
Weight	3440g	895g	1388g
Material	Composite shell of Mg and Al, Carbon fiber	Polycarbonate, Carbon fiber reinforced nylon	Magnesium alloy
Diagonal length	65.0cm	38.1cm	35.0cm

II. MEASUREMENT SETUP

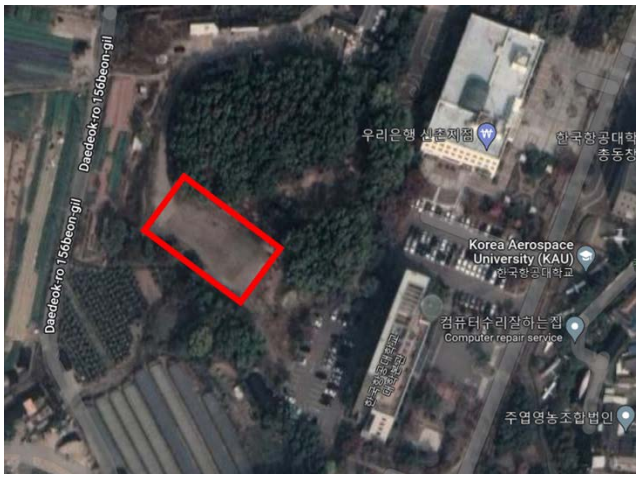
This section provides a description of the location and surrounding environments where the experiments were conducted, the specification of drones, and the setup of sensing devices for detection and classification of drones.

A. LOCATIONS FOR MEASUREMENT

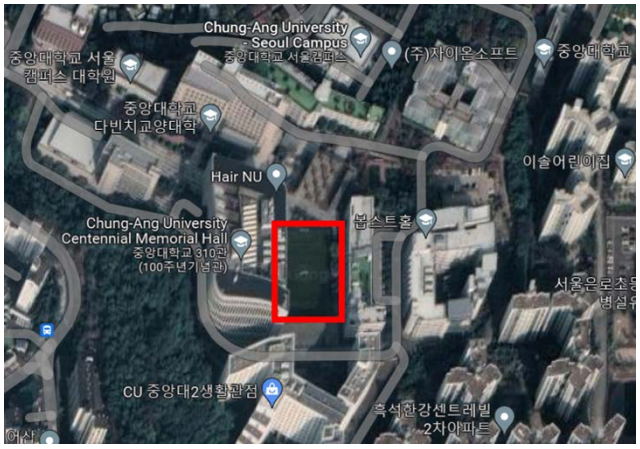
Experiments were mainly conducted in the Drone Airfield at Korea Aerospace University (KAU) located in Goyang-si, Republic of Korea, as shown in Fig. 1(a). The measurement data in this place is influenced by trees and grass surrounding the location. To obtain measurement data from various environments, field experiments were performed in the Futsal Field at Chung-Ang University (CAU) located in Dongjak-gu, Seoul, where the place is surrounded by buildings with more than five stories on three sides, as shown in Fig. 1(b). We collected non-UAV measurement data such as moving cars, people passing the crosswalk, and running people in the street with a 40 m width in South Seoul.

B. DRONES FOR FIELD EXPERIMENTS

In the experiments, three types of drones are used, namely DJI Inspire2, Mavic3, and Phantom4. Each drone is different in size, shape, color, and material, as shown in Table 1. Inspire2 is the largest and heaviest in size and weight and has black and gray colors. The body is made of plastic and magnesium-aluminum alloy, and the arm is made of carbon fiber. Mavic3 is the thinnest and lightest and has black and gray colors. The material comprises plastic, polycarbonate and carbon fiber reinforced nylon. The body of Phantom4 has the same length



(a)



(b)

FIGURE 1. Locations for field measurement: (a) Drone Airfiled at KAU, (b) Futsal Field at CAU.

and width with a medium size and weight, whose color is white. The material is made of plastic and magnesium alloy.

C. SENSING EQUIPMENT

TABLE 2. Specifications of the optical camera used in experiments.

Parameter	Value
Angle of view	77
Effective resolution	4032 × 3024
Focal length	26 mm
Size of image sensor	1/2.55 inch
Pixel size	1.4 um
Wide dynamic range (WDR)	Smart WDR
Color filter	RGB Bayer pattern

Three types of sensors were used in the experiments: optical, radar, and acoustic sensors. For optical sensing, a camera attached to a commercial smartphone was used. As shown in Table 2, the focal length is 26 mm, the resolution is 12 Mpixels (the image size is 4032 × 3024), the image sensor

(a)

(b)

(c)

FIGURE 2. Drones used in experiments: (a) Inspire2, (b) Mavic3, (c) Phantom4.

FIGURE 3. Setup for field measurements using the smartphone camera, the FMCW radar, and the microphone.

size is 1/2.55 inch, and the size per pixel is 1.4 μm.

Table 3 presents the specifications of the radar sensor with multiple-input multiple-output (MIMO) FMCW waveforms which are transmitted and received in the range of [24 GHz, 24.25GHz] frequency band. Two transmit antennas and four receive antennas are located in front of the radar. The radar can detect the range, speed, and angle of multiple targets simultaneously. As an acoustic sensor, a Cardioid microphone was used with a polar pattern. The sensitivity of the microphone is -45 dB, the sampling rate is 48 kHz, and the bit depth is 16 bits, as shown in Table 4.

As shown in Fig. 3, each sensor is attached to a tripod and placed at an identical height. In the case of the radar and microphone, the sensors are controlled by built-in softwares

TABLE 3. Specifications of the radar used in experiments.

Parameter	Value
Waveform	FMCW
MIMO	2 TX × 4 RX antennas
RF output power	8 dBm
Antenna gain	12.6 dBi
Range Resolution	60 cm
3 dB beam width	75° in azimuth, 15° in elevation
Lower/upper frequencies	24.0/24.25 GHz
Chirp repetition interval	1 ms
Upchirp duration	512 μs
Sampling frequency	1 MHz
Number of samples per chirp	512

TABLE 4. Specifications of the acoustic microphone used in experiments.

Parameter	Value
Polar pattern	Cardioid
Sensitivity	-45 dB
Sampling rate	48 kHz
Bit per sample	16 bits
Time interval per recording	5 ∼ 30 sec

and the measured data are saved in Laptop1 and Laptop2 through USB cables. Since the data obtained using the optical sensor is an image, a conversion procedure is unnecessary. The data obtained through the radar and the acoustic sensor are converted into images through signal processing. The echoes received from radar sensors are converted to a range-Doppler map through signal rearrangement and fast Fourier transform (FFT), and the waveform obtained by the acoustic sensor is converted into a spectrogram via the STFT. The signal processing procedure will be explained in the following section.

III. UAV DETECTION AND CLASSIFICATION USING EACH INDIVIDUAL SENSOR

In this section, we explain the procedures for obtaining the optical images, range-Doppler maps, and spectrograms from the field measurement data obtained by the camera, FMCW radar, and microphone, respectively. Then, we present example results corresponding to the optical images, range-Doppler maps, and spectrograms. Moreover, CNN models are employed for detecting and identifying drones using the optical, radar, and audio data, separately.

A. OPTICAL SENSING

The built-in camera of a commercial smartphone was used to obtain the optimal images of drones and non-drone objects such as helicopters, sky, surrounding buildings, background trees, and so on. The size of the image data taken through a smartphone is $4032 \times 3024 \times 3$. Since the drone can be operated until the sun goes down, the experiments were conducted during the daytime, and the images were taken with the sun behind. The focal length was set to 27 mm and 52 mm corresponding to the 2x zoom mode, and the continuous shooting mode of the smartphone was exploited to take as many pictures as possible. The actual drone images were obtained by taking pictures of three kinds of drones in Table 1 during flight, and the non-drone images were achieved by taking pictures of the sky, the helicopters in flight near KAU, the trees around the hill in KAU Drone Airfield, and the surrounding buildings near the CAU Futsal Field.

FIGURE 4. Images obtained by the built-in smartphone camera: (a) Phantom4 drone in flight, (b) Inspire2 drone in forest background with a similar color, (c) Inspire2 drone in building background with a similar color, (d) Military helicopter.

Moreover, additional external images were acquired for drones and non-drone objects from open datasets in [56]. The non-drone images include airplanes, warplanes, helicopters, rockets, and other objects which look like drones seen from a distance. Notice that the external datasets provide a variety of images that is difficult to obtain through measurements.

Figs. 4(a), 4(b), and 4(c) show the drone images taken by the built-in smartphone camera, and Fig. 4(b) presents an image of a military helicopter acquired from the open dataset in [56]. While the Phantom4 drone is clearly recognized in Fig. 4(a), the Inspire2 drone is not well distinguished in Figs. 4(b) and 4(c) due to the background colors similar to the drone. These example images illustrate the disadvantage of drone detection based on optical imaging.

B. RADAR SENSING

We obtain the radar sensing data by *EV-TINYRAD24G* [59]. This radar transmits the rapid chirps waveform, and the received echoes are used to construct the range-Doppler map representing the target range and velocity that can be used for drone detection and classification.

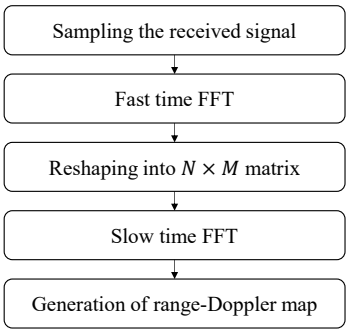


FIGURE 5. Overall procedure for generating the range-Doppler image from FMCW received signals.

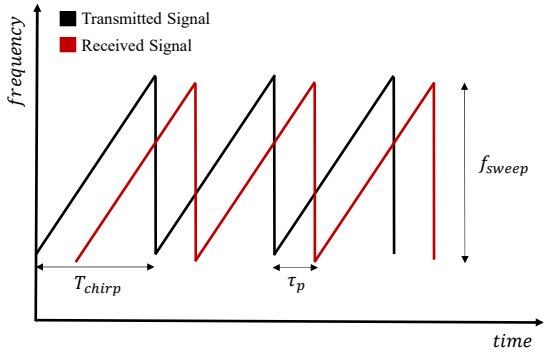


FIGURE 6. FMCW waveform with rapid chirps.

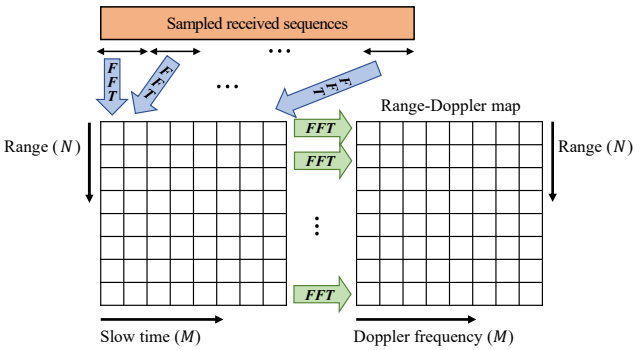


FIGURE 7. Generation of the range-Doppler map using two FFT operations.

Fig. 5 shows the overall procedure for generating the range-Doppler map from the received echoes of FMCW signals. Firstly, the received signal is down-converted to baseband signals and then sampled to form a matrix composed of complex samples. Secondly, the fast time FFT is performed on the columns of the sample matrix, and then the slow time FFT is conducted on the rows in order to create a range-Doppler map corresponding to the sample matrix. The range-Doppler map includes the range and velocity information of the targets, which can be used for drone detection and classification.

Fig. 6 presents a conventional FMCW waveform with rapid chirps which have a very short duration T_{chirp}. By reducing this duration, the frequency components regarding the distance and velocity can be independently estimated, enabling low-complex and high-accuracy radar signal processing. Specifically, the received signal is composed of reflected echoes from multiple targets as follows:

$$r(t) = \sum_{p=1}^{P} r_p(t), \tag{1}$$

where $r_p(t)$ is the received FMCW echoes reflected from the pth target and P is the total number of targets. Suppose that the chirp is expressed as a frequency-modulated signal with instantaneous phase ϕ_i and duration T_{chirp}. By neglecting the noise and clutters, the received signal can be expressed as [19]

$$r(t) = \sum_{p=1}^{P} A_p \sum_{m=0}^{M-1} \cos(\varphi_i(t - mT_{chirp} - \tau_p)) \exp(j2\pi v_p t), \tag{2}$$

where A_p, τ_p, and v_p denote the amplitude, time delay, and Doppler frequency shift corresponding to the pth target, respectively, and M is the number of chirps in the received signal. Here, the instantaneous phase φ_i is given by

$$\varphi_i(t) = 2\pi f_0 t + 2\pi k_f \alpha \frac{t^2}{2}, \tag{3}$$

where f_0 is the lower carrier frequency, k_f is the frequency deviation, and α is the modulator signal amplitude.

The frequency down conversion is separately performed for each in-phase and quadrature component of the received signal. Then, after the lowpass filtering, the baseband signals are accumulated to obtain the beat signal as follows:

$$b(t) = \sum_{m=0}^{M-1} \sum_{p=1}^{P} A_p \exp(j\varphi_{bmp}(t)), \tag{4}$$

where the instantaneous phase of the received beat signal, $\varphi_{bmp}(t)$, is given by

$$\varphi_{bmp}(t) = \varphi_i(t - T_{chirp}) - \varphi_i(t - mT_{chirp} - \tau_p) + 2\pi v_p t. \tag{5}$$

By substituting (3) into (5), we have

$$\varphi_{bmp}(t) = \varphi_0 + 2\pi k_f \alpha \tau_p t + 2\pi v_p t, \tag{6}$$

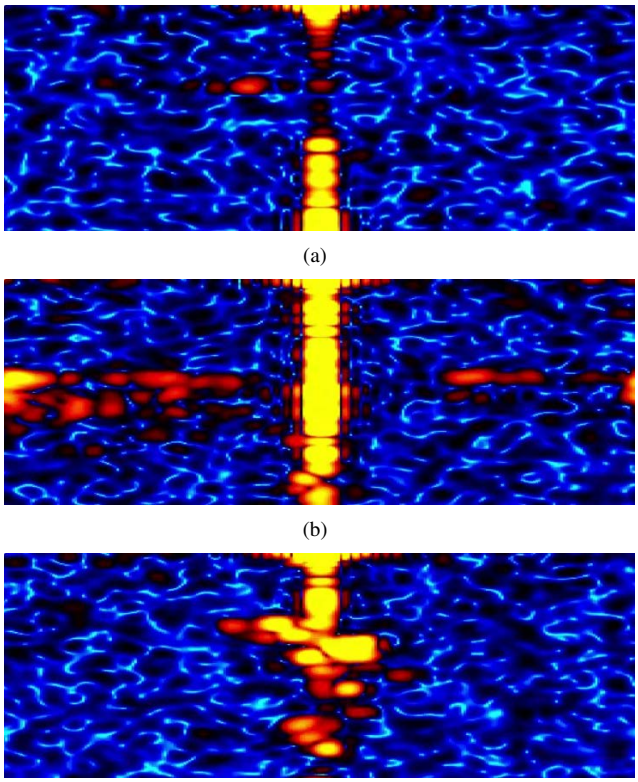
where φ_0 is a constant phase term independent of the time t. By taking the first derivative of $\varphi_{bmp}(t)$ with respect to t, the instantaneous frequency of the beat signal is obtained as

$$f_{bmp} = \frac{1}{2\pi} \frac{d\varphi_{bmp}(t)}{dt} = k_f \alpha \tau_p + v_p. \tag{7}$$

Here, f_{bmp} consists of two components. The first component is proportional to the delay τ_p which can be used for range estimation, and the second component is equal to the Doppler frequency v_p used for target velocity estimation. Thus, (7) can be rewritten as

$$f_{bmp} = \frac{f_{sweep}}{T_{chirp}} \tau_p + v_p = f_{Rp} + f_{Dp}, \tag{8}$$

where f_{Rp} and f_{Dp} are the instantaneous frequencies related



(a) (b) (c)

FIGURE 8. Range-Doppler map obtained from the FMCW radar signals: (a) Inspire2, (b) Vehicles driving on an boulevard, (c) People playing soccer.

to the target range and velocity, respectively. Since the rapid chirp waveform is used, T_{chirp} is very short and $f_{Rp} \gg f_{Dp}$. Thus, we can approximate as $f_{bmp} \approx f_{Rp}$.

The received signal is sampled and the range-matched filtering is performed. Then, the samples are arranged as an $N \times M$ complex matrix, where N is the number of samples during the fast time and M is the number of chirps. The frequency f_{Rp} can be estimated by taking the fast time FFT, i.e., the N-point FFT is carried out for each column to the range direction, and the results are stored in the columns of the $N \times M$ matrix as shown in Fig. 7. The FFT magnitudes are proportional to the amplitudes of targets (if a target exists at the considered frequency). After the FFT, the peaks of columns correspond to the target ranges, and the phases at the peaks of columns are denoted as

$$\varphi_{m,p} = \varphi_{0p} + 2\pi f_{Dp} m T_{chirp}, \tag{9}$$

where $\varphi_{0p} = 2\pi f_0 \tau_p - \pi k_f \alpha \tau_p^2$ is a constant phase independent of the fast time and slow time indexes. The velocity for the pth target is represented as

$$v_p = \frac{c f_{Dp}}{f_0}. \tag{10}$$

f_{Dp} can be estimated by taking the slow time FFT, i.e., the FFT is performed to each row of the $N \times M$ matrix. Note that the phases after the fast time FFT depend on the chirp index m as shown in (9). After the slow time FFT, the resulting peak values are mapped to the Doppler frequencies f_{Dp}, as seen on the right side of Fig. 7. After the fast time and slow time FFTs, the final matrix represents the range-Doppler map with the range and velocity information of targets.

Fig. 8 presents the range-Doppler maps obtained from the measured FMCW radar signals. The intense yellow lines around the zero Doppler frequency are a sort of clutters caused by the leakage of the transmit FMCW signal. In Fig. 8(a), the red spot indicates the Inspire2 drone moving away from the radar sensor at a 20 m distance (i.e., having a negative Doppler frequency), and Fig. 8(b) denotes several vehicles driving on an eight-lane boulevard. Moreover, Fig. 8(c) shows the range-Doppler map obtained from people playing soccer so that several spots are located near the yellow center line. This example implies that the range-Doppler map can be utilized for drone detection.

C. ACOUSTIC SENSING

In the field test, acoustic signals are measured by the microphone of Fig. 3 using actual drone sounds in flight and non-drone sounds such as helicopters, vehicles, human voices, background noises, and so on. Additional non-drone sounds are obtained from the open dataset in [57] including sounds from engines, propellers, aircraft, rain and thunder, air conditioners, and background noises. The measured acoustic data is stored as `.wav` file with 48 kHz sampling rate and 16-bit quantization per sample.

As shown in Fig. 9, the audio file is converted to a mel spectrogram through audio signal processing. A spectrogram is a method for analyzing a sound waveform whose frequency characteristics change over time, which is derived by arranging the spectrum of the acoustic data in the time-frequency domain. Firstly, from the recorded audio waveform, we extract the one-second interval with the highest entropy. When the number of quantization bits is B, the entropy of the audio sequences starting at ℓ is defined as

$$H(\ell) = -\sum_{m=1}^{2^B} p_{m,\ell} \log_2(p_{m,\ell}), \tag{11}$$

where $p_{m,\ell}$ is the empirical probability corresponding to the quantization level q_m, i.e.,

$$p_{m,\ell} = E[x_n = q_m],\ n = \ell, \ell+1, \cdots, \ell+L-1. \tag{12}$$

Here, x_n is the nth input audio sample, $q_m = (m - 0.5 - 2^{B-1})/2^B$, and L is the number of samples during the 1-second interval. From the definition in (11), the audio interval with the highest entropy can be selected as

$$\ell_o = \arg \max_{\ell \in \{1, 1+\Delta\ell, 1+2\Delta\ell, \cdots\}} H(\ell), \tag{13}$$

where $\Delta\ell$ is the index spacing. Note that the candidate starting index ℓ is adjusted by $\Delta\ell$ to reduce complexity.

As a next step, the Hamming window is applied to the extracted audio signals, and then 1440-point STFT is performed with 960 samples of analysis window overlap length. The

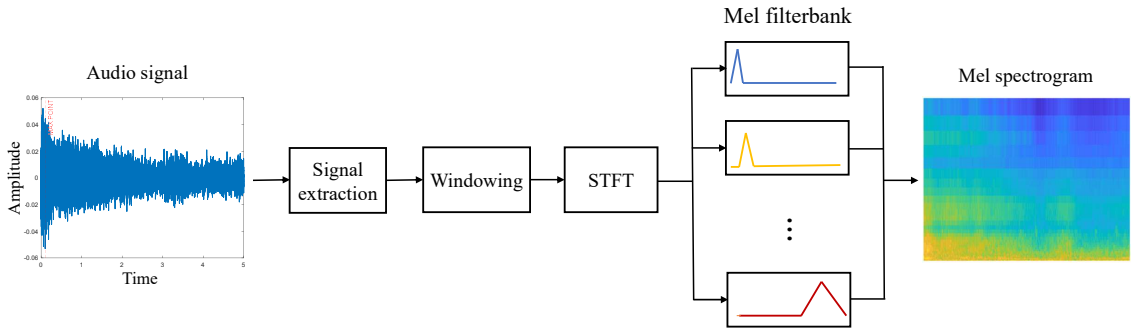


FIGURE 9. Overall procedure for generating the spectrogram from measured acoustic signals.

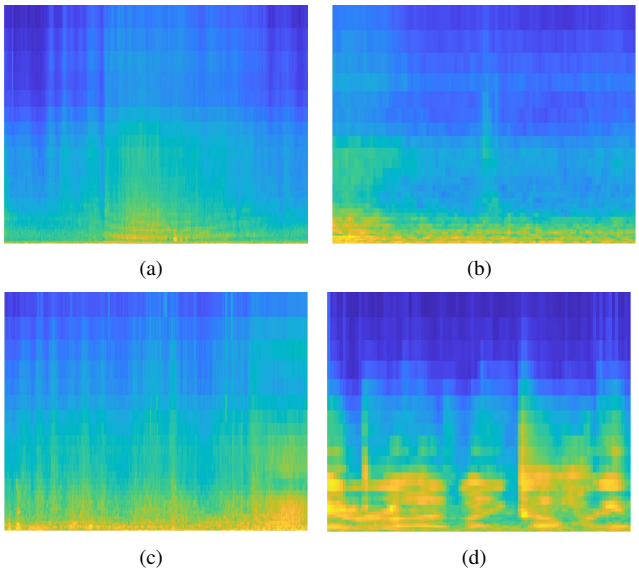


FIGURE 10. Spectrogram obtained from the audio signals: (a) Inspire2, (b) Mavic3, (c) Phantom4, (d) Sound of people conversation.

number of bands for mel filtering is 32 and the number of frames is 89, when the sampling rate is 48 kHz. Finally, the size of the audio mel spectrogram is adjusted to fit the input size for the CNN model that conducts the drone detection or classification. Notice that the mel filter (or mel-scale triangle filter) describes low-frequency bands in more detail while denoting high-frequency bands in less detail. Therefore, the mel filter is used to emphasize the acoustic characteristics of drones in low-frequency bands.

Fig. 10 shows example spectrograms obtained by the audio signal processing in Fig. 9. Figs. 10(a), 10(b), and 10(c) present the spectrograms corresponding to Inspire2, Mavic3, and Phantom4, respectively. It is clearly seen that the spectrogram is different according to the type of drone. Fig. 10(d) is the spectrogram obtained from the sounds of people's conversation, which is totally different from the spectrogram of drones.

D. CNN-BASED DETECTION AND CLASSIFICATION

In this paper, we consider two-step approach composed of drone detection and classification. In the first step, we determine whether it is a drone or a non-drone object from a given image. When a drone is detected in the first step, we identify the type of drone in the second step from the same image. The input image can be one of the optical image, the range-Doppler map, and the audio spectrogram. Drone detection and classification are conducted by exploiting six CNN models that individually adjust the coefficients of neural networks from the training data. Three CNN models are used for drone detection from the optical image, the range-Doppler map, and the audio spectrogram, respectively, and the other three models are utilized for drone classification based on the same input images obtained from three sensors. In a CNN model shown in Fig. 11, convolution layers and pooling layers that perform convolutional operations are repeatedly arranged to extract features of an input image, and the features are used as input data for image processing and sent to the fully connected layer for classification. In this paper, GoogLeNet [60], AlexNet [61], and ResNet-18 [62] are employed for drone detection and classification among the pre-trained CNN models.

As mentioned before, measurement data obtained by the camera, radar, and microphone are used in combination with open datasets for drone detection and classification. Though, the number of images is not enough to train the CNN models, because the CNN models such as GoogLeNet, AlexNet, and ResNet-18 include a lot of parameters for feature extraction[1], image processing, and metric computation for classification. To overcome this problem, we employ transfer learning that partially modifies a pre-trained CNN model for other purposes. As shown in Fig. 12, a pre-trained model is imported, and then some layers are newly configured and modified to produce an output suitable for new tasks. In this paper, the final layers of a pre-trained CNN model are replaced for drone surveillance, and the modified CNN model is trained using the corresponding training images. Through this procedure, six trained CNN models are developed for drone detection

[1]GoogLeNet is composed of 22 layers with 6.8 million parameters, AlexNet has 8 layers with 61 million parameters, and ResNet-18 consists of 18 layers with 11.7 million parameters [60]–[62].

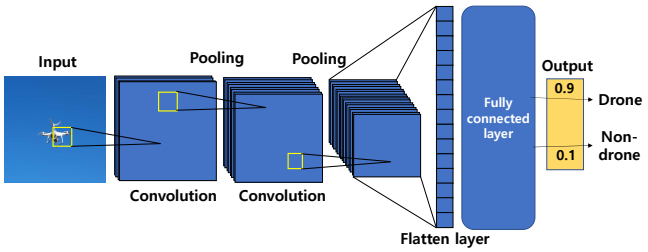


FIGURE 11. Overall processing architecture of a convolutional neural network.

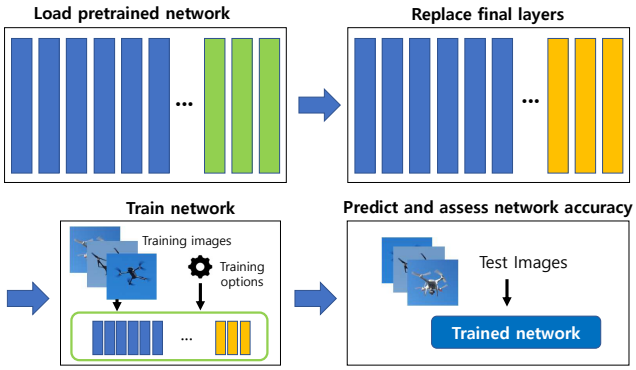


FIGURE 12. Modification of a CNN model for transfer learning.

TABLE 5. Parameters for transfer learning with optical images, radar images, and spectrogram

Item	Optical Image	Radar Image & Spectrogram
Image augmentation	Reflection around x-axis Translation in x-axis and y-axis over [-30, 30]	No reflection, Translation in y-axis over [-30, 30]
Baseline CNN model	GoogLeNet, AlexNet, ResNet-18	
Input image size	GoogLeNet: $224 \times 224 \times 3$ AlexNet: $227 \times 227 \times 3$ ResNet-18: $224 \times 224 \times 3$	
Minimum batch size	128	64
Number of epochs	40	40
Initial learning rate	0.0001	0.0001
Training/Validation ratio	70% / 30%	80% / 20%

and classification with three kinds of sensing images (i.e., an optical image, range-Doppler map, and audio spectrogram).

As shown in Table 5, three kinds of sensing images are converted to $224 \times 224 \times 3$ for GoogLeNet and ResNet-18 and $227 \times 227 \times 3$ for AlexNet to fit the input image size of the pre-trained models. In the case of optical images, we use 70% of the data for training and 30% for verification of the trained CNN model, while we split the radar images and audio spectrograms into 80% and 20% for training and verification, respectively. Image augmentation is used to create more training examples from the measurement data and open datasets. Optical images are reflected around x-axis as well as translated in both x-axis and y-axis over $[-30, 30]$, and radar images and audio spectrograms are translated in y-axis over $[-30, 30]$ without reflection. In addition, the minimum batch size is set to 128 for optical images and 64 for range-Doppler maps and audio spectrograms considering the number of training data, the number of epochs is set to 40, and the initial learning rate is set to 0.0001.

IV. PROPOSED SENSOR FUSION METHOD FOR UAV DETECTION AND CLASSIFICATION

In this section, to improve the surveillance performance, we propose a new drone detection and classification technique that combines multiple sensing schemes. When using the optical, radar, and acoustic sensing data, we can combine the optical image, range-Doppler map, and audio spectrogram to make a decision. For notational convenience, the optical image, radar sensing data, and audio signals are henceforth referred to as *image*, *radar*, and *audio* throughout the paper. For example, the proposed sensor fusion method can combine two sensing data like *image* + *radar*, *image* + *audio*, and *radar* + *audio* as well as three sensing data like *image* + *radar* + *audio*. In the following of this section, we explain the proposed sensor fusion method combining three kinds of sensing data, i.e., *image* + *radar* + *audio*, because two-sensor fusion is a special case of three-sensor fusion.

Fig. 13 presents the overall block diagram for drone detection and classification through sensor fusion of the image, radar, and audio data. As described in Section III-D, three kinds of sensing data are converted to the resized optical image, range-Doppler map, and audio spectrogram through pre-processing, respectively, and an initial drone detection procedure is conducted by the CNN model with individual sensing data. By utilizing the logistic regression model, we combine the initial detection probabilities obtained from three CNN models corresponding to the image, radar, and audio data, and then determine whether a drone is present or not. If a drone is detected, we perform the drone classification procedure. The CNN models for drone classification separately compute the probabilities for Inspire2, Mavic3, and Phantom4 using the same input data as the CNN models for drone detection, i.e., the optical image, range-Doppler map, and audio spectrogram. Finally, the multinomial regression model is employed to obtaine the combined probabilities for drone classification through sensor fusion.

The drone detection based on the sensor fusion is accomplished by the logistic regression model. Given training datasets, the logistic regression model is given by

$$\hat{\boldsymbol{y}} = g(a_0 + a_1\boldsymbol{p}_1 + a_2\boldsymbol{p}_2 + a_3\boldsymbol{p}_3), \tag{14}$$

where $\hat{\boldsymbol{y}}$ is an $N \times 1$ vector predicting the probability for drone presence; $\boldsymbol{p}_1$, $\boldsymbol{p}_2$, and $\boldsymbol{p}_3$ represent the $N \times 1$ probability vectors for training obtained from the CNN models with the image, radar, and audio sensing data, respectively; a_0 is a bias term; a_1, a_2, and a_3 are weight coefficients for the probabilities obtained from the image, radar, and audio CNN models; and N is the number of training datasets. Here, $g(\boldsymbol{x})$

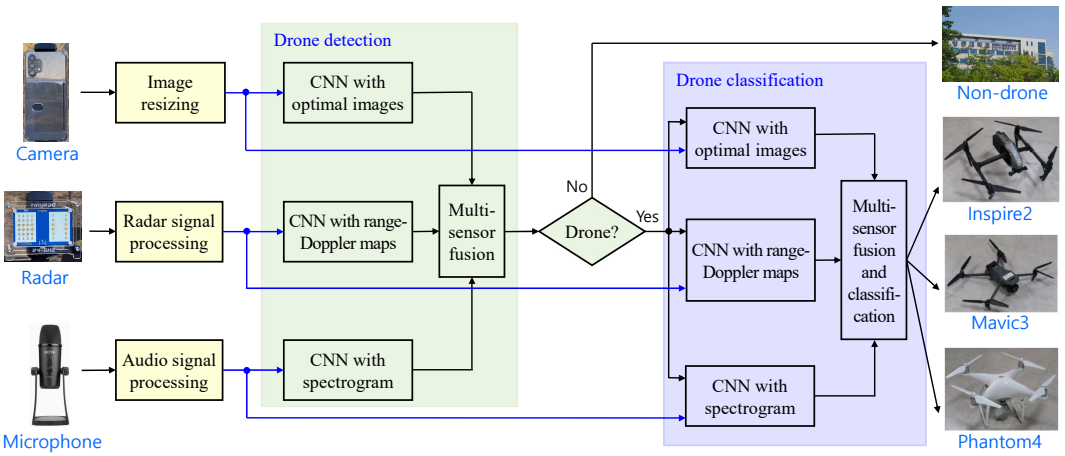


FIGURE 13. Block diagram for drone detection followed by classification through sensor fusion of the image, radar, and audio.

is the sigmoid function defined as

$$g(\boldsymbol{x}) = \frac{\mathbf{1}}{1+e^{-\boldsymbol{x}}}. \tag{15}$$

The logistic regression model in (14) can be rewritten in a vector-matrix form as follows:

$$\hat{\boldsymbol{y}} = g(\boldsymbol{P}\boldsymbol{a}), \tag{16}$$

where $\boldsymbol{P} = [\mathbf{1}, \boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3]$ and $\boldsymbol{a} = [a_0, a_1, a_2, a_3]^T$. To find the optimal coefficients $\{a_0, a_1, a_2, a_3\}$ for sensor fusion, we define the cost function as

$$J(\boldsymbol{a}) = -\frac{1}{N}\left[\boldsymbol{y}^T \log(\hat{\boldsymbol{y}}) + (\mathbf{1}-\boldsymbol{y})^T \log(\mathbf{1}-\hat{\boldsymbol{y}})\right] + \frac{\lambda}{2N}\boldsymbol{a}_0^T\boldsymbol{a}_0, \tag{17}$$

where $\boldsymbol{a}_0 = [0, a_1, a_2, a_3]^T$ and λ is the regularization parameter. By differentiating $J(\boldsymbol{a})$ with respect to $\boldsymbol{a}$, we obtain the gradient as

$$\frac{\partial J(\boldsymbol{a})}{\partial \boldsymbol{a}} = \frac{1}{N}\boldsymbol{D}^{-1}(\hat{\boldsymbol{y}})\frac{\partial \hat{\boldsymbol{y}}^T}{\partial \boldsymbol{a}}\boldsymbol{y} + \frac{1}{N}\boldsymbol{D}^{-1}(\mathbf{1}-\hat{\boldsymbol{y}})\frac{\partial \hat{\boldsymbol{y}}^T}{\partial \boldsymbol{a}}(\boldsymbol{y}-\mathbf{1}) + \frac{\lambda}{N}\boldsymbol{a}_0, \tag{18}$$

where $\boldsymbol{D}(\boldsymbol{x}) = \mathrm{diag}([x_1, x_2, \cdots, x_N])$ for $\boldsymbol{x} = [x_1, x_2, \cdots, x_N]^T$. Here, from the logistic regression model in (16), we have

$$\frac{\partial \hat{\boldsymbol{y}}^T}{\partial \boldsymbol{a}} = \boldsymbol{P}^T\boldsymbol{D}^2(\hat{\boldsymbol{y}})\boldsymbol{D}(\exp(-\boldsymbol{P}\boldsymbol{a})). \tag{19}$$

By substituting (19) into (18), the gradient is expressed as

$$\frac{\partial J(\boldsymbol{a})}{\partial \boldsymbol{a}} = \frac{1}{N}\left[\boldsymbol{P}^T(\hat{\boldsymbol{y}}-\boldsymbol{y}) + \lambda\boldsymbol{a}_0\right]. \tag{20}$$

Using the gradient method, the coefficient vector $\boldsymbol{a}(j)$ at the jth iteration can be updated as

$$\boldsymbol{a}(j) = \boldsymbol{a}(j-1) - \mu\frac{\partial J(\boldsymbol{a})}{\partial \boldsymbol{a}} = \boldsymbol{a}(j-1) - \mu\frac{1}{N}\left[\boldsymbol{P}^T(\hat{\boldsymbol{y}}-\boldsymbol{y}) + \lambda\boldsymbol{a}_0\right]. \tag{21}$$

If an optimal coefficient for sensor fusion, $\boldsymbol{a}^o = [a_0^o, a_1^o, a_2^o, a_3^o]^T$, is determined by (21) from the training datasets, the drone detection probability is computed by using the test datasets as below:

$$\hat{\boldsymbol{y}}^{test} = g(a_0^o + a_1^o\boldsymbol{p}_1^{test} + a_2^o\boldsymbol{p}_2^{test} + a_3^o\boldsymbol{p}_3^{test}), \tag{22}$$

where $\hat{\boldsymbol{y}}^{test}$ is an $M \times 1$ vector representing the probability of drone presence; $\boldsymbol{p}_1^{test}$, $\boldsymbol{p}_2^{test}$, and $\boldsymbol{p}_3^{test}$ are the $M \times 1$ probability vectors for test obtained from the CNN models with the image, radar, and audio sensing data, respectively; and M is the number of test datasets.

For the drone classification based on sensor fusion, we exploit the multinomial logistic regression with the logit model. Given the training datasets, the model for the relative risk is denoted as [58]

$$\log(\boldsymbol{r}_{13}) = b_0 + b_{1,M}\boldsymbol{q}_{1,M} + b_{1,P}\boldsymbol{q}_{1,P} + b_{2,M}\boldsymbol{q}_{2,M} + b_{2,P}\boldsymbol{q}_{2,P} + b_{3,M}\boldsymbol{q}_{3,M} + b_{3,P}\boldsymbol{q}_{3,P} \tag{23a}$$

$$\log(\boldsymbol{r}_{23}) = c_0 + c_{1,M}\boldsymbol{q}_{1,M} + c_{1,P}\boldsymbol{q}_{1,P} + c_{2,M}\boldsymbol{q}_{2,M} + c_{2,P}\boldsymbol{q}_{2,P} + c_{3,M}\boldsymbol{q}_{3,M} + c_{3,P}\boldsymbol{q}_{3,P}, \tag{23b}$$

where $\boldsymbol{q}_{1,*}$, $\boldsymbol{q}_{2,*}$, and $\boldsymbol{q}_{3,*}$ represent the $N_c \times 1$ probability vectors obtained from the CNN models for drone classification with the image, radar, and audio sensing data, respectively; $\boldsymbol{q}_{j,M}$, $\boldsymbol{q}_{j,P}$ mean the probability vectors corresponding to Mavic3 and Phantom4; b_0 and c_0 are bias terms; $\{b_{j,*}\}$ and $\{c_{j,*}\}$ are weight coefficients for logistic regression; N_c is the number of training datasets for drone classification; and $\boldsymbol{r}_{k\ell}$ is given by

$$\boldsymbol{r}_{k\ell} = \left[\frac{P(y_1=k)}{P(y_1=\ell)}, \frac{P(y_2=k)}{P(y_2=\ell)}, \cdots, \frac{P(y_{N_c}=k)}{P(y_{N_c}=\ell)}\right]^T. \tag{24}$$

Here, $k, \ell \in \{1, 2, 3\}$, and $P(y_i = 1)$, $P(y_i = 2)$, and $P(y_i = 3)$ denote the probabilities that the ith observation is Inspire2, Mavic3, and Phantom4, respectively. The equations in (23)

can be rewritten in a vector-matrix form as follows:

$$\log \frac{P(\boldsymbol{y}=1)}{P(\boldsymbol{y}=3)} = \boldsymbol{Q}\boldsymbol{b} \tag{25a}$$

$$\log \frac{P(\boldsymbol{y}=2)}{P(\boldsymbol{y}=3)} = \boldsymbol{Q}\boldsymbol{c}, \tag{25b}$$

where $\boldsymbol{Q} = \begin{bmatrix}\mathbf{1} & \boldsymbol{q}_{1,M} & \boldsymbol{q}_{1,P} & \boldsymbol{q}_{2,M} & \boldsymbol{q}_{2,P} & \boldsymbol{q}_{3,M} & \boldsymbol{q}_{3,P}\end{bmatrix}$, $\boldsymbol{b} = [b_0, b_{1,M}, b_{1,P}, b_{2,M}, b_{2,P}, b_{3,M}, b_{3,P}]^T$, and $\boldsymbol{c} = [c_0, c_{1,M}, c_{1,P}, c_{2,M}, c_{2,P}, c_{3,M}, c_{3,P}]^T$. Following the approach in [58], the problem for finding the optimal coefficients is formulated as the maximum a posterior (MAP) estimation, and can be solved by an iterative procedure such as the gradient-based optimization algorithm [58] and the coordinate descent algorithm [63].

Using $\boldsymbol{b}^o$ and $\boldsymbol{c}^o$, we can predict the probabilities for drone classification given test datasets. From (25), we may write

$$P(\boldsymbol{y}^{test}=1) = P(\boldsymbol{y}^{test}=3)\exp(\boldsymbol{Q}^{test}\boldsymbol{b}^o) \tag{26a}$$

$$P(\boldsymbol{y}^{test}=2) = P(\boldsymbol{y}^{test}=3)\exp(\boldsymbol{Q}^{test}\boldsymbol{c}^o), \tag{26b}$$

where $\boldsymbol{Q}^{test}$ is an $M_c \times 7$ matrix composed of the probabilities obtained from the CNN models using the test datasets corresponding to $\boldsymbol{Q}$ for training. Using (26) and the fact that $P(\boldsymbol{y}=1)+P(\boldsymbol{y}=2)+P(\boldsymbol{y}=3)=\mathbf{1}$, we can predict the probabilities for drone classification as below:

$$P(\boldsymbol{y}^{test}=1) = \frac{\exp(\boldsymbol{Q}^{test}\boldsymbol{b}^o)}{\mathbf{1}+\exp(\boldsymbol{Q}^{test}\boldsymbol{b}^o)+\exp(\boldsymbol{Q}^{test}\boldsymbol{c}^o)} \tag{27a}$$

$$P(\boldsymbol{y}^{test}=2) = \frac{\exp(\boldsymbol{Q}^{test}\boldsymbol{c}^o)}{\mathbf{1}+\exp(\boldsymbol{Q}^{test}\boldsymbol{b}^o)+\exp(\boldsymbol{Q}^{test}\boldsymbol{c}^o)} \tag{27b}$$

$$P(\boldsymbol{y}^{test}=3) = \frac{1}{\mathbf{1}+\exp(\boldsymbol{Q}^{test}\boldsymbol{b}^o)+\exp(\boldsymbol{Q}^{test}\boldsymbol{c}^o)}. \tag{27c}$$

V. NUMERICAL RESULTS

In this section, we evaluate the drone surveillance performance of the proposed sensor fusion method and compare the performance with the schemes based on individual sensors. Specifically, we account for the following methods for drone detection and classification.

- *Image* [10]: Based on the optical images, CNN models are used for drone detection and classification as in [10]. For training and verification, the measured optical images in Section III-A are used along with the open datasets available in [56].
- *Radar* [21]: Based on the range-Doppler maps obtained from the FMCW radar, CNN models are used for drone detection and classification as in [21]. For training and verification, the measured radar signals are converted to range-Doppler maps as explained in Section III-B.
- *Audio* [36]: Based on the audio spectrograms, CNN models are used for drone detection and classification as in [36]. For training and verification, the measured audio signals are converted to spectrograms as in Section III-C and the open datasets in [57] are used as well.
- *Image + Radar*: The proposed sensor fusion method is designed by combining the optical images and the range-Doppler maps.

TABLE 6. Overall datasets for training and verification of the CNNs models corresponding to the optical images, radar range-Doppler maps, and audio spectrograms.

Stage	Sensor	Object	Online Data	Measured Data
Detection	Image	Drone	462	8470
		Non-drone	4353	400
	Radar	Drone	-	13620
		Non-drone	-	10728
	Audio	Drone	1225	1233
		Non-drone	4478	300
Classification	Image	Inspire2	-	3576
		Mavic3	-	2665
		Phantom4	-	2229
	Radar	Inspire2	-	7085
		Mavic3	-	2777
		Phantom4	-	3758
	Audio	Inspire2	-	426
		Mavic3	-	433
		Phantom4	-	374

TABLE 7. Open datasets obtained from [56], [57].

Type	Drone	Non-Drone Object	
Image	462	Airplane	1611
		Helicopter	1730
		Warplane	751
		Rocket	261
Audio	1225	Vehicle engine	837
		Aircraft propeller	402
		Rain & Thunder	320
		Air conditioner	873
		Background noise	2346

- *Image + Audio*: The proposed sensor fusion method is utilized by combining the optical images and the audio spectrograms.
- *Radar + Audio*: The proposed sensor fusion method is used by combining the range-Doppler maps and the audio spectrograms.
- *Image + Radar + Audio*: The proposed sensor fusion method in Section IV is fully implemented by combining the optical images, the range-Doppler maps, and the audio spectrograms.

We employed pre-trained CNN models provided by MATLAB deep learning toolbox, and modified the CNN models via transfer learning as described in Section III-D. Table 6 presents the overall datasets for training the CNN models with the optical images, range-Doppler maps, and audio spectrograms. In the case of the optical sensing, a total of 13685 datasets were used including 8870 field measurement images and 4815 online datasets in [56]. Through actual measurements, we obtained 400 non-drone images and 8470 drone images composed of 3576, 2665, and 2229 datasets for Inspire2, Mavic3, and phantom4, respectively. In the case of the radar sensing, a total of 24348 datasets were obtained through field measurements, i.e., 10728 range-Doppler maps for non-drone objects, 7085 datasets for Inspire2, 2777 datasets for Mavic3, and 3758 datasets for Phantom4. Note

TABLE 8. Number of datasets for coefficient training and final test in the proposed sensor fusion method based on the multinomial logistic regression.

Type	Class	Coeff. Training	Final Test
Detection	drone	1200	900
	non-drone	450	300
Classification	Inspire2	450	300
	Mavic3	450	300
	Phantom4	450	300

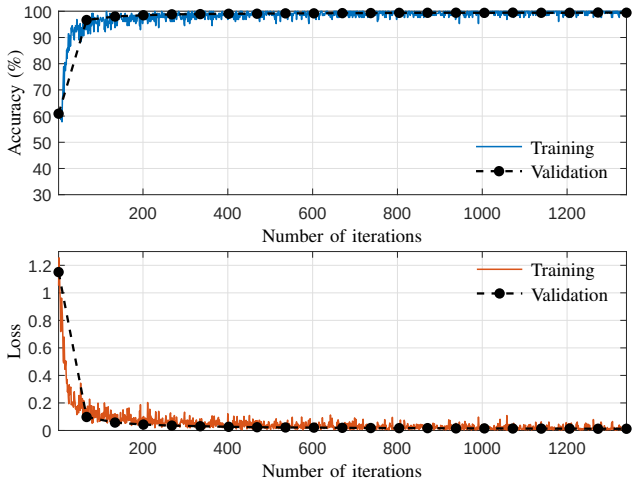


FIGURE 14. Learning curves of the GoogLeNet used for drone detection with optical images.

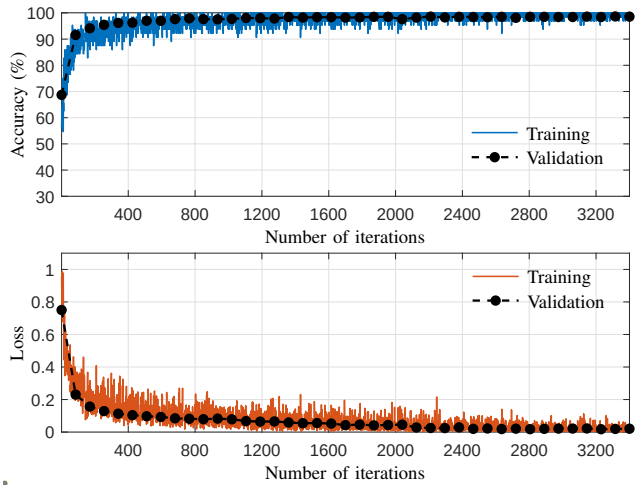


FIGURE 15. Learning curves of the GoogLeNet used for drone detection with radar range-Doppler maps.

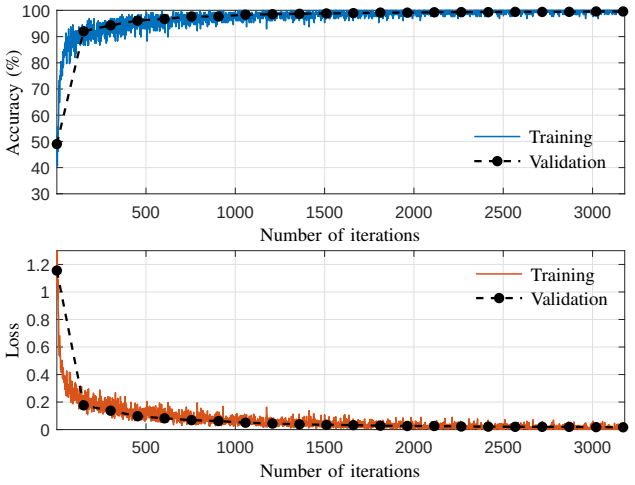


FIGURE 16. Learning curves of the GoogLeNet used for drone detection with audio spectrograms.

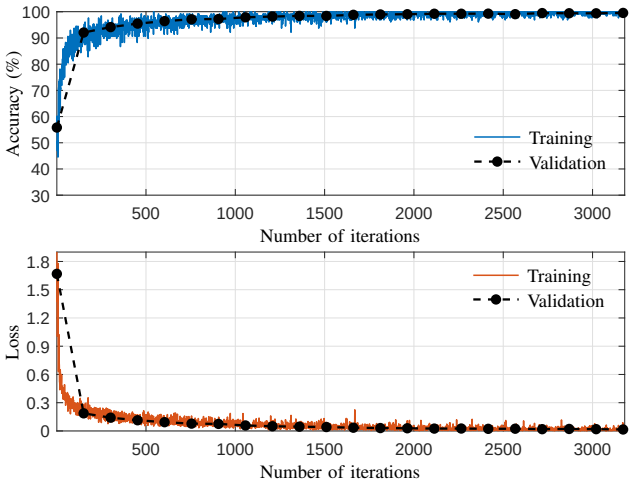


FIGURE 17. Learning curves of the GoogLeNet used for drone classification with optical images.

that open datasets were not used for radar sensing because it is difficult to find range-Doppler images matching the FMCW radar specifications used in our experiments. In the case of the acoustic sensing, a total of 7236 spectrograms were used including 1533 actual measurement datasets and 5703 online datasets in [57]. In the field measurements, we acquired 300 audio datasets for non-drone objects, 426 datasets for Inspire2, 433 datasets for Mavic3, and 374 datasets for Phantom4. As shown in Table 7, the open datasets for optical images consist of images for airplanes, helicopters, warplanes, and rockets similar to drone images in flight. Also, the open datasets for acoustic signals include the sounds of vehicle engines, aircraft propellers, rain and thunder, air conditioners, and various background noises. It is noticeable that all datasets for drones and non-drone objects are utilized when the CNN models are applied to drone detection and only the datasets for drones are used to the CNN models for drone classification.

The proposed sensor fusion method requires datasets concurrently measured from the camera, radar, and microphone. Table 8 describes the number of datasets for coefficient training and the final test in the multinomial logistic regression model obtained by the field measurements. For drone detection, we used 900 drone datasets and 300 non-drone datasets in the training mode to find the optimal coefficients, and performed the final test for 450 drone datasets and 150 non-drone datasets. For drone classification, we used the same

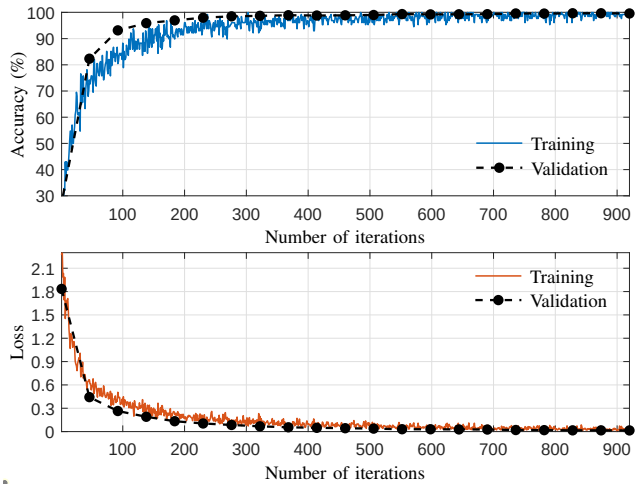


FIGURE 18. Learning curves of the GoogLeNet used for drone classification with radar range-Doppler maps.

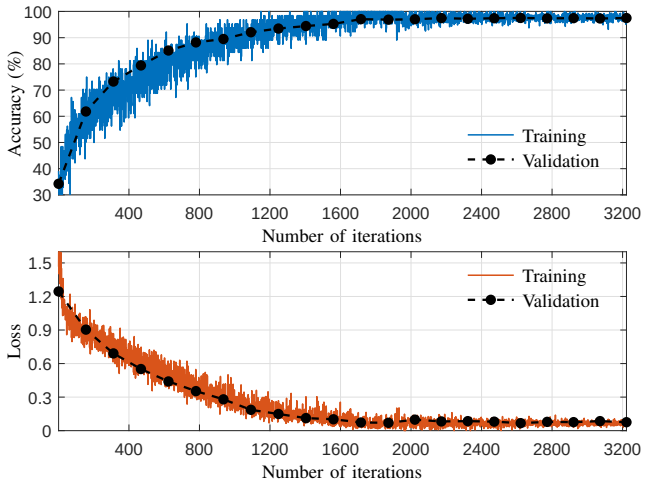


FIGURE 19. Learning curves of the GoogLeNet used for drone classification with audio spectrograms.

drone datasets as those for drone detection. So, we exploited 300 datasets for each type of drone in the training and 150 datasets for each type of drone in the final test.

Figs. 14–16 show the learning curves of the CNN models, which are applied to drone detection using optical images, radar range-Doppler maps, and audio spectrograms, respectively, and Figs. 17–19 present the learning curves of the CNN models used for drone classification. In the simulations, GoogLeNet was used as a pre-trained CNN model and the parameters were set as in Table 5. Overall, the accuracy gradually increases while the loss function gradually decreases as the number of iterations increases. The converging speed is somewhat different depending on the type of sensors and the sort of surveillance (detection or classification), yet the accuracy and the loss function converge to the steady-state values when the number of iterations is greater than 2000 in all cases.

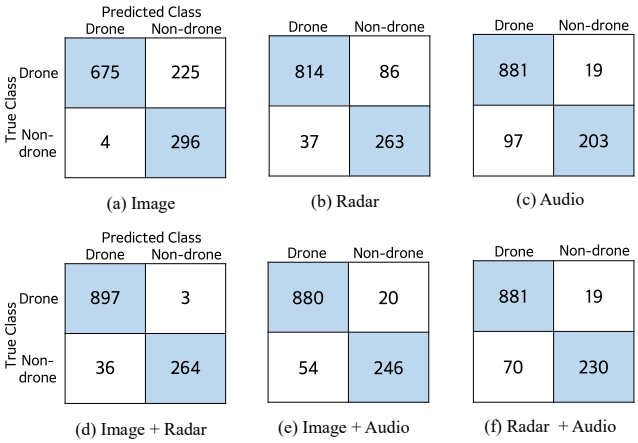


FIGURE 20. Drone detection results using the GoogLeNet with individual sensors and two-sensor fusion.

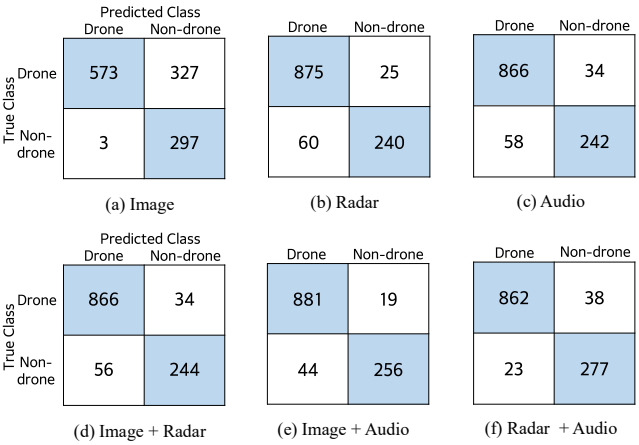


FIGURE 21. Drone detection results using the AlexNet with individual sensors and two-sensor fusion.

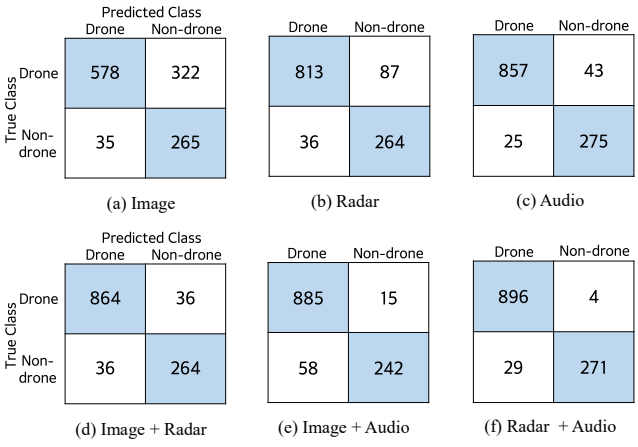


FIGURE 22. Drone detection results using the ResNet-18 with individual sensors and two-sensor fusion.

A. UAV DETECTION RESULTS

Figs. 20–22 show the confusion matrices for drone detection using GoogLeNet, AlexNet, and ResNet-18, respectively, with the input images obtained from the individual sensors and the

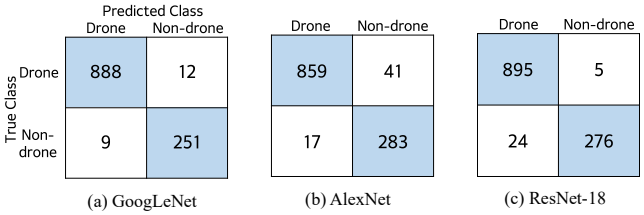


FIGURE 23. Drone detection results using various CNN models with three-sensor fusion (Image+Radar+Audio).

TABLE 9. Detection accuracy of individual and combined sensing methods for drones and non-drone objects, where the bold numbers indicate the highest value in each column (PPV = positive predictive value, TPR = true positive rate).

Method	Detection Accuracy (%)								
	GoogLeNet			AlexNet			ResNet-18		
	PPV	TPR	F-score	PPV	TPR	F-score	PPV	TPR	F-score
Image	**99.4**	75.0	85.5	**99.5**	63.7	77.6	94.3	64.2	76.4
Radar	95.7	90.4	93.0	93.6	97.2	95.4	95.8	90.3	93.0
Audio	90.1	97.9	93.8	93.7	96.2	95.0	97.2	95.2	96.2
Image + Radar	96.1	**99.7**	97.9	93.9	96.2	95.1	96.0	96.0	96.0
Image + Audio	94.2	97.8	96.0	95.2	**97.9**	96.5	93.8	98.3	96.0
Radar + Audio	92.6	97.9	95.2	97.4	95.8	96.6	96.9	**99.6**	98.2
Image + Radar + Audio	99.0	98.7	**98.8**	98.1	95.4	**96.7**	**97.4**	99.4	**98.4**

two-sensor fusion techniques. Fig. 23 presents the results for drone detection using the proposed three-sensor fusion method. Moreover, Table 9 denotes the detection accuracy of individual and combined sensing methods for drones and non-drone objects, where the positive predictive value (PPV) and the true positive rate (TPR) are also called the precision and the recall, respectively. The F-score, F_1, is defined as

$$F_1 = 2\,\frac{PPV \times TPR}{PPV + TPR}. \tag{28}$$

Among individual sensing methods, the optical image has the lowest detection accuracy and the audio sensor achieves the highest detection accuracy in terms of the F-score, because the datasets for verification include drone images with a background color similar to a drone and non-drone images difficult to distinguish from a drone (see Fig. 4). The F-score tends to increase as the number of combined sensors increases, and thus the proposed three-sensor fusion method presents the highest F-score for all CNN models. Specifically, the F-score is improved by $1.4\% \sim 24.6\%$ by the proposed three-sensor fusion method compared to the individual sensing schemes. In the proposed three-sensor fusion method, GoogLeNet achieves the best F-score, while AlexNet has the worst value.

B. UAV CLASSIFICATION RESULTS

Figs. 24–26 denote the confusion matrices for drone classification using GoogLeNet, AlexNet, and ResNet-18, respectively, with the input images obtained from the individual sensors and the two-sensor fusion schemes. Fig. 27 shows the results for drone classification using the proposed three-sensor fusion method. Also, Table 10 presents the accuracy of drone classification for individual and combined sensing techniques.

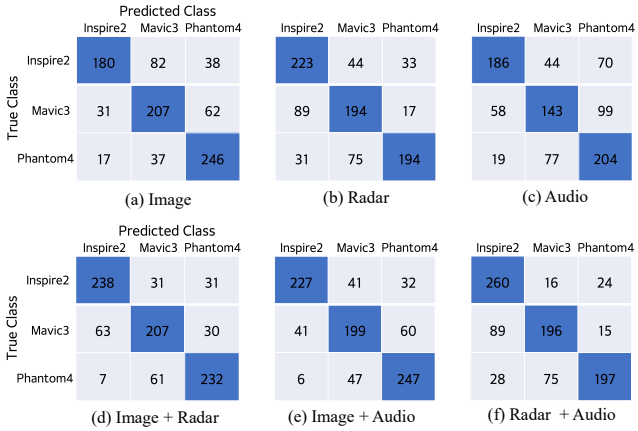


FIGURE 24. Drone classification results using the GoogLeNet with individual sensors and two-sensor fusion.

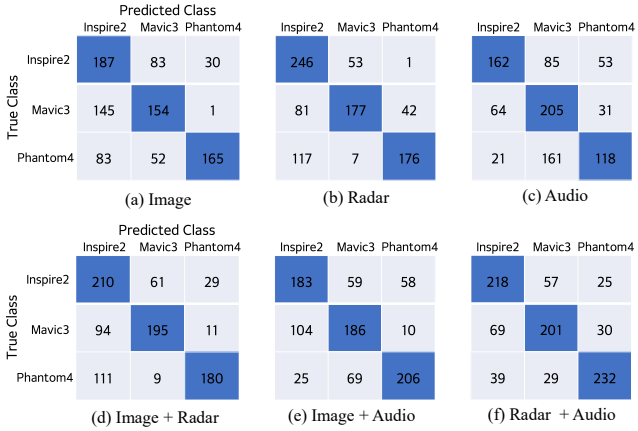


FIGURE 25. Drone classification results using the AlexNet with individual sensors and two-sensor fusion.

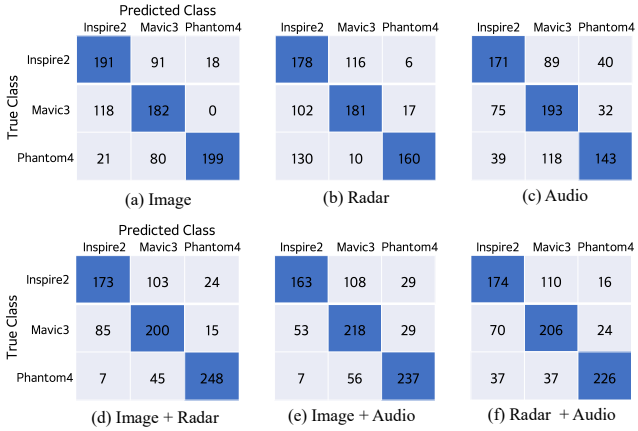
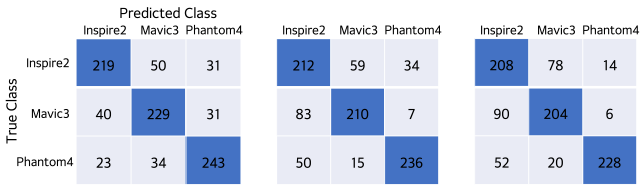


FIGURE 26. Drone classification results using the ResNet-18 with individual sensors and two-sensor fusion.



(a) GoogLeNet (b) AlexNet (c) ResNet-18

FIGURE 27. Drone classification results using various CNN models with three-sensor fusion (Image+Radar+Audio).

TABLE 10. Classification accuracy of individual and combined sensing methods for entire datasets, where the bold numbers indicate the highest value in each column (PPV = average positive predictive value, TPR = average true positive rate, F-score = average of class-wise F-scores).

Method	Classification Accuracy (%)								
	GoogLeNet			AlexNet			ResNet-18		
	PPV	TPR	F-score	PPV	TPR	F-score	PPV	TPR	F-score
Image	71.2	70.3	70.2	60.8	56.2	57.0	67.0	63.6	64.5
Radar	68.8	67.9	68.0	70.2	66.6	66.6	63.3	57.7	58.7
Audio	59.9	59.2	59.1	56.5	53.9	53.6	58.3	56.3	56.4
Image + Radar	75.2	75.2	75.2	68.7	65.0	65.7	69.7	69.0	69.2
Image + Audio	75.0	74.8	74.7	64.4	63.9	64.1	70.2	68.7	68.6
Radar + Audio	73.6	72.6	72.4	72.6	72.3	72.4	68.4	67.3	67.6
Image + Radar + Audio	**76.8**	**76.8**	**76.8**	**73.5**	**72.6**	**72.9**	**73.0**	**71.1**	**71.7**

Here, the average PPV, the average TPR, and the average of class-wise F-scores are defined as

$$P_{avg} = \frac{1}{3}\sum_{k=1}^{3} P_k, \quad R_{avg} = \frac{1}{3}\sum_{k=1}^{3} R_k \tag{29a}$$

$$F_{1,avg} = \frac{1}{3}\sum_{k=1}^{3} F_{1,k}, \tag{29b}$$

respectively, where P_k, R_k, and $F_{1,k}$ are given by

$$P_k = \frac{c_{k,k}}{c_{1,k} + c_{2,k} + c_{3,k}} \tag{30a}$$

$$R_k = \frac{c_{k,k}}{c_{k,1} + c_{k,2} + c_{k,3}} \tag{30b}$$

$$F_{1,k} = 2\,\frac{PPV_k \times TPR_k}{PPV_k + TPR_k}. \tag{30c}$$

Here, $c_{m,n}$ is the (m,n)th element of the confusion matrix for drone classification. Overall, the classification accuracy in Table 10 is lower than the detection accuracy in Table 9, because distinguishing the type of drone is more difficult problem than determining the presence of a drone. Considering the individual sensors, the three kinds of sensors exhibit similar performance. The proposed two-sensor fusion methods such as Image+Radar, Image+Audio, and Radar+Audio achieve higher F-scores than individual sensing schemes, and the proposed three-sensor fusion method obtains the highest F-score among all drone classification techniques. Specifically,

TABLE 11. Classification accuracy of individual and combined sensing methods for Dataset-A, where the bold numbers indicate the highest value in each column.

Method	Classification Accuracy (%)								
	GoogLeNet			AlexNet			ResNet-18		
	PPV	TPR	F-score	PPV	TPR	F-score	PPV	TPR	F-score
Image	90.8	87.8	87.5	67.2	61.3	62.4	73.8	72.9	73.1
Radar	75.6	72.9	73.2	74.1	69.6	70.4	67.4	66.0	66.0
Audio	69.3	67.8	67.9	69.8	67.3	67.1	68.6	64.0	65.1
Image + Radar	90.6	90.7	90.6	82.3	80.4	80.9	81.3	79.8	79.8
Image + Audio	92.8	91.6	91.3	72.0	67.8	68.2	79.3	79.3	79.3
Radar + Audio	80.6	80.2	80.2	85.1	84.9	84.9	79.0	77.3	77.7
Image + Radar + Audio	**95.1**	**95.1**	**95.1**	**86.3**	**85.1**	**85.4**	**84.3**	**83.3**	**83.5**

TABLE 12. Classification accuracy of individual and combined sensing methods for Dataset-B, where the bold numbers indicate the highest value in each column.

Method	Classification Accuracy (%)								
	GoogLeNet			AlexNet			ResNet-18		
	PPV	TPR	F-score	PPV	TPR	F-score	PPV	TPR	F-score
Image	53.3	52.9	52.4	56.5	51.1	51.9	62.2	54.2	56.2
Radar	60.5	60.2	60.3	56.9	55.3	55.4	57.6	51.3	52.0
Audio	55.9	51.8	52.4	55.0	53.8	54.1	51.0	48.9	49.2
Image + Radar	**65.2**	**65.6**	**65.3**	60.2	56.7	57.2	**67.2**	**65.1**	**65.8**
Image + Audio	60.7	60.9	60.6	61.3	61.6	60.7	61.3	54.2	55.8
Radar + Audio	62.0	60.7	61.0	**61.5**	**62.2**	**61.7**	57.5	56.2	56.3
Image + Radar + Audio	58.5	58.4	58.4	61.2	60.3	60.4	64.0	58.9	60.1

the F-score is improved by $9.4\% \sim 36.0\%$ by the proposed three-sensor fusion method compared to the individual sensing schemes. In the proposed three-sensor fusion method, GoogLeNet achieves the best F-score.

To further investigate the results of drone classification, we separate the test datasets into two groups referred to as *Dataset-A* and *Dataset-B*. Dataset-A consists of test datasets with high classification accuracy. Specifically, the distance between the sensor and the drone is less than half the maximum distance for optical, radar, and audio sensing. In optical sensing, the drone altitude is lower than 10 m, and the background is relatively simple like a blue sky. Also, the audio signal is recorded in situations with low background noises. In contrast, Dataset-B is composed of test datasets with low classification accuracy. The distance between the sensor and the drone is greater than half the maximum distance for optical, radar, and audio sensing. In optical sensing, the drone altitude is higher than 10 m, and the background is relatively complicated like many trees and buildings. Moreover, relatively high background noises are included in the audio signals. Tables 11 and 12 denote the classification accuracies for Dataset-A and Dataset-B, respectively. As expected, in

Dataset-A, the individual sensing methods have the worst performance, and the performance is improved as the number of combined sensors increases. Thus, the proposed three-sensor fusion method achieves the best F-score irrespective of pre-trained CNN models. However, in Dataset-B, the proposed three-sensor fusion scheme does not guarantee the best performance. For example, the Image+Radar method has higher F-scores than the Image+Radar+Audio scheme in GoogLeNet and ResNet-18, and the Radar+Audio method presents the highest F-score in AlexNet. These results imply that the initial classification performance obtained by the individual sensors needs to be higher than a certain level to enhance the classification accuracy through sensor fusion.

VI. CONCLUSION

In this paper, we proposed a sensor fusion method for drone detection and classification based on the CNN models for individual sensing and the multinomial logistic regression for combining the optical, radar, and audio sensing data. Through field experiments and numerical simulations, it was verified that the proposed sensor fusion scheme improves drone surveillance performance compared to individual sensing methods. It was also shown that the sensor fusion approach does not guarantee performance enhancement when the individual sensing accuracy is low. Combining multiple sensors for drone surveillance is essential because individual sensing schemes, such as the optical camera, radar, and audio microphone, have complementary advantages and disadvantages. The results presented in this paper can be exploited to optimize the combining algorithm for sensor fusion when designing an anti-UAV defense system to protect security areas.

REFERENCES

[1] D. Floreano and R. J. Wood, "Science, technology and the future of small autonomous drones," *Nature*, vol. 521, no. 7553, pp. 460–466, May 2015.
[2] 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, Jan. 2018.
[3] 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 Commun. Mag.*, vol. 56, no. 4, pp. 68–74, Apr. 2018.
[4] I. Guvenc, F. Koohifar, S. Singh, M. L. Sichitiu, and D. Matolak, "Detection, tracking, and interdiction for amateur drones," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 75–81, Apr. 2018.
[5] H. Kang, J. Joung, J. Kim, J. Kang, and Y. S. Cho, "Protect your sky: A survey of counter unmanned aerial vehicle systems," *IEEE Access*, vol. 8, pp. 168 671–168 710, Sep. 2020.
[6] R. Ben Netanel, B. Nassi, A. Shamir, and Y. Elovici, "Detecting spying drones," *IEEE Security & Privacy*, vol. 19, no. 1, pp. 65–73, Jan. 2021.
[7] S. Park, H. T. Kim, S. Lee, H. Joo, and H. Kim, "Survey on anti-drone systems: Components, designs, and challenges," *IEEE Access*, vol. 9, pp. 42 635–42 659, Mar. 2021.
[8] M. A. Khan, H. Menouar, A. Eldeeb, A. Abu-Dayya, and F. D. Salim, "On the detection of unauthorized drones—techniques and future perspectives: A review," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 11 439–11 455, June 2022.
[9] Z. Zhang, Y. Cao, M. Ding, L. Zhuang, and W. Yao, "An intruder detection algorithm for vision based sense and avoid system," in *Proc. IEEE Int. Conf. Unmanned Aircraft Systems (ICUAS)*, June 2016, pp. 550–556.
[10] D. Lee, W. Gyu La, and H. Kim, "Drone detection and identification system using artificial intelligence," in *Proc. IEEE Int. Conf. Inform. Commun. Technol. Conv. (ICTC)*, Oct. 2018, pp. 1131–1133.
[11] W. Zhou, S. Gao, L. Zhang, and X. Lou, "Histogram of oriented gradients feature extraction from raw bayer pattern images," *IEEE Trans. Circuits Syst. II: Express Briefs*, vol. 67, no. 5, pp. 946–950, May 2020.
[12] K. Kim, J. Kim, H.-G. Lee, J. Choi, J. Fan, and J. Joung, "Uav chasing based on YOLOv3 and object tracker for counter UAV systems," *under review for pulication in IEEE Access*.
[13] F. Gökçe, G. Üçoluk, E. Şahin, and S. Kalkan, "Vision-based detection and distance estimation of micro unmanned aerial vehicles," *IEEE Sensors J.*, vol. 15, no. 9, pp. 23 805–23 846, June 2015.
[14] P. Tang, C. Wang, X. Wang, W. Liu, W. Zeng, and J. Wang, "Object detection in videos by high quality object linking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1272–1278, May 2020.
[15] N. J. Sie, S. Srigrarom, and S. Huang, "Field test validations of vision-based multi-camera multi-drone tracking and 3D localizing with concurrent camera pose estimation," in *Proc. IEEE Int. Conf. Control Robot. Eng. (ICCRE)*, Apr. 2021, pp. 139–144.
[16] P. Andraši, T. Radišić, M. Muštra, and J. Ivošević, "Night-time detection of UAVs using thermal infrared camera," *Transp. Res. Procedia*, vol. 28, pp. 183–190, Feb. 2017.
[17] Y. Wang, Y. Chen, J. Choi, and C.-C. J. Kuo, "Towards visible and thermal drone monitoring with convolutional neural networks," *APSIPA Trans. Signal Inform. Process.*, vol. 8, pp. 1–13, Jan. 2019.
[18] P. Wellig, P. Speirs, C. Schuepbach, R. Oechslin, M. Renker, U. Boeniger, and H. Pratisto, "Radar systems and challenges for C-UAV," in *Proc. IEEE Int. Radar Symp. (IRS)*, June 2018, pp. 1–8.
[19] A. Macaveiu, C. Nafornita, A. Isar, A. Campeanu, and I. Nafornita, "A method for building the range-Doppler map for multiple automotive radar targets," in *Proc. Int. Symp. Electron. Telecommun. (ISETC)*, Nov. 2014, pp. 1–6.
[20] J. Farlik, M. Kratky, J. Casar, and V. Stary, "Radar cross section and detection of small unmanned aerial vehicles," in *Proc. IEEE Int. Conf. Mechatronics - Mechatronika (ME)*, Dec. 2016, pp. 1–7.
[21] E. Kaya and G. B. Kaplan, "Neural network based drone recognition techniques with non-coherent S-band radar," in *Proc. IEEE Radar Conf.*, May 2021, pp. 1–6.
[22] J. Liu, Q. Y. Xu, and W. S. Chen, "Classification of bird and drone targets based on motion characteristics and random forest model using surveillance radar data," *IEEE Access*, vol. 9, pp. 160 135–160 144, Nov. 2021.
[23] V. Semkin, M. Yin, Y. Hu, M. Mezzavilla, and S. Rangan, "Drone detection and classification based on radar cross section signatures," in *Proc. IEEE Int. Symp. Antennas Propag. (ISAP)*, Jan. 2021, pp. 223–224.
[24] S. Björklund, "Target detection and classification of small drones by boosting on radar micro-Doppler," in *Proc. European Radar Conf. (EuRAD)*, Sep. 2018, pp. 182–185.
[25] Y. D. Zhang, X. Xiang, Y. Li, and G. Chen, "Enhanced micro-Doppler feature analysis for drone detection," in *IEEE Radar Conf.*, May 2021, pp. 1–4.
[26] H. Kuschel, D. Cristallini, and K. E. Olsen, "Tutorial: Passive radar tutorial," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 2, pp. 2–19, Feb. 2019.
[27] N. Souli, I. Theodorou, P. Kolios, and G. Ellinas, "Detection and tracking of rogue UASs using a novel real-time passive radar system," in *Proc. Int. Conf. Unmanned Aircraft Syst. (ICUAS)*, June 2022, pp. 576–582.
[28] J. Drozdowicz, M. Wielgo, P. Samczynski, K. Kulpa, J. Krzonkalla, M. Mordzonek, M. Bryl, and Z. Jakielaszek, "35 GHz FMCW drone detection system," in *Proc. IEEE Int. Radar Symp. (IRS)*, May 2016, pp. 1–4.
[29] M. Jian, Z. Lu, and V. C. Chen, "Drone detection and tracking based on phase-interferometric Doppler radar," in *Proc. IEEE Radar Conf.*, Apr. 2018, pp. 1146–1149.
[30] P. K. Rai, H. Idsøe, R. R. Yakkati, A. Kumar, M. Z. Ali Khan, P. K. Yalavarthy, and L. R. Cenkeramaddi, "Localization and activity classification of unmanned aerial vehicle using mmWave FMCW radars," *IEEE Sensors J.*, vol. 21, no. 14, pp. 16 043–16 053, July 2021.
[31] Z. Shi, X. Chang, C. Yang, Z. Wu, and J. Wu, "An acoustic-based surveillance system for amateur drones detection and localization," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2731–2739, Mar. 2020.
[32] J. Guo, I. Ahmad, and K. Chang, "Classification, positioning, and tracking of drones by HMM using acoustic circular microphone array beamforming," *EURASIP J. Wireless Commun. Netw.*, vol. 9, pp. 63 456–63 462, Jan. 2020.
[33] B. Kang, H. Ahn, and H. Choo, "A software platform for noise reduction in sound sensor equipped drones," *IEEE Sensors J.*, vol. 19, no. 21, pp. 10 121–10 130, Nov. 2019.

[34] 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.
[35] S. Al-Emadi, A. Al-Ali, A. Mohammad, and A. Al-Ali, "Audio based drone detection and identification using deep learning," in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, June 2019, pp. 459–464.
[36] Y. Seo, B. Jang, and S. Im, "Drone detection using convolutional neural networks with acoustic STFT features," in *Proc. IEEE Int. Conf. Advanced Video and Signal Based Surveillance (AVSS)*, Nov. 2018, pp. 1–6.
[37] L. Wang and A. Cavallaro, "Acoustic sensing from a multi-rotor drone," *IEEE Sensors J.*, vol. 18, no. 11, pp. 4570–4582, June 2018.
[38] S. V. Sibanyoni, D. T. Ramotsoela, B. J. Silva, and G. P. Hancke, "A 2-D acoustic source localization system for drones in search and rescue missions," *IEEE Sensors J.*, vol. 19, no. 1, pp. 332–341, Jan. 2019.
[39] Z. Uddin, J. Nebhen, M. Altaf, and F. A. Orakzai, "Independent vector analysis inspired amateur drone detection through acoustic signals," *IEEE Access*, vol. 9, pp. 63 456–63 462, May 2021.
[40] P. Nguyen, M. Ravindranatha, A. Nguyen, R. Han, and T. Vu, "Investigating cost-effective RF-based detection of drones," in *Proc. 2nd Workshop Micro Aerial Veh. Netw., Syst., Appl. Civilian Use*, June 2016, p. 17–22.
[41] 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.
[42] P. Flak, "Drone detection sensor with continuous 2.4 ghz ISM band coverage based on cost-effective SDR platform," *IEEE Access*, vol. 9, pp. 114 574–114 586, Aug. 2021.
[43] B. Kaplan, I. Kahraman, A. R. Ekti, S. Yarkan, A. Görçin, M. K. Özdemir, and H. A. Çirpan, "Detection, identification, and direction of arrival estimation of drone FHSS signals with uniform linear antenna array," *IEEE Access*, vol. 9, pp. 152 057–152 069, Nov. 2021.
[44] W. Nie, Z.-C. Han, M. Zhou, L.-B. Xie, and Q. Jiang, "UAV detection and identification based on WiFi signal and RF fingerprint," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13 540–13 550, June 2021.
[45] W. Nie, Z.-C. Han, Y. Li, W. He, L.-B. Xie, X.-L. Yang, and M. Zhou, "UAV detection and localization based on multi-dimensional signal features," *IEEE Sensors J.*, vol. 22, no. 6, pp. 5150–5162, Mar. 2022.
[46] S. Yang, Y. Luo, W. Miao, C. Ge, W. Sun, and C. Luo, "RF signal-based UAV detection and mode classification: A joint feature engineering generator and multi-channel deep neural network approach," *Entropy*, vol. 23, no. 12, pp. 1–20, Dec. 2021.
[47] D.-I. Noh, S.-G. Jeong, H.-T. Hoang, Q.-V. Pham, T. Huynh-The, M. Hasegawa, H. Sekiya, S.-Y. Kwon, S.-H. Chung, and W.-J. Hwang, "Signal preprocessing technique with noise-tolerant for RF-based UAV signal classification," *IEEE Access*, vol. 10, pp. 134 785–134 798, Dec. 2022.
[48] S. Samaras, E. Diamantidou, D. Ataloglou, N. Sakellariou, A. Vafeiadis, V. Magoulianitis, A. Lalas, A. Dimou, D. Zarpalas, K. Votis, P. Daras, and D. Tzovaras, "Deep learning on multi sensor data for counter uav applications—a systematic review," *Sensors*, vol. 19, no. 22, Nov. 2019.
[49] F. Christnacher, S. Hengy, M. Laurenzis, A. Matwyschuk, P. Naz, S. Schertzer, and G. Schmitt, "Optical and acoustical UAV detection," in *Proc. SPIE Electro-Optical Remote Sensing X*, vol. 9988, Oct. 2016, p. 99880B.
[50] H. Liu, Z. Wei, Y. Chen, J. Pan, L. Lin, and Y. Ren, "Drone detection based on an audio-assisted camera array," in *Proc. IEEE Int. Conf. Multimedia Big Data (BigMM)*, Apr. 2017, pp. 402–406.
[51] S. Jamil, Fawad, M. Rahman, A. Ullah, S. Badnava, M. Forsat, and S. S. Mirjavadi, "Malicious UAV detection using integrated audio and visual features for public safety applications," *Sensors*, vol. 20, no. 14, July 2020.
[52] A. Hommes, A. Shoykhetbrod, D. Noetel, S. Stanko, M. Laurenzis, S. Hengy, and F. Christnacher, "Detection of acoustic, electro-optical and RADAR signatures of small unmanned aerial vehicles," in *Proc. SPIE Target and Background Signatures II*, vol. 9997, Oct. 2016, p. 999701.
[53] M. Aledhari, R. Razzak, R. M. Parizi, and G. Srivastava, "Sensor fusion for drone detection," in *Proc. IEEE Vehi. Technol. Conf. (VTC2021-Spring)*, Apr. 2021, pp. 1–7.
[54] F. Svanström, C. Englund, and F. Alonso-Fernandez, "Real-time drone detection and tracking with visible, thermal and acoustic sensors," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 7265–7272.
[55] M. Caris, S. Stanko, W. Johannes, S. Sieger, and N. Pohl, "Detection and tracking of micro aerial vehicles with millimeter wave radar," in *Proc. European Radar Conf. (EuRAD)*, Oct. 2016, pp. 406–408.
[56] [Online]. Available: https://www.kaggle.com/datasets/dasmehdixtr/drone-dataset-uav
[57] [Online]. Available: https://www.kaggle.com/sonain/malicious-uavs-detection
[58] S. Menard, *Applied logistic regression analysis, 2nd ed.* London, United Kingdom: SAGE Publications Ltd., 2002.
[59] Analog Devices, "EV-TINYRAD24G user guide," *UG-1709*, 2020. [Online]. Available: https://www.analog.com
[60] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, June 2015, pp. 1–9.
[61] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, May 2012.
[62] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *10.48550/ARXIV.1512.03385*, Dec. 2015.
[63] H.-F. Yu, F.-L. Huang, and C.-J. Lin, "Dual coordinate descent methods for logistic regression and maximum entropy models," *Machine Learning*, vol. 85, pp. 41–75, Oct. 2011.

HUNJE LEE is supposed to receive the B.S. degree from Korea Aerospace University (KAU), Goyang-si, Republic of Korea, in Feb. 2023. He is with the Intelligent Signal Processing Laboratory (ISPL), the School of Electronics and Information Engineering at KAU, as an undergraduate research assistant, where he has been focusing on field measurement for air-to-ground channels and signal processing based on deep learning techniques for wireless communication systems and UAV detection. His research interests include air-to-ground channel modeling for low altitude UAVs, UAV detection and classification using machine learning and deep learning, and signal processing for learning-based design of future wireless communication systems.

SUJEONG HAN is supposed to receive the B.S. degree from KAU, Goyang-si, Repubic of Korea, in Feb. 2023. She has been focusing on mobile communication and signal processing for wireless communication during her B.S. degree in the School of Electronics and Information Engineering. She is working as an undergraduate research assistant in the ISPL at KAU. Her research interests include air-to-ground channel modeling for low altitude UAVs, UAV detection and classficication using radar and acoustic sensors, and UAV trajectory optimization for delivery and transportation.

JEONG-IL BYEON is supposed to receive the B.S. degree from KAU, Goyang-si, Repubic of Korea, in Feb. 2023, and going to pursue the M.S. degree in KAU from Mar. 2023. Since 2021, he joined the ISPL, the School of Electronics and Information Engineering at KAU, as an undergraduate research assistant, where he performed research on signal processing for synthetic aperture radar (SAR) imaging and compressive sensing algorithms. His research interests include imaging techniques for drone/airborne/spaceborne SAR systems, design of wireless communication systems aided by intellignent reflecting surface (IRS), and signal processing for satellite communications.

SEOLGYU HAN is supposed to receive the B.S degree from KAU, Goyang-si, Republic of Korea, in Aug. 2023. He has been focusing on radar signal processing and image signal processing for computer vision during his B.S. degree in the School of Electronics and Information Engineering. He is working as an undergraduate research assistant in the Media Processing Laboratory at KAU. His research interests include radar signal processing, computer vision, and signal processing for machine learning and deep learning algorithms.

RANGUN MYUNG is supposed to receive the B.S. degree from KAU, Goyang-si, Republic of Korea, in Feb. 2024. She has been focusing on mobile communication and signal processing for wireless communication during her B.S. degree in the School of Electronics and Information Engineering. Her research interests include speech and acoustic signal processing, UAV detection and classification, and signal processing based on machine learning and deep learning.

JINGON JOUNG (S'03–M'07–SM'15) received the B.S. degree in Radio Communication Engineering from Yonsei University, Seoul, South Korea, in 2001, and the M.S. and Ph.D. degrees in Electrical Engineering and Computer Science from KAIST, Daejeon, South Korea, in 2003 and 2007, respectively.

He was a Postdoctoral Fellow with KAIST, South Korea and UCLA, CA, USA, in 2007 and 2008, respectively. He was a Scientist with the Institute for Infocomm Research, Singapore, from 2009 to 2015, and joined Chung-Ang University (CAU), Seoul, South Korea, in 2016, as a faculty member. He is currently an Associate Professor with the School of Electrical and Electronics Engineering, CAU, where he is also the Principal Investigator of the Intelligent Wireless Systems Laboratory. His research interests include wireless communication signal processing, numerical analysis, algorithms, and machine learning.

Dr. Joung was recognized as the Exemplary Reviewers of the *IEEE Communications Letters* in 2012 and the *IEEE Wireless Communications Letters* from 2012 to 2014 and in 2019. He served as the Guest Editor for the *IEEE ACCESS* in 2016. He served on the Editorial Board of the *APSIPA Transactions on Signal and Information Processing* from 2014 to 2019, and served as a Guest Editor for the *MDPI Electronics* in 2019, and served as an Associate Editor for the *IEEE Transactions on Vehicular Technology* from 2018 to 2023. He is an inventor of a *Space-Time Line Code (STLC)* that is a fully symmetric scheme to a Space-Time Block Code, a.k.a. Alamouti code.

JIHOON CHOI (S'99–M'04–SM'18) received the B.S., M.S., and Ph.D. degrees from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1997, 1999, and 2003, respectively.

From 2003 to 2004, he was with the Department of Electrical and Computer Engineering at The University of Texas at Austin, where he performed research on MIMO-OFDM systems as a Postdoctoral Fellow. From 2004 to 2008, he was with the Samsung Electronics, Korea, where he worked on developments of radio access stations for M-WiMAX and base stations for CDMA 1xEV-DO Rev.A/B. In 2008, he joined KAU, Goyang, South Korea, as a faculty member. He is currently a professor with the School of Electronics and Information Engineering, KAU, where he is also the Chief Investigator of the ISPL. His research interests include MIMO communications and signal processing algorithms, secure transmission in the physical layer, radar signal processing, UAV trajectory optimization, mobile edge computing, and modem design for future cellular networks, wireless LANs, IoT devices, and digital broadcasting systems.