# Deep Learning Techniques for Advanced Drone Detection Systems: A Comprehensive Review of Techniques, Challenges and Future Directions

**Fatin Najihah Muhamad Zamri[1], Teddy Surya Gunawan[2], Mira Kartiwi[3], Agus Pratondo[4], Siti Hajar Yusoff[5], Yasir Mohd. Mustafah[6]**

[1,2,5]Department of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia
[3]Department of Information Systems, International Islamic University Malaysia, Malaysia
[4]Department of Multimedia Engineering Technology, School of Applied Science, Telkom University, Indonesia
[6]Department of Mechatronics, International Islamic University Malaysia, Malaysia

**Article Info**

***Article history:***

Received Nov 10, 2024
Revised Nov 14, 2024
Accepted Nov 15, 2024

***Keywords:***

Drone detection
Deep learning
Sensor fusion
Real-time processing
UAV security
Thermal infrared imaging
Radio frequency detection
Radar-based detection
Acoustic-based detection

**ABSTRACT**

The widespread use of Unmanned Aerial Vehicles (UAVs), commonly known as drones, across various sectors, such as civilian, commercial, and military operations, has created significant challenges in ensuring security, safety, and privacy. This paper provides a comprehensive review of the latest advancements in drone detection systems leveraging deep learning techniques, covering the period from 2020 to 2024. It critically evaluates both optical (visible light and thermal infrared) and non-optical (radio frequency, radar, and acoustic) detection methodologies. The analysis includes cutting-edge models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs), focusing on their application in drone detection. Key challenges like real-time processing, environmental interference, and differentiation between drones and similar objects are examined. Potential solutions, including sensor fusion, attention mechanisms, and the integration of emerging technologies such as the Internet of Things (IoT) and 5G networks, are discussed in detail. The paper concludes with future research directions to enhance drone detection systems' robustness, scalability, and accuracy, particularly in complex and dynamic environments. This review offers valuable insights for researchers and industry professionals working towards next-generation drone detection technologies.



***Corresponding Author:***

Teddy Surya Gunawan
Department of Electrical and Computer Engineering
International Islamic University Malaysia
Kuala Lumpur, Malaysia
Email: tsgunawan@iium.edu.my

## 1. INTRODUCTION

Unmanned Aerial Vehicles (UAVs), commonly known as drones, have evolved far beyond their initial military applications to become essential tools in various civilian and commercial sectors. Technological advancements have significantly enhanced their flexibility, efficiency, and affordability, making drones indispensable in aerial surveillance, precision agriculture, disaster management, and infrastructure inspection. UAVs enable efficient crop monitoring, targeted pesticide application, and the delivery of high-resolution imagery and real-time data, greatly enhancing agriculture productivity and disaster recovery efforts. Additionally, drones are becoming increasingly vital in scientific research, environmental monitoring, and delivery services—particularly in remote or congested urban areas—offering a safer and more cost-effective alternative to traditional methods.

However, the growing prevalence of drones has also raised significant security and safety concerns. Drones are increasingly misused for unauthorized surveillance, airspace violations, and smuggling activities, posing threats to privacy, public safety, and national security. As a result, effective drone detection systems have become crucial to mitigate these risks, safeguard restricted airspaces, and protect sensitive environments. This review addresses these concerns by examining recent advancements in drone detection technologies, focusing on deep learning approaches. Figure 1 illustrates the core components involved in drone detection systems, including key inputs (radio frequency, radar, acoustic, digital, and thermal infrared cameras), deep learning models (CNNs, RNNs, and GANs), and advanced detection techniques (sensor fusion and attention mechanisms). It also highlights the significant challenges (real-time detection, environmental effects) and emerging drone threats (target attacks, privacy invasion). This framework provides a roadmap for addressing these challenges by integrating advanced deep-learning techniques and sensor technologies. Furthermore, the paper explores techniques to improve detection accuracy and robustness, offering insights into the future of drone detection technologies.

This review follows an integrative evaluation method guided by several key research questions: What methods are used to detect drones? What are the advantages and drawbacks of each method? How do environmental conditions impact the performance of detection models? What are the pathways to enhancing model accuracy and robustness? Moreover, what recommendations can be made for the future development of advanced drone detection technologies to create reliable and scalable systems?

To address these questions, this review analyzes studies published between 2020 and 2024, ensuring a focus on the most recent advancements in deep learning applications for drone detection. Sources were drawn from reputable databases, including IEEE Xplore, ScienceDirect, and SpringerLink. This timeframe comprehensively analyzes cutting-edge technologies and methodologies, representing state-of-the-art drone detection. Our review also evaluates various optical and non-optical detection methods, considering their advantages, limitations, and potential for improvement through sensor fusion, attention mechanisms, and other deep learning innovations.

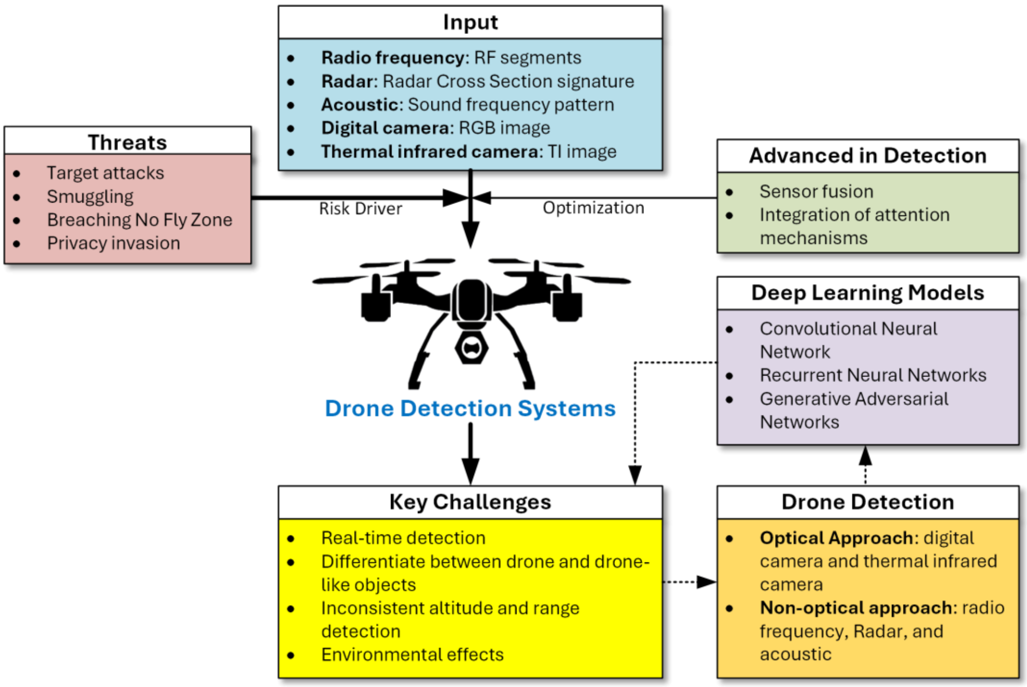


Figure 1. A Comprehensive Framework for Drone Detection Systems

This review offers a fresh perspective on the evolving landscape of drone detection by focusing on deep learning advancements since 2020. It highlights significant real-world drone threat cases and examines how these incidents have influenced current research trends. The review identifies persistent challenges in drone detection, such as differentiating drones from similar objects, operating in complex environments, and achieving real-time performance. To tackle these challenges, we explore cutting-edge deep learning models and publicly available datasets, offering a detailed examination of optical (visible light and thermal infrared) and non-optical (radar, radio frequency, and acoustic) approaches. Moreover, advanced techniques such as sensor fusion and attention mechanisms are explored for their potential to improve detection accuracy and system robustness. By outlining future research directions, this review serves as a guide for researchers and practitioners looking to advance the field of drone detection in increasingly complex environments.

Table 1 provides a comparative overview of the contributions made by recent review papers in drone detection systems. The authors of [1] reviewed anti-drone technologies up to 2020, including detection, identification, and neutralization methods. The work in [2] reviewed drone detection systems up to 2021, focusing on acoustic, radio frequency (RF), radar, visual, and sensor fusion methods, examining each method's advantages, limitations, and potential improvements. The reviews in [3-5] extended the scope to 2023, encompassing various machine learning and deep learning techniques across radar, visual, acoustic, RF, and hybrid sensors. In contrast, our review covers the most recent advancements from 2020 to 2024, offering a more up-to-date and comprehensive analysis. We focus specifically on deep learning models for drone detection, concentrating on foundational approaches such as CNNs, RNNs, and GANs, which are particularly relevant to this field.

Table 1. Summary of the contribution of this review paper to other existing review papers on drone systems.

| References | Year | Contribution |
|---|---|---|
| [1] | 2021 | Reviewed anti-drone technologies, including detection, identification, and neutralization methods up to 2020. |
| [2] | 2022 | Focused on acoustic, RF, radar, visual, and sensor fusion methods for drone detection up to 2021, highlighting each method's advantages, limitations, and improvements using machine learning and deep learning techniques. |
| [3] | 2024 | Reviewed detection and classification of drones using machine learning and deep learning techniques across radar, visual, acoustic, RF, and hybrid sensors up to 2023. |
| [4] | 2024 | A comprehensive review of drone detection systems based on deep learning, covering radar, RF, acoustic, and visual approaches from 2014 to 2023. |
| [5] | 2024 | Reviewed various drone detection techniques using machine learning and deep learning, including radar, RF, acoustic, visual, and sensor fusion, WiFi fingerprinting, 5G networks, and IoT technologies up to 2023. |
| Ours | 2024 | This paper focuses on deep learning-based drone detection methods from 2020 to 2024 and categorizes detection approaches into optical (digital and thermal cameras) and non-optical (radar, RF, and acoustic) approaches. It highlights advancements in sensor fusion, attention mechanisms, and 5G and IoT integration for improving detection accuracy, real-time processing, and scalability. |

Unlike previous reviews, which provided broader overviews of deep learning models, our paper focuses on how these models are applied to drone detection tasks. We categorize detection methods into optical (digital and thermal infrared cameras) and non-optical (radar, RF, and acoustic) approaches, offering a clearer framework for analyzing these technologies. Additionally, we review studies on advanced techniques like sensor fusion and attention mechanisms, highlighting their potential to improve detection accuracy and real-time processing. This review introduces the common challenges in drone detection and suggests innovative solutions and future directions for improving real-time, high-quality detection, including integrating 5G and IoT technologies.

The remainder of this paper is structured as follows: Sections 2 and 3 discuss the evolving threats posed by drones and the critical challenges in drone detection. Section 4 overviews the fundamental deep learning models applied in drone detection systems. Section 5 surveys existing studies on deep learning-based drone detection, including publicly available datasets, while Section 6 explores methods for enhancing the reliability and accuracy of detection models. Section 7 highlights future research directions and critical areas for improvement in drone detection technologies, and Section 8 concludes the review by summarizing key findings and recommendations.

## 2. UNMASKING DRONE THREATS

As drones continue to proliferate beyond their original military applications, their use has expanded to various civilian sectors such as agriculture, logistics, and entertainment. However, as drones become more commonplace, the potential for illegal and hazardous activities has also increased. The growing threats posed by drones highlight the urgent need for effective detection systems. This section delves into the critical risks of unauthorized drone activities and their far-reaching consequences.

### 2.1. Targeted Attacks

Drones' versatility, ease of use, and affordability make them highly attractive tools for malicious actors. A drone can be equipped with explosives, biological and chemical weapons which can be dropped or detonated over a targeted area, or it can be modified to carry firearms or other weapons, enabling remote shooting or stabbing, causing significant harm or fatalities on specific individuals, public institutions, business organizations or even entire nations. According to [5], target attacks became the first category of threats under drone attacks. In Mexico in 2021, two police officers were injured due to an attack received from armed drones that brought improvised explosive devices (IEDs), as reported by [6]. In 2023, 88 people died during a religious celebration when drones mistakenly dropped a bomb at the celebration area in Nigeria [7]. In 2024, 13 civilians, including seven children, were killed by a drone strike in Mali [8]. These incidents underscore the growing risk of drone-enabled targeted attacks, which are becoming a significant threat to public safety and national security.

### 2.2. Smuggling

Drones are increasingly being used to facilitate smuggling, enabling the transport of illegal contraband across borders, into prisons, and other restricted areas. Their ability to fly long distances, carry payloads, and evade traditional security systems makes them a formidable challenge for law enforcement and border control. Drones are commonly used to smuggle drugs, weapons, and other prohibited items [5]. According to recent reports, over 20 smuggling incidents involving drones have been recorded since January 2024 [9]. These incidents include the transportation of drugs across international borders and the delivery of contraband into correctional facilities, illustrating the growing scale of drone-enabled smuggling operations.

### 2.3. Breaching No-Fly Zones

Breaching restricted airspace around airports is one of the most dangerous drone-related threats. Unauthorized drones in no-fly zones, such as airports and prisons, have led to numerous incidents of flight delays, diversions, and cancellations, endangering the lives of thousands of passengers. For example, the runway at Heathrow Airport had to be temporarily closed after a drone, flying at nearly 200 mph, came within 3 feet of an aircraft [10]. Similarly, Pittsburgh International Airport halted operations after unauthorized drones were spotted near the airport [11]. These breaches create significant safety concerns, as drone-aircraft collisions could result in catastrophic loss of life and massive economic damage.

### 2.4. Privacy Invasion

Drones equipped with high-resolution cameras and advanced sensors can be used for covert surveillance, invading the privacy of individuals, businesses, and governments. These drones can gather sensitive data without detection, posing severe personal privacy and security risks. In December 2022, an organization in North Carolina providing equine care and therapy for traumatized children reported that a drone was harassing staff, children, and therapy animals, causing distress and safety concerns [12]. Similarly, in September 2022, a parent in California reported that a drone was hovering near her daughter's balcony, seemingly spying on her [13]. These incidents illustrate the growing risk of drones being used to violate personal privacy and create fear and intimidation without direct physical interaction.

## 3. KEY CHALLENGES OF DETECTING DRONES

With the rapid proliferation of drones across industries and public spaces, drone detection systems have become critical tools for ensuring security and safety. Despite significant advancements, drones' inherent complexity and variability present numerous challenges to developing effective detection systems. These challenges must be addressed to enhance the accuracy and reliability of such systems in real-world applications.

### 3.1. Real-Time Detection

Real-time drone detection is essential for maintaining continuous surveillance and swiftly responding to potential threats. However, it presents significant challenges due to the dynamic and unpredictable nature of drone movements and variations in size and speed [14, 15]. For instance, small drones often travel at speeds of up to 15 m/s, while larger drones can reach speeds exceeding 100 m/s [16]This wide range in drone size and velocity complicates real-time detection efforts and requires fast and precise processing capabilities.

The complexity of real-time detection can be modeled, as shown in Eq. (1), by considering the required processing time $t_p$ for detecting a drone. The detection system must operate within a fixed time window $T$, where:

$$t_p \leq T \tag{1}$$

If the processing time $t_p$ exceeds $T$, the system fails to detect the drone in real time. Optimization of algorithms and hardware, such as reducing the model complexity while maintaining accuracy, helps ensure that $t_p$ remains below the threshold.

One approach to achieving this is optimizing model architectures to improve computational efficiency. Future research should focus on creating robust, lightweight models that can handle various real-world contexts and dynamic drone behaviors without sacrificing speed or accuracy.

### 3.2. Differentiating Drones from Drone-Like Objects

A significant challenge in drone detection arises when the system must differentiate between drones and drone-like objects such as birds or airplanes. This challenge is particularly prominent in optical-based approaches, where drones may appear as small indistinct pixels, especially at long distances. The lack of information hinders the detection system's accuracy in identifying drones, potentially leading to misidentifications with similar objects [17]. When drones and birds are in the same visual frame, the problem becomes more complex, leading to a higher chance of false alarms [18]. Misidentifications can influence the accuracy of detection systems and increase the chances of security oversights. Researchers give this critical issue significant emphasis. As a result, the Drone vs. Bird Detection Challenge [19-22] was established to address this issue collaboratively. Misidentification between drones and birds increases the likelihood of false positives and security oversights. The false positive rate $FPR$ in a detection system can be expressed as follows:

$$FPR = \frac{FP}{FP+TN} \tag{2}$$

where $FP$ is the false positive, and $TN$ is the false negative. To minimize $FPR$, advancements in detection algorithms are needed, such as improved resolution and pattern recognition techniques that can incorporate additional contextual information beyond the visual spectrum. Deep learning algorithms trained on comprehensive datasets, including scenarios with birds and drones in the same frame, can significantly reduce false alarms. By enhancing the learning process through diverse datasets, the system becomes more adept at distinguishing between drones and similar objects.

### 3.3. Altitude and Range Detection

Detecting drones at different altitudes is difficult for visual detection systems because they become smaller as they fly higher, eventually looking like pixels on the camera's screen [23]. The decreased size and level of detail make it difficult to accurately identify objects since the visual information is limited, hindering the detection system's ability to distinguish drones from other small flying objects or background interference. Radio frequency-based and acoustic signals-based detection will also be affected as RF signals and acoustic signals weaken over long distances, thus limiting the range of effective drone-detecting systems [3]. Researchers have extensively reported detecting drones at high altitudes, emphasizing the challenges in identifying small drones against extensive aerial backgrounds [24]. In this context, the signal attenuation over distance can be modeled using the inverse square law for both RF and acoustic signals, as follows:

$$\boldsymbol{S = \frac{S_0}{d^2}} \tag{3}$$

where $S_0$ is the initial signal strength and $d$ is the distance from the source. As distance increases, the signal strength $S$ decreases significantly, limiting the effectiveness of detection systems over large areas. To address this, detection algorithms must incorporate enhanced feature extraction techniques to reliably identify small drones at high altitudes or long ranges. High-sensitivity sensors and training models on datasets, including distant and small drones, can improve detection accuracy.

### 3.4. Environmental Effects

Drones frequently operate in various environments where weather, background noise, and lighting conditions can significantly affect detection performance. These environmental factors pose a significant challenge to the efficiency and accuracy of drone detection systems. Detecting drones in adverse weather conditions, such as rain, fog, or snow, becomes difficult because these factors can distort camera visual data and interfere with radar and RF signals [14]. Fluctuating lighting conditions, including low-light environments or bright sunshine, can conceal drones or create shadows, altering their apparent shape and size and making identification more complex.

Adverse weather conditions such as rain or fog reduce the clarity of visual signals by introducing noise into the system. This degradation can be modeled using a simple image degradation model:

$$I_{degraded}(x,y) = I_{original}(x,y) \cdot H(x,y) + N(x,y) \tag{4}$$

where $I_{degraded}(x,y)$ is the observed image, $I_{original}(x,y)$ is the original or noise-free image, $H(x,y)$ represents the degradation function due to weather effects, e.g., rain or fog, and $N(x,y)$ is the additive noise caused by environmental factors. In this model, weather acts as a blurring or noise factor, reducing the signal quality and thus impairing the ability of detection systems to recognize drones accurately.

Radar-based detection is also affected by adverse weather conditions, where signal attenuation occurs due to precipitation, such as rain. The radar signal attenuation can be quantified by:

$$A_r = k_r \cdot R^{\alpha} \tag{5}$$

where $A_r$ is the radar signal attenuation in dB/km, $k_r$ and $\alpha$ are constants dependent on the radar frequency and $R$ is the rain rate in mm/h. This relationship indicates that radar signals degrade rapidly as rainfall intensity increases, reducing the detection range and accuracy. As a result, it becomes more difficult to detect drones accurately under heavy precipitation.

Environmental noise also impacts RF-based and acoustic-based detection systems. The presence of noise and interference can degrade the quality of signals and reduce the system's ability to differentiate drone signals from background noise. The performance of these systems under noisy conditions can be modeled by the signal-to-noise ratio (SNR):

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{6}$$

where $P_{signal}$ is the power of the drone signal, and $P_{noise}$ is the power of environmental noise. When $P_{noise}$ is high, as in the case of strong background interference, the $SNR$ decreases, making it more difficult for detection systems to maintain reliable performance.

Advanced detection algorithms are needed to compensate for visual distortions to adapt to environmental changes [25]. One such technique is the adaptive Wiener filter, which adjusts based on local image statistics to reduce noise while preserving essential image details. The Wiener filter is particularly effective for restoring image quality in the presence of environmental noise and can be expressed as:

$$\widehat{H}(u,v) = \frac{S_s(u,v)}{S_s(u,v)+\hat{S}_n(u,v)} \tag{7}$$

where $\widehat{H}(u,v)$ is the estimate of Wiener gain in the frequency domain, $S_s(u,v)$ is the power spectral density of the original image, and $\hat{S}_n(u,v)$ is the estimate of the noise's power spectral density. By applying the Wiener filter, the system can better compensate for the distortions introduced by environmental factors, improving the visual quality of drone detection even under challenging conditions.

In addition to adaptive filtering techniques, integrating thermal or infrared imaging can enhance detection capabilities when visual signals are degraded. Thermal imaging is less affected by lighting conditions, providing an alternative detection method when visible light cameras struggle due to low visibility or extreme brightness. Machine learning models trained on diverse datasets that include weather conditions, lighting variations, and other environmental factors can further improve the robustness of detection systems, ensuring consistent performance across a wide range of operational settings.

## 4. FUNDAMENTAL OF DEEP LEARNING FOR DRONE DETECTION

Artificial Intelligence (AI) is a vast and rapidly evolving field, with its applications permeating various industries, including drone technology [26]. Within AI, Machine Learning (ML) is a critical subset that enables machines to learn from data, enhancing their ability to imitate human decision-making processes [27]. ML can be categorized into three main types: supervised learning, unsupervised learning, and reinforcement learning. The nature of the data distinguishes each type used and the type of problem being addressed [28]Given the nature of drone detection, supervised learning is the most suitable ML approach due to its reliance on labeled data, allowing the model to predict drone presence accurately.

Table 2. Type of Machine Learning

| Type | Description | Problems | Algorithms |
|---|---|---|---|
| Supervised Learning | Utilizes labeled datasets (input-output pairs) to train the model, allowing it to make accurate predictions on new, unseen data. Supervised learning is well-suited for tasks such as drone detection, where labeled images of drones are available. | Classification, Regression | Linear regression, Support Vector Machines (SVM), Decision Trees, Random Forests, Naïve Bayes, Neural Networks, and CNNs |
| Unsupervised Learning | Identifies patterns and structures within unlabeled data. This method clusters data based on similarities without prior knowledge of the data categories, which can be useful in discovering patterns within sensor data or camera feeds from drones. | Clustering, Association | K-Means, DBSCAN |
| Reinforcement Learning | Learns by interacting with an environment and receiving feedback through rewards or penalties. Although less commonly applied to drone detection, it can be used in dynamic path optimization and autonomous navigation of detection systems. | Positive/Negative Reinforcement | Q-Learning, Markov Decision forms |

Unlike traditional methods that rely on manually designed features, deep learning (a specialized subset of ML) automatically extracts targeted features directly from the data, offering more sophisticated and accurate outcomes [29]. The increasing popularity of deep learning is fueled by the growing availability of large datasets, enhanced computational power, and advances in hardware [30]. These advances have enabled deep learning models to perform exceptionally well in object detection tasks, which is crucial for effective drone detection systems. The architecture of Convolutional Neural Networks (CNNs) has emerged as the dominant choice for object detection tasks. CNNs excel at recognizing patterns within images, making them particularly well-suited for identifying drones in visual data [30]. Deep learning models are typically trained on large, labeled datasets that allow them to learn and recognize the unique features of drones across different environments. Table 2 outlines the key characteristics of each type of ML, providing insights into their applications for solving drone detection challenges.

The decision to focus on CNN, RNN, and GAN for drone detection is driven by the specific capabilities these deep learning algorithms offer, each well-suited for the unique challenges of drone detection tasks. CNNs are exceptional at handling spatial data, making them the go-to choice for analyzing images and videos, which are crucial in drone detection for identifying objects in visual data. CNNs' ability to automatically extract features from images enables them to effectively distinguish drones from other objects, even in complex backgrounds [30]. RNNs, on the other hand, are particularly adept at processing sequential data, such as acoustic signals produced by drones. Their internal memory and recurrent connections allow RNNs, especially LSTMs, to capture temporal dependencies, making them ideal for analyzing time series data, such as the sound generated by a drone's propellers [31]. Finally, GANs offer a unique advantage in scenarios where training data is limited, such as thermal infrared (TIR) images of drones. GANs can generate synthetic data that mimic real-world scenarios, thus expanding the training dataset and improving the model's generalization capability [32]. By integrating these three algorithms, drone detection systems can effectively handle visual, acoustic, and data augmentation tasks, ensuring robust and accurate performance across various environments and conditions.

### 4.1. Convolutional Neural Network (CNN)

CNNs are the backbone of modern drone detection systems because they can efficiently process and learn from large visual datasets. Through convolutional layers, pooling layers, activation functions, and fully connected layers, CNNs can extract critical features and make accurate predictions about the presence of drones. These architectures, combined with optimization techniques such as backpropagation, enable CNNs to continually improve their performance in detecting drones, even in challenging environments [33]. Figure 2 shows the basic architecture of CNN. This figure illustrates the key components of a CNN used in drone detection systems. The architecture begins with an input layer that receives the image data, which is processed through multiple convolutional layers to extract hierarchical features from low-level edges to high-level object semantics. Pooling layers reduce the dimensionality, retaining the most significant features, while activation layers introduce non-linearity through functions like ReLU. Finally, fully connected layers combine the learned features to produce the final classification or detection result, with output determined by activation functions such as Softmax or Sigmoid. This architecture enables accurate and efficient detection of drones in complex environments by leveraging deep learning methods.

In the case of drone detection, the input to the CNN is typically an image represented as a 3D tensor. This tensor contains three key pieces of information: the height of the image (number of rows of pixels), the width of the image (number of columns of pixels), and the number of color channels (usually 3 for RGB images). Thus, the input layer encodes the image's dimensions, allowing the CNN to process its spatial and color information.

The convolutional layer is responsible for most of the computations in a CNN and is crucial for feature extraction. The operation can be represented mathematically as:

$$Z_{ij} = (X * W)_{ij} = \sum_m \sum_n X[i+m, j+n] \cdot W[m,n] + b \tag{8}$$

where $X$ is the input image (or feature map) of the size $M \times N$, $W$ is the convolution filter (kernel) of size $k \times k$ , $b$ is the bias term, $Z$ is the output feature map, and $*$ denotes the convolution operation. In the context of drone detection, the convolutional layers extract features such as edges, shapes, and textures from drone images. These features are then passed through multiple layers to extract higher-order representations, allowing the model to distinguish drones from other objects, such as birds or airplanes.

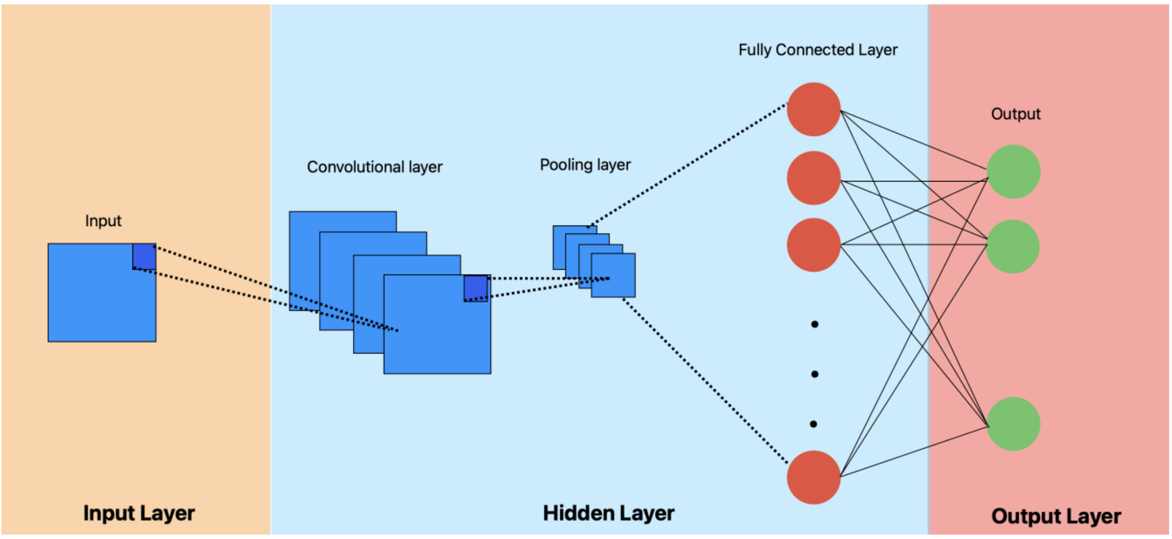


Figure 2. Basic Architecture of Convolutional Neural Network for Drone Detection

To reduce the computational load and avoid overfitting, CNNs often incorporate pooling layers, which perform downsampling operations on the feature maps. The most common pooling method is max pooling, where the largest value in a window (typically $2 \times 2$) is retained. Max pooling can be expressed as follows:

$$Z_{ij}^{pool} = \max_{(m,n)\in W} Z[i+m, j+n] \tag{9}$$

where $Z_{ij}^{pool}$ is the pooled feature map, $Z$ is the input feature map, and $W$ is the pooling window. For drone detection, pooling layers help reduce the dimensionality of the feature maps while retaining the most important features, ensuring that the model can operate efficiently while still capturing the necessary information for classification.

After pooling, the data is passed through an activation layer to introduce non-linearity into the model, enabling it to learn complex patterns. The most commonly used activation function in CNNs is the Rectified Linear Unit (ReLU), which is defined as:

$$f(x) = \max(0, x) \tag{10}$$

This function sets all negative values to zero while retaining positive values, helping the network learn complex patterns and speeding up convergence during training. The ReLU activation function is particularly useful for drone detection because it prevents the vanishing gradient problem and allows the network to learn efficiently from large datasets [33].

The fully connected layer connects all neurons from the previous layer to every neuron in the next layer, effectively combining the features learned throughout the CNN into a final prediction. The output layer then provides the classification or detection result. For drone detection, this could involve predicting whether an object is a drone or not or classifying multiple drone types.

If we are performing binary classification (drone vs. non-drone), the sigmoid activation function is typically used as follows:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{11}$$

For multi-class classification, such as detecting different types of drones, the softmax function is employed, which converts the output of the network into a probability distribution:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{12}$$

where $z_i$ represents the score for the class $i$, and the softmax function ensures that the predicted probabilities sum to 1 across all classes ($j$).

The backpropagation algorithm is used to optimize the weights and biases of the CNN during training. In backpropagation, the network's prediction is compared to the ground truth, and the error is propagated backward through the layers to update the weights. The error is computed using a loss function (e.g., cross-entropy for classification tasks). The goal is to minimize this loss by adjusting the weights using gradient descent:

$$w_{t+1} = w_t - \eta \cdot \frac{\partial L}{\partial w_t} \tag{13}$$

where $w_{t+1}$ and $w_t$ are the weights at iterations $t+1$ and $t$, $\eta$ is the learning rate, and $\frac{\partial L}{\partial w_t}$ is the gradient of the loss function to the weights. For drone detection, backpropagation enables CNN to learn and refine its feature extraction process iteratively, improving the network's ability to detect drones in various environments accurately.

**4.2. Recurrent Neural Network (RNN)**

Recurrent Neural Networks (RNNs) are a class of neural networks specifically designed to process sequential data, such as time series, speech, or text. Unlike traditional feedforward networks, RNNs have recurrent connections that enable them to maintain an internal state and capture temporal dependencies in data [4]. This makes RNNs well-suited for tasks where information order is critical, such as handwriting recognition, speech processing, and acoustic-based drone detection systems. Figure 3 shows the basic architecture of RNNs.

RNNs are particularly effective in drone detection tasks that involve analyzing sequential inputs, such as acoustic signals generated by drones. Drone engines, propellers, and other mechanical components produce unique sound signatures that can be captured and analyzed to detect and classify drone activity. Long Short-Term Memory (LSTM), a variant of RNNs, is especially useful in this context due to its ability to capture long-range dependencies in time-series data and overcome the vanishing gradient problem commonly associated with traditional RNNs [31].

In a standard RNN, the hidden state at the time step $t$, denoted as $h_t$, is computed based on the input $x_t$ at that time step and the hidden state from the previous time step $h_{t-1}$. The update rule for the hidden state can be expressed as follows:

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b_h) \tag{14}$$

where $h_t$ is the hidden state at time step $t$, $W_h$ is the weight matrix for the hidden state, $W_x$ is the weight matrix for the input, $x_t$ is the input at a time step $t$, $b_h$ is the bias term, and $\sigma$ is the activation function, such as the sigmoid or tanh function. The output at the time step $t$, denoted as $y_t$, is computed as:

$$y_t = \sigma\left(W_y h_t + b_y\right) \tag{15}$$

where $W_y$ is the weight matrix for the output layer, and $b_y$ is the bias term for the output layer.

For drone detection, the input sequence $x_t$ could be a time series of acoustic data where each $x_t$ represents the acoustic features captured at a specific time point. By maintaining the hidden state across time steps, the RNN can learn evolving patterns, making it highly effective for detecting drones based on their unique acoustic signatures.

While RNNs are powerful, they suffer from the vanishing gradient problem, which hampers their ability to learn long-term dependencies in sequences. LSTM networks address this issue by incorporating memory cells that selectively retain or forget information over longer sequences. The critical components of an LSTM cell include the input gate, forget gate, and output gate, which regulate the flow of information through the network.

The core functions governing an LSTM cell at a time step $t$ include the forget gate ($f_t$), input gate ($i_t$), cell state update ($c_t$), output gate ($o_t$), and hidden state update ($h_t$), as shown in Eq. (16) to (20), respectively. The forget gate controls what proportion of the previous cell state $c_{t-1}$ to retain. The input gate determines how much new information will be added to the cell state. The cell state $c_t$ is updated by a combination of the previous state and the new candidate values. The output gate controls what information from the cell state will be output at the time step $t$. Finally, the hidden state $h_t$ is updated based on the new cell state and the output gate's modulation.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{16}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{17}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{18}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{19}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{20}$$

In the context of drone detection, LSTM networks are highly effective at modeling the sequential nature of acoustic data, where each time step contains information about the drone's sound patterns. By maintaining and updating the cell state across time steps, LSTM can capture both short-term and long-term dependencies, making it an ideal solution for detecting drones based on fluctuating acoustic signals.

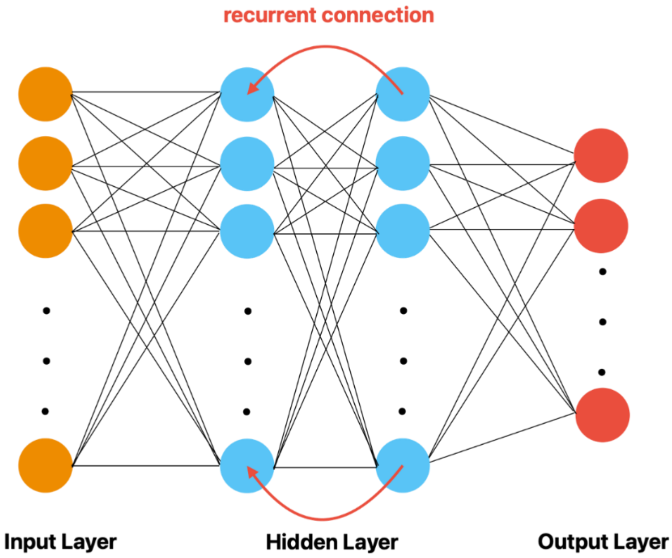


Figure 3. Basic Architecture of Recurrent Neural Network

The loss function for training an RNN or LSTM on drone detection tasks can be represented as follows:

$$L(\theta) = \frac{1}{T}\sum_{t=1}^{T} Loss(y_t, \hat{y}_t) \tag{21}$$

where $T$ is the total number of time steps, $y_t$ is the true label, e.g., drone or non-drone, at time step $t$, $\hat{y}_t$ is the predicted output at time step $t$, $\theta$ represents the parameters of the network (weights and biases), and $Loss(y_t, \hat{y}_t)$ is the loss function used to quantify the difference between the true and predicted labels, e.g., cross-entropy loss. The training process minimizes this loss by adjusting the network's parameters using backpropagation through time (BPTT). The goal is to minimize the classification error across all time steps, allowing the model to learn the temporal patterns in drone acoustics and make accurate predictions about drone activity.

### 4.3. Generative Adversarial Network (GAN)

GANs represent a robust framework in deep learning, particularly in unsupervised learning and data generation [34]. A GAN consists of two neural networks, the generator and the discriminator, which compete against each other in a zero-sum game. The generator attempts to produce realistic synthetic data (e.g., images or audio), while the discriminator tries to distinguish between real data and synthetic data generated by the generator [32]. In drone detection applications, GANs are particularly useful for generating additional training data, such as thermal infrared (TIR) images or drone sound datasets, which are often limited due to the high costs of data collection or experimental setup [35]. By leveraging GANs, researchers can enhance the diversity of training data, improving model performance, especially in scenarios where acquiring labeled data is expensive or infeasible.

As described in [36], many variants of GANs have been proposed. Figure 4 depicts the mechanism of GANs, illustrating the interaction between the generator and the discriminator. The process starts by feeding a random noise vector $z$ into the generator $G$. The generator attempts to create a synthetic sample $G(z)$ that resembles the real data distribution $p_{data}$. Simultaneously, the discriminator $D$ receives either a real sample $x$ or a generated sample $G(z)$. The discriminator's task is to classify the sample as real (label 1) or fake (label 0). As both networks train, the generator improves its ability to generate realistic samples while the discriminator becomes better at detecting fakes. The goal is to reach Nash equilibrium, where neither network can improve further without the other improving.

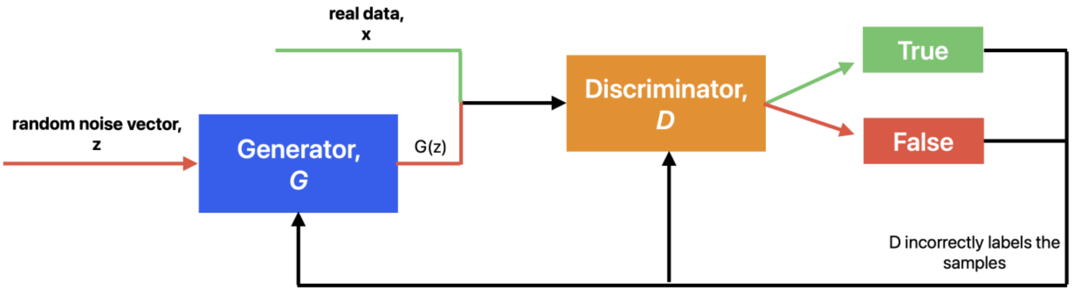

Figure 4. Mechanism of Generative Adversarial Network

GANs operate using a min-max optimization framework, where the generator $G$ and the discriminator $D$ are engaged in an adversarial game. The objective of a GAN can be represented as follows:

$$\min_G \max_D V(D,G) = \mathbf{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbf{E}_{z\sim p_z(z)}\left[\log\left(1 - D\big(G(z)\big)\right)\right] \tag{22}$$

where $x$ represents the real data sampled from the true data distribution $p_{data}(x)$, $z$ is a random noise vector sampled from a noise distribution $p_z(z)$, $G(z)$ is the synthetic data generated by the generator, $D(z)$ is the discriminator's estimate of the probability that the data is real, and $D\big(G(z)\big)$ is the discriminator's estimate that the synthetic data is real. The generator aims to minimize the discriminator's ability to correctly classify real versus fake data, while the discriminator maximizes its accuracy in making this distinction. This competitive dynamic forces both networks to improve over time until the generator produces data indistinguishable from real data, and the discriminator can no longer differentiate between real and fake.

One practical application of GANs in drone detection is generating synthetic thermal infrared (TIR) images. Acquiring large TIR datasets can be challenging due to the high costs of equipment and controlled environments required to capture such data. By using GANs, particularly image-to-image translation models such as pix2pix and CycleGAN, synthetic TIR data can be generated from easily available RGB images [37].

In the context of TIR image generation, the goal is to translate RGB images into TIR images by learning to map between these two domains. The pix2pix framework, for instance, uses paired training data (RGB and corresponding TIR images) and a U-Net generator with skip connections to perform image translation [38]. The PatchGAN discriminator classifies local patches of the image, which is effective for texture and style transformation tasks [39, 40]. The objective of the pix2pix model is to minimize the Euclidean distance between the generated TIR image and the real TIR image, ensuring that the synthesized data closely matches the real data.

The CycleGAN framework, on the other hand, can be used when paired training data is not available. Instead, CycleGAN performs cycle-consistent image translation, ensuring that the translation from RGB to TIR and back to RGB preserves the essential features of the image [41]. In this case, the model's loss function includes a cycle consistency loss term that ensures the translations maintain semantic consistency:

$$L_{cyc}(G,F) = \mathbf{E}_{x\sim p_{data}(x)}\left[\left\|F\big(G(x)\big) - x\right\|_1\right] + \mathbf{E}_{y\sim pd_{data}(y)}\left[\left\|G\big(F(y)\big) - y\right\|_1\right] \tag{23}$$

where $G$ translates images from the RGB domain to the TIR domain, $F$ translates images from the TIR domain back to the RGB domain, and the cycle consistency loss encourages $F\big(G(x)\big)$ to resemble $x$, and $G\big(F(y)\big)$ to resemble $y$. Recent GAN variants, such as BicycleGAN, enable the generation of diverse outputs from a single input, which is particularly valuable for generating various synthetic TIR images. This capability expands the training data's diversity and improves drone detection models' robustness [42, 43]. In acoustic-based drone detection, WaveGAN generates synthetic drone audio clips. These synthetic audio datasets can help overcome the scarcity of labeled drone sound data, improving the performance of acoustic drone detection systems by providing more training samples [44, 45].

## 5. DEEP LEARNING APPROACHES FOR DRONE DETECTION

Drone detection systems leveraging deep learning can be broadly classified into optical and non-optical approaches. Each approach focuses on different types of data input, using various sensors and techniques to identify and track drones in diverse environments. The effectiveness of these approaches depends on factors such as the type of environment, available sensor technology, and the specific challenges related to drone behavior and movement. This section will explore optical and non-optical approaches' fundamental principles and methodologies, highlighting their advantages, limitations, and applications in real-world drone detection systems.

As shown in Table 3, drone detection is a multi-faceted challenge that demands a combination of sensors to ensure effectiveness in diverse conditions. While Millimeter-Wave Radar and Lidar excel in long-range detection and adverse weather, their high cost makes them less accessible for widespread deployment. Conversely, acoustic detectors and ultrasonic sensors offer lower-cost solutions but are limited in range and performance under certain conditions, such as noisy environments or high-speed drones. Thermal infrared sensors and hyperspectral imaging stand out for their effectiveness in low-visibility scenarios, yet their accuracy decreases with distance, particularly for small drones. Multispectral cameras and optical systems are versatile but can be hampered by environmental factors like rain and fog. To overcome the limitations of individual sensors, sensor fusion offers the most promising path forward by leveraging multiple data sources to enhance accuracy and robustness. However, the high costs and complexity associated with integrating multiple sensors and processing their data in real time can be a significant barrier to adoption. The trade-offs between cost, performance, and integration complexity will continue to shape the evolution of drone detection technologies.

Table 3. Comparative Evaluation of Drone Detection Sensors Based on Performance, Cost, and Application Scenarios

| Sensor Type | Pros | Cons | Price |
|---|---|---|---|
| **Millimeter-Wave Radar** | • Excellent long-range detection.<br>• Effective in adverse weather conditions like fog, rain, and snow.<br>• High accuracy. | • Expensive.<br>• Struggles with detecting small drones at close range. | **High** |
| **Lidar (Light Detection and Ranging)** | • High-resolution 3D mapping.<br>• Effective in low light.<br>• Real-time detection. | • Limited performance in heavy fog or rain.<br>• Expensive and limited detection range. | **High** |
| **Hyperspectral Imaging Sensors** | • Identifies drones based on spectral signatures.<br>• High detection accuracy. | • Very expensive.<br>• Heavy and power-consuming.<br>• May struggle with fast-moving drones. | **High** |
| **Sensor Fusion (Multi-sensor Integration)** | • Combines strengths of various sensors.<br>• Provides robust detection under all conditions. | • High cost and power consumption.<br>• Complex integration and processing required. | **High** |
| **Thermal Infrared Sensors** | • Ideal for night-time and low-visibility conditions.<br>• Detects heat signatures of drones. | • Limited range.<br>• Inability to detect non-heat-emitting objects.<br>• Costlier for long-range detection. | **Medium** |
| **Multispectral Cameras** | • Detects across multiple light spectrums.<br>• Effective in diverse conditions. | • Limited range.<br>• May struggle to distinguish drones from environmental clutter. | **Medium** |
| **Passive RF Sensors** | • Detects drone-controller communication.<br>• Effective over long detection ranges.<br>• Unaffected by weather. | • Cannot detect drones without communication signals.<br>• Signal interference issues. | **Medium** |
| **Optical Cameras with Adaptive Optics** | • High-resolution detection.<br>• Effective in normal light and clear conditions.<br>• Cost-effective for basic applications. | • Poor performance in bad weather or low light.<br>• Requires powerful processing for real-time use. | **Medium** |
| **Acoustic Sensors** | • Inexpensive.<br>• Effective in detecting propeller sounds.<br>• Performs well in low-visibility conditions. | • Limited range.<br>• Susceptible to environmental noise. | **Low** |
| **Ultrasonic Sensors** | • Inexpensive.<br>• Simple to implement for close-range detection. | • Very short-range.<br>• Limited usefulness for high-speed or long-range detection. | **Low** |
| **Magnetometers** | • Detects magnetic fields from drone motors.<br>• Passive (no need for active emissions). | • Only effective for drones with large magnetic fields.<br>• Limited to short-range detection. | **Low** |

### 5.1. Non-Optical Approach

The non-optical approach to drone detection is distinguished by its ability to detect drones without visual information. It is advantageous when limited or obscured visibility, such as fog, darkness, or harsh weather conditions. Unlike optical systems that rely on cameras and imaging sensors, non-optical systems utilize alternative sensors like acoustic, radar, and radio frequency (RF) to identify drones based on their emitted signals or the disturbances they create in their surroundings. This approach's primary advantage is its robustness in environments impairing optical detection methods, such as when a drone operates beyond visual range or under cover of obstructions like trees or buildings. Acoustic sensors detect the unique sound signatures produced by the drone's rotors.

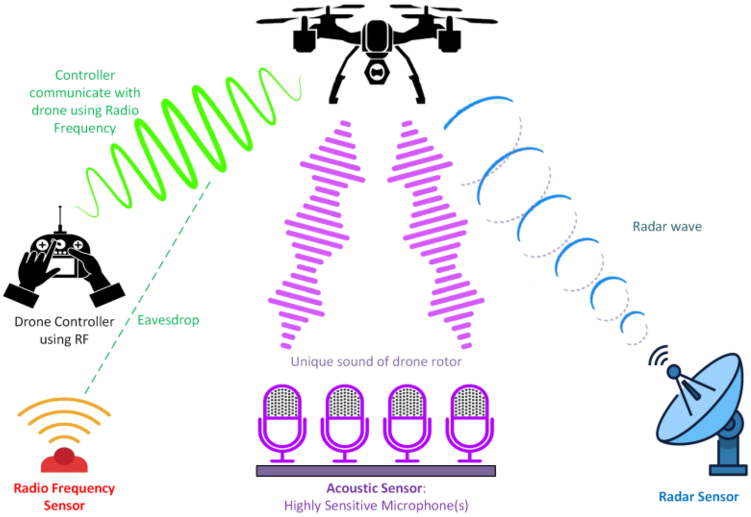


Figure 5. Drone Detection Based on a Non-Optical Approach

In contrast, RF sensors intercept the communication signals between the drone and its controller, offering a reliable way to detect drones that rely on radio communication. Radar sensors, on the other hand, emit electromagnetic waves and detect drones based on the reflection of these waves, making them particularly effective in detecting drones at longer ranges and in poor visibility conditions. Figure 5 demonstrates the various mechanisms of non-optical drone detection, highlighting how each sensor type interacts with drone activities. These systems are invaluable for applications requiring 24/7 drone monitoring across multiple environments and scenarios where visual sensors may struggle. Consequently, non-optical approaches provide a robust and versatile solution for detecting drones under diverse operational conditions.

#### *5.1.1. Radio Frequency-based Drone Detection*

Radio Frequency (RF) signals are the primary communication between a drone and its ground-based controller. These signals allow for real-time updates, such as live video feeds or telemetry data, and can be exploited for drone detection. By capturing and analyzing RF signals, an RF sensor can trigger deep-learning models to detect the presence of drones in real-time. One of the main advantages of RF-based detection systems is their ability to operate effectively in various lighting conditions and adverse weather, making them highly adaptable for day and night operations. However, these systems can be susceptible to interference from other devices that operate on the same 2.4 GHz frequency band, such as WiFi routers or Bluetooth devices. Implementing deep learning algorithms on RF data allows the system to extract specific features from the RF signals, improving detection accuracy and resilience to noise or interference from non-drone signals.

To optimize the feature extraction module at different resolutions, the end-to-end deep learning-based model built with stacked convolutional layers and multiscale architecture has been proposed by [46] to detect the presence of drones through RF signature. The proposed system can classify signals from both UAVs and the controller, and the communication is established at the 2.4 GHz frequency band. To create a challenging signal classification environment, other devices operating in the same range, such as WiFi signals and Bluetooth, are also considered in CardRF [47]. An additive white Gaussian noise (AWGN) is added to the dataset to create noisy samples of different SNRs. Those signals have varying noise levels and were trained on the proposed model, which helps to produce a robust model. The training result obtained an average accuracy of 97.53% for drones. The precision, sensitivity, and F1 score for both drone and Bluetooth were high, meaning the proposed model can detect and classify them well. The proposed system is then evaluated using unseen noise levels to validate its performance, and the overall accuracy achieved is over 94%. Although the proposed model has the highest number of parameters, inference time is the fastest compared to [48], and it is believed that eliminating manual feature extraction techniques to identify the drone’s signal makes it computationally effective. The proposed model is suitable for real-time detection and can address the limitations in distinguishing drone RF signals from other RF signals, especially in noisy environments.

A multi-channel 1-dimensional convolution neural network (multi-channel 1DCNN) based on a deep-learning approach was proposed by [49] to train the DroneRF dataset [50]. To capture the presence of RF signals in surroundings such as WiFi, Bluetooth, and the RF signals between drone and controller, a USRP device [51] was utilized along with the presence of WiFi RF signals. Then, the obtained signals are stored and sent to the processing unit to extract the RF frequency segments from the obtained data. Before feeding them to the classification model, data channelization is performed using a full WiFi frequency spectrum to divide into multiple channels and feed each to the classifier. Rather than analyzing the frequency spectrum simultaneously, breaking the whole frequency into smaller channels can prevent misclassification because the classifier can focus more on the smaller spectrum. Multi-channel 1DCNN is an upgraded version of basic 1DCNN that can manage multivariate time series classification at one time, and this network will process these sequences together to identify the pattern. This model will extract and learn features from RF signals to classify. The proposed model is not only able to detect the presence of drones, but it also can recognize what type of the detected drones (Parrot Bebop, Parrot AR, and DJI Phantom) and their operation mode as well (off, on, connected, flying, hovering and video recording). These three situations demonstrate the reliability of the proposed model in detecting drones and ensuring their safety and security from drone attacks. Overall, the proposed model performed excellently during model training. By comparing model performance in terms of accuracy with [52]. The proposed model has outperformed overall accuracy in three different tasks: 100% for drone detection, 94.6% for drone detection and type identification, and 87.4 % for drone detection, type, and state identification. It demonstrates the reliability of the proposed model in detecting drones with other RF signals as disturbances. However, their research does not include the detection's inference time, so the detection system's reliability to detect in real-time cannot be verified.

Another model is introduced by [53] to detect the presence of drones and classify their types and operational modes. By aiming computationally efficient, RF-NeuralNet, a deep learning-based network, is designed to identify drones using its RF signatures. RF dataset was utilized [50], and background interference from multiple sources, such as WiFi and Bluetooth, was added to create a challenging environment. The proposed system mainly detects the presence of RF signals first, and then it will proceed with the drone classification once it is detected. To address the vanishing gradient problems, the proposed network incorporated multiple-level skip connections, and multiple-level pooling layers are involved in the deep-level feature extraction. Three monitoring tasks were performed, which involved identifying RF signals, classifying drones, and monitoring a variety of drone operating modes to ensure the robustness of the proposed model. RF-NeuralNet is compared with other state-of-the-art in accuracy, GFLOPs, and model parameters to validate model performance. RF-NeuralNet obtained the highest accuracy, 89%, with the smallest model parameters, 8k, and fewer GFLOPs, 5M, compared to [54-56]. The proposed model used the same dataset as the Multi-Channel 1DCNN model proposed by [49] during model training, and both models obtained excellent results. The reliability of the dataset used in the RF-based approach can be highlighted. Like the Multi-Channel 1DCNN model, the proposed model performs drone detection and classification tasks well. With fewer parameters than [54-56], the proposed model offers reduced complexity, which usually indicates low memory usage and faster inference times, making it more efficient for real-time detection. However, inference time was not included in the model evaluation, so the real-time performance of this proposed model cannot be guaranteed.

Several machine-learning algorithms were evaluated by [57] to detect drones using RF signals using self-made datasets [58]. The proposed residual CNN model, called Deep Residual Neural Network (DRNN), has achieved the highest accuracy based on several experiments conducted on spectrogram datasets under AWGN conditions and multipath environments. The more depth of CNN architecture there is, the more it could lead to vanishing or exploding gradients during backpropagation. Therefore, to solve the degradation problem in this situation, the classifier, without skipping connections from the residual block, could not learn the features well under noisy conditions (-60 dBm to 10 dBm), while the one that incorporated the skipping connection, which is the proposed model, able to show good performance from noisy dataset. However, a classifier without a skip connection can learn essential features for low noise levels (-60 dBm to -15 dBm). It can also distinguish and classify drone signals and multiple drones simultaneously, up to 7 drones, even at lower SNR regions, even with WiFi signals. Almost 99% accuracy is obtained during 0 dB SNR during single and multi-drone situations, and at -10 dB, 5% higher than the F1 score of the existing model was obtained during the single drone detection scenario. However, model evaluation does not consider inference time, so the real-time effectiveness of the proposed system cannot be ascertained.

**Table 4.** Available Datasets for RF-Based Drone Detection

| Reference | Details of Datasets |
|---|---|
| [47] | The dataset consists of four types of signals: five UAVs, five UAV flight controllers, five Bluetooth devices, and two WiFi routers. Each signal includes five million sampling points at 30 dB SNR. The dataset provides a challenging environment by incorporating interference from non-drone RF signals, such as WiFi and Bluetooth. |
| [50] | The DroneRF dataset includes RF signals from three types of drones (AR, Bebop, Phantom) in five operational modes (off, on, connected, hovering, flying, and video recording). The dataset contains 186 drone segments and 41 no-drone segments for detection and 227k RF signals across low and high-frequency bands for classification. The diverse operational modes enhance the dataset’s utility for training classification models in various scenarios. |

| | |
|---|---|
| [58] | This dataset contains nine drone signals, including remote control and video signals operating at 2.4 GHz and WiFi signals. It is particularly useful for studying the effects of interference from WiFi networks on drone signal detection and classification. |

Table 4 provides an overview of crucial datasets used for RF-based drone detection. These datasets encompass various drone types, operational modes, and interference scenarios, making them invaluable for developing robust detection systems. The datasets include millions of sampling points, providing a substantial basis for training deep learning models. The DroneRF dataset [50], for instance, contains data for different drone types and operational modes, which is essential for classifying not just the presence of drones but also their specific activities. Meanwhile, the dataset from [47] offers a complex environment with interference from Wi-Fi and Bluetooth signals, challenging models to effectively distinguish between drone and non-drone RF signals. Similarly, the dataset from [58] includes a range of drone signals, such as remote control and video signals at 2.4 GHz and WiFi signals, which allow for an extensive evaluation of interference effects. These datasets contribute to advancing deep learning techniques for accurate and efficient drone detection in diverse and challenging RF environments.

#### *5.1.2. Radar-based Drone Detection*

Radar-based drone detection is an electromagnetic method that uses radar waves to detect and locate drones, often outperforming other methods like RF and acoustic detection in terms of accuracy and robustness under various environmental conditions. Radar can reliably operate in noisy environments and is mainly unaffected by visual challenges such as rain, fog, or dust. Initially, detecting drones posed challenges for traditional radars due to the small radar cross-sections (RCS) of drones compared to larger objects like airplanes. However, introducing the micro-Doppler signature (MDS) with time-domain analysis has greatly improved radar's ability to distinguish between noise and drone targets, outperforming traditional Doppler-shift signature methods in this context [59].

Radar detection systems emit radio waves from a transmitter, reflecting off any object within the radar's range. The receiver captures these reflected signals and sends them to a processor for analysis. By examining these reflections, the system can gather essential data on the detected object's characteristics, such as its size, velocity, and distance from the radar sensor. Active radar systems, which transmit their signals, are preferred for drone detection over passive radars, which depend on external signal sources. Two common types of active radar used for drone detection are pulse radar and Frequency Modulated Continuous Wave (FMCW) radar. Pulse radar excels at long-range detection but struggles with short-range accuracy. In contrast, FMCW radar offers superior short-range performance (50-100 meters) and excellent range resolution, making it more suitable for drone detection [2].

Deep learning has enhanced radar-based drone detection by reducing noise and interference in radar signals [60]. Deep learning models can reconstruct original signals and filter out noise before converting the time-domain signals into the frequency domain using Fast Fourier Transform (FFT) methods. Several deep-learning techniques have been employed to classify radar signals more accurately, as shown in Table 5.

RCS signatures that were collected using millimeter-wave (mmWave) radars from [61] were used by [62] to perform drone detection and classification using the deep learning technique. The benefit of using mmWave frequencies [63] in radar systems is that they can provide a high resolution, which benefits detecting smaller drones. Traditionally, after the RCS signature is captured by radar, it is converted to an image for CNN to process and make a classification. Since this conversion technique increases the computational overhead and the model can be trained using a fixed learning rate, this research implemented a new approach: a long short-term memory (LSTM) network with weight optimization to reduce the computational overhead and an adaptive learning rate optimization (ALRO). It is used during training to enhance the ability of the model to adapt well to unpredictable and dynamic environments while classifying various types of drones. The performance of LSTM-ALRO is compared and outperformed the model performance of GoogLeNet and CNN-based model [64]. With 99.88% accuracy, it has successfully obtained and produced excellent performance in detecting both large and small drones. However, this research does not include obstacles in the dataset, such as bird samples or challenging weather. The ability to perform real-time detection also cannot be verified since this research did not consider inference time during model evaluation.

Table 5. Available Datasets for Radar-Based Drone Detection

| Reference | Details of Datasets |
|---|---|
| [65] | Measurements from 77 GHz FMCW radar, including data on six drones, birds, and humans. The dataset provides rich micro-Doppler signatures for training classifiers to distinguish between various objects in short-range detection. |
| [66] | The Real Doppler RAD-DAR dataset collected by the Microwave and Radar Group includes over 17,000 samples of drones, cars, and people. The radar operates at 8.75 GHz and has a bandwidth of 500 MHz FMCW. |
| [67] | The dataset from IRIS FMCW radar contains data on drones, birds, wind turbines, and other ground targets, with drones flying on moving backgrounds, providing a realistic and diverse environment for model training and testing. |

Deep convolutional neural (DCNN) was utilized by [68] as a drone detection model. FMCW radar dataset [65] was used for the training and testing model, which contains various sources such as birds, drones, and humans. The collected signatures will be transformed using Short Time Fourier Transform (STFT) to produce a matrix representing the received signature. This matrix forms a Micro-doppler RGB image in PNG, JPG, or JPEG format. This Micro-doppler image will be input by the DCNN classifier to extract features. The performance of the proposed DCNN-based model is compared with [69-71]. The size of the micro-Doppler image used was only 32x29, which cost less computation compared to others. Although the proposed model did not achieve the highest accuracy, 97.4 % accuracy was obtained with 15ms of signal dwell time, which is the fastest compared to others, and radar operating at 77 GHz is considered good as the proposed model used dataset that has more than three classes (drones, birds, people) compared to others. However, 77 GHz frequency is effective at short-range detection, which is equivalent to less than 1 km only, which means the proposed model may not detect targets at greater distances.

Radar data by microwave and Radar Group named Real Doppler RAD-DAR (Radar with Digital Array Receiver) [66], which contains drones, cars, and people, was used by [72]. To detect the presence of drones, the radar used is FMCW, which operates on an 8.75 GHz based frequency band with a $BW_{max}$ of 500 MHz. The received radars were processed to produce a large matrix, but then the size was reduced to 11×61 matrices representing the distance and Doppler frequencies in dBm. The proposed detection network is based on CNN, which is called CNN-32DC, with 32 filters along with a depth concatenation layer obtained good results where the highest accuracy of 96.85% with low time needed to detect and the least number of parameters compared to other models, ResNet-18, SqueezeNet, SVM, K-NN and LDA. The low time consumed shows that the model can perform fast detection that is suitable for real-time application. Obtaining high accuracy using a dataset containing objects other than drones shows that the model learns to recognize drones and distinguish them from others well. However, it would be more advantageous if flying objects like drone-like objects like birds were also included in the dataset. Their similar sizes will be more challenging for the detection system, but it is still a good approach to include it in the dataset so the model can be trained to learn and differentiate between drones and birds.

A moving surveillance radar system was used by [73] to sense the presence of a drone. The received signal is then processed to generate a range-doppler image in real-time, where each image contains information about the detected object, such as speed and distance. YOLOv5s is selected to analyze the range-Doppler plots images to make a classification. In this research, it is believed that implementing this surveillance radar system can address limitations in volumetric spatial coverage if focusing on the time and frequency of radar signatures to extract features. The dataset used is from IRIS FMCW radar by Robin Radar Systems B.V. [67], which contains drones, birds, wind turbines, and other ground objects. Some augmentation techniques were introduced in the dataset to create noises and variation. The involvement of objects other than drones can give good exposure to the model to analyze characteristics owned by them so the model can perform accurate detection. The effectiveness of this proposed system is proved by the mean Average Precision (mAP) obtained was over 99%. However, the results do not reflect the model's capability to detect drones at long distances, as the dataset contains drones that fly 500m from radar.

#### *5.1.3. Acoustic-based Drone Detection*

Acoustic-based detection leverages the unique sound produced by the rotating blades of drones. This distinct sound typically has a higher amplitude, distinguishing it from background noise. Since this approach relies on sound waves, it is unaffected by low-light or adverse weather conditions, such as fog or dust. However, noisy environments can reduce the detection system's effectiveness [74]. Another advantage of acoustic-based detection is its independence from the drone’s speed, making it suitable for real-time applications [2].

The process of acoustic detection involves three key steps: sound wave detection, feature extraction, and classification. Sound waves generated by drone rotors are captured by highly sensitive microphones, acting as acoustic sensors. The captured sound is matched against pre-identified drone signatures stored in a database (fingerprinting). The acoustic signature can be processed using both traditional and deep learning methods. Traditional techniques extract features such as Mel-frequency cepstral coefficients (MFCCs) [74, 75], Gammatone cepstral coefficients, linear prediction coefficients, and spectral roll-off [76], which helps identify important acoustic characteristics. Deep learning models like CNNs have substantially improved drone detection accuracy [75-77]. These models are trained using labeled datasets to extract important features from acoustic signatures, allowing them to recognize drones in real-world scenarios automatically. Table 6 lists various datasets available for training models using the acoustic-based approach.

Several experiments have been conducted to assess the performance of machine learning algorithms like Support Vector Machines (SVM) and CNNs in drone detection through acoustic signals [75]. The dataset used in these studies includes internally generated audio data of drones at varying altitudes and publicly available datasets, enhancing the reliability of the detection models. MFCCs are typically extracted from the audio clips to feed into the CNN models for drone identification. A field test using 220 microphones captured 10-second audio clips, which were analyzed in real-time on a Raspberry Pi. A survey of 35 participants showed that humans could distinguish drone sounds with 92.47% accuracy, while the CNN model achieved 80% accuracy, highlighting the potential for further improving machine learning models. In addition, CNNs also outperformed studies that used SVM [83], significantly reducing false positives.

Table 6. Available Datasets for Acoustic-Based Drone Detection

| Reference | Details of Datasets |
|---|---|
| [78] | Contains 90 audio clips and 650 video data comprising airplanes, birds, drones, and helicopters. |
| [79] | Includes two 11-minute, 6-second audio clips of Parrot Bebop and Parrot Mambo drones. |
| [80] | Includes audio clips from five drone types: Parrot Bebop 2, DJI Mavic Pro, DJI Matrice 100, DJI Spark, and DJI Phantom 4 Advanced, each with 12 signals of approximately 30 seconds. |
| [81] | It contains drone audio clips recorded using propeller noise in an indoor environment. |
| [82] | Features sounds from various drone classes with multiple background noises. |

The CNN14 model is a pre-trained audio neural network (PANN) [84] on the AudioSet dataset [85] that was used for drone detection and classification by [77]. By using the dataset of [78], which contains 90 audio clips from three classes (drone, helicopter, and background noise). The model achieved a mean Average Precision (mAP) of 97.2% on validation data and 88% on unseen test data. However, the model misclassified some helicopter samples as noise, suggesting the need for further refinement in distinguishing drone sounds from similar airborne vehicles.

In a study by [74], Random Forest (RF) and Multi-Layer Perceptron (MLP) classifiers were evaluated using public datasets [79, 80]. Acoustic features, including 26 MFCCs, were extracted to help distinguish different drone noises. MLP achieved an accuracy of 83% on unseen test data, outperforming RF, which achieved 75%. It suggests that neural networks like MLP can handle complex and non-linear acoustic features better than traditional algorithms. However, evaluating detection performance at varying distances from the microphone remains essential for optimizing detection range.

To address the lack of acoustic datasets, the authors [44] generated an artificial dataset based on this dataset [81] using Generative Adversarial Networks (GANs) [86]. This GAN-based approach, modeled on WaveGAN architecture [45], combined drone sounds with background noises from publicly available datasets. The generated audio clips were converted into spectrograms to serve as input for deep learning models, including CNNs, RNNs, and CRNNs. CNN demonstrated the best performance in detecting drones even in noisy environments, showing that GAN-generated data could enhance detection accuracy and serve as effective data augmentation. The benefits of data augmentation for acoustic-based detection were demonstrated by [87], who used a dataset combining [82] with no-drone sound data from YouTube. Various types of noise distortions, including harmonic distortion, environmental noise, pitch shifting, and delay, were introduced to simulate real-world conditions. The VGGish network, inspired by VGGNet for image classification, achieved 99.1% accuracy in detecting drone sounds and 97.2% in recognizing non-drone sounds. These results suggest that data augmentation techniques can significantly improve the performance of acoustic-based drone detection models by mimicking real-world noise distortions.

Acoustic-based detection holds promise due to its independence from visual conditions and effectiveness in real-time applications. Deep learning models, particularly CNNs, have proven capable of improving detection accuracy. However, the availability of high-quality acoustic datasets and the inclusion of real-world noise and environmental conditions remain critical for advancing these systems' performance.

### 5.2. Optical Approach

The optical approach to drone detection employs cameras to identify and track drones by analyzing the images or video footage captured. This method requires high-quality visual data for optimal performance, as image clarity significantly impacts detection accuracy. While various types of cameras are available for optical detection, this section focuses on two fundamental types: digital and thermal infrared. Digital cameras offer high-resolution imagery, making them ideal for detecting drones in clear weather conditions. On the other hand, thermal infrared cameras excel in detecting drones based on heat signatures, making them useful in low-light or adverse weather conditions. Figure 6 illustrates the various mechanisms involved in optical drone detection, showcasing how different types of cameras capture and process drone activities.

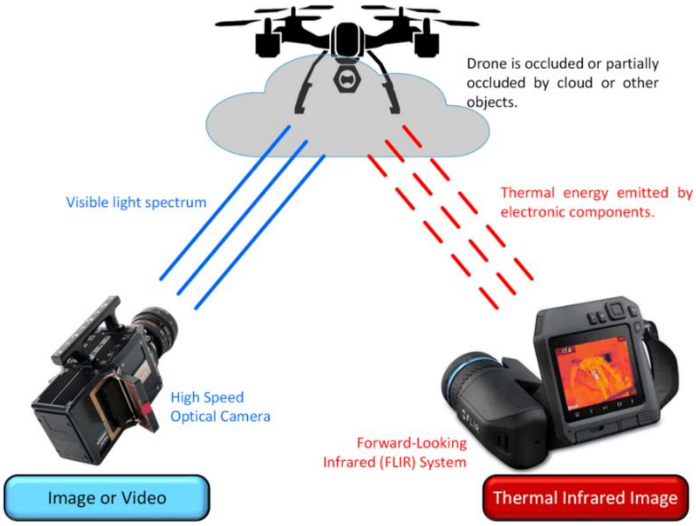

Figure 6. Drone Detection Based on Optical Approach

***5.2.1. Visible Image-based Drone Detection***

The optical or visual drone detection system focuses on identifying and locating drones through images or videos. The performance of these systems is significantly impacted by environmental challenges such as heavy rain, dust, fog, and long-range distances, as well as objects that may visually resemble drones. Cameras capture the images or video from the target area, after which the drone's features are extracted to identify its presence. To enhance system effectiveness, real-time detection provides early warning and tracking. With technological advancements, integrating deep learning has become a popular and effective approach for drone detection, particularly in object detection tasks [88].

CNNs are widely employed for drone detection due to their superior performance in image-based tasks. Compared to traditional radar systems, object detection models have shown greater accuracy and adaptability in drone detection [89, 90]. Object detection systems typically involve two main tasks: localization and classification. Localization determines where the target object is located within the image or video, while classification identifies the object. Object detection methods are categorized into one-stage and two-stage detectors. Examples of one-stage detectors include You Only Look Once (YOLO) [91] and Single Shot Multibox Detector (SSD) [92]. In contrast, two-stage detectors include Region-based Convolutional Neural Network (R-CNN) [93], Fast R-CNN [94], and Faster R-CNN [95]. Table 7 lists various datasets utilized in the visual approach to drone detection.

Table 7. Available Datasets for Visual-Based Drone Detection

| Reference | Details of Datasets |
|---|---|
| [96] | A series of annotated videos with a resolution of 1270×720 pixels for detecting small objects, including UAVs, boats, vehicles, people, and birds. |
| [97] | Four types of UAVs are included in this dataset, with images extracted from videos at 1280×720 resolution and a frequency of 10 fps. |
| [98] | 479 images of birds from 300 species and 1916 images of drones, sourced from Google and Kaggle. |
| [99] | 4635 images of drones with diverse backgrounds at a resolution of 1920×1080. |
| [100] | The dataset contains five classes: airplane, bird, drone, helicopter, and malicious drone. |
| [101] | Small drones and birds are captured in 77 training videos and 30 testing videos. |
| [102] | 2860 images of small drones at 1920×1080 resolution, captured from a distance of approximately 500 meters from a fixed ground camera. |

Research by [103] evaluated pre-trained SSD-based models to determine which model was most suitable for deployment in web applications. The models were trained on RGB datasets containing four classes—drones, helicopters, kites, and birds—sourced from YouTube to provide detailed pixel-level insights. SSD MobileNet v2 FPN demonstrated superior performance among the chosen lightweight models, achieving the lowest loss value, highest precision, and best recall for large, medium, and small target detection. It also had the lowest average detection time when deployed on a web application, showing the model's reliability for real-time detection. However, this study did not consider the effects of weather or multi-target environments during evaluation.

A lightweight and accurate detection model for real-time detection was proposed by [97]. This research made its dataset, which contains four types of drones, to train on the modified YOLOv4 model. Instead of using the default backbone of YOLOv4, MobileNet is used, and all convolution blocks are replaced with depth-wise separable convolutions to reduce model parameters and enhance feature extraction. By comparing with Faster-RCNN, SSD, EfficientDet, and original YOLOv4, proposed a lightweight model, combination of YOLOv4 and MobileNetv1 as the backbone, obtained high mAP, 93.14 %, fast detection of 82 fps in performing small and multi-target real-time detection. However, the results do not cover challenging scenarios like drone-like object detection, harsh weather, or inadequate lighting. In continuation of this research, research by [104] aimed to address challenges such as the struggle to differentiate drones and birds, the powerful GPU needed for real-time detection, and the limitations of the dataset. The dataset used is the same as [98] for a fair comparison. The authors trained and fine-tuned YOLOv5's architecture on Google Collab to ensure it was compatible with a custom dataset and tuned the value of the hyperparameter during training to increase detection accuracy. As a result, this study demonstrated the superior performance of YOLOv5 over YOLOv4 in terms of mean Average Precision (mAP), F1 score, and recall. However, the inference speed is slightly increased due to the increment in several parameters, but still suitable for real-time detection.

Research by [99] has tried to solve the challenge of detecting small targets, and as an effort, a new detection method was introduced using a YOLOv5-based model. This involved extending the input layer of YOLOv5s from a single-frame image to multiple frames and incorporating inter-frame optical flow to enhance small target detection. The model is trained on a dataset comprising 4625 images of drones with various backgrounds, each with a size of 1920×1080 pixels. 7:2:1 is a ratio of splitting datasets for training, validation, and testing. The model achieved a mAP of 0.8687, outperforming YOLOv4 with a 6.89% improvement in average precision while maintaining a speed above 30 frames per second. It shows that the proposed model can perform real-time detection on multi-targets and detect small drones even in complex backgrounds. However, drone-like objects are not included in the dataset, which poses a challenge for the model in distinguishing them from drones, and it is also limited in detecting drones using video as input.

By improving the architecture of the YOLOv8m model, a fast and reliable drone detection system was developed [105]. With the addition of P2 Layer, Multi-Scale Image Fusion (MSIF), and utilizing copy and paste augmentation technique, this improved version provides a fair balance between speed and accuracy, especially in targeting small objects. The proposed model is trained on the dataset [101] containing drones and birds with various lighting conditions. As a result, YOLOv8m with P2 Layer and MSIF can detect small and multi-target objects in 640-pixel images at 45.7 fps while in 1280-pixel images at 17.6 fps. However, the research did not state the model size, but due to the addition of the P2 layer, the number of parameters of the proposed model is expected to increase.

#### *5.2.2. Thermal Infrared Image*

Thermal infrared (TIR) imaging utilizes heat signatures emitted by objects, including drones, to detect their presence. Infrared cameras, equipped with thermal energy sensors, convert these heat signatures into images. Drones emit thermal energy through components such as batteries, rotors, and propellers, enabling infrared sensors to identify them. Studies have shown that batteries produce the most heat, with motors and speed controllers contributing less [106]. This approach offers significant advantages in environments with low visibility, such as fog, smoke, or nighttime conditions, making it ideal for drone detection in a wide range of scenarios [107]. However, detecting small targets remains challenging due to TIR images' lack of shape and texture detail [108]. Traditional methods have addressed these challenges by enhancing features and suppressing background noise, but these methods require high computational resources and are time-consuming. While deep learning models offer faster detection speeds, they are often optimized for visible images, necessitating additional adjustments to improve small infrared target detection [109]. Table 8 lists various public datasets available for TIR-based drone detection.

Table 8. Available Datasets for Thermal Infrared Image-Based Drone Detection

| Reference | Details of Datasets |
|---|---|
| [110] | Contains 10,000 thermal images with 160×120 resolution, with multiple drones per image, recorded in an indoor lab. |
| [111] | Includes 4737 IR images of a small quadrotor drone with 640×512 pixels across various backgrounds such as mountains, cities, and seas. |
| [112] | 7908 IR images containing speckle noise, salt and pepper noise, and uneven illumination, with small drone and bird targets at 640×640 resolution. |
| [113] | 318 RGB-thermal video pairs at 25 FPS with various drone types flying in different scenarios, such as urban areas and forests. |
| [114] | 410 thermal infrared videos containing 438k bounding boxes of drones were recorded in varied environments, with over half the targets being less than 50 pixels. |

Research by [110] developed a custom CNN model to detect multiple drones simultaneously by training on a self-generated thermal dataset. The thermal signatures of various drones, such as DJI and SYMA models, were captured using a FLIR Lepton 3.5 camera in indoor and outdoor environments. The model demonstrated an average accuracy of 99% in estimating the number of UAVs present, simultaneously detecting up to 10 drones. While the FLIR camera's lower resolution of 160×120 pixels facilitated faster computation, its frame rate of 8.7 Hz allowed for real-time detection. However, the study did not address detecting small targets or drone-like objects, such as birds, which may have similar thermal signatures.

The IRSDD-YOLOv5 model was proposed by [111] to address the challenge of detecting small infrared targets. Using YOLOv5s as a base model, an Infrared Small Target Detection Module (IRSTDM) is designed and added at the neck of YOLOv5 to provide better extraction for rich semantic information of small infrared targets. Then, an additional prediction head is added at the head part specifically for detecting small targets, which makes the current total prediction head four. A new loss function called NWD is introduced to enhance the positioning loss function detection and increase the accuracy and reliability of the model, specifically when involving small targets. The proposed model is trained using a self-made single-frame infrared small target drone dataset (SIDD) and evaluated its performance with other methods: BlendMask [115], CondInst [116], Solov2 [117], BoxInst [118], Yolact++ [119], Mask R-CNN [120], YOLOv5 and YOLOv7 [121]. Overall, the proposed model achieves good results, which are maintained in the top three results in average accuracy (AP), which can prove the reliability of the proposed model in detecting small drones. The small size of the model shows it is a lightweight model, and the high FPS shows the ability to detect in real-time detection. When comparing the detection results with four different scenes (city, mountain, sea surface, and sky scenario), it can be seen that the detection in the mountain scenario is quite challenging because the overall result of AP for all models in this scenario is the lowest among other scenarios. For segmentation results, the proposed model is more consistent in detecting real targets than other methods. With complex backgrounds and varying shapes and sizes of targets, the proposed model still achieves excellent performance with lower model parameters and higher real-time detection. However, this research still faces unresolved issues, particularly in multi-target detection, such as when multiple drones are in a single frame.

By targeting to solve drone detection and tracking in challenging scenarios for real-time detection, several techniques were proposed to develop accurate and reliable detection and tracking drones [122]. The techniques include drone detection that utilizes YOLOv7, Simple Online, and real-time tracking (SORT) algorithm for drone tracking, and noise distortion detection that involves a vision transformer (ViT) paralleled with customized CNNs. The model has been trained using the Drone Detection Dataset [112], which contains infrared images of small drones and birds in challenging environments like cloudy sky, fog, mist, and thick forest cover to increase the algorithm's robustness. Some noise distortions, such as speckle noise, salt and pepper noise, and uneven illumination noise, have also been introduced into the dataset. Overall, this proposed method achieved excellent performance in distinguishing between drones and birds during training and proves the robustness of the detection model even under noise distortions. However, the evaluation process is performed using a testing dataset that only involves new images with drone bounding boxes, and the result is that the YOLOv7 model obtained 94.26% of mAP. Since bird bounding boxes were not included during the testing phase, their presence was not adequately assessed to determine whether the model could distinguish between drones and birds.

Meanwhile, the proposed drone tracking algorithm can detect drones in all frames during the testing phase on videos without tracking labels [112]. The SORT algorithm can track multiple objects simultaneously, but this research did not cover infrared images containing multi-object scenarios. In terms of real-time detection, this research did not provide any details about the inference time for drone tracking in IR videos, which is one of the essential criteria to be evaluated for real-time detection and tracking. The evaluation has been decided using a confusion matrix for noise distortion detection. It can be seen that the proposed algorithm produced high precision in classifying each kind of noise distortion introduced in the dataset. Despite the absence of IR images from extreme weather conditions such as heavy rainfall in the dataset, the noise distortion introduced comparable challenges, making it an effective proxy for testing in challenging weather.

### 5.3. Review Summary of Deep Learning Approaches for Drone Detection

Table 9 presents various non-optical drone detection approaches utilizing datasets and deep learning configurations. Each study contributes uniquely to the field, with several demonstrating impressive accuracy levels—the RF-based detection models, such as those by Alam et al. [46] and Misbah et al. [53], benefit from deep learning architectures like stacked convolutional layers and lightweight networks. These models perform well in RF signal detection, even under noisy conditions, showcasing their potential for real-time applications. However, real-time inference times are often not reported, leaving a gap in assessing their practicality for deployment in operational settings. The reliance on RF signals also makes them vulnerable to interference from common signals such as WiFi and Bluetooth, suggesting the need for robust interference mitigation techniques in future models.

Table 9. Review of Non-optical Drone Detection

| Paper | Dataset | Deep Learning Configurations | Contribution | Future Recommendation |
|---|---|---|---|---|
| Alam et al., 2023 [46] | Cardinal RF Dataset [47] | End-to-end deep learning with stacked convolutional layers and multiscale architecture | Proposed a robust model for RF-based drone detection using multiscale architecture, achieving 97.53% accuracy with high precision and sensitivity across noisy environments. | Improve interference handling from non-drone RF signals; explore real-time performance enhancements. |
| Allahham et al., 2020 [49] | DroneRF Dataset [50] | Multi-Channel 1D Convolutional Neural Network (CNN) | Proposed a model to classify drones and their operational states from RF signals, achieving 100% drone detection accuracy and 87.4% state identification accuracy. | Investigate real-time detection performance; add further RF signal variations for better generalization. |
| Misbah et al., 2023 [53] | DroneRF Dataset [50] | RF-NeuralNet, lightweight deep learning with skip connections and multi-level pooling | Developed a computationally efficient RF detection model achieving 89% accuracy with smaller parameters and fewer GFLOPs than other state-of-the-art models. | Study the real-time performance and extend the dataset to include more challenging interference scenarios. |
| Basak et al., 2021 [57] | Self-made dataset [58] | Deep Residual Neural Network (DRNN) | Demonstrated effective drone classification under AWGN conditions and multipath environments using residual CNN architecture, achieving 99% accuracy in noisy environments. | Improve robustness in multipath environments with more diverse datasets; validate in real-time conditions. |
| Fu et al., 2021 [62] | Experimental mmWave radar dataset [61] | Long Short-Term Memory (LSTM) with adaptive learning rate optimization (ALRO) | Achieved 99.88% drone detection accuracy using LSTM for mmWave radar signals, improving detection efficiency by reducing computational overhead. | Extend the dataset with obstacles like birds or adverse weather; evaluate real-time performance. |
| Gomez et al., 2023 [68] | FMCW Radar dataset [65] | Deep Convolutional Neural Network (DCNN)-based model as a classifier to extract features. | With the fastest dwell time, 15 ms, the proposed model obtained a 97.4% accuracy rate in identifying three targets (drones, birds, and people). | Improve detection at long-range detection. |
| Garcia et al., 2022 [72] | Real Doppler RAD-DAR (Radar with Digital Array Receiver) [66] | A CNN-based model with 32 filters along with a dept concatenation layer, called CNN-32DC | With the least number of parameters, the proposed model obtained the highest accuracy, 96.85%, with fast inference speed compared to other models. | Extend the dataset with other flying objects, such as drone-like objects like birds. |
| Haifawi et al., 2023 [73] | IRIS FMCW Radar dataset [67] | YOLOv5s for classification of range-Doppler plot images | Achieved over 99% mean Average Precision (mAP) for drone detection using real-time surveillance radar systems combined with YOLOv5s for range-Doppler analysis. | Evaluate performance at greater distances; expand the dataset to include more flying objects for increased robustness. |
| Alaparthy et al., 2021 [75] | Internally generated drone audio data and public datasets | CNNs and SVMs for drone acoustic detection | Demonstrated improved drone detection accuracy using CNN over SVM with MFCC features and validated the performance with a real-time system based on Raspberry Pi. | Investigate the effects of weather conditions on detection performance; include more diverse drone audio datasets. |
| Yaacoub et al., 2022 [77] | Drone detection dataset [78] | Fine-tuned CNN14 model for audio detection | Achieved 97.2% mAP for drone detection with CNN14, demonstrating the effectiveness of transfer learning in acoustic drone detection tasks. | Improve the model's ability to distinguish helicopter sounds from noise; increase robustness against more complex sounds. |
| Ahmed et al., 2022 [123] | Public datasets [79, 80] | Random Forest (RF) and Multi-Layer Perceptron (MLP) algorithms for drone classifier | MLP outperformed RF with 83% accuracy, demonstrating the effectiveness of neural networks compared to traditional algorithms in handling complex and nonlinear features like MFCCS. | Investigate the model performance based on the detection distance from the sensor. |
| Al-Emadi et al., 2021 [44] | Generated artificial drone acoustic dataset using Generative Adversarial | CNN is built with two convolutional layers and a hidden fully connected layer, RNN is built based on an | Demonstrated the ability of CNN over RNN and CRNN to identify the sound of drones even in noisy backgrounds and proved the effectiveness of using | Extend the dataset with the sound of flying objects. |

| Paper | Dataset | Deep Learning Configurations | Contribution | Future Recommendation |
|---|---|---|---|---|
| | Network (GAN) based on drone audio dataset [81] | LSTM layer, and CRNN is built with a convolutional layer, two RNN-GRU-based layers, and a fully connected layer. | GAN in generating datasets where to solve the issue of dataset limitation. | |
| Kümmritz, 2024 [87] | YouTube no-drone sound dataset and dataset [82] | VGGish network for drone sound detection with data augmentation | Achieved 99.1% drone detection accuracy and 97.2% non-drone detection accuracy using the VGGish network with augmented audio data, simulating real-world conditions. | Further explore data augmentation techniques to simulate varying detection distances and more real-world environmental factors. |

The radar-based studies, particularly those utilizing millimeter-wave (mmWave) radars [63], highlight the advantages of radar systems in challenging environmental conditions. These systems can accurately detect smaller drones even at longer ranges and are unaffected by visual obstacles like fog or dust. Incorporating deep learning, such as LSTM networks, enables efficient processing and classification of radar signals, further enhancing detection accuracy. Nonetheless, many radar-based systems lack evaluations under complex real-world scenarios, such as interference from birds or other small objects. It underscores the need for future research to include more diverse datasets and incorporate real-time performance measures to ensure these models can function effectively in operational environments.

Acoustic-based detection models, as explored by Alaparthy et al. [75] and Yaacoub et al. [77], demonstrate the effectiveness of using sound as a distinguishing feature for drone detection. The use of deep learning models, such as CNNs and pre-trained audio neural networks, significantly improves the classification of drone sounds, even in noisy environments. These approaches are promising for situations where visual or RF-based methods may fail. However, challenges remain in distinguishing drone sounds from other airborne noises, like helicopters or birds, particularly in real-world environments with high ambient noise levels. Future studies should focus on enhancing the robustness of these models through better data augmentation techniques and the development of larger, more varied acoustic datasets that can improve detection accuracy in diverse settings.

Table 10 highlights the advancements in optical and thermal infrared image-based drone detection, showcasing the contributions of various studies in this field. Optical detection, mainly through digital and thermal infrared cameras, is crucial for identifying and tracking drones. This approach leverages high-quality visual data, and the integration of machine learning and deep learning models, such as CNNs, YOLO, and SSD, enhances the performance of these systems. The studies cited, including Wastupranata & Munir (2021) [103], Cai et al. (2022) [97], and Aydin & Singha (2023) [104], demonstrate how models such as YOLOv4 and YOLOv5 have evolved to improve accuracy in drone detection, even in complex environments. The datasets used in these studies, such as annotated RGB datasets, highlight the variety of data sources leveraged to train the models effectively, allowing them to differentiate between drones and other objects like birds or helicopters.

Table 10. Review of Optical Drone Detection

| Paper | Dataset | Deep Learning Configurations | Contribution | Future Recommendation |
|---|---|---|---|---|
| Wastupranata & Munir (2021) [103] | RGB dataset (YouTube): drones, helicopters, kites, birds | Pre-trained SSD models: SSD MobileNet v2, SSD ResNet50 | Demonstrated real-time drone detection on a web application using SSD MobileNet v2 with the best precision and recall | Suggested incorporating weather and multi-target scenarios for a more robust evaluation |
| Cai et al. (2022) [97] | Custom dataset with four types of drones | Modified YOLOv4 with MobileNet backbone | Developed a lightweight YOLOv4-MobileNet model with 93.14% mAP and 82 fps for real-time detection | Recommended addressing drone-like object detection in harsh weather and low lighting conditions |
| Aydin & Singha (2023) [104] | Same dataset as Singha & Aydin (2021): 479 bird images, 1916 drone images | YOLOv5 fine-tuned on a custom dataset | Achieved higher mAP, F1 score, and recall compared to YOLOv4 for small drone detection | Suggested improvements for inference speed and real-time detection efficiency |
| Sun et al. (2023) [99] | 4625 images of drones at 1920×1080 | YOLOv5 with optical flow and multi-frame input | Improved detection of small drones with complex backgrounds, achieving 0.8687 mAP while maintaining speed detection above 30 fps. | Recommended including drone-like objects in the dataset to enhance model differentiation capabilities |

| Paper | Dataset | Deep Learning Configurations | Contribution | Future Recommendation |
|---|---|---|---|---|
| Kim, Won (2023) [105] | Drones and birds with various lighting (640x, 1280x) | Improved YOLOv8m with P2 Layer and Multi-Scale Image Fusion | Balanced speed and accuracy, detecting small targets at 45.7 fps (640-pixel) and 17.6 fps (1280-pixel) | Future work could focus on optimizing model size and performance in different weather conditions. |
| Wilson et al. (2023) [110] | Self-generated thermal dataset (FLIR Lepton 3.5) | Custom CNN for multi-drone detection | Achieved 99% accuracy in estimating the number of drones in thermal images, with real-time capabilities | Proposed expanding detection capabilities to outdoor and long-range environments |
| Yuan et al. (2023) [111] | 4737 IR images of small drones | Infrared Small Target Detection Module is added at the neck of YOLOv5, IRSDD-YOLOv5, with aiming to enhance small target detection | Improved semantic extraction for small infrared targets and achieved high accuracy across diverse environments. | Recommended addressing multi-target scenarios and expanding dataset diversity |
| Bentamou et al. (2023) [122] | Drone Detection Dataset [112] | YOLOv7 with ViT and customized CNNs | Demonstrated robustness of drone detection under noisy conditions, with 94.26% mAP | Recommended extending evaluation to extreme weather conditions, such as heavy rainfall |
| Huang et al. (2024) [114] | 410 thermal infrared videos | Custom framework with optimized YOLO for tracking drones | Achieved 438k accurate bounding box detections in varied scenarios with over 50% targets under 50 pixels | Suggested expanding model generalization across different drone types and weather conditions |
| Kim & Hwang (2022) [42] | Acoustic and image data for UAV detection | GAN-based enhancement for drone detection | Improved detection and classification of malicious drones using multimodal data sources | Proposed further studies integrating multi-modal sensor fusion for improved detection in complex environments |
| Zamri et al. (2024) [124] | Small drone dataset with IR and visible images | YOLOv8 with attention mechanisms | Enhanced small drone detection using optimized YOLOv8, achieving high mAP and low latency | Suggested extending research into real-world applications with multi-target scenarios and adverse weather |
| Andraši et al. (2017) [106] | Thermal UAV images | TIR detection model using CNNs | Demonstrated night-time detection of UAVs using thermal infrared cameras, effective in low-visibility conditions | Proposed integrating deep learning models to improve detection under cluttered thermal backgrounds |
| Lin et al. (2023) [108] | Small UAV infrared images | YOLOv4-based TIR detection model | Optimized detection of small UAV targets with TIR data, demonstrating higher precision | Recommended extending work to real-time applications in harsh environments such as fog and dense forests |

The contributions of these previous studies reflect the rapid advancements in real-time drone detection capabilities, particularly through deep learning techniques such as YOLOv8 and SSD MobileNet, which achieve high precision and recall while maintaining efficiency in computational environments. However, several studies [102, 111] identify gaps in the current systems, such as difficulty detecting small targets in cluttered backgrounds and the challenges posed by adverse weather conditions like fog and rain. These limitations are evident in the datasets, many of which do not account for drone-like objects or environmental disturbances that affect detection accuracy. It highlights the need for more comprehensive datasets and models to adapt to these conditions.

Future directions proposed by the analyzed studies emphasize the need for further research into multi-target detection, adverse weather performance, and real-time applications in outdoor environments. For instance, integrating multi-modal sensor fusion to enhance detection accuracy in complex scenarios was suggested by [42], while expanding thermal infrared detection capabilities for longer-range targets was recommended by [110]. Additionally, the increasing popularity of lightweight models, such as YOLOv8, with attention mechanisms [124], suggests that future research should focus on optimizing model size and efficiency for deployment in real-world applications, where resource constraints may limit the use of larger models. It demonstrates the need to balance drone detection systems' accuracy, computational demands, and environmental adaptability.

## 6. ADVANCES IN DRONE DETECTION TECHNIQUES

Section 6 delves into recent advances in drone detection techniques, focusing on integrating multiple sensor types and applying deep learning models with attention mechanisms. Given the limitations of optical and non-optical approaches in various challenging environments, sensor fusion has emerged as a powerful method to enhance the reliability and accuracy of detection systems. This section explores how combining data from multiple sensors, such as acoustic, radar, and thermal infrared sensors, with optical systems helps to address the complexities of detecting drones in adverse conditions, such as low visibility, noise, and cluttered environments. In addition, attention mechanisms integrated into deep learning models further refine detection performance by allowing the models to focus on the most relevant parts of the data. The combination of sensor fusion and attention-based deep learning is proving to be a robust solution in overcoming the challenges posed by dynamic and diverse environments, enhancing the overall efficacy of drone detection systems across different operational scenarios.

### 6.1. Sensor Fusion

As discussed in the previous subsection regarding the two types of approaches, optical and non-optical approaches, it is evident that each has its limitations in solving challenges in drone detection. Sensor fusion is another way that can be employed to increase the effectiveness of the detection system, especially under adverse conditions, by adopting several approaches at one time to complement one another. While this may increase the complexity of the process due to the different types of datasets that need to be managed, the method's reliability makes it an appealing option. As discussed in [5], two approaches can be divided: early sensor fusion, as shown in Figure 7(a), and late sensor fusion, as shown in Figure 7(b). Early sensor fusion involves combining data of various types from different sources to create one harmonized dataset to be processed by the model. Meanwhile, different datasets will be processed separately by different models to make predictions for the late sensor fusion approach. Then, fusion layers merge the prediction outputs from different models using combination techniques such as concatenation, averaging, weighting, or attention-based. To produce the final output, ensemble algorithms connected to fusion layers combine the predictions made from each model using several methods, such as boosting, stacking, or voting.

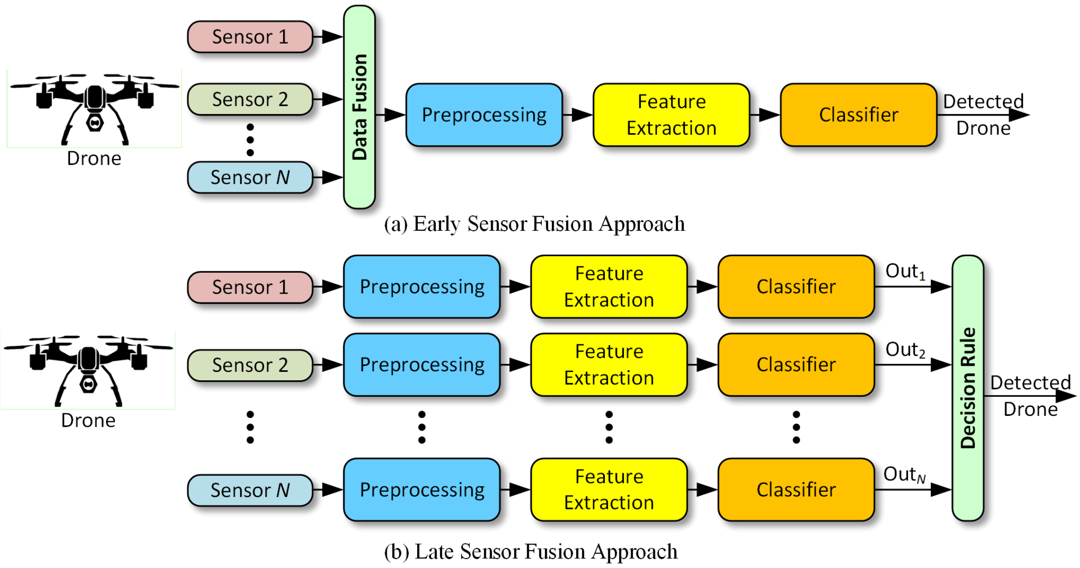


Figure 7. Sensor Fusion Approaches

Early and late sensor fusion each offer distinct advantages and drawbacks, making them suitable for different applications. Early sensor fusion combines data from multiple sensors into a single, unified dataset before processing, allowing for richer feature extraction by deep learning models. This approach can enhance performance in real-time applications by leveraging integrated data, making it particularly useful in complex environments. However, it comes with challenges, such as increased data management complexity and potential synchronization issues between heterogeneous sensor types, which can lead to higher computational costs. Additionally, if one sensor fails or underperforms, the overall dataset may suffer, diminishing system reliability. On the other hand, late sensor fusion processes data from each sensor independently and merges the results at the decision-making stage. This method allows for greater flexibility, as individual models can be optimized for specific sensor types, making it more resilient to sensor failures. Moreover, late fusion can better adapt to varying environmental conditions by allowing different sensors to specialize in specific tasks. However, it may introduce higher latency and often misses out on the deep data interaction that early fusion capitalizes on. Despite these drawbacks, late fusion is particularly beneficial for applications where sensor independence is crucial, as it minimizes the impact of errors from any single source. Ultimately, the choice between early and late fusion depends on the application's specific needs, balancing performance, complexity, and flexibility.

#### *6.1.1. Acoustic and Visual Sensor Fusion*

In [125], the authors developed a hybrid drone detection system that combines acoustic and visible light image data to improve the accuracy and reliability of drone detection, particularly in noisy environments. The system utilizes MFCC to classify drone acoustic signatures, employing CNNs for feature extraction and classification. Simultaneously, YOLOv5 is implemented for image-based detection, extracting visual features and predicting the presence of drones. The system then merges the outputs from both the acoustic and optical detectors using a logical OR function, improving overall accuracy by leveraging the strengths of each approach. While the visual detection system alone achieved an accuracy of 90.26% and the acoustic detection system achieved 88.96%, the fusion of these two methods led to a combined accuracy of 92.53%. This hybrid approach demonstrates the potential of multi-sensor fusion in enhancing drone detection accuracy, reducing the likelihood of false positives or missed detections by compensating for the limitations of individual sensor modalities. Integrating acoustic and visual data proves effective in environments where either method might struggle alone, such as in low-visibility conditions or ambient noise, highlighting the benefits of multi-modal fusion in improving system robustness.

#### *6.1.2. Radar and Visual Sensor Fusion*

A system utilizing a fusion of radar and visual detection methods was proposed by [126] to enhance drone detection accuracy. Initially, the radar-based detection operates using 3D K-band radar, with signals received by an antenna and processed using a CNN that analyzes Short-Time Fourier Transform (STFT) spectrograms. The system switches to visual detection when radar detection does not classify an object. For the visual detection network, ResNet-50 is employed to extract image features and classify the object. These two approaches work sequentially rather than concurrently. To validate this model's effectiveness, the authors compared it with existing studies, including [127-131]. While methods in [130] and [128] achieved higher accuracy at 97% and 98.70%, respectively. The proposed system achieved 71.43%, which was still deemed satisfactory. Notably, the datasets in previous studies did not include birds, which was a crucial challenge in this study. 3D K-band radar effectively extended the detection range and enabled the model to detect birds and drones. However, the accuracy, precision, recall, and F1 scores analysis revealed that visual detection outperformed radar in several scenarios, particularly when radar faced challenges detecting objects at long distances or in noisy environments. It underscores the value of combining both methods, as each compensates for the other's weaknesses. However, the dataset did not include varied environmental conditions, limiting the results' generalizability.

A real-time drone detection system was proposed by [132] that leverages radar and camera fusion to address challenges in detecting small drones at long distances and in adverse weather conditions. The dataset includes long-range data (up to 1.2 km) on drones, birds, and other objects like humans and vehicles. Radar detection manages time-series data with an algorithm [133] integrating target kinematics and tracking features. For visual detection, YOLOv5 was utilized to extract image features [134]. The radar and visual outputs are fused using a fully connected layer to generate combined features. Results demonstrated that the two approaches complement each other: when radar scores are low, the visual approach enhances classification accuracy, particularly in complex backgrounds and dynamic environments. It highlights the advantages of sensor fusion, where one sensor compensates for the weaknesses of the other, making the system more robust and versatile across different scenarios.

#### *6.1.3. Radio Frequency and Visual Sensor Fusion*

A novel sensor fusion approach combining Radio Frequency (RF) data and visual images was introduced by [135] to enhance real-time drone detection. The system leverages dual cameras and an RF analyzer to detect and classify drones by processing data from both sources. RF data was sourced from [50], while images were separately captured to build a fused dataset. Each visual image is paired with its corresponding RF signal, and the images are converted to greyscale to reduce processing complexity. The fused data is processed by two artificial neural networks (ANNs) for high and low bands of RF data, while a convolutional neural network (CNN) processes the images. A deep neural network combines the RF and visual data to produce a final classification. The model demonstrated excellent performance, especially compared to [52], which used a single sensor. The fusion system effectively handled multi-target scenarios and accurately distinguished between drones and helicopters with high confidence.

However, this fusion of RF and visual data introduces significant complexity due to the need to simultaneously process two distinct data types. The large dataset and the integration of multiple data streams require substantial computational infrastructure, particularly powerful GPUs, to maintain smooth real-time detection. The model's performance highlights the effectiveness of sensor fusion in multi-target environments, as it addresses limitations inherent in single-sensor systems. Nevertheless, the increased complexity raises concerns about scalability and the feasibility of deploying such systems in resource-constrained environments. Further research could explore optimization techniques to reduce computational demands or alternative methods of fusion that streamline processing without sacrificing accuracy. Additionally, testing the model under varied environmental conditions would provide insights into its robustness and generalizability, particularly in challenging scenarios such as adverse weather or complex urban environments.

#### *6.1.4. Multisensory Fusion Using Thermal Infrared, Visual, and Acoustic Approaches*

A comprehensive multisensory fusion system combining thermal infrared (TIR), visible light cameras (VCAM), and acoustic sensors was proposed by [78] to improve drone detection and classification accuracy. Instead of relying on a simple OR function for decision-making, this system assigns varying weights to each sensor based on their reliability. The dataset from [112] was used for training, containing diverse objects such as drones, birds, airplanes, and helicopters. YOLOv2 was employed to process and classify images captured by IRCAM and VCAM. At the same time, MFCCs were extracted from the acoustic data and classified using an LSTM network.

The performance of each sensor was evaluated under different scenarios, including short, medium, and long-distance detection. Results indicated that the TIR sensor (IRCAM) excelled at close-range detection. At the same time, the VCAM was more effective for longer distances due to its ability to capture richer object detail and shape. However, IRCAM struggled with certain elements, such as clouds, while the VCAM had issues with autofocus when multiple objects appeared in the frame. The acoustic sensor, evaluated based on overall F1-score rather than distance, outperformed IRCAM and VCAM, demonstrating a higher reliability in object detection. This sensor fusion approach significantly minimized false positives, achieving a drone classification accuracy rate of 78%, a notable improvement over the VCAM-only system's 67%. It demonstrates the power of sensor fusion in overcoming individual sensor limitations, particularly in diverse and challenging environments. Nevertheless, the system's complexity raises concerns about real-time processing and operational costs, which could be explored further to determine scalability and applicability in broader, real-world scenarios.

#### *6.1.5. Radio Frequency and Acoustic Sensor Fusion*

A fusion system combining radio frequency (RF) and acoustic sensors was proposed by [136] to enhance drone detection, particularly under noisy conditions with low signal-to-noise ratio (SNR). This system leverages the XBee Self-Built RF dataset [137], the DroneRF dataset [52], and the Drone Audio dataset [138], which contain diverse drone types, operating modes, and various noise conditions. Data preparation involved extracting key features such as Power Spectral Density (PSD), Short-Time Fourier Transform (STFT), MFCC, and Wavelet Packet Decomposition (WPD) from both the RF and audio datasets. These features were fused using an RNN-based method with LSTM layers, effectively integrating the temporal dynamics of both RF and audio signals.

Deep learning techniques, including CNN, RNN, LSTM, and SVM, were employed to enhance classification performance. Among them, the proposed fusion RNN-based model demonstrated superior accuracy, outperforming previous studies by an 8% improvement. This gain in accuracy is particularly significant in noisy environments, showcasing the robustness of the fusion approach in mitigating the effects of low SNR. This method highlights the strength of using sensor fusion to improve noise immunity. While individual RF or acoustic sensors may struggle with signal interference, their combined use enhances reliability and performance in real-time applications. However, the complexity of integrating features from distinct datasets and the computational load of deep learning models like LSTM and CNN pose challenges for real-time deployment.

### 6.2. Enhancing Drone Detection with Deep Learning Attention Mechanisms

Integrating attention mechanisms into deep learning models has become a powerful tool for enhancing drone detection systems. Drawing from the human brain's ability to selectively focus on relevant stimuli while ignoring distractions, attention mechanisms in neural networks similarly allow models to concentrate on the most essential features of the input data. In complex environments with noisy backgrounds, this selective focus enables the model to efficiently capture the essential structures, reducing the complexity of processing all the input simultaneously. Attention modules adaptively assign weights to different parts of the input data based on their relevance. It allows the model to prioritize critical features such as edges, textures, or spatial patterns crucial for accurate drone detection [139, 140].

Different attention mechanisms—such as SENet [141], CBAM [142], GAM [143], ECA [144], SA [145], and self-attention, commonly used in transformer architectures [146]—focus on various aspects of the input. These mechanisms can be categorized into four primary types: channel attention, which determines "what" to focus on; spatial attention, which identifies "where" to focus; temporal attention, which decides "when" to focus; and branch attention, which determines "which" part of the network to emphasize. Hybrid approaches combining two or more attention types can further improve performance. This flexibility in attention mechanisms makes them highly versatile and applicable across different neural network architectures and problem domains. The growing body of research confirms that attention mechanisms significantly improve drone detection models' precision, recall, and robustness by enabling them to filter out irrelevant information and focus on the key features distinguishing drones from other objects.

The core idea behind attention mechanisms is to compute a weighted sum of input features, where the weights are determined by a compatibility function that measures the relevance of each feature to the task at hand. Mathematically, this can be expressed as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (24)$$

where $Q$ is the query matrix derived from the input, $K$ is the key matrix, also derived from the input, $V$ is the value matrix, which represents the input features and $d_k$ is the dimensionality of the key vectors.

The attention mechanism first calculates the dot product between the query and key matrices, then scales it by $\sqrt{d_k}$, and applies a softmax function to normalize the result. It produces a set of attention weights to compute a weighted sum of the value matrix $V$. This process allows the model to dynamically focus on different parts of the input data based on their relevance to the task, improving its ability to detect and classify drones amidst complex environments. By incorporating such attention mechanisms, drone detection models can achieve higher accuracy, particularly in challenging scenarios where the distinction between drones and other objects (such as birds or background noise) is subtle and requires more nuanced analysis.

#### *6.2.1. Enhancing RF-Based Drone Detection with Multi-Dimensional Attention Mechanisms*

In RF-based drone detection, attention mechanisms have proven effective in improving model performance, particularly in complex electromagnetic environments. A notable example is implementing a Temporal-Channel-Spatial Joint Attention (TCSJA) module in the Spiking-EfficientNet architecture, as proposed by [147]. This attention-enhanced model replaces the traditional SE attention module in the MB-Conv blocks with a more sophisticated TCSJA module that combines three distinct attention mechanisms to address different dimensions of input data. Temporal attention focuses on important time steps using Global Average Pooling (GAP) to aggregate features over time, followed by Multi-Layer Perceptron (MLP) operations for weight assignments. Channel attention similarly uses MLP and squeeze-and-excitation to process features across channels at each time step. Spatial attention, inspired by the BAM mechanism, employs dilated convolutions to capture contextual information from the spatial layout of the RF signals. These combined modules enable the model to selectively focus on relevant features across time, channels, and spatial dimensions, enhancing its ability to classify drone signals accurately.

To evaluate the performance of this attention-enhanced model, datasets such as ZK_RF and DroneDetectV2 [148] were used, featuring RF signals from various drone types and interference sources like WiFi and Bluetooth. The attention-enhanced Spiking-EfficientNet was compared to models such as VGG11 [149], ResNet18 [150], and MobileNetV2 [151]. Results showed that the TCSJA module significantly improved classification accuracy while maintaining low energy consumption, a critical factor for SNN models prioritizing energy efficiency. This process makes the model ideal for high accuracy and low power consumption applications. The introduction of TCSJA offers several advantages, including enhanced adaptability to varying electromagnetic conditions and better detection rates in noisy environments. However, the complexity of the model also introduces challenges, such as increased computational load and potential overfitting in cases where the input data is less diverse. Additionally, while the model excels in controlled environments, its robustness in highly dynamic real-world scenarios requires further validation. The TCSJA-enhanced Spiking-EfficientNet is a promising direction for energy-efficient, high-accuracy RF-based drone detection.

#### *6.2.2. Attention-Enhanced Radar-Based Drone Detection*

In radar-based drone detection, attention mechanisms have been instrumental in improving the precision of distinguishing between drones and other objects, such as birds. In the network proposed by [152], channel and spatial attention modules are integrated into 1D-CNNs to allocate attention dynamically based on the characteristics of the input radar signals. These attention modules allow the model to focus on relevant features by computing weights, which helps filter out background noise and enhance the classification of drone signals. The model's core consists of Multi-Frequency Multi-Scale Deformable Convolutional (MFMSDC) Networks, which extract features at various frequencies and scales, a key factor in differentiating drones from birds—drones emit higher frequency signals while birds emit lower frequencies. Additionally, transition layers, equipped with attention modules, bridge the feature extraction process and further refine the output. Slow-time radar signals were collected and used as input to test the model, demonstrating significant improvements in accuracy, recall, and F1 scores when attention mechanisms were applied at the transition layers. These results highlight the attention mechanism's role in optimizing feature extraction, ensuring that the model better distinguishes between drones and background elements, particularly in noisy environments.

Another example is the YOLOv4-tiny-based model proposed by [153], which incorporates a hybrid attention mechanism, CBAM, combining channel and spatial attention. The CBAM was applied to radar data transformed into R-D spectrograms to enhance detection accuracy, particularly for small and fast-moving objects like drones [154]. Experiments with different attention mechanisms, including SE [141] and CA [155], showed that CBAM produced the highest mAP, demonstrating its effectiveness. However, the increased model size resulted in slightly reduced frames per second (fps), showing a trade-off between accuracy and speed. Attention mechanisms have improved detection accuracy across various radar-based models. However, the challenge lies in balancing accuracy gains with computational complexity, as adding attention layers can increase the model's size and reduce inference speed.

#### *6.2.3. Attention-Enhanced Acoustic-Based Drone Detection*

In acoustic-based drone detection, the Wavelet Packet Transform (WPT) was employed by [156] to address the challenge of detecting drones in environments with low signal-to-noise ratios (SNR). WPT enables the analysis of audio data at multiple resolutions, allowing the capture of both high- and low-frequency components in the signal. This zoomed-in feature extraction process provides detailed monitoring of specific signal parts, improving the model's ability to detect drone-specific acoustic signatures. A 1D-CNN was then utilized to extract relevant signal features, while a CNN detected patterns associated with drone sounds. A transformer model incorporating self-attention mechanisms was applied to enhance detection accuracy further. The self-attention mechanism selectively increases the weight of important features. It decreases the weight of irrelevant parts, allowing the model to focus more effectively on the acoustic signatures that correspond to drones.

The attention mechanism is critical in understanding and processing complex audio signals, particularly those with high sample rates. It enables the model to capture long-range dependencies in the audio data that might be missed by traditional convolutional layers, leading to improved performance. Comparing models with and without attention layers showed that including attention mechanisms significantly enhanced performance metrics such as mAP, precision, recall, and F1 score. However, an excessive number of stacked attention layers can harm performance, possibly because 1D-CNNs already capture many local features that do not require extensive attention processing. Overfitting due to redundant attention layers could also reduce the model's ability to generalize new data. While attention mechanisms offer substantial improvements in acoustic drone detection, their implementation requires careful tuning. Excessive stacking of attention layers may lead to diminishing returns and increased computational cost.

#### *6.2.4. Attention-Based Visual Drone Detection*

In the visible light image detection domain for drones, several advancements have been achieved by incorporating attention mechanisms into deep learning models. One significant enhancement was the integration of a Global Attention Mechanism (GAM) at the neck of the YOLOv8 architecture [89]. This attention mechanism improves the feature fusion process by focusing on critical parts of the image while minimizing information loss. A high-resolution detection head was also incorporated to enhance small target detection while reducing the model's parameters to boost speed. Using multi-scale feature extraction through SPD-Conv, rather than traditional convolutional layers, further refines the model's ability to capture essential features. Experiments on the TIB-Net dataset [102], which contains images of drones in various environmental conditions, demonstrated slight but meaningful improvements in precision, recall, and mAP, especially in scenarios involving multi-targets, blurry images, and small objects. However, this approach led to an increase in model size, which caused a slight reduction in frame rates (fps) but remained within acceptable limits for real-time applications.

Another approach employed spatial attention within the Tiny Iterative Backbone Network (TIB-Net), which helped filter out the noise and focus on small object localization [102]. This method enhanced the model's performance in detecting drones in complex environments, such as poor lighting or cluttered backgrounds. However, despite its effectiveness in improving the mAP by 1.4%, the added model complexity introduced delays in processing time, making it slower for real-time use. A similar observation was made in TGC-YOLOv5 [157], where GAM and a coordinate attention mechanism (CAM) were utilized. These mechanisms enhanced the detection of small targets by capturing richer semantic features and improving global feature interactions, particularly under challenging conditions such as fog or low light. Despite its success, the absence of drone-like objects in the dataset limited the model's ability to differentiate between similar objects.

Finally, an optimized YOLOv8n model, integrated with the ResCBAM attention mechanism, was designed to enhance the detection of small drones and birds [124] using the BirDrone dataset [158]. This optimization added an extra detection head to refine small object detection. Though this modification improved accuracy by 2.3% over baseline models, the increased model complexity led to decreased fps. Despite this drawback, the optimized model remained suitable for real-time detection, making it a promising approach for applications requiring the identification of small flying objects. While attention mechanisms can significantly improve detection accuracy, carefully balancing model complexity and real-time performance is crucial to ensure optimal deployment in practical scenarios.

#### *6.2.5. Attention-Enhanced Thermal Infrared Image Drone Detection*

In response to the challenges of detecting dim and small drones in thermal infrared imagery, [159] introduced an innovative approach with a modified RetinaNet model called A-RetinaNet. To enhance target texture features, the model incorporates a Super-Resolution Texture Enhancement (SRTE) module, which pre-processes images to improve their clarity before entering the network's backbone. An Asymmetric Attention Fusion Mechanism (AAFM) was deployed to refine detection accuracy further. This fusion mechanism integrates three attention modules—Pixel-by-Pixel Spatial Attention (PAM), Global-Channel Attention (GAM), and Effective Channel Attention (ECA)—to enable a bi-directional flow of rich contextual information between high- and low-level feature maps. This structure ensures that detailed features from high-resolution maps are effectively communicated to lower layers, enhancing the model's ability to detect even small and hard-to-see drones.

The A-RetinaNet replaces fully connected layers with Global Average Pooling (GAP), which reduces the number of parameters and increases efficiency without sacrificing accuracy. To address challenges in anchor box selection, the model utilizes the K-means algorithm to generate adaptive anchor boxes, significantly boosting the recall rate. The dataset used in testing this model contains drone images of varying sizes, from small to medium, against backgrounds such as sky and ground, making it a diverse and challenging benchmark. The ablation studies revealed a 1.33% improvement in mAP after integrating AAFM, and GAP further contributed a 0.36% boost in accuracy. The SRTE module significantly enhanced image quality, achieving a 1.33% accuracy gain. The adaptive anchor box generation also yielded substantial gains, increasing the average precision by 12.89%. Overall, the A-RetinaNet achieved an impressive 95.43% average precision and an 80.6% recall, outperforming standard RetinaNet by 16.32%. However, a notable drawback is the decrease in frames per second (fps) caused by the increased model complexity. Nevertheless, the balance between detection accuracy and processing speed makes this model suitable for real-time applications, particularly in complex thermal environments where small drone detection is critical. This solution illustrates the power of attention mechanisms in thermal infrared image processing. Despite the trade-off in speed, the performance gains in precision and recall reflect the significant potential of attention-based methods for drone detection in challenging scenarios. However, the decreased fps suggests that further optimization is needed to ensure practical real-time deployment, especially in scenarios requiring high-speed processing.

### 6.3. Review Summary of Advances in Drone Detection Techniques

Table 11 provides a comprehensive overview of the latest advances in drone detection techniques, highlighting the various datasets used, advanced methodologies employed, the contributions of each study, and recommendations for future work. This detailed summary offers insights into the effectiveness of different approaches, such as attention mechanisms, sensor fusion, and deep learning, across various modalities, including radar, radio frequency, visual, thermal, and acoustic data. The review of recent advancements in drone detection techniques shows that integrating attention mechanisms has significantly improved the performance of models across multiple sensing modalities. For instance, the use of multi-dimensional attention in RF-based detection [148] and the CBAM attention mechanism [124] demonstrate the power of attention modules in enhancing classification accuracy while mitigating background noise and irrelevant features. These methods prioritize the most salient information, leading to more efficient drone detection in complex environments. Similarly, self-attention mechanisms in acoustic-based detection [156] help models better capture long-range dependencies within audio signals, ensuring improved recognition even in low-SNR conditions.

Table 11. Review of Advanced Techniques in Drone Detection

| Paper | Dataset | Advanced Techniques Used | Contribution | Future Recommendation |
|---|---|---|---|---|
| Si et al. (2024) [147] | ZK_RF, DroneDetectV2 [148] | Spiking-EfficientNet based on Spiking Neural Networks (SNNs) with Temporal-Channel-Spatial Joint Attention (TCSJA) | Introduced multi-dimensional attention for RF signal processing; achieved the highest accuracy with the lowest energy consumption and without increasing the computational cost. | Further optimization for dynamic environmental conditions and interference |
| Liang et al. (2024) [152] | Slow-time Radar Signals | Channel and Spatial Attention in 1D-CNN | Improved classification of drones vs. birds in noisy radar environments; enhanced feature extraction | Explore more robust datasets and real-world scenarios for radar detection. |
| Liang et al. (2023) [153] | Radar Echo Data [154] | CBAM Attention Mechanism in YOLOv4-tiny. | Achieved high mAP with lightweight and compact architecture for radar detection | Focus on increasing fps for real-time applications. |
| Dong et al. (2023) [156] | Audio Dataset | Wavelet Packet Transform (WPT), 1-DCNN, and Transformer with Self-Attention | Improved overall result in mAP, precision, and recall of drone detection from audio data in low-SNR environments | Incorporate environmental noise variations in future experiments. |
| Zhao et al., 2023 [157] | SUAV-DATA, self-build dataset | Transformer Encoder Module, Global Attention Mechanism (GAM), and Coordinate Attention (CA) | Enhanced detection of small drones across various weather conditions | Test on drone-like objects to improve reliability |
| Xu et al. (2023) [159] | Public Infrared Drone Dataset | A-RetinaNet with Asymmetric Attention Fusion Mechanism (AAFM) and Super-Resolution Texture Enhancement (SRTE) module. | Achieved high precision and recall in detecting dim and small drones in thermal infrared images | Further reduce fps loss for real-time deployment. |
| Zamri et al. (2024) [124] | BirDrone Dataset [158] | ResCBAM Attention in YOLOv8n and additional detection head. | Improved accuracy in distinguishing between small drones and birds | Address fps reduction in complex environments |
| Svanström et al. (2020) [78] | IRCAM, VCAM, and Audio Sensor Dataset | Thermal, Visible Light, and Acoustic Fusion with Weighted Decision-Making | Minimized false alarms and enhanced drone classification | Explore more sophisticated fusion techniques for diverse sensor types |

Despite these advancements, challenges remain in optimizing models for real-time performance. Many studies [124, 153] emphasize the trade-off between accuracy and processing speed. As drone detection systems become more sophisticated with the addition of attention mechanisms, the computational complexity often increases, potentially hindering their application in real-time scenarios. This limitation calls for future research to focus on lightweight architectures that can balance performance and efficiency, ensuring that systems are accurate and responsive. Furthermore, while multi-sensor fusion techniques have proven effective, such as in [78], there is still a need for more robust methods capable of handling diverse environmental conditions. The current datasets often lack the variability to stress-test models across different weather, lighting, and background scenarios. Expanding the scope of datasets and developing adaptive fusion techniques could lead to more resilient drone detection systems, particularly in real-world, unpredictable environments.

## 7. CHALLENGES

The proliferation of drones in a variety of sectors has emphasized the existence of numerous obstacles in the development of detection systems that are both efficient and robust. This section critically evaluates the primary obstacles in drone detection, analyzing the underlying challenges and suggesting potential solutions.

### 7.1. Real-Time Detection

One of the most urgent obstacles in drone detection is the guarantee of real-time performance. Real-time detection is essential for effectively mitigating potential drone threats, particularly in high-risk and dynamic environments such as airports, military zones, and densely populated urban areas. Nevertheless, the variability in drone size, speed, and movement patterns substantially complicate the real-time detection process. Each drone requires a distinct processing and detection capability, ranging from small, fast-moving devices to larger, slower ones. Optimizing models for computational efficiency without compromising accuracy is necessary to alleviate this issue. For instance, lightweight deep learning models, such as architectures based on MobileNet, can significantly improve processing speed. Furthermore, real-time demands can be satisfied by incorporating hardware acceleration, such as GPUs, or specialized hardware, such as FPGAs. The optimization of algorithms that can maintain robustness across a variety of drone types while balancing detection speed and accuracy should be the primary focus of future research.

**7.2. Environmental Interference**

Background noise, weather conditions (rain, fog, snow), and lighting are all environmental factors that can significantly affect the precision of drone detection systems. For example, radar and RF-based systems are susceptible to interference from other electronic devices, while optical systems frequently encounter difficulties in low-light conditions. Detecting small drones in adverse weather conditions continues to be a critical concern. Sensor fusion solutions, which combine multiple detection systems (e.g., optical, thermal infrared, radar), present promising opportunities for addressing environmental challenges. Nevertheless, sensor fusion introduces its level of complexity, necessitating the use of advanced integration techniques and a corresponding increase in computational resources. Additionally, environmental disturbances may necessitate the investigation of sophisticated noise-filtering methodologies, including adaptive filtering and machine learning-based denoising.

**7.3. Differentiation from Drone-Like Objects**

Distinguishing drones from other aerial objects, such as birds or aircraft, presents a substantial obstacle in drone detection. This is especially challenging in optical-based methods, as drones may appear as small, indistinct objects. Misidentification can result in many false positives, undermining trust in the detection system and necessitating unnecessary security responses. Training machine learning models on extensive and diverse datasets encompassing drone images and drone-like objects enhances their ability to distinguish between them. It is possible to enhance the accuracy of differentiation by incorporating contextual information beyond visual cues, such as acoustic signatures or motion patterns. Additionally, integrating attention mechanisms into deep learning models can assist the system in concentrating on the most distinctive characteristics of drones, thereby decreasing the probability of misidentifications.

**7.4. Range and Altitude Limitations**

Drones can operate at various altitudes, and detection systems frequently encounter challenges when drones are too high or too distant from the sensor. Drones are more challenging to detect at longer distances due to their smaller size, which renders visual detection systems particularly restricted in terms of range. Similarly, signal attenuation over extended distances challenges radar and acoustic-based detection systems. Improved feature extraction techniques, such as deep learning-based super-resolution, and the implementation of high-sensitivity sensors could potentially improve detection capabilities at extended distances. Future advancements should also be investigated using advanced radar technologies, such as millimeter-wave radars, which provide superior resolution and range for detecting small aerial objects at considerable distances.

**7.5. Data Scarcity and Model Generalization**

Another significant challenge is the scarcity of high-quality training datasets, particularly for drone behaviors and diverse environmental conditions. To achieve high accuracy, deep learning models typically necessitate extensive labeled datasets. However, collecting real-world data under various conditions (e.g., adverse weather and multiple drone types) can be time-consuming and expensive. One potential resolution to this problem is the utilization of Generative Adversarial Networks to produce synthetic datasets that closely resemble real-world scenarios. Synthetic data has the potential to bridge the gaps in training datasets, thereby facilitating the generalization of models to previously unobserved environments. Nevertheless, it is imperative to exercise caution to guarantee that synthetic data accurately represents the intricacies of the real world.

**7.6. Sensor Fusion and Computational Complexity**

Although sensor fusion presents a promising solution to numerous challenges, it also introduces substantial computational complexity. Sophisticated processing techniques are necessary to synthesize data from various sensors, including radars, cameras, and acoustic detectors. Although late sensor fusion can enhance detection accuracy by allowing each sensor to process data independently before combining it at a later stage, it also increases the computational burden. Early sensor fusion, which involves combining and processing sensor data as a single input, is computationally simpler but more challenging to implement. It is imperative to resolve this trade-off to create cost-effective drone detection systems that can effectively manage real-time threats.

**7.7. Integration of IoT and 5G in Drone Detection Systems**

Integrating 5G networks and Internet of Things (IoT) infrastructure into drone detection systems is a substantial technological advancement that addresses numerous challenges encountered in current detection systems. The Internet of Things (IoT) facilitates the deployment of interconnected sensors across vast areas, thereby ensuring real-time detection with comprehensive coverage. Utilizing edge devices enables the decentralization of data processing, thereby enhancing the system's responsiveness and reducing latency. The interconnectedness of IoT also enables the continuous integration of new sensors without the need for substantial reconfiguration, rendering it highly adaptable to the expansion of applications such as autonomous drone operations and smart cities.

Introducing 5G networks further enhances drone detection systems by enabling high-bandwidth data transmission and ultra-low latency. These capabilities enable fluidly managing substantial quantities of high-quality data, including radar signals and high-resolution video, without bottlenecks. In high-risk environments, the low latency of 5G networks is particularly advantageous, as it facilitates rapid response times for threat mitigation. 5G's dependability guarantees uninterrupted operation in dense, urban environments with high network traffic, essential for preserving drone detection's accuracy and efficacy.

Although the integration of 5G and IoT presents many benefits, there are still obstacles to overcome, including network security, high deployment costs, and standardization. Establishing a large-scale IoT network and 5G infrastructure can be prohibitively expensive, particularly for smaller applications. Furthermore, IoT networks are susceptible to cyberattacks, necessitating consistent software updates and secure communication. It will also be essential to ensure these technologies' interoperability and seamless integration by standardizing protocols and communication frameworks across various systems. Future research should concentrate on overcoming these obstacles while simultaneously improving the security, efficiency, and scalability of drone detection systems powered by 5G and IoT.

Drone detection systems encounter numerous critical obstacles, such as the differentiation of drones from other objects, data scarcity, real-time processing requirements, and environmental interference. Deep learning and sensor fusion have demonstrated potential to address these challenges; however, they introduce system scalability and computational efficiency complexities. Innovative technologies like 5G and IoT present opportunities to improve these systems' scalability and real-time capabilities. However, they also introduce new standardization, security, and deployment cost challenges. Future research should concentrate on optimizing these technologies and developing drone detection systems that are more adaptable, efficient, robust, and capable of operating in complex environments.

## 8. CONCLUSION

This paper comprehensively examines drone detection systems that employ deep learning techniques, emphasizing the substantial modifications between 2020 and 2024. As drones' utilization in various sectors continues to expand, so do the associated security risks, such as privacy infringements, airspace violations, and illicit acts. The importance of detection systems that are both precise and dependable has never been greater. This review critically analyzed the primary obstacles to drone detection, including the absence of diverse and extensive datasets, sensor limitations, and environmental interference. We identified each approach's strengths and weaknesses by evaluating optical and non-optical methods, including radar, radio frequency (RF), acoustic detection methods, and visible and thermal imaging techniques. The primary contribution of this paper is the comprehensive examination of sensor fusion techniques and the integration of attention mechanisms into deep learning models, which have the potential to improve detection accuracy significantly. Attention mechanisms assist models in concentrating on the most pertinent features of the input data, thereby enhancing performance. At the same time, sensor fusion, in particular, mitigates the deficiencies of individual sensors by integrating complementary data sources. Furthermore, integrating emerging technologies, including the Internet of Things (IoT) infrastructure and 5G networks, is a promising approach to developing real-time, scalable, resilient drone detection systems. This paper urges the continuation of research to address the current challenges by improving sensor fusion methods, integrating synthetic datasets generated by GANs to improve training, and advancing the integration of IoT and 5G technologies. These developments will be essential in creating drone detection systems that are more dependable and durable, capable of reducing the increasing security risks associated with drones in a variety of applications.

## ACKNOWLEDGMENT

This publication has been funded by Telkom University grant number PU.007/AKD2.4/PPM/2022. The first author sincerely appreciates the Kulliyyah of Engineering, International Islamic University Malaysia, for providing the IIUM Engineering Merit Scholarship.

**REFERENCES**


[1] 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, 2021, doi: 10.1109/ACCESS.2021.3065926.

[2] 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. 11439-11455, 2022, doi: 10.1109/JSEN.2022.3171293.

[3] M. H. Rahman, M. A. S. Sejan, M. A. Aziz, R. Tabassum, J. I. Baik, and H. K. Song, "A Comprehensive Survey of Unmanned Aerial Vehicles Detection and Classification Using Machine Learning Approach: Challenges, Solutions, and Future Directions," *Remote Sensing*, vol. 16, 2024, doi: 10.3390/rs16050879.

[4] N. Al-lQubaydhi *et al.*, "Deep learning for unmanned aerial vehicles detection: A review," *Computer Science Review*, vol. 51, 2024, doi: 10.1016/j.cosrev.2023.100614.

[5] U. Seidaliyeva, L. Ilipbayeva, K. Taissariyeva, N. Smailov, and E. T. Matson, "Advances and Challenges in Drone Detection and Classification Techniques: A State-of-the-Art Review," *Sensors*, vol. 24, no. 1, 2024, doi: 10.3390/s24010125.

[6] R. J. Bunker and J. P. Sullivan, "Additional Weaponized Consumer Drone Incidents in Michoacán and Puebla, MX," in *Small Wars Journal*, ed, 2021.

[7] E. Ogao, "88 dead after military drone mistakenly strikes festival in Nigeria's Kaduna State, local officials say," in *ABC News*, ed, 2023.

[8] Amnesty, "Mali: Drone strikes killed 13 civilians including seven children in Amasrakad," in *Amnesty International*, ed, 2024.

[9] "Worldwide Drone Incidents." https://www.dedrone.com/resources/incidents-new/all (accessed 6 April, 2024).

[10] Y. Andrew. "Heathrow's runway had to be closed after illegally flown drone was 3ft away from smashing into Finnair plane flying at nearly 200mph, report reveals." https://www.dailymail.co.uk/news/article-12263721/Heathrow-runway-closed-illegally-flown-drone-3ft-hitting-200mph-Finnair-plane.html (accessed 6 April, 2024).

[11] C. Pittsburgh, "Unauthorized drone causes temporary ground stoppage at Pittsburgh International Airport," in *CBS News*, ed, 2023.

[12] M. Stephanie, "Drone being used to harass students undergoing trauma therapy in Asheville, organization says," in *WYFF4*, ed, 2022.

[13] G. Michele, "Drone hovers over house, peeps into daughter's room, mother says," in *KPLC*, ed, 2022.

[14] U. Seidaliyeva, D. Akhmetov, L. Ilipbayeva, and E. T. Matson, "Real-time and accurate drone detection in a video with a static background," *Sensors (Switzerland)*, vol. 20, no. 14, pp. 1-18, 2020, doi: 10.3390/s20143856.

[15] H. Liu, K. Fan, Q. Ouyang, and N. Li, "Real-time small drones detection based on pruned yolov4," *Sensors*, vol. 21, no. 10, 2021, doi: 10.3390/s21103374.

[16] S. A. H. Mohsan, N. Q. H. Othman, Y. Li, M. H. Alsharif, and M. A. Khan, "Unmanned aerial vehicles (UAVs): practical aspects, applications, open challenges, security issues, and future trends," *Intelligent Service Robotics*, vol. 16, 2023, doi: 10.1007/s11370-022-00452-4.

[17] C. Craye and S. Ardjoune, "Spatio-temporal semantic segmentation for drone detection," in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2019*, 2019, doi: 10.1109/AVSS.2019.8909854.

[18] M. Nalamati, A. Kapoor, M. Saqib, N. Sharma, and M. Blumenstein, "Drone detection in long-range surveillance videos," in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2019*, 2019/9// 2019: Institute of Electrical and Electronics Engineers Inc., doi: 10.1109/AVSS.2019.8909830.

[19] A. Coluccia *et al.*, "Drone-vs-Bird detection challenge at IEEE AVSS2017," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2017*, 2017, doi: 10.1109/AVSS.2017.8078464.

[20] A. Coluccia *et al.*, "Drone-vs-Bird Detection Challenge at IEEE AVSS2019," in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2019/9// 2019: IEEE, pp. 1-7, doi: 10.1109/AVSS.2019.8909876.

[21] A. Coluccia *et al.*, "Drone-vs-Bird Detection Challenge at IEEE AVSS2021," in *AVSS 2021 - 17th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2021: Institute of Electrical and Electronics Engineers Inc., doi: 10.1109/AVSS52988.2021.9663844.

[22] A. Coluccia, A. Fascista, L. Sommer, A. Schumann, A. Dimou, and D. Zarpalas, "The Drone-vs-Bird Detection Grand Challenge at ICASSP 2023: A Review of Methods and Results," *IEEE Open Journal of Signal Processing*, 2024, doi: 10.1109/OJSP.2024.3379073.

[23] Y. Zhang, N. Suda, L. Lai, and V. Chandra, "Hello Edge: Keyword Spotting on Microcontrollers," *arXiv preprint arXiv:1711.07128*, 2017.

[24] M. Zhou, "Research Advanced in Deep Learning Object Detection," in *2022 IEEE Conference on Telecommunications, Optics and Computer Science, TOCS 2022*, 2022, doi: 10.1109/TOCS56154.2022.10016018.

[25] D. K. Behera and A. B. Raj, "Drone Detection and Classification using Deep Learning," in *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*, 2020, doi: 10.1109/ICICCS48265.2020.9121150.

[26] T. Yigitcanlar *et al.*, "Artificial Intelligence Technologies and Related Urban Planning and Development Concepts: How Are They Perceived and Utilized in Australia?," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 6, no. 4, pp. 187-187, 2020/12// 2020, doi: 10.3390/joitmc6040187.

[27] P. P. Shinde and S. Shah, "A Review of Machine Learning and Deep Learning Applications," in *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*, 2018/7// 2018: Institute of Electrical and Electronics Engineers Inc., doi: 10.1109/ICCUBEA.2018.8697857.
[28] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets,* vol. 31, no. 3, 2021, doi: 10.1007/s12525-021-00475-2.
[29] L. Liu *et al.*, "Deep Learning for Generic Object Detection: A Survey," *International Journal of Computer Vision,* vol. 128, no. 2, pp. 261-318, 2020, doi: 10.1007/s11263-019-01247-4.
[30] P. Shruti and R. Rekha, "A Review of Convolutional Neural Networks, its Variants and Applications," in *2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS)*, Coimbatore, India, 2023: IEEE, pp. 31-36, doi: 10.1109/ICISCoIS56541.2023.10100412.
[31] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation,* vol. 9, no. 8, pp. 1735-1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[32] I. Goodfellow *et al.*, "Generative Adversarial Nets," in *Advances in neural information processing systems*, 2014 2014, vol. 27: Curran Associates, Inc.
[33] M. Jayasree and L. K. Rao, "A Deep Insight into Deep Learning Architectures, Algorithms and Applications," in *Proceedings of the International Conference on Electronics and Renewable Systems, ICEARS 2022*, Tuticorin, India, 2022: IEEE, pp. 1134-1142, doi: 10.1109/ICEARS53579.2022.9752225.
[34] E. Suryawati *et al.*, "Unsupervised feature learning-based encoder and adversarial networks," *Journal of Big Data,* vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00508-9.
[35] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F. Y. Wang, "Generative adversarial networks: Introduction and outlook," *EEE/CAA Journal of Automatica Sinica,* vol. 4, pp. 588-598, 2017, doi: 10.1109/JAS.2017.7510583.
[36] Y. Hong, U. Hwang, J. Yoo, and S. Yoon, "How generative adversarial networks and their variants work: An overview," *ACM Computing Surveys (CSUR),* vol. 52, 2019, doi: 10.1145/3301282.
[37] L. Zhang, A. Gonzalez-Garcia, J. Van De Weijer, M. Danelljan, and F. S. Khan, "Synthetic Data Generation for End-To-End Thermal Infrared Tracking," *IEEE Transactions on Image Processing,* vol. 28, no. 4, 2019, doi: 10.1109/TIP.2018.2879249.
[38] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference*, Munich, Germany, 2015: Springer International Publishing, doi: 10.1007/978-3-319-24574-4_28.
[39] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *IEEE conference on computer vision and pattern recognition*, 2017 2017: IEEE, doi: 10.1109/CVPR.2017.632.
[40] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," presented at the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 2016, 2016.
[41] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017 2017: IEEE, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.
[42] J. H. Kim and Y. Hwang, "GAN-Based Synthetic Data Augmentation for Infrared Small Target Detection," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 60, 2022, doi: 10.1109/TGRS.2022.3179891.
[43] J. Y. Zhu *et al.*, "Toward multimodal image-to-image translation," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, 2017 2017, vol. 30: Advances in neural information processing systems.
[44] S. Al-Emadi, A. Al-Ali, and A. Al-Ali, "Audio-based drone detection and identification using deep learning techniques with dataset enhancement through generative adversarial networks†," *Sensors,* vol. 21, no. 15, 2021, doi: 10.3390/s21154953.
[45] C. Donahue, J. McAuley, and M. Puckette, "Adversarial audio synthesis," in *7th International Conference on Learning Representations, ICLR 2019*, 2019.
[46] S. S. Alam *et al.*, "RF-Enabled Deep-Learning-Assisted Drone Detection and Identification: An End-to-End Approach," *Sensors,* vol. 23, no. 9, pp. 4202-4202, 2023/4// 2023, doi: 10.3390/s23094202.
[47] O. Medaiyese, M. Ezuma, A. Lauf, and A. Adeniran. *Cardinal RF (CardRF): An Outdoor UAV/UAS/Drone RF Signals with Bluetooth and WiFi Signals Dataset*, doi: https://dx.doi.org/10.21227/1xp7-ge95.
[48] O. O. Medaiyese, M. Ezuma, A. P. Lauf, and I. Guvenc, "Wavelet transform analytics for RF-based UAV detection and identification system using machine learning," *Pervasive and Mobile Computing,* vol. 82, 2022, doi: 10.1016/j.pmcj.2022.101569.
[49] M. S. Allahham, T. Khattab, and A. Mohamed, "Deep Learning for RF-Based Drone Detection and Identification: A Multi-Channel 1-D Convolutional Neural Networks Approach," in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies, ICIoT 2020*, Doha, Qatar, 2020: IEEE, doi: 10.1109/ICIoT48696.2020.9089657.
[50] M. Al-Sa'd, M. S. Allahham, A. Mohamed, A. Al-Ali, T. Khattab, and A. Erbad. *DroneRF dataset: A dataset of drones for RF-based detection, classification, and identification*, doi: 10.17632/f4c2b4n755.1.
[51] "USRP Software Defined Radio Device - National Instruments." National Instruments Corp. https://www.ni.com/en/shop/wireless-design-test/what-is-a-usrp-software-defined-radio.html?srsltid=AfmBOoqqczjOMwFo-uqMjP-14Re2SeiZUZ0imDf9WP2ni9ZpXHdEclv5 (accessed 6 April, 2024).

[52] M. F. Al-Sa’d, A. Al-Ali, A. Mohamed, T. Khattab, and A. Erbad, "RF-based drone detection and identification using deep learning approaches: An initiative towards a large open source drone database," *Future Generation Computer Systems*, vol. 100, pp. 86-97, 2019/11// 2019, doi: 10.1016/j.future.2019.05.007.
[53] M. Misbah, M. Dil, W. Khalid, and Z. Kaleem, "RF-NeuralNet: Lightweight Deep Learning Framework for Detecting Rogue Drones from Radio Frequency Signatures," in *2023 7th International Conference on Automation, Control and Robots (ICACR)*, Kuala Lumpur, Malaysia, 2023: IEEE, pp. 163-167, doi: 10.1109/ICACR59381.2023.10314637.
[54] S. Abeywickrama, L. Jayasinghe, H. Fu, S. Nissanka, and C. Yuen, "RF-based Direction Finding of UAVs Using DNN," in *2018 IEEE International Conference on Communication Systems, ICCS 2018*, Chengdu, China, 2018: IEEE, doi: 10.1109/ICCS.2018.8689177.
[55] R. Akter, V. S. Doan, G. B. Tunze, J. M. Lee, and D. S. Kim, "RF-Based UAV Surveillance System: A Sequential Convolution Neural Networks Approach," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju, Korea (South), 2020: IEEE, doi: 10.1109/ICTC49870.2020.9289281.
[56] T. Huynh-The, Q. V. Pham, T. V. Nguyen, D. B. D. Costa, and D. S. Kim, "RF-UAVNet: High-Performance Convolutional Network for RF-Based Drone Surveillance Systems," *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3172787.
[57] S. Basak, S. Rajendran, S. Pollin, and B. Scheers, "Drone classification from RF fingerprints using deep residual nets," in *2021 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, Bangalore, India, 2021: IEEE, doi: 10.1109/COMSNETS51098.2021.9352891.
[58] S. Basak. "Drone signals." https://github.com/sanjoy-basak/dronesignals (accessed 2024).
[59] F. Hoffmann, M. Ritchie, F. Fioranelli, A. Charlish, and H. Griffiths, "Micro-Doppler based detection and tracking of UAVs with multistatic radar," in *2016 IEEE Radar Conference (RadarConf)*, Philadelphia, PA, USA, 2016: IEEE, pp. 1-6, doi: 10.1109/RADAR.2016.7485236.
[60] D. Yu, L. Deng, I. Jang, P. Kudumakis, M. Sandler, and K. Kang, "Deep learning and its applications to signal and information processing," *IEEE Signal Processing Magazine*, vol. 28, no. 1, 2011, doi: 10.1109/MSP.2010.939038.
[61] P. Hügler, F. Roos, M. Schartel, M. Geiger, and C. Waldschmidt, "Radar Taking Off: New Capabilities for UAVs," *IEEE Microwave Magazine*, vol. 19, no. 7, pp. 43-53, 2018, doi: 10.1109/MMM.2018.2862558.
[62] R. Fu, M. A. Al-Absi, K. H. Kim, Y. S. Lee, A. A. Al-Absi, and H. J. Lee, "Deep Learning-Based Drone Classification Using Radar Cross Section Signatures at mmWave Frequencies," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3115805.
[63] M. Polese, L. Bertizzolo, L. Bonati, A. Gosain, and T. Melodia, "An Experimental mmWave Channel Model for UAV-to-UAV Communications," in *Proceedings of the 4th ACM Workshop on Millimeter-Wave Networks and Sensing Systems, mmNets 2020*, New York, NY, USA, 2020: Association for Computing Machinery, doi: 10.1145/3412060.3418431.
[64] B. K. Kim, H. S. Kang, S. Lee, and S. O. Park, "Improved Drone Classification Using Polarimetric Merged-Doppler Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 11, 2021, doi: 10.1109/LGRS.2020.3011114.
[65] A. Karlsson, M. Jansson, and M. Hämäläinen, "Model-Aided Drone Classification Using Convolutional Neural Networks," in *2022 IEEE Radar Conference (RadarConf22*, New York City, NY, USA, 2022: IEEE, pp. 1-6, doi: 10.1109/RadarConf2248738.2022.9764194.
[66] I. Roldan *et al.*, "DopplerNet: a convolutional neural network for recognising targets in real scenarios using a persistent range–Doppler radar," *IET Radar, Sonar & Navigation*, vol. 14, no. 4, pp. 593-600, 2020/04/01 2020, doi: https://doi.org/10.1049/iet-rsn.2019.0307.
[67] "IRIS Drone Radar.": https://www.robinradar.com/iris-counter-drone-rada (accessed 6 April, 2024).
[68] S. Gomez, A. Johnson, and P. Ramesh Babu, "Classification of Radar Targets Based on Micro-Doppler Features Using High Frequency High Resolution Radar Signatures," in *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, 2023, pp. 1-5, doi: 10.1109/NMITCON58196.2023.10276375.
[69] Y. Kim and T. Moon, "Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, 2016, doi: 10.1109/LGRS.2015.2491329.
[70] A. Angelov, A. Robertson, R. Murray-Smith, and F. Fioranelli, "Practical classification of different moving targets using automotive radar and deep neural networks," *IET Radar, Sonar and Navigation*, vol. 12, no. 10, 2018, doi: 10.1049/iet-rsn.2018.0103.
[71] S. Rahman and D. A. Robertson, "Classification of drones and birds using convolutional neural networks applied to radar micro-Doppler spectrogram images," *IET Radar, Sonar and Navigation*, vol. 14, no. 5, 2020, doi: 10.1049/iet-rsn.2019.0493.
[72] A. J. Garcia, A. Aouto, J. M. Lee, and D. S. Kim, "CNN-32DC: An improved radar-based drone recognition system based on Convolutional Neural Network," *ICT Express*, vol. 8, no. 4, 2022, doi: 10.1016/j.icte.2022.04.012.
[73] H. Haifawi, F. Fioranelli, A. Yarovoy, and R. Van Der Meer, "Drone Detection & Classification with Surveillance 'Radar On-The-Move' and YOLO," in *2023 IEEE Radar Conference (RadarConf23)*, San Antonio, TX, USA, 2023: IEEE, pp. 1-6, doi: 10.1109/RadarConf2351548.2023.10149588.
[74] C. A. Ahmed, F. Batool, W. Haider, M. Asad, and S. H. Raza Hamdani, "Acoustic Based Drone Detection Via Machine Learning," in *2022 International Conference on IT and Industrial Technologies, ICIT 2022*, 2022, doi: 10.1109/ICIT56493.2022.9989229.

[75] V. Alaparthy, S. Mandal, and M. Cummings, "Machine Learning vs. Human Performance in the Realtime Acoustic Detection of Drones," in *2021 IEEE Aerospace Conference (50100)*, Big Sky, MT, USA, 2021: IEEE, pp. 1-7, doi: 10.1109/AERO50100.2021.9438533.
[76] S. Salman, J. Mir, M. T. Farooq, A. N. Malik, and R. Haleemdeen, "Machine Learning Inspired Efficient Audio Drone Detection using Acoustic Features," in *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, Islamabad, Pakistan, 2021: IEEE, pp. 335-339, doi: 10.1109/IBCAST51254.2021.9393232.
[77] M. Yaacoub, H. Younes, and M. Rizk, "Acoustic Drone Detection Based on Transfer Learning and Frequency Domain Features," in *2022 International Conference on Smart Systems and Power Management (IC2SPM)*, Beirut, Lebanon, 2022: IEEE, pp. 47-51, doi: 10.1109/IC2SPM56638.2022.9988816.
[78] F. Svanström, C. Englund, and F. Alonso-Fernandez, "Real-time drone detection and tracking with visible, thermal and acoustic sensors," in *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy, 2020: IEEE, doi: 10.1109/ICPR48806.2021.9413241.
[79] S. Al-Emadi, A. Al-Ali, A. Mohammad, and A. Al-Ali, "Audio based drone detection and identification using deep learning," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Tangier, Morocco, 2019: IEEE, pp. 459-464, doi: 10.1109/IWCMC.2019.8766732.
[80] H. Kolamunna *et al.*, "AcousticPrint: Acoustic signature based open set drone identification," in *13th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2020: Association for Computing Machinery, doi: 10.1145/3395351.3401700.
[81] S. Al-Emadi. "DroneAudioDataset." https://github.com/saraalemadi/DroneAudioDataset (accessed 2024).
[82] S. Kümmritz and L. Paul, "COMPREHENSIVE DATABASE OF UAV SOUNDS FOR MACHINE LEARNING," in *10th Convention of the European Acoustics Association*, Turin, Italy, 2023: Forum Acusticum 2023, doi: 10.61782/fa.2023.0049.
[83] S. Mandal, L. Chen, V. Alaparthy, and M. Cummings, "Acoustic detection of drones through real-time audio attribute prediction," in *AIAA Scitech 2020 Forum*, 2020, vol. 1 PartF, doi: 10.2514/6.2020-0491.
[84] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition," *IEEE/ACM Transactions on Audio Speech and Language Processing,* vol. 28, 2020, doi: 10.1109/TASLP.2020.3030497.
[85] J. F. Gemmeke *et al.*, "Audio Set: An ontology and human-labeled dataset for audio events," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, 2017: IEEE, doi: 10.1109/ICASSP.2017.7952261.
[86] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM,* vol. 63, no. 11, pp. 139-144, 2020/10// 2020, doi: 10.1145/3422622.
[87] S. Kümmritz, "The Sound of Surveillance: Enhancing Machine Learning-Driven Drone Detection with Advanced Acoustic Augmentation," *Drones,* vol. 8, no. 3, 2024, doi: 10.3390/drones8030105.
[88] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 30, 2019, doi: 10.1109/TNNLS.2018.2876865.
[89] X. Zhai, Z. Huang, T. Li, H. Liu, and S. Wang, "YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection," *Electronics (Switzerland),* vol. 12, no. 17, 2023/9// 2023, doi: 10.3390/electronics12173664.
[90] H. J. Al Dawasari, M. Bilal, M. Moinuddin, K. Arshad, and K. Assaleh, "DeepVision: Enhanced Drone Detection and Recognition in Visible Imagery through Deep Learning Networks," *Sensors (Basel, Switzerland),* vol. 23, no. 21, 2023, doi: 10.3390/s23218711.
[91] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016/6// 2016: IEEE, pp. 779-788, doi: 10.1109/CVPR.2016.91.
[92] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9905 LNCS: Springer Verlag, pp. 21-37, doi: 10.1007/978-3-319-46448-0_2.
[93] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014/6// 2014: IEEE, pp. 580-587, doi: 10.1109/CVPR.2014.81.
[94] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015: IEEE, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.
[95] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 6, pp. 1137-1149, 2017/6// 2017, doi: 10.1109/TPAMI.2016.2577031.
[96] B. Bosquet, M. Mucientes, and V. M. Brea, "STDNet: A convnet for small target detection," in *British Machine Vision Conference (BMVC) 2018*, 2018.
[97] H. Cai, Y. Xie, J. Xu, and Z. Xiong, "A Lightweight and Accurate UAV Detection Method Based on YOLOv4," *Sensors,* vol. 22, no. 18, 2022, doi: 10.3390/s22186874.
[98] S. Singha and B. Aydin, "Automated drone detection using YOLOv4," *Drones,* vol. 5, no. 3, 2021, doi: 10.3390/drones5030095.
[99] Y. Sun *et al.*, "Enhancing UAV Detection in Surveillance Camera Videos through Spatiotemporal Information and Optical Flow," *Sensors,* vol. 23, no. 13, 2023/7// 2023, doi: 10.3390/s23136037.
[100] S. Jamil, M. S. Abbas, and A. M. Roy, "Distinguishing Malicious Drones Using Vision Transformer," *AI (Switzerland),* vol. 3, no. 2, 2022, doi: 10.3390/ai3020016.
[101] "Drone-vs-Bird Detection Challenge." https://wosdetc2023.wordpress.com (accessed 2024).

[102] H. Sun, J. Yang, J. Shen, D. Liang, L. Ning-Zhong, and H. Zhou, "TIB-Net: Drone Detection Network with Tiny Iterative Backbone," *IEEE Access,* vol. 8, pp. 130697-130707, 2020, doi: 10.1109/ACCESS.2020.3009518.
[103] L. M. Wastupranata and R. Munir, "UAV Detection using Web Application Approach based on SSD Pre-Trained Model," in *Proceedings of the 2021 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology, ICARES 2021*, Bali, Indonesia, 2021: IEEE, doi: 10.1109/ICARES53960.2021.9665191.
[104] B. Aydin and S. Singha, "Drone Detection Using YOLOv5," *Eng,* vol. 4, no. 1, pp. 416-433, 2023/2// 2023, doi: 10.3390/eng4010025.
[105] J.-H. Kim, N. Kim, and C. S. Won, "High-Speed Drone Detection Based On Yolo-V8," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, 2023: IEEE, pp. 1-2, doi: 10.1109/icassp49357.2023.10095516.
[106] P. Andraši, T. Radišić, M. Muštra, and J. Ivošević, "Night-time Detection of UAVs using Thermal Infrared Camera," *Transportation Research Procedia,* vol. 28, pp. 183-190, 2017/01/01/ 2017, doi: https://doi.org/10.1016/j.trpro.2017.12.184.
[107] Z. Yang, J. Lian, and J. Liu, "Infrared UAV Target Detection Based on Continuous-Coupled Neural Network," *Micromachines,* vol. 14, no. 11, 2023, doi: 10.3390/mi14112113.
[108] Z. Lin, M. Huang, and Q. Zhou, "Infrared small target detection based on YOLO v4," 2023, vol. 2450, 1 ed., doi: 10.1088/1742-6596/2450/1/012019.
[109] Z. Lin, M. Huang, and Q. Zhou, "Infrared small target detection based on YOLOv4," *Journal of Physics: Conference Series,* vol. 2450, 2023 2023, doi: 10.1088/1742-6596/2450/1/012019.
[110] N. A. Wilson, A. Jha, A. Kumar, and L. R. Cenkeramaddi, "Estimation of UAV Count Using Thermal Imaging and Lightweight CNN," in *2023 11th International Conference on Control, Mechatronics and Automation (ICCMA)*, Grimstad, Norway, 2023 2023: IEEE, pp. 92-96, doi: 10.1109/ICCMA59762.2023.10374791.
[111] S. Yuan *et al.*, "IRSDD-YOLOv5: Focusing on the Infrared Detection of Small Drones," *Drones,* vol. 7, no. 6, 2023, doi: 10.3390/drones7060393.
[112] F. Svanström, F. Alonso-Fernandez, and C. Englund, "A dataset for multi-sensor drone detection," *Data in Brief,* vol. 39, 2021, doi: 10.1016/j.dib.2021.107521.
[113] N. Jiang *et al.*, "Anti-UAV: A Large Multi-Modal Benchmark for UAV Tracking," *arXiv preprint arXiv:2101.08466 (2021),* 2021.
[114] B. Huang, J. Li, J. Chen, G. Wang, J. Zhao, and T. Xu, "Anti-UAV410: A Thermal Infrared Benchmark and Customized Scheme for Tracking Drones in the Wild," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 5, 2024, doi: 10.1109/TPAMI.2023.3335338.
[115] H. Chen, K. Sun, Z. Tian, C. Shen, Y. Huang, and Y. Yan, "Blendmask: Top-down meets bottom-up for instance segmentation," in *IEEE/CVF conference on computer vision and pattern recognition*, 2020 2020, pp. 8573-8581, doi: 10.1109/CVPR42600.2020.00860.
[116] Z. Tian, B. Zhang, H. Chen, and C. Shen, "Instance and Panoptic Segmentation Using Conditional Convolutions," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 45, no. 1, 2022, doi: 10.1109/TPAMI.2022.3145407.
[117] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, "SOLOv2: Dynamic and fast instance segmentation," in *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, Canada, 2020 2020.
[118] D. Jifeng, K. He, and J. Sun, "BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015 2015: IEEE, pp. 1635-1643, doi: 10.1109/ICCV.2015.191.
[119] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT++ Better Real-Time Instance Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 44, no. 2, 2022, doi: 10.1109/TPAMI.2020.3014297.
[120] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 42, no. 2, 2020, doi: 10.1109/TPAMI.2018.2844175.
[121] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 7464-7475.
[122] A. Bentamou, A. Zein-Eddine, O. Messai, and Y. Gavet, "Real-Time Drone Detection and Tracking in Distorted Infrared Images," in *2023 IEEE International Conference on Image Processing* Kuala Lumpur, Malaysia, 2023 2023: IEEE, doi: 10.1109/ICIPC59416.2023.10328374.
[123] C. A. Ahmed, F. Batool, W. Haider, M. Asad, and S. H. Raza Hamdani, "Acoustic Based Drone Detection Via Machine Learning," in *2022 International Conference on IT and Industrial Technologies (ICIT)*, Chiniot, Pakistan, 2022: IEEE, doi: 10.1109/ICIT56493.2022.9989229.
[124] F. Najihah Muhamad Zamri, T. S. Gunawan, S. Hajar Yusoff, A. A. Alzahrani, A. Bramantoro, and M. Kartiwi, "Enhanced Small Drone Detection Using Optimized YOLOv8 With Attention Mechanisms," *IEEE Access,* vol. 12, no. May, pp. 90629-90643, 2024, doi: 10.1109/ACCESS.2024.3420730.
[125] J. Kim *et al.*, "Deep Learning Based Malicious Drone Detection Using Acoustic and Image Data," in *2022 Sixth IEEE International Conference on Robotic Computing (IRC)*, Italy, 2022: IEEE, doi: 10.1109/IRC55401.2022.00024.
[126] S. E. Abdelsamad *et al.*, "Vision-Based Support for the Detection and Recognition of Drones with Small Radar Cross Sections," *Electronics (Switzerland),* vol. 12, no. 10, 2023, doi: 10.3390/electronics12102235.
[127] C. Wang, J. Tian, J. Cao, and X. Wang, "Deep Learning-Based UAV Detection in Pulse-Doppler Radar," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 60, 2022, doi: 10.1109/TGRS.2021.3104907.

[128] Y. Sun, S. Abeywickrama, L. Jayasinghe, C. Yuen, J. Chen, and M. Zhang, "Micro-Doppler Signature-Based Detection, Classification, and Localization of Small UAV with Long Short-Term Memory Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 8, 2021, doi: 10.1109/TGRS.2020.3028654.
[129] X. Cai and K. Sarabandi, "A machine learning based 77 GHz radar target classification for autonomous vehicles," in *2019 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, Atlanta, GA, USA, 2019: IEEE, doi: 10.1109/APUSNCURSINRSM.2019.8888647.
[130] W. Kim, H. Cho, J. Kim, B. Kim, and S. Lee, "Target Classification Using Combined YOLO-SVM in High-Resolution Automotive FMCW Radar," in *2020 IEEE Radar Conference (RadarConf20)*, IEEE, 2020: Florence, Italy, doi: 10.1109/RadarConf2043947.2020.9266477.
[131] W. Zhang, S. Li, C. Zhu, and C. Wang, "Classification of combination target's components based on deep learning," in *2019 International Applied Computational Electromagnetics Society Symposium-China, ACES 2019*, 2019, doi: 10.23919/ACES48530.2019.9060725.
[132] V. Mehta, F. Dadboud, M. Bolic, and I. Mantegh, "A Deep Learning Approach for Drone Detection and Classification Using Radar and Camera Sensor Fusion," in *2023 IEEE Sensors Applications Symposium, SAS 2023 - Proceedings*, 2023, doi: 10.1109/SAS58821.2023.10254123.
[133] V. Mehta, M. Bolic, I. Mantegh, and C. Vidal, "Tracking and Classification of Drones and Birds at a Far Distance Using Radar Data," in *Drone Detectability : Modelling the Relevant Signature*, 2021.
[134] F. Dadboud, V. Patel, V. Mehta, M. Bolic, and I. Mantegh, "Single-Stage UAV Detection and Classification with YOLOV5: Mosaic Data Augmentation and PANet," in *AVSS 2021 - 17th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2021: Institute of Electrical and Electronics Engineers Inc., doi: 10.1109/AVSS52988.2021.9663841.
[135] M. Aledhari, R. Razzak, R. M. Parizi, and G. Srivastava, "Sensor Fusion for Drone Detection," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, 2021: IEEE, pp. 1-7, doi: 10.1109/VTC2021-Spring51267.2021.9448699.
[136] A. Frid, Y. Ben-Shimol, E. Manor, and S. Greenberg, "Drones Detection Using a Fusion of RF and Acoustic Features and Deep Neural Networks," *Sensors*, vol. 24, no. 8, pp. 2427-2427, 2024/4// 2024, doi: 10.3390/s24082427.
[137] N. Ashush, S. Greenberg, E. Manor, and Y. Ben-Shimol, "Unsupervised Drones Swarm Characterization Using RF Signals Analysis and Machine Learning Methods," *Sensors*, vol. 23, no. 3, 2023, doi: 10.3390/s23031589.
[138] A. Deleforge, D. Di Carlo, M. Strauss, R. Serizel, and L. Marcenaro, "Audio-Based Search and Rescue with a Drone: Highlights from the IEEE Signal Processing Cup 2019 Student Competition [SP Competitions]," *IEEE Signal Processing Magazine*, vol. 36, 2019, doi: 10.1109/MSP.2019.2924687.
[139] G. Brauwers and F. Frasincar, "A General Survey on Attention Mechanisms in Deep Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, 2023, doi: 10.1109/TKDE.2021.3126456.
[140] M. H. Guo *et al.*, "Attention mechanisms in computer vision: A survey," *Computational Visual Media*, vol. 8, 2022, doi: 10.1007/s41095-022-0271-y.
[141] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, 2020, doi: 10.1109/TPAMI.2019.2913372.
[142] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11211 LNCS, doi: 10.1007/978-3-030-01234-2_1.
[143] Y. Liu, Z. Shao, and N. Hoffmann, "Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions," *ArXiv*, vol. abs/2112.05561, 2021.
[144] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, doi: 10.1109/CVPR42600.2020.01155.
[145] Q. L. Zhang and Y. B. Yang, "SA-Net: Shuffle attention for deep convolutional neural networks," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2021, vol. 2021-June: IEEE, doi: 10.1109/ICASSP39728.2021.9414568.
[146] A. Vaswani *et al.*, "Attention is all you need," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, 2017: Curran Associates Inc.
[147] Z. Si, C. Liu, J. Liu, and Y. Zhou, "Application of SNNS Model Based On Multi-Dimensional Attention In Drone Radio Frequency Signal Classification," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024/4// 2024: IEEE, pp. 231-235, doi: 10.1109/ICASSP48485.2024.10446694.
[148] J. S. Carolyn and C. W. John. *DroneDetect Dataset: A Radio Frequency dataset of Unmanned Aerial System (UAS) Signals for Machine Learning Detection & Classification*, doi: https://dx.doi.org/10.21227/5jjj-1m32.
[149] A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy, "Going Deeper in Spiking Neural Networks: VGG and Residual Architectures," *Frontiers in Neuroscience*, vol. 13, 2019, doi: 10.3389/fnins.2019.00095.
[150] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016: IEEE, doi: 10.1109/CVPR.2016.90.
[151] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, doi: 10.1109/CVPR.2018.00474.
[152] R. Liang and Y. Cen, "Radar Signal Classification with Multi-Frequency Multi-Scale Deformable Convolutional Networks and Attention Mechanisms," *Remote Sensing*, vol. 16, no. 8, pp. 1431-1431, 2024/4// 2024, doi: 10.3390/rs16081431.

[153] S. Liang, R. Chen, G. Duan, and J. Du, "Deep learning-based lightweight radar target detection method," *Journal of Real-Time Image Processing,* vol. 20, no. 4, 2023, doi: 10.1007/s11554-023-01316-5.

[154] Z. Song, B. Hui, and H. Fan. *A dataset for detection and tracking of dim aircraft targets through radar echo sequences*, doi: https://doi.org/10.11922/csdata.2019.0075.zh.

[155] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *IEEE/CVF conference on computer vision and pattern recognition. 2021*, 2021, pp. 13713-13722, doi: 10.1109/CVPR46437.2021.01350.

[156] H. Dong, J. Liu, C. Wang, H. Cao, C. Shen, and J. Tang, "Drone Detection Method Based on the Time-Frequency Complementary Enhancement Model," *IEEE Transactions on Instrumentation and Measurement,* vol. 72, 2023, doi: 10.1109/TIM.2023.3328072.

[157] Y. Zhao *et al.*, "TGC-YOLOv5: An Enhanced YOLOv5 Drone Detection Model Based on Transformer, GAM & CA Attention Mechanism," *Drones,* vol. 7, no. 7, 2023, doi: 10.3390/drones7070446.

[158] F. N. Muhammad Zamri, Gunawan, T.S. *BirDrone*, doi: https://dx.doi.org/10.21227/ettb-0w28.

[159] Z. Xu, J. Su, and K. Huang, "A-RetinaNet: A novel RetinaNet with an asymmetric attention fusion mechanism for dim and small drone detection in infrared images," *Mathematical Biosciences and Engineering,* vol. 20, no. 4, 2023, doi: 10.3934/mbe.2023285.


## BIOGRAPHY OF AUTHORS

Fatin Najihah Muhamad Zamri received her Bachelor’s degree in Electronics-Computer and Information from the International Islamic University Malaysia (IIUM), Kuala Lumpur, in 2023 and is currently pursuing a Master’s degree in Computer and Information Engineering at IIUM. Her research focuses on Artificial Intelligence, deep learning, and object detection, with a particular emphasis on practical implementations and optimization of algorithms. She is highly skilled in object detection frameworks, including Ultralytics and Python, and is proficient with the YOLO (You Only Look Once) model for real-time object detection tasks. Fatin’s expertise encompasses model training, evaluation, and deployment, particularly in high-accuracy applications such as drone detection systems, enabling her to design efficient, scalable solutions for complex engineering challenges. Her strong technical acumen and dedication position her as an emerging AI and machine learning talent.

Teddy Surya Gunawan (Senior Member, IEEE) is a Professor in the Electrical and Computer Engineering Department at the International Islamic University Malaysia (IIUM). He obtained his B.Eng. degree with cum laude honors in Electrical Engineering from Institut Teknologi Bandung (ITB), Indonesia, in 1998, his M.Eng. from Nanyang Technological University, Singapore, in 2001, and his Ph.D. from the University of New South Wales (UNSW), Australia, in 2007. Prof. Gunawan has held esteemed roles, including Visiting Research Fellow at UNSW (2010–2021) and Adjunct Professor at Telkom University (2022–2023), and he currently serves as a Senior Member of IEEE since 2012, previously chairing the IEEE Instrumentation and Measurement Society – Malaysia Section. Recognized for his contributions to speech and audio processing, biomedical signal processing, image and video processing, and parallel computing, he received IIUM's Best Researcher Award in 2018 and is listed among the World's Top 2% Scientists in Artificial Intelligence and Image Processing by Elsevier for 2023 and 2024. In addition to his academic and research accomplishments, Prof. Gunawan holds multiple professional engineering certifications, including Chartered Engineer (IET, UK), Insinyur Profesional Utama (PII, Indonesia), ASEAN Engineer, ASEAN Chartered Professional Engineer, and APEC Engineer, reflecting his commitment to professional excellence. Within IIUM, he has also served as Head of Department (2015–2016) and Head of Programme Accreditation and Quality Assurance (2017–2018) at the Faculty of Engineering, reinforcing his leadership and expertise in the field.

Mira Kartiwi (Member, IEEE) is a Professor in the Department of Information Systems, Kulliyyah of Information and Communication Technology, and is currently Director of Professional Development at International Islamic University Malaysia (IIUM). She was one of the recipients of the Australia Postgraduate Award (APA) in 2004. She was awarded the Higher Degree Research Award for Excellence in 2007 for her research achievement. She has also been appointed as an Editorial Board member in local and international journals to acknowledge her expertise. Dr. Kartiwi is also an experienced consultant specializing in the health, financial, and manufacturing sectors. Her areas of expertise include health informatics, e-commerce, data mining, information systems strategy, business process improvement, product development, marketing, delivery strategy, workshop facilitation, training, and communications.

Agus Pratondo (Senior Member, IEEE), an Associate Professor in the Department of Multimedia Engineering at Telkom University, was recognized in 2024 among Elsevier's World Top 2% Scientists in Artificial Intelligence and Image Processing. He obtained his Bachelor's in Informatics Engineering and Master's in Electrical Engineering from the Institut Teknologi Bandung. He advanced his studies with a Ph.D. in Electrical and Computer Engineering from the National University of Singapore. A senior IEEE member, Agus's research interests span artificial intelligence, machine learning, computer vision, data analytics, and multimedia applications. His notable contributions and expertise make him a distinguished AI research and innovation figure.

Siti Hajar Yusoff (Member, IEEE) received the M.Eng. degree (Hons.) in electrical engineering and the Ph.D. in electrical engineering from the University of Nottingham, U.K., in 2009 and 2014, respectively. In 2015, she became an Assistant Professor at the Department of Electrical and Computer Engineering at the International Islamic University Malaysia, Gombak. She is currently a Lecturer in control of power electronics and electrical power systems. Her research interests include controlling power converters and drives, matrix and multilevel converters, the IoT, smart meters, wireless power transfer for dynamic charging in electric vehicles (EVs), and renewable energy.

Yasir Mohd. Mustafah Yasir Mohd Mustafah obtained his Bachelor's in Electronics Engineering from the University of Southampton, UK, in 2004. After graduation, he worked as an engineer in Flextronics Technology (Malaysia) Sdn Bhd and Toyota Motor (Malaysia) for a while before joining International Islamic University Malaysia (IIUM). Later, he furthered his studies and earned his Ph.D from the University of Queensland, UK, in 2011. His Ph.D. research was in intelligent surveillance systems and was done in collaboration with National ICT Australia. He is an active advisor to IIUM Robotic Club and advises MARA in their project to set up a robotic-based school. He has authored and co-authored publications in international and local refereed journals and conferences. He has acquired several research projects/grants as a principal and co-investigator.