Classification of Small Drones Using Low-Uncertainty Micro-Doppler Signature Images and Ultra-Lightweight Convolutional Neural Network

Junhyeong Park, Member, IEEE, and Jun-Sung Park

Abstract—Many studies have attempted to classify small drones in response to threats posed by the technical progress of small drones. Recently, small drones have been classified utilizing convolutional neural networks (CNNs) with micro-Doppler signature (MDS) images generated from frequency-modulated continuous-wave (FMCW) radars. This study proposes a comprehensive method for classifying small drones in real-time using high-quality MDS images and an ultra-lightweight CNN. The proposed comprehensive method comprises an MDS image generation technique, which can improve the accuracy of MDS images generated via FMCW radars, and the ultra-lightweight CNN with high accuracy performance despite its remarkable lightness. Experimental results show that the proposed MDS image generation technique increases the accuracy of CNNs by enhancing the MDS image quality. This is further verified using the results of uncertainty quantification. The proposed ultra-lightweight CNN significantly decreases the computational cost while achieving high accuracy. Finally, we demonstrate that the proposed comprehensive method successfully classifies small drones from far distances with high efficiency and accuracy: the maximum and average accuracies for classification are 100% and 99.21%, respectively, and the numbers of parameters, nodes, and maximum and average accuracies for classification are 4.88K, 21.51K, and 31.52M, respectively.

Index Terms—Classification, convolutional neural network (CNN), frequency-modulated continuous-wave (FMCW) radar, leakage mitigation, lightweight CNN, micro-Doppler signature (MDS) image, small drones, uncertainty.

I. INTRODUCTION

Owing to the technological advances in drones, drones have become miniaturized and smarter. Accordingly, access to small drones has increased with numerous advantages for humans. However, small drones have also resulted in numerous problems. Today, there are issues where privacy is invaded by causing small drones to hover in the sky [1], [2]. A woman experienced threats to the safety and privacy of both her and her children due to a drone that persistently hovered near her dwelling [1]. Residents expressed concerns over invasions of privacy due to drones hovering near their homes and allegedly spying through windows [2]. A woman and her daughter experienced a privacy invasion from a hovering drone spying near their home [3]. Recent advancements in drone technology have significantly simplified the operation of small drones, facilitating their ability to maintain stable hover with minimal effort. Furthermore, most small drones these days are equipped with high-performance cameras. Consequently, anyone with malicious intent can easily, quietly, and secretly shoot illegally from considerable distances, potentially leading to repercussions more severe than those outlined in the preceding instances. To prevent these threats, small drones should be accurately detected and classified. Particularly, the classification is essential to identify their models and apply appropriate countermeasures.

Various methods, including radars, optical cameras, and acoustic systems, are employed for the detection and classification of small drones [4]. Among these, radar stands out due to its advantages, including the capability to operate day and night, its all-weather capability, its ability to ascertain the distance and velocity information of small drones, and its long-range detection capabilities. Recently, the utilization of convolutional neural networks (CNNs) with micro-Doppler signature (MDS) images generated from frequency-modulated continuous-wave (FMCW) radars is emerging as a promising approach for the classification of small drones [5], [6], [7], [8], [9]. In this approach, a CNN is trained to recognize the features of small drones in MDS images. Well-trained CNNs can classify small drones corresponding to arbitrary MDS images.

This study proposes a comprehensive method to implement a cost-effective real-time small drone classification system. The comprehensive method comprises two key components: The first is a novel MDS image generation technique that enhances the quality of MDS images obtained via FMCW radar. The second is an innovative ultra-lightweight CNN designed through a strategic approach to achieve high performance in classifying small drones based on MDS images, despite its remarkable lightness. We provide detailed descriptions of the proposed technique and strategy, substantiating these with reliable experiments and validation.
methods. In particular, the proposed MDS image generation technique is demonstrated through uncertainty quantification as well as classification performance comparisons. The main contributions of this study can be summarized as follows.

1) Development of the real-time, high-quality MDS image generation technique through digital signal processing, utilizing FMCW radar without the need for additional hardware components.

2) Development of the ultra-lightweight CNN based on the proposed lightweight CNN design strategy:
   a) Reduction of the input size for MDS image data.
   b) Utilization of minimal convolutional blocks containing very few relatively large-sized convolution filters for adjusting the receptive field to efficiently focus on capturing key features of the MDS of small drones, and a relatively large-sized pooling filter for efficient spatial dimension reduction.
   c) Utilization of only one fully connected layer, which is naturally required at the output layer.

3) Reliable verification of the proposed methods through the uncertainty quantification and the performance comparison of small drone classification using Monte Carlo cross-validation (MCCV).

The remainder of this paper is organized as follows. In Section II, reviews of previous related studies are provided, along with explanations of the gap between this work and those studies. In Section III, analyses of the conventional MDS image generation technique are provided with problems in FMCW radar systems, and the proposed MDS image generation technique that can resolve these problems is introduced. Additionally, analyses of MDS for the propellers of the small hovering drone are presented. In Section IV, the proposed design strategies for lightweight CNN are explained. In Section V, various experiments are described, which demonstrate the proposed MDS image generation technique, ultra-lightweight CNN, and comprehensive method that applies both of them. The results and discussion of the experiments are presented in Section VI. Finally, the conclusions are drawn in Section VII.

II. RELATED WORKS

Compared with pulse radars, FMCW radars are inexpensive, have a low peak power, and have no minimum detectable range. However, they have an inherent problem called leakage. A dominant leakage signal has enormous power and poor phase noise; thus, it can decrease the sensitivity of FMCW radars [10], [11], [12], [13], [14], [15], [16]. Therefore, the quality of MDS images generated by FMCW radars can also be degraded. Various studies to mitigate the leakage have been conducted [10], [11], [12], [13], [14], [15], [16]. Among them, a technique referred to as stationary point concentration (SPC), which we previously proposed, utilized a novel approach that concentrates the phase noise of the dominant leakage signal on the stationary point of the sinusoidal function, unlike other existing techniques [13], [14]. However, the SPC technique has some limitations, such as the requirement of strategic frequency planning and oversampling, and limitations in the applicable radar architecture. Recently, several techniques to remedy these limitations have been proposed [15], [16]. In [15], we proposed the advanced SPC (A-SPC) technique and demonstrated its ability to resolve those limitations, while improving its advantages. In [16], the authors showed that those limitations can be resolved, and proposed a leakage attenuation method that considered both the phase noise and amplitude noise.

However, these recent techniques were proposed and validated only for simple target detection, and are insufficient for further applications beyond that, such as target classification. This study proposes a novel MDS image generation technique that advances the A-SPC technique to generate high-quality MDS images through FMCW radar for classifying small drones. We analyze the problems of the conventional MDS image generation technique and introduce the proposed technique to resolve these problems. Subsequently, we demonstrate that the classification with high-quality MDS images generated using the proposed technique outperforms the classification with MDS images generated using the conventional technique.

Regarding CNNs, widely utilized CNN structures that had shown optimal performance in famous classification tasks in the field of computer vision were utilized in previous studies [5], [6], [7], [8], [9], [17]. However, these CNN structures were originally designed for the classification of natural images, not MDS images of small drones. In addition, such structures are extremely heavy; thus, small drones cannot be practically classified quickly, or they can increase the cost of radar systems to perform fast classification. Other studies proposed lightweight CNNs for efficient classification of small drones [6], [7], [18]. These CNNs were designed such that many small-sized convolution filters were applied to a hidden layer and many such hidden layers were added. Although these CNNs were lighter than the widely utilized CNNs, the design approaches originated from or were very similar to those applied to popular CNNs for classifying natural images.

Unlike previous studies, we propose a design strategy for lightweight CNN through the analyses of CNNs and MDS images of small drones. Contrary to the conventional approach, the proposed approach encourages the design of a CNN with minimal convolutional blocks containing very few large-sized convolution filters. Based on the proposed design strategy, we propose the ultra-lightweight CNN. The proposed CNN shows high accuracy in classifying small drones using MDS images, despite its remarkable lightness. For more reliable performance evaluation and comparison, we provide sufficient MCCV results, unlike previous studies that relied on k-fold cross-validation (KFCV) results or presented only the best outcome after a few trials.

In [19], we introduced the initial concept of this study. However, at that stage, theoretical analysis was not provided, and the proposed techniques and strategies were incomplete. Furthermore, the verification methods and results were deficient. In this paper, we present a significantly advanced version of the study in all aspects. In-depth theoretical analyses are provided, the completed techniques and strategies are explained in detail, and robust verification results are presented through reliable experiments and verification methods. The MDS image generation technique and the ultra-lightweight
CNN are validated respectively, and ultimately, the comprehensive method integrating both is demonstrated to provide the best performance in classifying small drones.

In addition to the results of classification accuracy improvement, uncertainty quantification results are provided to further prove that the quality of the MDS image is improved by utilizing the proposed MDS image generation technique. Two types of uncertainties exist in machine learning: epistemic uncertainty and is caused by the noise inherent in the data. The AU does not decrease with increasing training data. It can be reduced by improving the data extraction process to decrease the inherent noise and clarify the features in the data [20], [21]. Hence, the effectiveness of the proposed technique in enhancing image data quality can be verified if the AU values from the MDS images generated utilizing the proposed technique are lower than those from the MDS images generated utilizing the conventional technique. In other words, AU quantification is utilized in this study to assess image data quality from the perspective of data to perform artificial intelligence-based missions.

III. THEORETICAL ANALYSES AND PROPOSALS ON MDS IMAGE GENERATION TECHNIQUES

A. Analyses of Conventional MDS Image Generation Technique Considering Problems in FMCW Radars

The theoretical analyses assume a scenario where the radar is installed on the roof of a building and facing the sky, receiving the leakage signal and signals reflected from a small hovering drone. Furthermore, it is assumed that the leakage beat signal is definitively dominant within the beat signals and the stability of the radar system is very poor. After passing through an analog-to-digital converter (ADC) on the FMCW radar, the complex beat signals can be expressed as follows:

\[
x_m[n] = x_{1,m}[n] + jx_{Q,m}[n] = A_{\text{leakage}} \cos(2\pi(f_{\text{carrier}} + f_{\text{offset}} + f_{\text{random,ft}} + f_{\text{beat leakage}})T_sn + \theta_{\text{leakage}})
\]

where \( T_s \) is the sampling interval for the fast-time domain denoted by \( n \), and \( m \) represents the slow-time domain. \( T \) is the sweep period of the linear frequency-modulated (LFM) signal, also known as the ramp or chirp signal. \( A_{\text{leakage}} \) and \( \phi_{\text{leakage, m}}(T_sn) \) are the amplitude and phase noise of the leakage beat signal, respectively, which are the main components that significantly reduce the signal-to-noise ratio (SNR). \( A_{\text{scatterer, i}} \), \( \theta_{\text{D, } i}(m) \), and \( \phi_{\text{scatterer, i, m}}(T_sn) \) are the amplitude, micro-Doppler phase, and phase noise of the beat signals for the \( i \)th point scatterer in the small drone, respectively. \( f_{\text{beat leakage}} \) is the beat frequency of the leakage beat signal, which mainly results from internal delays in FMCW radar systems [13], [15]. \( f_{\text{beat scatterer, } i} \) is the beat frequency of the beat signal from the \( i \)th point scatterer, which contains the distance information. In (1), the homodyne and heterodyne architectures of FMCW radar systems were both considered. \( f_{\text{carrier}} \) is the final carrier frequency in the heterodyne FMCW radar system. This frequency value is determined by the frequency plan when designing the radar system. That is, because it is a known value, it can be removed with relative ease. However, an unwanted Doppler shift occurs when this frequency component is not properly removed. \( f_{\text{offset}} \) which is a problematic frequency component in the practical FMCW radar system, is the unwanted frequency offset of the carrier frequency. \( f_{\text{offset}} \) can occur owing to the instability of the reference oscillator in the phase-locked loop (PLL) for the local oscillator (LO) [28], and can also occur when PLL-based LOs in the radar do not share the same reference oscillator [15]. If the architecture of the FMCW radar system is homodyne, both \( f_{\text{carrier}} \) and \( f_{\text{offset}} \) are removed naturally because the direct down-conversion is performed simultaneously with the deramping process in the homodyne architecture. In (1), some problematic frequency components, \( f_{\text{random, ft}} \) and \( f_{\text{random, st}} \), were also considered for practical FMCW radars. These can be caused by the unlocked phase or frequency of the PLL as well as the instability of other RF components owing to various factors, such as temperature and humidity [29], [30]. \( f_{\text{random, ft}} \) and \( f_{\text{random, st}} \) vary randomly along the fast-time and slow-time domains, respectively. Because the beat signals of the scatterers also pass through radar systems, they contain all these unwanted frequency components. Therefore, these unwanted components cause errors in measuring the distances and velocities of the targets [15]. The last two problematic components of the beat signals in practical FMCW radars are the amplitude imbalance, \( A_{\text{E}} \), and phase imbalance, \( \theta_{\text{E}} \). These components called quadrature imbalance can induce unwanted image signals and reduce the SNR of desired signals. Further
Fourier transform (FT) is applied along the fast-time domain to the beat signals in each \( m \)th slow-time domain to detect the small drone. When the cosine-sum identity, small-angle approximation, and rearrangement according to the relationship between Euler’s formula and the trigonometric function are applied to (1), and if the desired terms and not the image terms are considered, the FT results are as follows:

\[
X_m[k] = X_{\text{beat leakage}, m}[k] + X_{\text{beat scatterers}, m}[k] 
\approx A_{\text{leakage}}(1 + AE^{j\phi_E})e^{j2\pi f_{\text{D, unwanted}} Tm} \delta_m(k - k_{\text{leakage}}) 
+ \frac{A_{\text{leakage}}(1 + AE^{j\phi_E})e^{j2\pi f_{\text{D, unwanted}} Tm}e^{j\frac{q}{2}}}{2} 
\cdot \Phi_{\text{leakage}, m}(k - k_{\text{leakage}}) 
+ \frac{\sum_{i=0}^{l-1} A_{\text{scatterer}, i}(1 + AE^{j\phi_E})e^{j2\pi f_{\text{D, unwanted}} Tm}e^{j\frac{q}{2}}}{2} 
\cdot \delta_m(k - k_{\text{scatterer CT}, i}) 
\]

where \( k \) represents the range index. The range index of the desired target is then selected to apply a short-time Fourier transform (STFT). In the most recent commercial small drones, the length between the two farthest point scatterers is less than 1 m. Therefore, unless the range resolution of the FMCW radar is extremely small, the beat signals of point scatterers would be represented as a single peak in a single range index. Therefore, the range index of a single peak within the desired target area, excluding the leakage peak, is selected to apply STFT along the slow-time domain to obtain the MDS image. In the conventional technique, we express this single range index as \( k_{\text{target CT}} = k_{\text{scatterer CT}, i} \), which represents \( f_{\text{leakage}} + f_{\text{beat scatterer}, i} \). When STFT is applied along the slow-time domain after \( k = k_{\text{target CT}} \) is assigned in (2) and the dc component owing to rigid bodies, such as the shell and frames of a small drone, is removed, the results can be as follows:

\[
\begin{align*}
\mathcal{X}[p, q] 
&= \sum_{m=pS}^{pS+W-1} X_m[k_{\text{target CT}}]e^{-j\frac{2\pi}{W}qm} 
\approx \frac{A_{\text{leakage}}(1 + AE^{j\phi_E})e^{j\frac{q}{2}}}{2} \Phi_{\text{leakage}, p(q - q_{\text{D, unwanted}})} 
\frac{\sum_{i=0}^{l-1} A_{\text{scatterer}, i}(1 + AE^{j\phi_E})X_{\text{D}, i, p(q - q_{\text{D, unwanted}})}}{2} 
\end{align*}
\]

A major cause of MDS image quality degradation

where \( p = 0, 1, 2, \ldots, P - 1 \). \( p \) and \( P \) are the horizontal index and width of the MDS image, \( q \) represents the Doppler index, \( S \) and \( W \) denote the stride and length of the sliding window for the STFT, respectively. \( X_{\text{D}, i, p(q - q_{\text{D, unwanted}})} \) is the STFT result of the micro-Doppler term of \( i \)th point scatterer, \( e^{j\phi_D, i(m)} \), at \( p \) index. Finally, the squared magnitude of (3), \( |\mathcal{X}[p, q]|^2 \), becomes an MDS image, which can also be referred to as a spectrogram. The values of each cell in \( |\mathcal{X}[p, q]|^2 \) are generally normalized into the range \([0, 1]\) by min-max feature scaling, or rescaled according to the selected colormap.

Through (3), various problems of the conventional MDS image generation technique can be analyzed. First, there is a noise term due to the phase noise of the leakage beat signal, \( A_{\text{leakage}}(1 + AE^{j\phi_E})e^{j\frac{q}{2}} \Phi_{\text{leakage}, p(q - q_{\text{D, unwanted}})} / 2 \). Because of its enormous power, the SNR is severely reduced. Therefore, this noise term is one of the main causes of MDS image quality degradation. Second, the quadrature imbalance components, \( AE \) and \( \phi_E \), exist in the desired signal terms. These components could also reduce the SNR of the desired signals in MDS images. Third, an unwanted Doppler shift, \( q_{\text{D, unwanted}} \), occurs because of the problematic frequency components, \( f_{\text{carrier}}, f_{\text{offset}}, \) and \( f_{\text{random}, s} \), mentioned earlier. Because \( q_{\text{D, unwanted}} \) is random owing to \( f_{\text{random}, s} \), the degree of unwanted Doppler shift can vary for each \( p \). If the degree of random variation is severe, it becomes impossible to generate the desired MDS, resulting in the absence of identifiable features for small drones within the image. Furthermore, during the processes in (2) and (3), there can be a problem caused by \( f_{\text{leakage}} \). Because \( f_{\text{leakage}} \) is random owing to \( f_{\text{random}, s} \), it can vary for every \( m \). This \( f_{\text{leakage}} \) is also included in \( f_{\text{scatterer CT}, i} \) as a frequency component of the target beat signal. Thus, the peaks of the target beat signals cannot be properly aligned along the slow-time domain if the degree of its random variation is not negligible, which can influence the SNR of the resulting STFT.

B. Proposed MDS Image Generation Technique

The proposed MDS image generation technique starts by applying the quadrature imbalance correction after the sampling process through the ADC. We utilize a combination of two data-based methods to correct the quadrature imbalance. One method estimates the quadrature imbalance components, \( AE \) and \( \phi_E \), using geometric ellipse-fitting based on the Taubin algebraic method and the Levenberg-Marquardt algorithm [31]. The other method corrects the quadrature imbalance with the estimated values of the quadrature imbalance components through the following transformation [32]:

\[
\begin{bmatrix}
y_{\text{im}}[n] 
y_{\text{Qm}}[n]
\end{bmatrix} = \begin{bmatrix}
1 & -\tan \phi_E & 0 \\
\frac{1}{AE \cos \phi_E} & 0
\end{bmatrix} \begin{bmatrix}
y_{\text{im}}[n] 
y_{\text{Qm}}[n]
\end{bmatrix}
\]
externally through additional equipment, such as a metal rod, metal sphere, motion controller, or milling machine, we utilize the disadvantage of leakage, which has enormous power, as an advantage to estimate the quadrature imbalance components internally. Furthermore, we implement zero-phase filtering using an infinite impulse response (IIR) digital filter before estimating quadrature imbalances. This can prevent the disruption of quadrature imbalance estimation that can arise from the mixing of the leakage signal with significant clutter signals. IIR filters have relatively low computational costs, and their phase distortion issue can be overcome through zero-phase filtering [33]. Consequently, no additional equipment is required, and quadrature imbalance can be estimated and corrected in real-time. When the quadrature imbalance correction in the proposed technique is performed, the resulting complex beat signals can be written as follows:

\[ y_m[n] = y_1,m[n] + jy_2,m[n] \]

\[ = A_{\text{leakage}} \exp(j(2\pi f_{\text{leakage}} T Sn + 2\pi f_D, \text{unwanted} T m + \phi_{\text{leakage}}(m) (T Sn))) \]

\[ + \sum_{i=0}^{l-1} A_{\text{scatterer}, i} \exp(j(2\pi f_{\text{scatterer}} CT, i T Sn + 2\pi f_D, \text{unwanted} T m + \theta_{D, i}(m) + \phi_{\text{scatterer}, i, m}(T Sn))) \]

(5)

The following process aims to achieve leakage mitigation, which includes the effects of frequency and phase correction. First, the range index of the leakage beat signal in \( m \)th slow-time domain is determined as follows:

\[ k_{\text{leakage}}(m) = \arg \max_k |Y_m[k]|^2 \]  

(6)

where \( Y_m[k] \) are the results of NFFT-point fast Fourier transform (FFT) for \( m \)th slow-time domain. Then, the frequency and phase of the leakage beat signal in \( m \)th slow-time domain are estimated as follows:

\[ f_{\text{leakage}}(m) = \frac{F_S}{N FFT} k_{\text{leakage}}(m) \]

\[ \theta_{\text{leakage}}(m) = \angle Y_m[k_{\text{leakage}}(m)] \]  

(7)

where \( F_S \) is the minimum available sampling frequency according to the Nyquist theorem, and \( N FFT \) is the sum of the number of samples and the number of zero pads for zero-padding. If the index number for \( k_{\text{leakage}}(m) \) starts at 1 rather than 0, the frequency of the leakage beat signal is calculated as \( F_S \times (k_{\text{leakage}}(m) - 1)/N FFT \). \( \angle Y_m[k] \) is the phase response of \( Y_m[k] \). Numerical LOs, whose frequency and phase are those of the leakage beat signal, are generated:

\[ LO_m[n] = e^{j(2\pi f_{\text{leakage}}(m) T Sn + \theta_{\text{leakage}}(m))} \]  

(8)

Down-conversion is subsequently performed by multiplying the complex conjugate of (8) with (5), and only the real part of the output of the down-conversion is extracted. The results can be expressed as follows:

\[ z_m[n] = A_{\text{leakage}} \cos(\phi_{\text{leakage}, m}(T Sn)) \]

\[ + \sum_{i=0}^{l-1} A_{\text{scatterer}, i} \cos(2\pi f_{\text{beat scatterer}, i} T Sn + \theta_{D, i}(m) + \phi_{\text{scatterer}, i, m}(T Sn)) \]  

(9)

As the phase noise of the leakage beat signal is concentrated at the stationary point of the sinusoidal function, the magnitude of this noise, which manifests as voltage or current noise, decreases significantly. FT is then applied along the fast-time domain to the beat signals in each \( m \)th slow-time domain to detect the small drone. When the cosine-sum identity, small-angle approximation, and rearrangement according to the relationship between Euler’s formula and the trigonometric function are applied to (9), the FT results are as follows:

\[ Z_m[k] \approx A_{\text{leakage}} \delta_m(k) \]

\[ + \sum_{i=0}^{l-1} A_{\text{scatterer}, i} e^{j\theta_{D, i}(m)} e^{j\frac{\pi}{2}} \delta_m(k - k_{\text{beat scatterer, i}}) \]

(10)

Through the process from (6) to (9), frequency correction, which includes the effect of internal delay compensation, has been performed by removing frequency components of the leakage signal included in the beat frequency of the target. Therefore, the accuracy of the distance measurement of the detected target can be improved, and the range index for the detected target now represents \( f_{\text{beat scatterer, i, not leakage}} + f_{\text{beat scatterer, i}} \). In the proposed technique, we express this range index as \( k_{\text{target PT}} = k_{\text{beat scatterer, i}} \). When STFT is applied along the slow-time domain after \( k = k_{\text{target PT}} \) is assigned in (10) and the dc component owing to the rigid bodies is removed, the results can be expressed as follows:

\[ Z[p, q] \approx \sum_{m=pS}^{pS+W-1} Z_m[k_{\text{target PT}}] e^{-j \frac{2\pi}{W} q m} \]

\[ + \sum_{i=0}^{l-1} A_{\text{scatterer}, i} X_D, i, p(q) \]

\[ + \sum_{i=0}^{l-1} A_{\text{scatterer}, i} e^{j\frac{\pi}{2}} \frac{1}{W} X_D, i, p(q) \oplus \phi_{\text{scatterer}, i, p(q)} \]  

(11)

The MDS image through the proposed MDS image generation technique can be obtained by utilizing the squared magnitude of (11). Then, the values of each cell can be normalized into the range \([0, 1]\) by min-max feature scaling or rescaled according to the selected colormap.

The effects of the proposed technique can be analyzed by comparing (3) and (11). Fig. 1 aids in understanding the limitations of the conventional technique and the effects of the proposed technique. First, the noise term due to the phase
noise of the leakage beat signal, which is primarily responsible for the deterioration of the MDS image quality, has been removed because of the leakage mitigation process in the proposed technique. Therefore, the SNR can be significantly improved across the entire MDS image. Second, the quadrature imbalance correction prevents image signals, caused by quadrature imbalance, from entering the desired target area due to the down-conversion in the proposed technique. Thus, the unnecessary generation of MDS images from unwanted image signals can be avoided. Third, all the problematic frequency and phase components in the target signals have been eliminated owing to the frequency and phase correction in the proposed technique. This improves the accuracy of target distance measurements, resolves the misalignment problem, and prevents unwanted Doppler shifts. Consequently, the proposed MDS image generation technique can enhance the quality of MDS images and improve the clarity of the features for small drone classification.

Despite our analyses of potential issues impacting MDS imaging and the detailed exposition of how the proposed technique addresses these concerns, it should be noted that the scenarios discussed represent extreme cases. In most practical radar systems, the quadrature imbalance, as well as the beat frequency and phase of the leakage, rarely change dramatically over the short time, approximately a few milliseconds to tens of milliseconds, required to collect the beat signals for generating a single MDS image of small drones. Therefore, in the actual implementation of the proposed technique, it is not necessary to estimate and account for those values for every mth beat signal. Fig. 2 shows the method of implementing the proposed technique for the real-time application. For the initial beat signal, the quadrature imbalance components, along with the frequency and phase of the leakage beat signal, are estimated, and a numerical LO is also generated. Then, utilizing these data, the remaining processes of the proposed technique are applied to each mth beat signal required for generating an MDS image. Depending on the stability of the radar system, the optimal cycle for applying the process of estimating necessary components and generating the numerical LO can be adjusted, provided it does not impair real-time performance.

The proposed technique is applicable across various FMCW radars that generate sinusoidal beat signals via the deramping process, achieved by mixing transmitted and received LFM signals. Furthermore, this technique holds potential utility in pulse radars as well. Due to its fundamental operating principle, pulse radar can circumvent leakage issues, rendering the proposed technique generally superfluous. However, there are specific circumstances under which its application may be beneficial. For instance, when applying the stretch processing
to LFM-based pulse radar for high range resolution, which analyzes signals in the frequency domain, substantial clutter within the range window can damage desired target signals in a manner akin to the leakage signal. Therefore, in such cases, the proposed technique may also be effective in pulse radar.

C. Analyses of Micro-Doppler Frequency From Propellers of Small Drone

A radar and a small hovering drone can be depicted as Fig. 3, according to which the distance between the radar and scatterer is related as $R^2 = a^2 + R_{\text{wh}}^2$. Because $a^2 = l^2 + R_{\text{wh}}^2 - 2lR_{\text{wh}}\cos(\omega_0t)$, $R_0^2 = R_{0h}^2 + R_{0v}^2$, and $R_{0h} = R_0\cos\theta$, the distance between the radar and scatterer can be written as follows:

$$R(t) = \sqrt{l^2 + R_0^2 - 2lR_0\cos\theta\cos(\omega_0t)}$$

(12)

The phase and frequency of the micro-Doppler effect in beat signals can be expressed as $\theta_D(t) = 4\pi R(t)/\lambda$, $\omega_D(t) = d\theta_D(t)/2\pi dt$, respectively. Considering (12), the micro-Doppler frequency of a scatterer in the propeller can be written as follows:

$$f_D(t) = \frac{1}{2\pi} \frac{d\theta_D(t)}{dt} = \frac{\pi}{2} \frac{dR(t)}{dt} = \frac{\lambda}{2\pi} \frac{dR(t)}{dt}$$

$$= \frac{\lambda}{2\pi} \frac{1}{\sqrt{l^2 + R_0^2 - 2lR_0\cos\theta\cos(\omega_0t)}}$$

$$\approx \frac{2lR_0\cos\theta\cos(\omega_0t)}{\sqrt{l^2 + R_0^2 - 2lR_0\cos\theta\cos(\omega_0t)}}$$

(13)

According to (13), the period and amplitude of the micro-Doppler frequency over time depend on the rotational frequency of the scatterer and distance from the center of the propeller to the scatterer. In other words, the MDS changes depending on the rotational frequency and propeller length of the small drone. In particular, the rotational frequency influences both the period and amplitude of MDS. Therefore, the classification of small drones can be accomplished by utilizing MDS images. Because the period and amplitude of MDS are the key features in MDS images, the radar parameters should be determined to properly reveal these features and a CNN should be designed to properly detect these features.

IV. PROPOSED LIGHTWEIGHT CNN DESIGN STRATEGY FOR CLASSIFYING SMALL DRONES USING MDS IMAGES

A. Input Layer

In studies in which CNNs are designed focusing on accuracy improvement, there has been a trend of increasing the input size to reveal the features that can discriminate classes as much as possible from the beginning of the neural network. [34], [35]. Accordingly, the width and height of the input data were set to more than 200 pixels, and the number of input data channels was set to at least three RGB in previous studies to classify small drones with CNN [5], [6], [7], [8], [17], [18]. However, as the size of the input increases, not only the number of nodes and floating-point operations (FLOPs) at the input layer but also the numbers of parameters, nodes, and FLOPs of hidden layers increase. Therefore, the size of the input data should be considered carefully to achieve a lightweight CNN. We suggest rethinking whether such a large input size is really necessary, despite the risk of making the neural network complex and heavy. The size of the input to obtain satisfactory accuracy depends on the task and data. In other words, features may already be sufficiently revealed when viewed by a model even with a considerably smaller size than expected for MDS image data, unlike natural image data. Consequently, the use of a small input size is recommended. In this study, we set the width and height of the input data to 64 pixels and assigned only a single channel for the proposed ultra-lightweight CNN; thus, the input size was $64 \times 64 \times 1$.

B. Hidden Layers and Output Layer

1) Convolution: To determine the size of a convolution filter, the following formula can be used:

$$F = \begin{cases} \text{int}(NR), & \text{if int}(NR) \text{ is odd} \\ \text{int}(NR) - 1, & \text{if int}(NR) \text{ is even} \end{cases}$$

(14)

where $F$ is the size of the convolution filter, specifically its width and height. $F$ is an odd number greater than 1. That is, an odd-sized filter is recommended to symmetrically consider the information of neighboring pixels around a pixel. $N$ is the size of the input in a convolution layer and $R$ is a ratio within $(0, 1)$. As analyzed in Section III-C, the key features to identify small drones with MDS images are the period and amplitude of micro-Doppler frequency over time. When considering a convolution filter, suppose small-sized filters, such as sizes from $1 \times 1$ to $7 \times 7$, are simply utilized as in previous studies. In that case, the key features in MDS images may not be adequately detected due to the insufficient spatial coverage of the small-sized filters. Even if many such convolution layers with many small-sized convolution filters are cascaded to increase the global receptive field, this may not effectively capture the key features of MDS, or if it does, this may capture them inefficiently, potentially undermining the lightness of the CNN. Therefore, we propose adjusting the size of the convolution filter to enlarge the receptive field without repeating or cascading convolution layers, thereby enabling...
the model to more efficiently recognize the key features of MDS. Generally, MDS images are generated by determining $P$ in (3) and (11) to contain several MDS cycles. Considering small drones with relatively low rotational frequency, $R$ may need to be greater than 0.5 to enable the model to capture at least one MDS cycle. However, excessively large $R$ makes the CNN heavy. Compromising this trade-off, we recommend setting $R = 0.5$ to determine the size of the convolution filter.

Regarding the number of convolution filters, the classification performance generally improves as the overall number of convolution filters in CNN increases [34]. Each convolution filter captures various detailed features that are difficult for humans to perceive. However, the number of relevant detailed features depends on the task and data. In other words, there may not be as many relevant detailed features in the MDS image as expected to enable the model to identify small drones. Furthermore, increasing the number of convolution filters significantly raises the numbers of parameters, nodes, and FLOPs. Consequently, we propose the utilization of fewer convolution filters for efficiency in the small drone classification task using MDS image data. For the proposed ultra-lightweight CNN, we utilized four convolution filters in the convolution layer.

2) Padding: Padding, which generally utilizes zero pads, enables the convolution to consider information around the edges and prevents the rapid loss of information [36]. Because MDS can be spread around the edges of the MDS image, padding is recommended for the classification of small drones. Zero pads have a value of zero; thus, the padding effect can be implemented without adding zeros around the edges or calculating them in practice. For the proposed ultra-lightweight CNN, we employed padding with zeros such that the output dimension matches the input dimension.

3) Pooling: Pooling reduces the size of feature maps while leaving representative pixels. Two typical pooling methods exist: max and average pooling. Each pooling method returns the maximum and average values of the pixels in the pooling filter, respectively, while reducing the spatial dimension of the feature map. Given that the MDS exhibits transformations such as slight shifts along the time axis in each image and is highlighted within the image, the max pooling can be suitable. Provided that the proposed convolution filters are sufficient to capture the key features of MDS, essential information could be transmitted to the next layer without being lost even if a pooling filter of relatively large size and stride is used. Therefore, while avoiding overly aggressive reductions, we suggest determining a relatively large size and stride for efficient pooling. In this study, we set the size and stride of the pooling filter to $8 \times 8$ and 8, respectively, for the proposed ultra-lightweight CNN.

4) Activation: Various activation functions exist for CNN. Among them, rectified linear unit (ReLU) is suitable for constructing lightweight CNN. Because ReLU passes positive inputs as they are and sets the others to zero, it can be implemented in practice without direct addition or multiplication operations during inference. Furthermore, ReLU suffers considerably less from the vanishing gradient problem compared with other activation functions, such as sigmoid or hyperbolic tangent that has gradients in the range of $(0, 1)$ [37]. Therefore, we recommend utilizing ReLU as an activation function to design a lightweight CNN for the targeted task.

5) Convolutional Block: In this study, a set of layers, which includes convolution, pooling, and activation layers is called a convolutional block. The analysis regarding the number of convolution filters can also be applied when determining the number of convolutional blocks. Most studies designed deep CNNs with numerous convolutional blocks to achieve high accuracy. However, the effectiveness and efficiency of that strategy also depend on the task and data, and it makes CNN complex and heavy. The model may be able to efficiently identify small drones with MDS images using much fewer convolutional blocks than expected. Consequently, we propose the utilization of fewer convolutional blocks for efficiency in the small drone classification task using MDS image data. We utilized only one convolutional block in the proposed ultra-lightweight CNN.

6) Fully Connected Layer and Dropout: Fully connected layers have functions of high-level reasoning and decision-making [37]. However, they can easily and significantly increase the numbers of parameters, nodes, and FLOPs. Therefore, we suggest placing only one fully connected layer after the minimal convolutional blocks. The proposed ultra-lightweight CNN has only one convolutional block and only one fully connected layer. That is, the fully connected layer becomes the output layer in the proposed ultra-lightweight CNN. In addition, we recommend the application of a dropout because it prevents overfitting and improves generalization capability without increasing the number of parameters.

C. Summary of Proposed Lightweight CNN Design Strategy

The proposed lightweight CNN design strategy, unlike previous studies that originate from CNN design strategies for classifying natural images in the field of computer vision, is developed through the analyses of MDS images from small drones. One of the key points of the proposed strategy is not to obsess over using small-sized convolution filters, as previous studies have done, but rather to increase the size of the convolution filters to efficiently enlarge the receptive field. This adjustment enables the model to directly capture the key features within MDS images from small drones. If this adjusted receptive field is sufficient to catch the key features of MDS, then the extensive input size, large number of convolution filters, multiple convolutional blocks, and numerous fully connected layers commonly utilized in previous studies might be deemed excessive, and a significant reduction in these components could be acceptable. Therefore, the proposed strategy significantly lightens the overall CNN by reducing the input size, minimizing the number of convolution filters, convolutional blocks, and fully connected layers, and using relatively large-sized filters and strides for pooling. In other words, the proposed strategy designs the CNN in a way that maximizes the efficiency of very few large-sized convolution filters by concentrating the feature detection functionality for MDS within these filters, which are included in minimal convolutional blocks.
TABLE I
SPECIFICATIONS AND PARAMETERS OF THE FMCW RADAR

| Parameter              | Value               |
|------------------------|---------------------|
| Radar configuration    | Quasi-monostatic    |
| System architecture    | Heterodyne          |
| Operating frequency    | 14.35-14.50 GHz     |
| Transmit power         | 30 dBm              |
| Antenna gain           | 16 dBi              |
| Sweep bandwidth        | 150 MHz             |
| True range resolution  | 1 m                 |
| Final carrier frequency| fcarrier MHz        |
| Sweep period (t)       | 200 µs              |
| Desired digital bandwidth| 2.5 MHz             |
| Sampling frequency (F_S)| 5 MHz               |
| Desired maximum detectable range | 500 m |
| NFFT to estimate f_leakage(m) and f0_leakage(m) | 256 |
| Window for FFT along the last-time domain | Hann |
| Window for STFT along the slow-time domain | Hann |
| NFFT for STFT         | 256                 |
| # of samples (M) to generate one MDS image | 271 (54 ms) |
| Length of sliding window for STFT (W) | 16 (3 ms) |
| Stride of sliding window (S) | 1 |
| Width of MDS image (P) | 256                 |

V. EXPERIMENTS
A. Specifications and Parameters of FMCW Radar

According to analyses in Section III-C, the specifications and parameters of the FMCW radar should be considered to enable the model to successfully detect the key features of the MDS image, which are the period and amplitude of the MDS. Referring to (13), the wavelength of the transmitted signal also influences the amplitude of the micro-Doppler frequency over time. Therefore, the operating frequency of the radar should be increased such that the amplitude changes sensitively according to the length and rotation frequency of the propeller. This renders the features in MDS images for each small drone more distinguishable. Moreover, the reciprocal of the sweep period is the sampling frequency of the slow-time domain, which can control the ambiguity of the Doppler frequency domain. Therefore, a sufficiently unambiguous range should be secured in the Doppler frequency domain by setting a small sweep period. Regarding the parameters of the sliding window for STFT in (3) and (11), W, S, P should be considered. If the sliding window length, W, is excessively long, the instantaneous Doppler frequency cannot be properly presented; whereas, if it is excessively short, the main lobe width of the resulting STFT becomes very large. As the stride of the sliding window, S, increases, the MDS becomes discontinuous. Therefore, S should be set such that the MDS is continuous. The width of the MDS image, P, should be determined such that at least several cycles of the MDS are contained within the image. The specifications and parameters of the FMCW radar set up with all these considerations are listed in Table I.

B. MDS Image Acquisition

The experimental setup is shown in Fig. 4. The FMCW radar, also utilized in [13], [14], and [15], was installed on the rooftop of a building, facing the sky. Commercial small drones, DJI Inspire 1, DJI Inspire 2, and DJI Spark, were utilized as targets. We performed a scenario in which a small drone was hovering for the purpose of spying secretly from far distances. Each small drone was set to hover at various distances and elevation angles. We successfully detected them through the FMCW radar and generated MDS images by utilizing both the conventional and proposed techniques. We utilized MATLAB 2021a for the generation of MDS images. Considering the situation of false detection, we also generated MDS images of random noise as an additional class. If noise can be accurately identified in the classification stage, even if the false detection problem occurs in the detection stage, where noise is detected as a target, it can be resolved. Therefore, each MDS image dataset was prepared to classify four classes. The number of MDS images in each class was 1,400, and the total number of MDS images in each dataset was 5,600. The ratio of training and test sets was set to 8:2.

We normalized the result of the spectrogram to bring all pixel values into the range [0,1], and no colormap was applied. We resized MDS images to 64×64 pixels. Therefore, the input size was 64×64×1 as described in Section IV-A. Some other CNN structures utilized to validate the proposed methods had assigned three RGB channels to the input data in their studies. Therefore, we also prepared MDS images with the three RGB channels and applied the jet colormap, which is commonly utilized for MDS images. We resized these MDS images to fit the respective input sizes used in the studies that proposed each CNN structure. In this study, the INTER_AREA from the cv2 package was used for resizing, and float32 was employed as the data type.

TABLE II
DETAILED STRUCTURE OF THE PROPOSED ULTRA-LIGHTWEIGHT CNN

| Layer      | Output Shape | Description | Value       |
|------------|--------------|-------------|-------------|
| Input      | 64×64×1      |             |             |
| Conv2D     | 64×64×4      | Filter Size | 31×31       |
|            |              | Stride      | 1           |
| MaxPooling2D | 8×8×4     | Filter Size | 8×8         |
| Activation | 8×8×4        | Stride      | 8           |
| Flatten    | 256          | Function    | ReLU        |
| Dropout    | 256          | Dropout Rate| 0.2         |
| Dense      | 4            | Activation  | Softmax     |
TABLE III
HYPERPARAMETERS IN EACH CNN FOR EXPERIMENTS

| CNN   | Optimizer | Learning Rate | Batch Size |
|-------|-----------|---------------|------------|
| VGG16 | Adam      | 1e-4          | 32         |
| ResNet50 | Nadam      | 1e-4          | 32         |
| ConvNeXtTiny | Nadam      | 1e-3          | 32         |
| CNN in [6] | Nadam      | 1e-3          | 32         |
| CNN in [7] | Adam      | 1e-3          | 64         |
| CNN in [18] | Adam      | 1e-3          | 32         |
| Proposed CNN | Nadam      | 1e-3          | 64         |

C. Verification of Proposed Methods

We conducted experiments to demonstrate the effectiveness of the proposed MDS image generation technique, ultra-lightweight CNN, and comprehensive method applying both. For comparison with the proposed ultra-lightweight CNN, we utilized VGG16 and ResNet50, which are the CNN structures widely used in the field of image classification and were also employed in previous studies [7], [17], [38]. We also utilized ConvNeXtTiny, one of the latest CNNs widely used in the field of image classification [40]. In addition, we implemented the CNN structures proposed for the classification of small drones in previous studies [6], [7], [18]. Finally, we implemented the proposed ultra-lightweight CNN. All the CNNs were trained and tested with each dataset generated utilizing the conventional and proposed MDS image generation techniques. TensorFlow 2.10, cuDNN 8.6.0, CUDA 11.8, and NVIDIA RTX 3090 GPU were utilized for the implementation, training, and testing of the CNNs.

For the widely used CNN structures, we utilized the modules in TensorFlow. The output layer was customized to classify the four classes, and pre-trained weights were not used. The remaining input parameters of the function followed the default settings of TensorFlow. For the CNN structures proposed in previous studies [6], [7], [18], we changed only the output layer to classify the four classes, while the rest of the structure was implemented as is. The proposed ultra-lightweight CNN comprised only one convolutional block and only one fully connected layer, as described in Section IV. Details on the proposed ultra-lightweight CNN can be found in Table II.

For training each CNN on our task, the optimizer, learning rate, and batch size, which are hyperparameters, were determined through the grid search based on KFCV, where K was set to five. The grid search ranges of the optimizer, learning rate, and batch size were [SGD, Adam, Nadam], [1e-2, 1e-3, 1e-4], and [32, 64], respectively. The best optimizer and the optimum values of the learning rate and batch size for each CNN, which resulted from the grid search, are listed in Table III. Softmax was commonly used for the activation and classification in the output layers, and categorical cross-entropy was commonly used as a loss function for training all the CNNs. The number of epochs for most CNNs was set to 150, by which point the validation accuracy had sufficiently reached saturation, and the validation loss was low enough or had converged. The epoch for VGG16 was set to 50, as while the validation accuracy had saturated, the validation loss continued to increase, clearly indicating overfitting. Regarding weight initializers, although He initializers are known to be particularly effective when using ReLU as an activation function [41], GlorotUniform produced slightly better results compared with HeNormal or HeUniform in our experiments. In addition, zero was better than a small constant value such as 0.01 for the bias initializer. Therefore, we used GlorotUniform and zero as the weight and bias initializers, respectively.

Unlike previous studies [5], [7], [8], [9], we applied MCCV to ensure reliable performance evaluation and comparison. KFCV outperformed MCCV in terms of time consumption. However, MCCV outperformed KFCV in terms of reliability [42], [43]. The number of iterations for MCCV was set to 100. The numbers of parameters, nodes, and FLOPs were calculated to check the complexity of each CNN. For the number of parameters, we followed the count results reported by the summary method in TensorFlow. We calculated the number of nodes based on the output shapes of all layers constituting the model. Finally, we computed the FLOPs utilizing the profiler module in TensorFlow.

To provide more comprehensive analyses of the efficiency of the proposed methods, we also measured time consumption. The experiment involved generating MDS images as many as the test set and conducting inference using each type of CNN implemented in this study, thereby calculating the average time required to generate and infer a single MDS image across all experimental cases. The generation time for MDS images was measured using the tic-toc function in MATLAB 2021a, and inference time was measured using the time function in Python 3.10. We measured inference times using CPU as well as GPU. This experiment was conducted on a computer equipped with an AMD Ryzen 5 5600X CPU, 32 GB (2 × 16 GB) G.SKILL Trident Z Neo RAM, and an NVIDIA RTX 3070 GPU.

D. Additional Verification for Proposed MDS Image Generation Technique Using Uncertainty Quantification

The uncertainty of the neural network output can be estimated by utilizing two widely used methods in recent time: the Monte Carlo dropout (MCD) method based on the Bayesian approach [20] and the deep ensemble (DE) method based on the non-Bayesian approach [44]. Notably, the DE method can arguably be interpreted as a form of Bayesian model averaging [22], [45]. Both methods have been used in many studies. In general, the MCD method outperforms the DE method in terms of computational resource burden and time consumption. However, many recent papers have shown that the DE method outperforms the MCD method in quantifying the uncertainty of neural network outputs [22], [23], [24], [45]. Therefore, we applied the DE method for reliable uncertainty quantification in further validating the proposed technique.

The proposed MDS image generation technique improves the quality of the MDS image by extremely attenuating the phase noise of the leakage, which is the main cause of MDS image quality degradation, thus increasing the SNR in the MDS image. In other words, provided target signals do not exist, the MDS image of noise is merely a normalized noisy image of the noise itself, making the improvement from the proposed technique pointless. Therefore, we conducted the experiment with three classes of small drones, excluding the noise class that does not fit in this additional experiment.
For training, we trained the proposed ultra-lightweight CNN with entire data that equally combined two MDS datasets generated utilizing the conventional and proposed techniques. In this manner, we can prevent the model from acquiring biased knowledge and enable it to quantify uncertainty on an equivalent scale to facilitate an appropriate comparison of the uncertainty. We randomly selected 20% of the entire data for the test and then randomly divided the remaining data into training and validation sets at a ratio of 8:2. We randomly split the remaining data by employing Monte Carlo random sampling at every moment when each model in the ensemble model was independently trained with random initialization to enable the ensemble model to provide a more comprehensive representation of uncertainty by sufficiently including the diversity of data and models. However, this process only checks the results of one test set selected by chance with one ensemble model when performed only once. Therefore, this entire process was also performed 100 times by employing MCCV to reliably demonstrate the degree to which the uncertainty is improved by utilizing the proposed technique. The AU was quantified based on Shannon entropy as follows [25], [26], and [27]:

$$\text{AU}(y|x) = \frac{1}{M} \sum_{m=0}^{M-1} \left( -\sum_{c=0}^{C-1} \hat{p}(y = c|x, \theta^{(m)}) \log_b \hat{p}(y = c|x, \theta^{(m)}) \right)$$  

(15)

where $M$ is the number of models constituting an ensemble model, $C$ is the number of classes, $\theta^{(m)}$ are the parameters of $m$th model. We set the base of the logarithm as the number of classes to make the uncertainty range [0,1].

VI. RESULTS AND DISCUSSION

A. MDS Images

The results of MDS images generated utilizing the conventional and proposed techniques are shown in Fig. 5. In Fig. 5(a), 5(c), and 5(e), which are MDS images generated utilizing the conventional technique, the MDS of small drones was influenced by noise, and its features were distorted. These images even look similar to an MDS image of random noise in Fig. 6. This implies the possibility of a trained model identifying small drones as noise or identifying falsely detected noise as small drones in real-life situations. By contrast, in Fig. 5(b), 5(d), and 5(f), which are MDS images generated utilizing the proposed technique, the noise was significantly reduced, and the features of MDS became clear. Therefore, it was confirmed that the proposed MDS technique improved the quality of MDS images, even with the human eye.

B. Results of Classification Performance

The MCCV results for each experimental case as box plots are shown in Fig. 7 and the performance metrics of the MCCV results along with the complexity of each CNN are listed in Table IV. In all CNNs, the results from MDS images generated utilizing the conventional technique surpassed those from images generated utilizing the proposed technique. The values for the overall performance metrics improved in all CNNs owing to the MDS images generated utilizing the proposed technique. Across all CNNs, the maximum, median, and mean values of accuracy improved by approximately 4.59%, 6.37%, and 6.79% on average, respectively. Furthermore, the observed improvements in both standard deviation (SD) and IQR confirm that training with MDS image data generated using the proposed technique also leads to more stable learning.

The proposed ultra-lightweight CNN was overwhelmingly lighter than the other CNNs. Its numbers of parameters, nodes, and FLOPs were dramatically lower, being at least 1/21, 1/78, and 1/5.7 times and at most 1/27,538, 1/1,726, and 1/982 times smaller, respectively, than those of other CNNs. When trained and tested with MDS image data generated by both conventional and proposed techniques, the proposed ultra-lightweight CNN consistently achieved the highest performance metrics, both relative to its complexity and in absolute terms, across
all experimental scenarios. Although the hyperparameters of each CNN were optimized in the same manner, we observed relatively unstable results in ResNet50 and the CNNs in [6] and [18], exhibiting a higher number of outliers, as well as increased values in both the SD and IQR. In summary, considering all experimental cases, the proposed comprehensive method demonstrated superior performance, achieving the maximum, median, and mean values of approximately 100%, 99.20%, and 99.21%, respectively, with a complexity of 4.88 K parameters, 21.51 K nodes, and 31.52 M FLOPs.

Despite its considerably lower complexity, the proposed ultra-lightweight CNN exhibited remarkably high accuracy in small drone classification, thereby validating the effectiveness of the proposed lightweight CNN design strategy. Even with the significant reduction in overall complexity—achieved by decreasing the input size and employing only a single convolutional block and a single fully connected layer—very few large-sized convolution filters, adjusted to efficiently enlarge the receptive field, were sufficient to directly capture the key features of MDS from small drones. Moreover, even with relatively large pooling leading to rapid shrinkage, key features were not lost and were efficiently transmitted to the output layer. Consequently, as anticipated, the functionality to detect key features of MDS from small drones was successfully concentrated in the very few large-sized convolution filters, thereby maximizing their efficiency.

Table V shows the results of the average time taken to generate a single MDS image and perform inference on the classification of small drones. Although the proposed MDS image generation technique takes more time compared to the conventional technique, it sufficiently supports real-time MDS image generation. Assuming the radar system is stable enough to skip the initial estimation process in the proposed technique and utilize data estimated earlier, the MDS generation time of the proposed technique was reduced to approximately 3.20 ms. The inference times on both CPU and GPU were not strictly proportional to the complexity of the CNNs. In particular, the GPU showed results in reducing the gap in inference times between models with high complexity and those with low complexity. Despite having a complexity similar to that of ResNet50, ConvNeXtTiny exhibited quite a slower performance compared to other CNNs. The several distinctions from other CNNs are the application of layer normalization and the use of Gaussian error linear unit as the activation function in
ConvNeXtTiny. The proposed ultra-lightweight CNN was the fastest among all CNNs. While the results in Table V are useful for estimating how quickly small drones can be classified in practice, assessments of model efficiency should be taken into account together with the results in Table IV. Considering not only the time consumption but also the complexity and classification accuracy performance presented in Table IV, applying the proposed comprehensive method emerges as the most reasonable choice.

The test results of confusion matrices for each model that achieved the maximum accuracy in each experimental case are shown in Fig. 8 and Fig. 9. As concerned in Section VI-A, when trained and tested with MDS images generated using the conventional technique, the CNN models occasionally misclassified small drones as noise and noise as small drones. In contrast, when trained and tested with MDS images generated using the proposed technique, the CNN models did not misclassify small drones as noise and noise as small drones. Therefore, it can be seen that by generating MDS images using the proposed technique and utilizing them for model training, it is possible to resolve false detection issues at the classification stage, even if such problems occur during the detection stage. The test performance of the classification for
Inspire 1 and Inspire 2, which had similar appearances, was also better when training and testing with MDS images generated utilizing the proposed technique than when training and testing with MDS images generated utilizing the conventional technique.

C. Results of Uncertainty Quantification

The AU quantification results of each MDS dataset generated utilizing the conventional and proposed techniques are shown in Fig. 10 and Table VI. In addition to the performance metrics of the total MDS images in each dataset, those of the MDS images for each small drone class in each dataset are listed. The AU values for the overall performance metrics significantly improved in MDS images generated utilizing the proposed technique. Moreover, most key performance metrics for total MDS images generated with the proposed technique were less than half of those for images generated with the conventional technique. This substantial reduction in AU, which is caused by the MDS image data during class inference, confirms that the proposed MDS image generation technique effectively improves the quality of MDS images. Additionally, the reduction in most key performance metrics for MDS images of each small drone class by more than half further demonstrates that the proposed technique not only improved the quality of MDS images for specific drone classes but also across all small drone classes.

D. Comparisons With Other Works

Table VII compares this study with other recent works that predominantly classify small drones using CNNs and MDS images generated from FMCW radars. Exceptionally, CW radars were utilized in [17] and [18]. Direct comparison poses challenges due to different objectives and experimental conditions across the studies; however, our work has produced noteworthy results in several aspects. Unlike other studies, which introduced novelty in either data or model, this paper proposes novelty in both. Moreover, despite the experiments being conducted with small drones at the most considerable distances, our work has yielded overwhelming classification accuracy, demonstrating the robustness of the proposed methods. Additionally, we have implemented most CNNs utilized in the other works in Table VII and conducted experiments on them. Considering Table VII alongside Table IV and Table V, it can be ascertained that our work achieved the best performance in both accuracy and efficiency.

E. Limitations and Future Works

Although the proposed comprehensive method has demonstrated effectiveness, it is not entirely devoid of limitations. The proposed technique can be applied even to FMCW radars with very poor stability, enabling them to benefit from the effectiveness of the proposed technique and thus turning such radars that would have been discarded into valuable assets. However, the severe instability of such FMCW radars necessitates the application of the estimation process in the proposed technique not only for the initial beat signal but also for every subsequent $n$th beat signal required to generate an MDS image. This requirement may limit its application in real-time scenarios.

We intend to conduct further research as future works to enhance the proposed methods to overcome the limitations. Additionally, the impact of STFT parameters on classification performance may present an intriguing analysis. Therefore,
TABLE VII

COMPARISON WITH OTHER WORKS

| Paper | Freq Band | Targeted Classes | Distances | Proposed Methods | Accuracy |
|-------|-----------|------------------|-----------|------------------|----------|
| [5]   | Ku        | 10 classes:      | 5 m       | GoogLeNet (InceptionV1) using proposed merged Doppler image data | 94.7% (Avg.) (4-Fold Cross-Validation) |
|       |           | (2 conditions of fixed state) × (5 conditions of motor operation) with HobbyFloyd F820 |           |                  |          |
| [6]   | K or W    | 4 classes:       | 30-120 m  | Proposed series network | 95.28% (Max.) |
|       |           | Drone, Bird, Clutter, Noise | (Outdoor) |                  |          |
| [7]   | W         | 2 classes:       | 1.5-10.2 m m | Proposed lightweight CNN | 92.75% (Avg.) (4-Fold Cross-Validation) |
|       |           | Rotating drone, Non-rotating drone | (Outdoor) |                  |          |
| [8]   | Ku        | 6 classes:       | 5 m       | GoogLeNet (InceptionV1) using proposed polarimetric merged Doppler image data | 99.83% (Avg.) (4-Fold Cross-Validation) |
|       |           | 2 conditions of fixed state with XTM Bird-like drone + (2 conditions of fixed state) × (2 conditions of propellers operation) with HobbyFloyd F820 | (Indoor anechoic chamber) |                  |          |
| [9]   | Ku        | 6 classes:       | 30 m      | VGG11 using proposed channel extension dataset based on four polarizations | 97.9% (Avg.) (10-Fold Cross-Validation) |
|       |           | 6 different elevation angles with DJI Inspire 2 | (Outdoor) |                  |          |
| [17]  | X         | 6 classes:       | 10-15 m   | VGG16 and VGG19 using proposed DIAT-µSAT dataset | 95.48% (VGG16, Max.) 96.71% (VGG19, Max.) |
|       |           | 2 blades rotor, 3 short blades rotor, 3 long blades rotor, Quadcopter, Bionic bird, 2 blades rotor and Bionic bird | (Outdoor) |                  |          |
| [18]  | X         | 6 classes:       | 10-15 m   | Proposed DIAT-RadSATNet | 97.33% (Max.) (Normal weather) 97.15% (Max.) (Adverse weather) |
|       |           | RC plane, 3 short blades rotor, 3 long blades rotor, Quadcopter, Bionic bird, Mini helicopter and Bionic bird | (Outdoor) |                  |          |
|       |           | 150-200 m        |           |                  |          |
| This work | Ku     | 4 classes:      | 221-412 m | Proposed ultra-lightweight CNN using MDS image data generated via proposed technique | 100% (Max.) 99.21% (Avg.) |
|       |           | DJI Inspire 1, DJI Inspire 2, DJI Spark, Noise | (Outdoor) |                  | (100 iterations of MCCV) |

we would like to include this as one of the future works as well.

VII. CONCLUSION

This study has proposed the comprehensive method to accurately classify small drones in real-time using high-quality MDS images generated through a novel MDS image generation technique and the ultra-lightweight CNN developed using a novel CNN design strategy based on the analyses of CNNs and MDS images of small drones.

The experimental results have shown that the MDS images generated using the proposed technique significantly improved the accuracy of the various CNNs employed in the experiments. Furthermore, AU quantification results based on the DE method have exhibited that most performance metrics for MDS images generated using the proposed technique were less than half of those for images generated using the conventional technique. Therefore, it has been verified that the proposed MDS image generation technique significantly enhances the quality of MDS image data and contributes to the improved accuracy of CNN-based small drone classification.

Regarding lightweight CNNs, in contrast to the conventional approaches, this study has proposed a new approach that designs lightweight CNNs with minimal convolutional blocks, containing very few large-sized convolution filters, to efficiently enlarge the receptive field for capturing key features of MDS from small drones. The proposed ultra-lightweight CNN developed using the proposed design strategy has demonstrated superior performance, despite being the lightest among all CNNs utilized in this study.

Finally, the proposed comprehensive method, employing both the MDS images generated via the proposed MDS image generation technique and the proposed ultra-lightweight CNN, has shown the best performance across all experimental cases. Consequently, it has been proven that the proposed comprehensive method can accurately identify small drones from far distances with outstanding efficiency.

ACKNOWLEDGMENT

The raw data for the MDS image generation in this study were acquired in the Microwave and Antenna Laboratory at Korea Advanced Institute of Science and Technology. The authors would like to thank Prof. Seong-Ook Park for his support.

REFERENCES

[1] K. Suarez. Lakeland Woman Concerned About Privacy After Multiple Visits From Hovering Drone. FOX 13 Tampa Bay. Accessed: Jan. 16, 2023. [Online]. Available: https://www.fox13news.com/news/lakeland-woman-calls-police-after-drone-hovers-outside-her-apartment
[2] S. Headley. Drones Hovering Near Windows Spark Concern in Leicestershire Village. Leicester Mercury. Accessed: Jan. 16, 2023. [Online]. Available: https://www.leicestermercury.co.uk/news/local-news/drones-hovering-near-windows-spark-6743808?utm_source=linkCopy&utm_medium=social&utm_campaign=sharebar
