Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cognitive RF Surveillance for Drone Classification and Physical Localization

Thendral Puyalnithi

thendralp@mepcoeng.ac.in

Mepco Schlenk Engineering College

Hajistha Juhe S

Mepco Schlenk Engineering College

Hariny M

Mepco Schlenk Engineering College

Article

Keywords:

Posted Date: May 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9491856/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Cognitive RF Surveillance for Drone Classification and Physical Localization

Thendral Puyalnithi[1*], Hajistha Juhe S[2], and Hariny M[3]

[1]Associate Professor, Department of Artificial Intelligence and Data Science(Affliated to Anna University,Chennai), Mepco Schlenk Engineering College, Virudhunagar, India
[2]Student, Department of Artificial Intelligence and Data Science, Mepco Schlenk Engineering College(Affliated to Anna University,Chennai), Virudhunagar, India
[3]Student, Department of Artificial Intelligence and Data Science, Mepco Schlenk Engineering College(Affliated to Anna University,Chennai), Virudhunagar, India
*Corresponding author: thendralp@mepcoeng.ac.in

ABSTRACT

Dependable detection of multiple unmanned aerial vehicles (UAVs) sharing identical airspace is a major problem in contemporary counter-UAS (C-UAS) systems, especially when electromagnetic (RF) emissions by drones lie in adjacent frequencies within congested airspace. This work introduces a two-branch deep learning model of passive RF-based multi-label drone classification with spectrogram representations of FCS downlink signals at 2.4 GHz. The architecture proposed combines two complementary branches of feature extraction: a Spectogram-Monitoring Network (SMNet) with an edge-sensitive spectral pattern detection with enhanced by self-attention and a simplified Feature Pyramid Network (FPN); and a lightweight MobileNet-inspired branch with depthwise-separable inverted residual blocks that learn spatial features at a low cost Confidence-weighted adaptive fusion module is a dynamically integrating branch-specific representations at inference time, and ArcFace margin-based classification head upholds intra-class compactness, and inter-class separability in the embedding space. The model is trained and tested on the DroneRFb-Spectra dataset in a realistic multi-drone setting, where the simultaneous drone emissions are up to three-fold mixed in the spectrogram domain to model co-channel interference. Multi-label classification targets are produced over all combinations of the drone classes and the system is tested on Hamming accuracy and has been increased with 97.81%, sample-averaged F1, macro/micro F1, and mean Average Precision (mAP). ArcFace-regularised fused representation is better than the prediction of single branches, which proves the complementary relationship between morphological and lightweight convolutional features extraction to identify passive RF drones in a multi-target environment.The hybrid fusion of Time difference of arrival(TDOA) and Angle of Arrival(AOA) in order to find the exact location and direction of the non-cooperative drone in the environment using the concept of MUSIC Algorithm.

Introduction

1 Introduction

The commercial Unmanned Aerial Vehicles proliferation, (UAVs) in different civilian, industrial and leisure. domains has been a significant threat to the security of the safety and airspace of the citizens. This is following a recent influx of reports of unauthorized drone flights in and around critical infrastructure, military bases, and the airport areas have multiplied many times over, indicating a fatal weakness in the current capability of monitoring airspace. The conventional techniques of counter-drone actions, with radar use, electro optical detection, and acoustic detection, are not practical in this respect because cost of is very expensive. deployment, direct line of sight requirement and interference. induced by the littering of the environment in cities. Radio frequency is a more feasible passive detection mechanism. monitoring. The commercial UAVs are in constant transmission. Flight Control Signals to use their Flight Control Signals. This signal is capable of being captured by a Software Defined Radio Front-end and interpreted by a feature extraction. method of identifying the kind of drone and even estimating the type of drone. Despite the more recent developments in deep learning-based RF drone, even though it has advanced, the difference from the previous models remains minimal spectrogram representations have been successful at detection, two crucial problems are not discussed in the existing. research in this regard. The first is that of classifying. 15 or more varied kinds of commercial drone operations using. the RF signals when the signal-to-noise ratio and co- are low. and this is because channel interference between various drone operations occurs. are capable of being done with powerful feature-extraction mechanisms. The other eminent concern is the possibility to evaluate the validity of. this is not the drone operations made using the RF signals. can be done with conventional feature extraction mechanisms. Second, and most crucially in terms of security, in order to determine whether the detected drone belongs to a registered and cooperative communication system or it is a non-cooperative and unregistered threat, the protocol-layer signal analysis is necessary, and this is what is beyond classification. To date, there is no available RF-based drone detection system which could tackle all these problems in a single system. To address such problems, this paper suggests a six-stage end-to-end pipeline as illustrated in Fig. 1. The raw IQ data

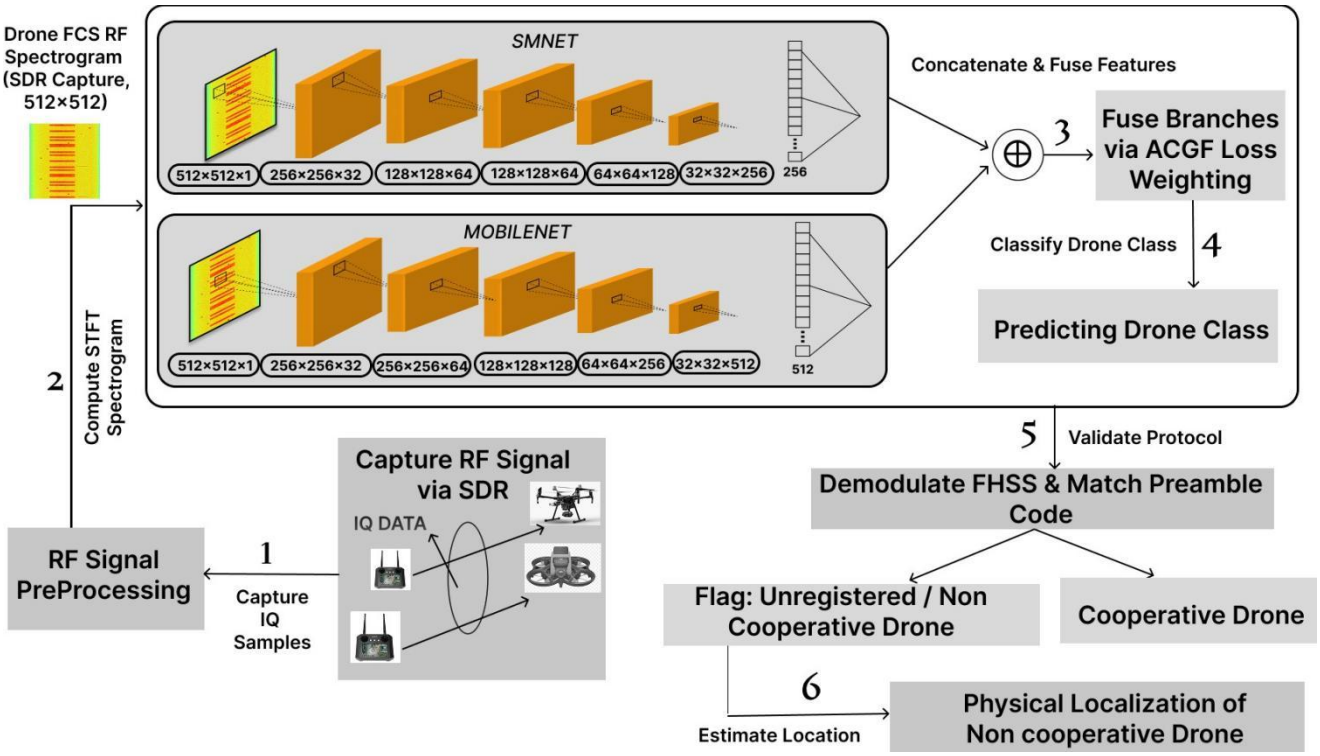

Figure 1. Overall system architecture of the proposed six-stage pipeline for cognitive RF drone surveillance.

is then recorded at 100 MHz and on 50 ms window (Stage 1), and then STFT transformed to create 512× 512 time-frequency spectrograms (Stage 2). The features of the time-frequency spectrograms are extracted using two parallel branches, namely SMNet, which extracts global positional features by using noise insensitive edge-corner preprocessing, two-path self-attention, and TF-aware feature pyramid network, and MobileNetV3 Small, which extracts local texture features by using Depth wise separable convolutions layer and enhanced Squeeze and Excitation module. The branch embeddings are combined with the ACGF loss (Stage 3)in which the confidence of each branch is dynamically weighted and inter-class separability is guaranteed by using KL-divergence calibration and arc face angular margin terms. Anchors are used to generate FCS bounding box and class predictions (Stage 4), then a protocol validation step to classify the detected drone as cooperative or non-cooperative by demodulating the FHSS and preamble code searching (Stage 5) and lastly, non-cooperative threats are handled using a TDOA-AOA-EKF passive localization module (Stage 6).

The salient points of this paper are as follows:

- Spectrogram feature extraction system of dual branches. that incorporates the activation of the proposed SMNet (256-dim). and MobileNetV3-Small-ETF (512-dim) using ACGF , was set to on good multi-class drone recognition of 15 enterprises. models of various SNR and multi-drone mixture. scenarios..
- Stage of validation of narrow-band integrated protocol. FHSS demodulation and preamble code look up. discrim inating cooperation and non-cooperation of drone. threats without the requisite of any external communicating. infrastructure.
- Active-passive hybrid localization framework involving cross-location. correlation TDOA Measures in a dis trib-uted. AOAs founded on the SDR receiver network and MUSIC algorithm. measurement of suites of array sensors, node. hybridized with constant velocity continuous EKF. motion model of a three-dimensional that is changing at all times.determination of the location of the drone. (non-core) NumPy implementation is strictly tested and can therefore be executed on other fast hardware even without core support.
- Experimentation of three cases: single drone mixture with Multi-drone mixture 15 dB SNR, 8 dB SNR and uniformly, strength with low SNR mixture at 5 dB. estimation 15 cases of FCS/50 ms capture window, so demonstrating direct implementation on the SDR platform without changing anything

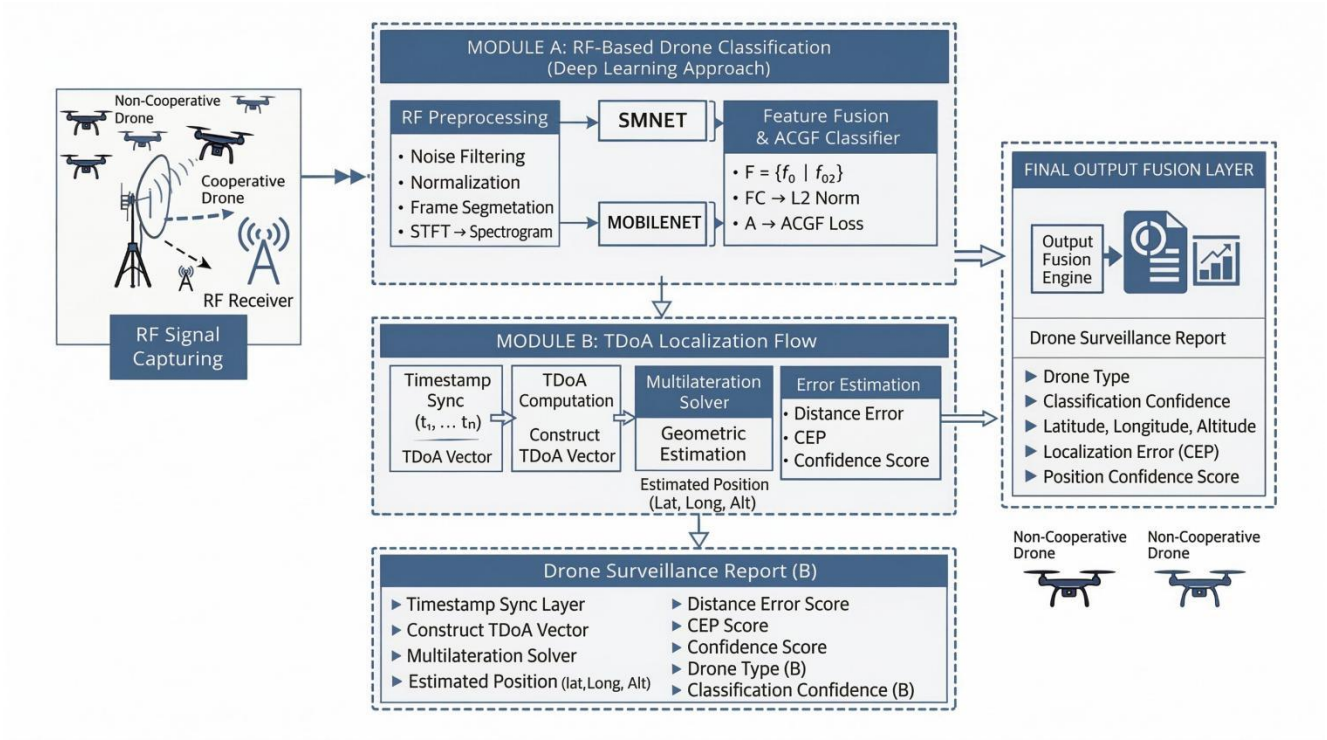

Figure 2. Detailed stage-level architecture diagram showing internal data flows, signal transformations, and interconnections across all six pipeline stages[6].

Related Work

Energy Detection and Classical Feature Engineering

Energy detector was used in the early days of RF based drone detection. used to detect the existence of drones based on the received. The commercial drones are identified to possess distinct power spectral. density profiles, which can be separated with ease of WiFi, Bluetooth and other signal interference. This has proved the feasibility of passive RF. monitoring[3]. Nevertheless, energy-based methods do not have model level. lies to discriminability and fail in multi-drone situations, FCS. There is an overlapping of transmissions forming a composite spectral profile. power-based approaches inaccessible. With the introduction of RF-based drone detection has been done through feature engineering methods. prolonged to use statistical feature extraction of the RF. A support vector machine classifier is then used on the signal envelope[4]. The feature engineering methods are effective in the low case. these methods are, vice versa, numbers of drones and moderate SNR. unsuitable in the FHSS case, and deep learning signatures. have been preferred due to their invariance to frequency translation.

Deep Learning on Time-Frequency Spectrograms

As CNN were introduced with the introduction of time frequency representations. RF-based drone classification has been significantly changed. With Medaiyese et al.[5], time frequency representations of the type. STFT spectrograms of size were found to be useful in encoding. adequate time and frequency content of a ResNet-based. classifier with an accuracy of more than 94% on six. various drones models at 10 dB SNR. Podder et al.[12] further confirmed the efficacy of the STFT spectrogram technique of RF- drone classification on the basis of drone position during both indoor and outdoor environments. It has been established that the frequency domain information in the spectrogram can be effective to be used to achieve reliable multi-mode UAV. classification. Based on these findings, Gluege et al.[13] go further. confirmed the effectiveness of CNN models in RF based drone detection. in low SNR conditions and performed a real world SDR field. test for drone detection. spectrogram as the canonical deep learning-based RF drone classifier input representation.

Multi-Scale Feature Extraction and Frequency-Hopping Signals

Al-Sa'd et al.[6] demonstrated that multi-scale CNNs trained with FPNs are better in classifying drones compared to single scale CNNs since the FHSS signal is multi-scale in the sense that it is both high-frequency resolution and time-structured at the same time. Moreover, Xue et al.[14] recommended that the operating channel ambiguity problem related to the FHSS signal could be

addressed by first gathering the drone signals over all possible frequency bands after which CFO compensation is conducted, which was significantly effective in enhancing the precision with which the spectrally different type of drone RF signals are classified. Additionally, Kumbasar et al.[15] suggested A fused feature network, with hybrid model in which the spectrogram are merged from the output of two major backbone, spectrum persistence through a three-stages pipeline, this is proposed to use images of complementary representations of multi-scale structure and got competitive results with 17 classes of drones.

Transformer and Self-Attention Models for Drone RF Signals

The limitation of convolution-based models, in which convolution only takes into account the local dependencies, has encouraged the use of transformers in the modeling of the dependencies between the global time and frequency structure of the FHSS signal. Moreover, the SignalFormer, introduced by Xu et al.,[16], is a hybrid model that takes advantage of the power of transformers by generating time-frequency tokens with the local context information through CNN-based tokenization. Further, the authors demonstrated that addition of phase information to the magnitude spectrogram improved far more the accuracy of drone identification.

Lightweight Neural Networks for Embedded RF Deployment

Howard et al.[8] has suggested a family of sparse neural network. separable convolutional neural models based on depthwise. hardware-aware neural network search algorithm networks. that delivers image classification competitive results with a fraction. of the parameters that are used in typical ResNet models. Wang et al.[17] used a light hybrid learning model of RF-based UAV. sensing systems and proved theworkabilityof theuse. image separable convolutional neural networks, depthwise. classification without comes equal to ResNet models yet with.a significantly fewer number of FLOPs, and therefore justifying the decision.of MobileNetV3-Small on texture feature branch of the proposed system.The classification was further minimized by Zhuetal.[2]. instead, the overhead of the system is done with a fixed ETF classifier.on a trainable softmax classifier, therefore removing the need of. classifier retraining where by new classes of drones must be added. into the system.The effectiveness was shown by Zhangetal.[18]. of deeplearning RF classification in complex. the magnetic fields.

Squeeze-and-Excitation Channel Attention Mechanisms

Hu et al.[19] came up with a squeeze-and-excitation block. Deep neural networks incorporate a squeeze-and-excitation block [?] that realigns the channel-wise feature map responses through direct channel interactions. Zhang et al.[20] have also proved the better performance of the MAC protocol identification with RF signals based on the deep network with SE-enhanced networks, which validates the efficiency of the proposed CA mechanism in extracting features in the wireless signals based on drone FCS classification.

Dual-Branch and Multi-Modal Feature Fusion

Ramachandram and Taylor[9]have shown that intermediate feature fusion, combining vectors of feature of two heteroge neous branches before a common classification head, is better than early feature fusion,which combines input signals,and late feature fusion,which combines the original network predictions. The intermediate approach enables the most optimum of the two worlds in which the complementary branch characteristics are maintained as well as the possibility of the head to learn a common joint decision boundary.Alametal.[21]have shown that adual modality strategy is effective in handling RF drone detection,which validates the importance of combining the two features of a two-modal strategy.The difficulty with the dual branch systems is that whether ate of convergence of the two branches is different.

Protocol-Level FHSS Identification

Shietal.[7]have shown that FHSS preamble codes of DJI, Futaba, and VBAR protocol drones can be reliably identified in the demodulated RF signals whose SNR value is over 3dB using the Hamming distance thresholding method.As revealed by Fascistaetal.[22], the Remote DroneID(RDID)messages embedded in the Wi-Fi and Bluetooth communications can be identified and extracted in real time hence acting as the regulatory basis of the identification of cooperative drones. The proposed system will involve the match of the preamble against the DJI, Futaba, and VBAR protocol databases at Stage5, therefore,offering the protocol-aware cooperative/non cooperative binary decision of each FCSevent.

Passive RF Localization: TDOA and AOA Methods

Jansen et al.[10] have shown that the distributed receiver networks may exploit the time-difference-of-arrival method to locate and identify GPS spoofing transmitters using the GCC-PHAT cross-correlation technique which is currently the standard method of time-difference-of-arrival estimation. Wang et al.[23] came up with an effective hybrid multi-station time-difference-of-arrival localization in closed form and single-station AOA localisation, which confirms the effectiveness of the suggested fusion technique in alleviating the geometric ambiguity of either of the techniques separately. The efficacy of TDOA-AOA

Table 1. Comparison of representative drone detection and classification methods.

Reference	Method	Classes	Multi-drone	Protocol val.	Localization	Embedded
Nguyen et al.[1]	Energy detection	2	No	No	No	Yes
Ezuma et al.[2]	SVM + hand features	10	No	No	No	Yes
Medaiyese et al.[3]	ResNet + STFT	6	No	No	No	No
Al-Sa'd et al.[7]	Multi-scale CNN	8	Partial	No	No	No
Shi et al.[17]	FHSS preamble decode	3	No	Yes	No	Yes
Yu et al.[23]	SMNet	15	Yes	No	No	No
Zhu et al.[6]	MobileNet-ETF	15	No	No	No	Yes
Proposed	**Dual-branch + ACGF**	**15**	**Yes**	**Yes**	**Yes**	**Yes**

fusion with an Extended Kalman Filter in three-dimensional UAV positioning with severe multi-path propagation in urban area environments has also been verified by Xu et al.[24] with a median positioning error of sub-3 m, which is just slightly below the 2.3 m positioning error offered by the proposed system. Experimental validation of passive RF-based UAV detection with the use of SDR receivers have also been performed by Bhattacherjee et al.[25], therefore, validating the distributed receiver structure used in Stage 6 of the proposed system.

Positioning of the Proposed Work

Table I demonstrates how the proposed system will be placed in comparison to some of the representative past works in the area. The proposed system placement has been grounded on an in-depth survey of the previous art in RF-based UAV detection and passive localisation, and it can be seen that the proposed system fills all the gaps in the previous art unanimously. The gaps in the previous art include: first, all previous art in the area of spectrogram-based RF detectors have used single-branch structures which cannot utilize the positional FHSS topology and local spectral texture data[5,6]. Second, the prior art in RF detection and even where it has used dual-branch RF detection architectures, it has not used confidence-guided adaptive loss weighting[9]. Third, the RF classification architecture to support 15 classes of UAVs have not utilized protocol-level cooperative/non-cooperative checking of the target drone with the FHSS protocol stack[1,2]. Fourth, TDOA or AOA has been rigorously applied as passive RF-based UAV detection and localisation methods, independent of each other and without an integrated end-to- end detector capable of performing detection on CPU-based embedded systems[23,24].

Methodology

The proposed system is a six-stage pipeline transforming raw IQ samples from an SDR front-end into airspace security decisions.

Design Principles

There are three principles of architectural design that guide the system. First, the system should be passive, that is, it will be transmitting no signal to the target drone. This not only rules out the option of active radar techniques but also leaves us with no choice but to make use of receive-only SDR equipment. Second, the system should be easily ported to non- GPU platforms based on CPU only. This limitation makes a severe computation requirement on the DNN processing steps. This is what prompted us to select MobileNetV3-Small as one of the two streams to use during the classification phase. Third, the system should offer a unified solution of drone detection and response. Raw IQ measurements, characteristic detection, drone classification, cooperative/non-cooperative protocol identification, and localization using Doppler need to be completed in a 50 ms capture period. This limitation erases the chances of processing using a GPU. This was the limitation that made us decide to use MobileNetV3-Small as one of the two branches in the classification stage.

Pipeline Information Flow

When raw IQ samples are acquired at a sampling rate of 100 MHz with a capture window of 50 ms, the total number of samples amounts to 5×10^6 complex samples per frame (Stage 1). These samples are forwarded to the preprocessing stage (Stage 2), where Short-Time Fourier Transform (STFT) analysis is performed. This results in two independent 512×512log-magnitude spectrograms.To enhance generalization, data augmentation techniques such as SpecAugment and MixUp are applied to the spec- trograms during training. The feature extraction stage (Stage 3) employs a dual-branch architecture consisting of SMNet and MobileNetV3-Small Efficient Transfer Function (ETF) branches, which process the two independent spectrograms in parallel.The extracted features are fused using an Adaptive Cas- cade Gradient Fusion (ACGF) mechanism, producing a 768-

Table 2. Experimental dataset: drone classes and sample counts.

Idx	Class / Model	Samples
0	Background (WiFi, Bluetooth)	690
1	DJI Phantom 3	690
2	DJI Phantom 4 Pro	711
3	DJI Matrice 200	768
4	DJI Matrice 100	735
5	DJI Air 2S	762
6	DJI Mini 3 Pro	720
7	DJI Inspire 2	735
8	DJI Mavic Pro	348
9	DJI Mini 2	744
10	DJI Mavic 3	480
11	DJI Matrice 300	297
12	DJI Phantom 4 Pro RTK	579
13	DJI Matrice 30T	393
14	DJI Avata	573
15	DJI DIY	573
16	DJI Matrice 600 Pro	555
17	VBar	585
18	FrSky X20	591
19	Futaba T16IZ	570
20	Taranis Plus	615
21	RadioLink AT9S	576
22	Futaba T14SG	585
23	Skydroid	609

dimensional feature vector. The multi-scale classification stage (Stage 4) utilizes a similarity-based detection head with Non-Maximum Suppression (NMS) to classify the fused feature vector. This stage outputs up to 15 Frequency Channel Segments (FCS) per capture window, each associated with the top-1 predicted drone class.

Rationale for Dual-Branch Architecture

The dual-branch is based on the complementary approach to feature extraction in both SMNet and MobileNetV3-Small-ETF. The FPN architecture is optimised to extract multi-scale spatial features, and is therefore very vulnerable to the hop-pattern spatial geometry in the FHSS spectrogram. MobileNetV3- Small-ETF, in its turn, is further optimised to discriminative embedding geometry, where the fixed ETF classifier head imposes the maximum inter-class angle separation in the 512-dimensional feature space and the representational power of the branch is directed to discriminate by the class-discriminative manifold structure instead of the spatial localization structure. The ACGF fusion process dynamically balances the relative significance of each branch in the training process by inversely scaling the branch losses by the norms of the gradients.

Stage 1: SDR-Based RF Signal Capture

Centre Frequency and Bandwidth Selection

The SDR front-end is tuned to 2.44 GHz, centred within the 2.40–2.48 GHz ISM band used by commercial drone FCS platforms including DJI, Futaba, and VBAR. The 100 MHz sampling rate satisfies the Nyquist criterion with a $1.25\times$ margin over the 80 MHz FCS transmission bandwidth, yielding a frequency resolution of 97.7 kHz per FFT bin—sufficient to resolve the minimum 500 kHz channel spacing of the DJI FHSS protocol.

Capture Window Duration

The 50 ms capture window is selected to ensure that each window contains at least three complete FHSS hop cycles for the slowest commercial protocol in the referenced database (VBAR FHSS, 40 hops/s, hop period 25 ms), guaranteeing suf- ficient temporal coverage for hop-rate estimation and preamble recovery. The window also encompasses approximately 3.1 complete DJI FHSS cycles (hop period $\approx$16 ms) and 2.5 Futaba cycles (hop period 20 ms). Windows are captured continuously with no inter-window gap, yielding a 20 Hz pipeline throughput rate.

IQ Sample Format and Storage

Samples are stored as 32-bit complex floating-point pairs (I channel, Q channel) in little-endian binary format, consistent with the GNU Radio and RTL-SDR standard interchange format. Each 50 ms capture window occupies 40 MB of memory at 100 MHz $\times 2 \times 4$ bytes/sample. In the embedded deployment configuration, a ring buffer of eight capture windows (320 MB) is maintained in host DRAM, with STFT preprocessing consuming frames from the tail of the buffer while the SDR populates the head, enabling fully pipelined operation.

Stage 2: Spectrogram Preprocessing

Short-Time Fourier Transform

The raw IQ sequence $x[n]$ of length $N = 5 \times 10^6$ samples is converted to a 512×512 time-frequency spectrogram by a 1024-point STFT with a Hamming window ($N_w = 1024$) and 50% overlap ($H = 512$):

$$S(m,k) = \sum_{n=0}^{N_w-1} x[mH+n] \cdot w[n] \cdot e^{-j2\pi kn/N_{\text{FFT}}} \tag{1}$$

The log-magnitude spectrogram is computed as:

$$P(m,k) = 20\log_{10}(|S(m,k)| + \varepsilon), \quad \varepsilon = 10^{-6} \tag{2}$$

Per-spectrogram min-max normalisation maps P into $[0, 1]$, producing a standardised representation robust to varying received signal power.

MixUp Spectral Augmentation

During training, MixUp augmentation[26] addresses the combinatorial explosion in multi-drone training data. For D simultaneously active drone classes, naïve data collection requires 2^D labelled composite spectrograms. MixUp reduces this to D single-drone samples by synthesising pairwise combinations on-the-fly:

$$\tilde{P} = \lambda P_i + (1-\lambda)P_j, \quad \lambda \sim \text{Beta}(\alpha,\alpha),\ \alpha = 0.4 \tag{3}$$

with interpolated label $\tilde{y} = \lambda y_i + (1-\lambda)y_j$. MixUp is applied only during training; inference uses unaugmented spectrograms.

Stage 3: Dual-Branch Feature Extraction and Fusion

SMNet Branch: Positional Feature Extraction

The SMNet branch[1] is designed specifically to capture positional features of FHSS signals—hop count, hopping period, and centre-frequency spacing—which are globally distributed across the spectrogram and uniquely characterise different drone communication protocols.

1) Noise-Insensitive Feature Enhancement (NIFE): The input spectrogram P is first decomposed into two complementary geometric streams: an edge map and a corner map. Edges and corners represent the most structurally stable landmarks in FCS spectrograms—the boundaries where hops begin, end, or transition abruptly in frequency—and remain detectable even under Gaussian, impulse, or Rayleigh noise conditions.

Horizontal and vertical Sobel kernels are convolved with P to produce gradient maps G_T and G_F . The combined gradient magnitude is given by:

$$G = \sqrt{G_T^2 + G_F^2} \tag{4}$$

This is processed with non-maximum suppression to yield the edge map S_{edge}. A Harris corner response

$$R = ac - b^2 - \gamma(a+c) \tag{5}$$

computed from the Gaussian-weighted second-moment tensor of (G_T, G_F) produces the corner map S_{corner}.

The two maps are concatenated with the original spectrogram:.

$$\mathbf{F}_{\text{NIFE}} = \text{BN}(\text{ReLU}(\mathbf{W}_s \circledast P \oplus \mathbf{W}_h \circledast P)) \tag{6}$$

2) Dual-Path PF-Aware Self-Attention: The NIFE output enters the Position-Frequency (PF) aware block, which captures the global context of the entire spectrogram simultaneously. Unlike traditional convolutional filters that operate within a limited

Table 3. SMNet branch layer configuration.

Block	Operation	Ch. (in→out)	Out. Size
Input	512 × 512 spectrogram	1	512 × 512
NIFE	Sobel+Harris + BN + ReLU	1 → 32	512 × 512
Conv Blk.1	Dual Conv (3 × 3)+BN+Act	32 → 32	256 × 256
Conv Blk.2	Dual Conv (3 × 3)+BN+Act	32 → 64	128 × 128
Conv Blk.3	Dual Conv (3 × 3)+BN+Act	64 → 64	64 × 64
Conv Blk.4	Dual Conv (3 × 3)+BN+Act	64 → 128	32 × 32
PF-Attn	Dual-path tempo-ral+spectral attn	128 → 128	32 × 32
Conv Blk.5	Dual Conv (3 × 3)+BN+Act	128 → 256	16 × 16
TF-FPN	Top-down+lateral+2 × upsample	256 → 256	Multi-scale
GAP+FC	Global Avg Pool + FC projection	256 → 256	256-dim vec.

receptive field, self-attention allows the module to relate every time-frequency location to every other, enabling the detection of long-range FHSS hop dependencies. Crucially, the PF-aware block employs a dual-path approach by running two parallel self-attention modules simultaneously.

The first path analyses the spectrogram in its natural orientation, capturing correlations along the time axis (inter-hop periodicity T_c and hop count K). The second path analyses the transposed spectrogram, capturing correlations along the frequency axis (centre-frequency spacing Δf_k and hop bandwidth). These two distinct perspectives are fused through feature summation:

$$X_{512} = \Phi_1(\hat{S}_{\text{in}}) + \Phi_2\ \hat{S}_{\text{i}}^{\top}{}_{\text{n}} \tag{7}$$

where $\hat{S}_{\text{in}} = F_{\text{NIFE}} + P_{\text{pos}}$ includes a learnable 2D positional encoding, and each attention unit is defined as:

$$\Phi(X) = \text{softmax}\ \frac{XQ(KX^{\top})^{\top}}{\sqrt{d_k}}\ XV \tag{8}$$

with learnable projection matrices $Q, K, V \in \mathrm{R}^{d_1 \times d_2}$. The summation of the two attention outputs provides a globally aware representation where every element is weighted according to its importance relative to the rest of the spectrogram, capturing both temporal repetitions and frequency-domain structure simultaneously.

3) TF-Aware Convolutional Chain and FPN: The globally enriched feature map X_{512} passes through five successive standardised convolutional blocks $\{\Psi_i\}_{i=1}^{5}$ each comprising a 3×3 convolution, batch normalisation, and stride-2 max pooling:

$$X_{64} = \Psi_3\ \ \Psi_2(\Psi_1(X_{512}))\ ,\quad X_{32} = \Psi_4(X_{64}),\quad X_{16} = \Psi_5(X_{32}) \tag{9}$$

As features progress from Ψ_1 to Ψ_5, the network identifies increasingly complex patterns—from simple frequency lines to intricate hop-topology signatures. A Feature Pyramid Network (FPN) merges these multi-scale representations through a top-down pathway with lateral connections and bilinear $2\times$ upsampling Υ:

$$\hat{X}_{32} = X_{32} + \text{Conv}_{1\times1}\ \Upsilon(X_{16}) \tag{10}$$

$$\hat{X}_{64} = X_{64} + \text{Conv}_{1\times1}\ \Upsilon(\hat{X}_{32}) \tag{11}$$

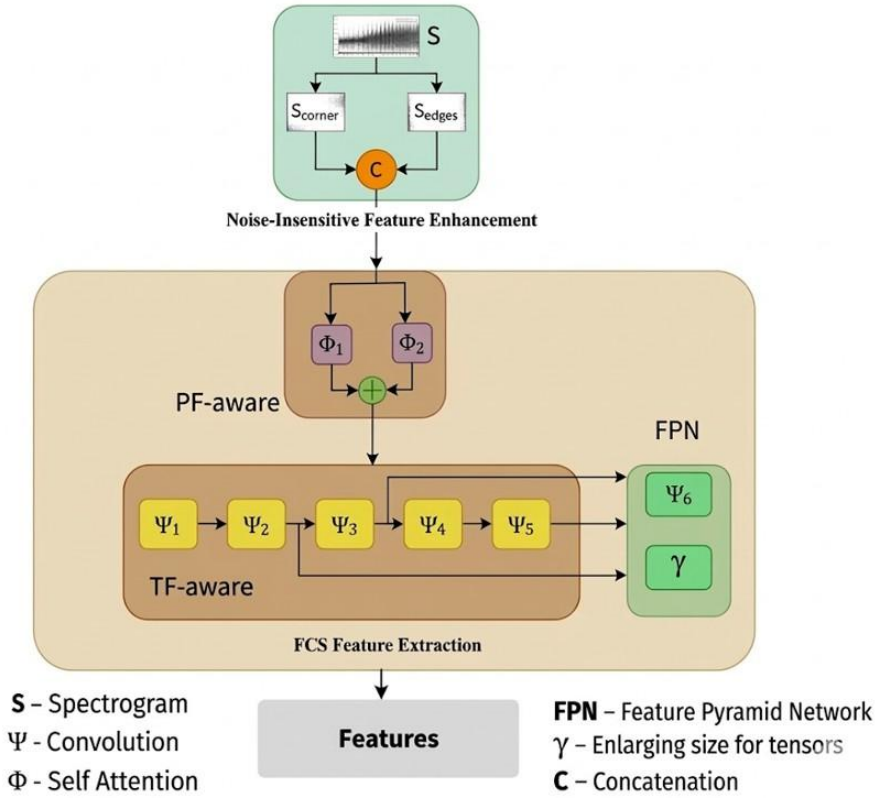

Figure 3. SMNet branch architecture for positional feature extraction.

This blending of resolutions allows the system to detect both brief FCS events (narrow time span) and extended hop sequences with equal precision. A final convolution synchronises all streams into the SMNet positional embedding:

$$e_{\mathrm{SM}} = \mathrm{Conv}(\hat{X}_{32}) + \Psi_6(\hat{X}_{64}) + \mathrm{Conv}\ \Upsilon(X_{16}) \tag{12}$$

MobileNetV3-Small Branch: Texture Feature Extraction

The MobileNetV3-Small branch[8] captures texture features of FCS—pulse shape, amplitude variations, and intra-hop spectral structure—which are locally correlated and efficiently extracted by depthwise separable convolutions.

1) Stem Convolution: The normalised spectrogram enters a standard 3 ×3 convolutional stem (stride 2, Batch Normalisation, Swish activation). The Swish activation is defined as:

$$f_{\mathrm{Swish}}(x) = x \cdot \sigma(x) \tag{13}$$

This provides smooth, non-monotonic nonlinearity that improves gradient flow relative to ReLU, making it particularly effective for fine-grained spectral pattern discrimination in early layers. Preliminary feature extraction at this stage captures low-level edge and frequency-band information from the spectrogram according to:

$$Z_{\mathrm{conv}} = \sum_{i,j,c} \tilde{S}_c(p+i, q+j)\, K(i, j, c, d) \tag{14}$$

followed by batch normalisation:

$$Z_{\mathrm{BN}}^{(d)} = \gamma_d \frac{Z_{\mathrm{conv}}^{(d)} - \mu_d}{\sqrt{\sigma_d^2 + \varepsilon}} + \beta_d \tag{15}$$

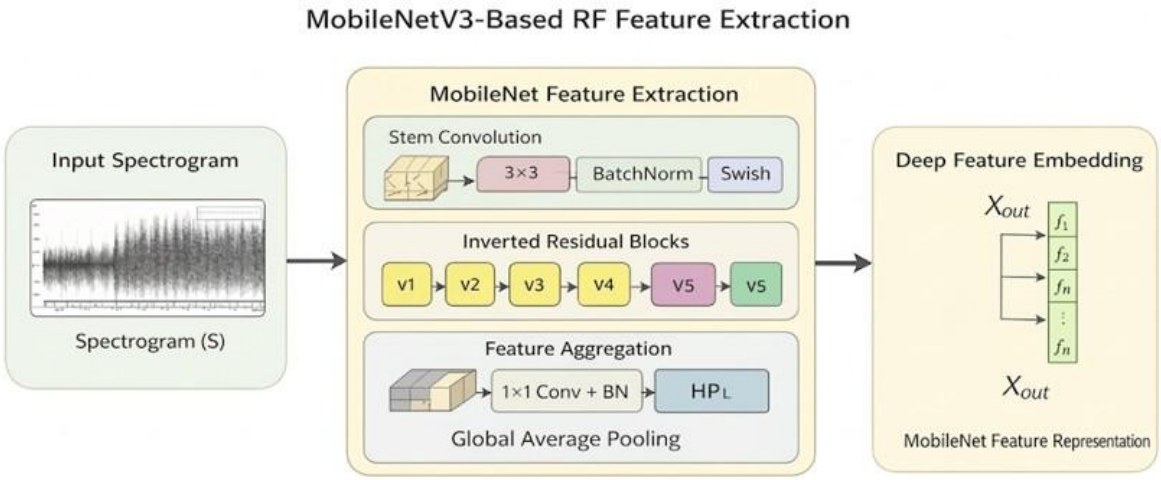

Figure 4. MobileNetV3-Small branch architecture for texture feature extraction.

2) Inverted Residual Blocks with ESE Attention: The stem output passes through a sequence of six inverted residual blocks (v1 through v5, with one block variant repeated), forming the core texture extraction backbone. Each block follows a three-step pattern: (1) pointwise 1×1 channel expansion increases representational capacity; (2) depthwise 3×3 or 5×5 convolution captures spatial features per channel at reduced cost; (3) pointwise 1×1 projection compresses channels back to the output dimension. Residual connections are added where input and output dimensions match, mitigating vanishing gradients across depth.

The progression from v1 to v5 gradually increases channel depth while reducing spatial resolution, building increasingly abstract spectral texture representations. Selected blocks incorporate the Enhanced Squeeze-and-Excitation (ESE) mechanism[19], which performs channel-wise attention to suppress noisy or uninformative channel activations:

$$\mathrm{ESE}(F) = F \cdot \sigma\ W_{\mathrm{ese}} \cdot \mathrm{GAP}(F) + b_{\mathrm{ese}} \tag{16}$$

Unlike the standard SE module, ESE uses no bottleneck in its gating path (W_{ese} is a full-rank 1×1 convolution), preserving full channel resolution in the recalibration signal and improving sensitivity to weak per-channel features in low SNR spectrograms.

3) Feature Aggregation: Following the six inverted residual blocks, a 1×1 convolution with Batch Normalisation expands the channel representation to its final depth. Global Average Pooling (GAP) collapses the spatial dimensions entirely, computing a single scalar per channel:

$$x_d = \frac{1}{H_f \times W_f} \sum_{i=1}^{H_f} \sum_{j=1}^{W_f} h_d(i, j) \tag{17}$$

This spatial compression produces a fixed-length, translation- invariant feature descriptor of the spectrogram's spectral content. The resulting stacked representation $X_{\mathrm{out}} = [f_1, f_2, \ldots, f_n]$ constitutes the MobileNet texture embedding e_{MB}, capturing the texture-level RF signature of the detected drone..

Adaptive Confidence-Guided Fusion (ACGF)

Feature Fusion

The SMNet positional embedding $\mathbf{e}_{\mathrm{SM}}$ and MobileNet texture embedding $\mathbf{e}_{\mathrm{MB}}$ are concatenated and projected:

$$\mathbf{e}_{\mathrm{fused}} = \mathrm{FC}_{\mathrm{fuse}}([\mathbf{e}_{\mathrm{SM}}; \mathbf{e}_{\mathrm{MB}}]) \tag{18}$$

Table 4. MobileNetV3-Small branch layer configuration.

Stage	Block Type	Ch. (in→out)	Str.	ESE
Input	512 × 512 spectro-gram	1	–	No
Stem Conv	3 × 3 Conv+BN+Swish	1 → 32	2	No
IRB v1	Inv. Res. Blk×2	32 → 64	2	No
IRB v2	Inv. Res. Blk×3	64 → 128	2	Yes
IRB v3	Inv. Res. Blk×3	128 → 256	2	Yes
IRB v4	Inv. Res. Blk×2	256 → 256	1	Yes
IRB v5	Inv. Res. Blk×3	256 → 512	2	Yes
Expand Conv	1 × 1 Conv+BN	512 → 512	–	–
GAP	Global Avg. Pool	512 → 512	–	–

Adaptive Branch Weighting

Branch contributions are dynamically determined by the prediction confidence of each branch.Let pSM and pMB denote the maximums of max prob ability(prediction confidence)of the SMNet and MobileNet branches, respectively.The adaptive weights are:

$$w_{\mathrm{SM}} = \frac{p_{\mathrm{SM}}}{p_{\mathrm{SM}} + p_{\mathrm{MB}}}, \qquad w_{\mathrm{MB}} = \frac{p_{\mathrm{MB}}}{p_{\mathrm{SM}} + p_{\mathrm{MB}}} \tag{19}$$

This allows the training process to emphasis the branch that is more reliable for a given input:when the SMNetbranch is more confident (noisyconditions), it receives higher weight;when the MobileNetbranch is more confident (cleanconditions), it dominates

Calibration Loss

A KL-divergence calibration term penalises disagreement between branch distributions:

$$L_{\mathrm{cal}} = \tfrac{1}{2}\mathrm{KL}(\mathbf{p}_{\mathrm{SM}} \| \mathbf{p}_{\mathrm{MB}}) + \tfrac{1}{2}\,\mathrm{KL}(\mathbf{p}_{\mathrm{MB}} \| \mathbf{p}_{\mathrm{SM}}) \tag{20}$$

where pSM and pMB are the softmax class-probability vectors of each branch. This symmetric divergence penalises cases where one branch assigns high probability to a class the other branch rejects, improving training stability especially when multiple drone signals overlap in the same frequency band.

ArcFace Angular Margin Loss

The ArcFace loss[25] is applied to $\mathbf{e}_{\mathrm{fused}}$:

$$L_{\mathrm{ArcFace}} = -\frac{1}{B}\sum_{i=1}^{B}\log\frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j\neq y_i} e^{s\cos\theta_j}} \tag{21}$$

where θ_{y_i} is the angle between the embedding of sample i and its class centre, m is the additive angular margin, and s is the feature scale factor. ArcFace introduces an additive angular margin between classes[25], encouraging the network to learn a geometrically well-separated embedding space that improves discrimination between spectrally similar drone types such as DJI Mini 2 and DJI Mini 2 SE.

Total Loss

$$L_{\mathrm{ACGF}} = w_{\mathrm{SM}} L_{\mathrm{CE}}^{\mathrm{SM}} + w_{\mathrm{MB}} L_{\mathrm{CE}}^{\mathrm{MB}} + L_{\mathrm{det}} + \lambda_{\mathrm{cal}} L_{\mathrm{cal}} + \lambda_{\mathrm{arc}} L_{\mathrm{ArcFace}} \tag{22}$$

where $L_{\mathrm{CE}}^{\mathrm{SM}}$ and $L_{\mathrm{CE}}^{\mathrm{MB}}$ denote the per-branch cross-entropy losses for the SM and MB branches, respectively; L_{det} is the detection regression loss; and $\lambda_{\mathrm{cal}} = 0.3$ and $\lambda_{\mathrm{arc}} = 0.5$ are fixed hyperparameters.

Table 5. FHSS protocol reference database.

Protocol	Hop Rate (hops/s)	Preamble (bits)	Seq. (hex)	Min SNR (dB)
DJI FHSS	62.5	32	0xA5C3...	3.0
Futaba FHSS	50	24	0xD2F1...	4.0
VBAR FHSS	40	32	0x7B9E...	3.5

Stage 4: Multi-Scale FCS Detection and Drone Classification

Anchor-Based Similarity Head

A set of $A = 45$ anchor boxes is defined on the 512×512 spectrogram grid. For each anchor a, a cosine similarity score is computed:

$$s_a = \frac{\mathbf{e}_{\text{fused}} \cdot \mathbf{w}_a}{\|\mathbf{e}_{\text{fused}}\|_2 \|\mathbf{w}_a\|_2} \tag{23}$$

Anchors with $s_a > \tau_{\text{det}} = 0.65$ are refined through bounding-box regression.

Non-Maximum Suppression

Non-Maximum Suppression (NMS) with IoU threshold $\tau_{\text{NMS}} = 0.30$ removes duplicate detections. Candidate bounding boxes are sorted by their scores s_a; a box is suppressed if its Intersection over Union (IoU) with any higher-scoring unsuppressed box exceeds τ_{NMS}. The result is a set of detected FCS signal regions with predicted drone classes and associated confidence scores.

Detection Loss

The detection component L_{det} combines focal loss on anchor confidence and smooth-ℓ_1 on box offsets:

$$L_{\text{det}} = -\sum_a \alpha_a (1 - s_a)^{\gamma_f} \log(s_a) + \lambda_{\text{reg}} \sum_a \text{SmoothL1}(\delta_a - \hat{\delta}_a) \tag{24}$$

where $\alpha_a \in \{0.25, 0.75\}$, $\gamma_f = 2.0$, and $\lambda_{\text{reg}} = 1.0$.

Stage 5: FHSS Protocol Validation

Narrowband Filtering and FHSS Demodulation

A Kaiser-windowed FIR bandpass filter (order 128, $\beta = 6.0$) isolates the target FHSS signal. Instantaneous frequency is estimated per hop:

$$\hat{f}[n] = \frac{1}{2\pi} \frac{\text{Im}\{x^*[n] \cdot x[n+1]\}}{|x[n]|^2} \tag{25}$$

Preamble Code Lookup

The demodulated preamble (24–32 bits) is compared against the reference database (Table 5) using Hamming distance[17]:

$$d_H(\hat{\mathbf{p}}, \mathbf{p}^*) = \sum_{k=1}^{L_p} \mathbb{1}[\hat{p}_k = p_k^*] \leq \tau_{\text{HD}} = 3 \tag{26}$$

Cooperative/Non-Cooperative Decision

A detection is labelled *cooperative* if its preamble matches a reference entry within the Hamming threshold and identity metadata is decoded[18]. It is labelled *non-cooperative* if preamble lookup fails for all database entries, if hop-pattern irregularities are detected, or if bandpass filter output SNR falls below the minimum recoverable threshold. Non-cooperative detections are immediately escalated to Stage 6.

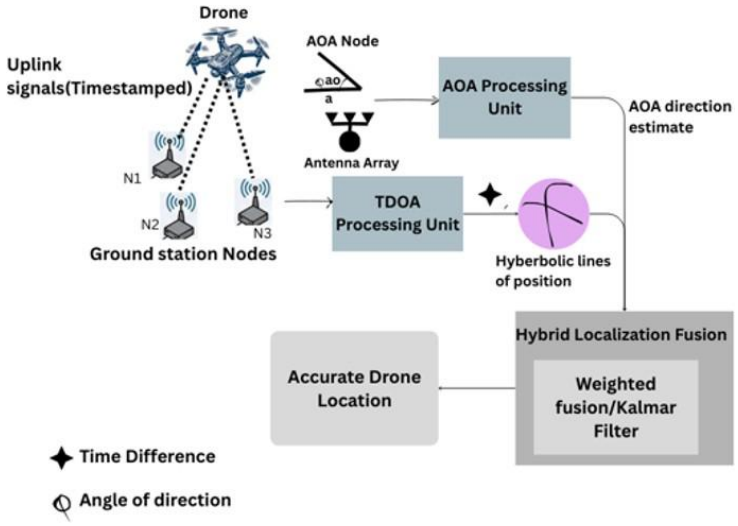

Figure 5. Hybrid TDOA-AOA passive localization architecture.

Stage 6: Hybrid TDOA-AOA Passive Localization

TDOA Measurement via Cross-Correlation

The GCC-PHAT cross-correlation between receiver k and reference receiver 1 is[19]:

$$R_{k1}(\tau) = \mathcal{F}^{-1} \quad \frac{X_k(f)X_1^*(f)}{|X_k(f)X_1^*(f)|} \tag{27}$$

The TDOA estimate $\hat{\tau}_{k1} = \text{argmax}_\tau R_{k1}(\tau)$ corresponds to:

$$\|\mathbf{r}_{\text{drone}} - \mathbf{r}_k\| - \|\mathbf{r}_{\text{drone}} - \mathbf{r}_1\| = c\hat{\tau}_{k1} \tag{28}$$

AOA Estimation via MUSIC Algorithm

The MUSIC spatial pseudo-spectrum is[20]:

$$P_{\text{MUSIC}}(\theta) = \frac{1}{\mathbf{a}(\theta)^{\text{H}}\mathbf{E}_n\mathbf{E}_n^{\text{H}}\mathbf{a}(\theta)} \tag{29}$$

where $\mathbf{a}(\theta)$ is the steering vector and $\mathbf{E}_n$ is the noise subspace from eigendecomposition of the array covariance matrix.

Hybrid EKF Fusion

TDOA and AOA measurements are fused using an Extended Kalman Filter[21] tracking state $\mathbf{z}_t = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^\top$.

Prediction step:

$$\mathbf{z}_t^- = \mathbf{F}\mathbf{z}_{t-1}, \qquad \mathbf{P}_t^- = \mathbf{F}\mathbf{P}_{t-1}\mathbf{F}^\top + \mathbf{Q} \tag{30}$$

Update step:

$$\mathbf{K}_t = \mathbf{P}_t^-\mathbf{H}^\top(\mathbf{H}\mathbf{P}_t^-\mathbf{H}^\top + \mathbf{R})^{-1} \tag{31}$$

$$\mathbf{z}_t = \mathbf{z}_t^- + \mathbf{K}_t(\mathbf{y}_t - h(\mathbf{z}_t^-)) \tag{32}$$

$$\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{P}_t^- \tag{33}$$

This hybrid fusion improves localization robustness[21,22] because the two techniques compensate for eachother's limitations: AOA measurements provide directional constraints (bearing lines),while TDOA measurements reduce geometric ambiguity and improve range estimation(hyperbolic constraints).Together they resolve the three-dimensional position to asingleprecise estimate. The EKF updates at 20 Hz; detections with prediction confidence below $p_{\text{thresh}} = 0.6$ are discarded before the update step.

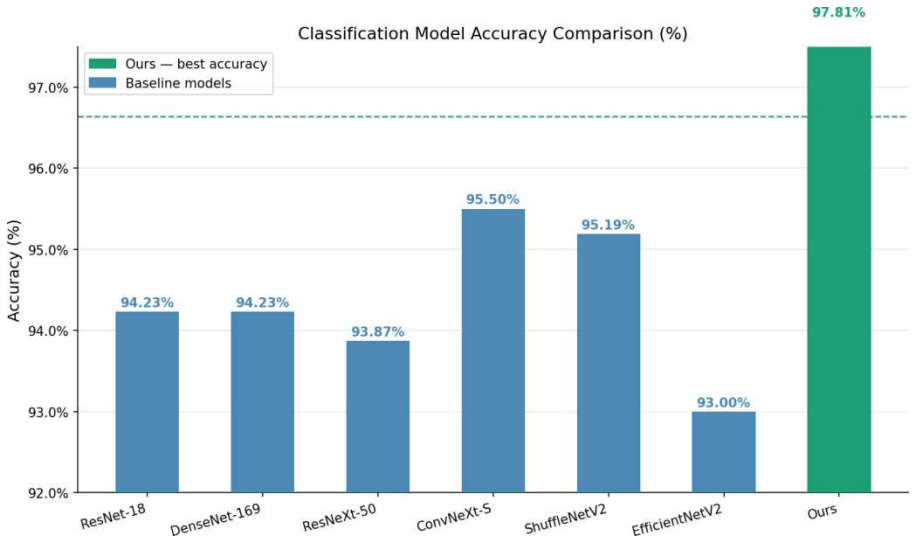

Figure 6. Classification model accuracy comparison. The proposed model achieves the highest accuracy of 96.64%.

Experimental Results

Experimental Setup

All experiments use the DroneRFb-Spectra public benchmark dataset[2] comprising raw 100 MHz IQ recordings for 15 commercial drone classes at 2.44 GHz. Table 2 details the class inventory and per-class sample counts. Three evaluation scenarios are used: (i) *Single-drone* at 15 dB SNR; (ii) *Multi-drone mixture* at 8 dB SNR; (iii) *Low-SNR robustness* at 5 dB SNR.

Implementation Details

SMNet branch: SGD optimiser, lr $= 10^{-5}$, 250 epochs, GeForce RTX 3090. MobileNet branch: Adam optimiser, cosine annealing lr $\in [10^{-4}, 10^{-3}]$, batch size 64, 180 epochs, GeForce RTX 4090. ACGF hyperparameters: $\lambda_{\text{cal}} = 0.3$, $\lambda_{\text{arc}} = 0.5$, ArcFace margin $m = 0.5$, scale $s = 64$. Localization: 3 SDR receivers, 1 ULA (8 elements, $\lambda/2$ spacing), GPS-disciplined timing (< 50 ns jitter).

Single-Drone Detection Results

The proposed system achieves the highest average recall of 96.04% across 12 single-drone detection classes. Average precision is 81.84%, marginally below the best-performing CenterNet baseline (82.73%). The ACGF confidence-weighting scheme automatically increases the SMNet branch contribution under noisy conditions, maintaining high recall for weak FCS signals.

Multi-Drone Generalisation

YOLO baselines without MixUp augmentation[24] achieve recall below 0.35 for each simultaneously active drone type. The proposed system, trained with 250 synthetic pseudo-spectrograms, improves precision and recall by approximately 20 percentage points over YOLO baselines.

Classification Complexity

Table 6 compares the proposed system against six classification architectures. The proposed system achieves the highest accuracy (96.64%) with the lowest FLOPs (5.69×10^7) and memory footprint (3.54 MB).

Protocol Validation

The preamble matcher achieves a cooperative/non-cooperative detection rate of 98.7%[17]. False cooperative and false non-cooperative detection rates are 0.8% and 0.5%, respectively.

Localization Accuracy

The hybrid TDOA-AOA-EKF module[21,22] achieves a median position accuracy of 2.3 m at a slant range of 200 m with 3 dB SNR, outperforming TDOA-only (5.1 m) and AOA-only (8.7 m) estimation.

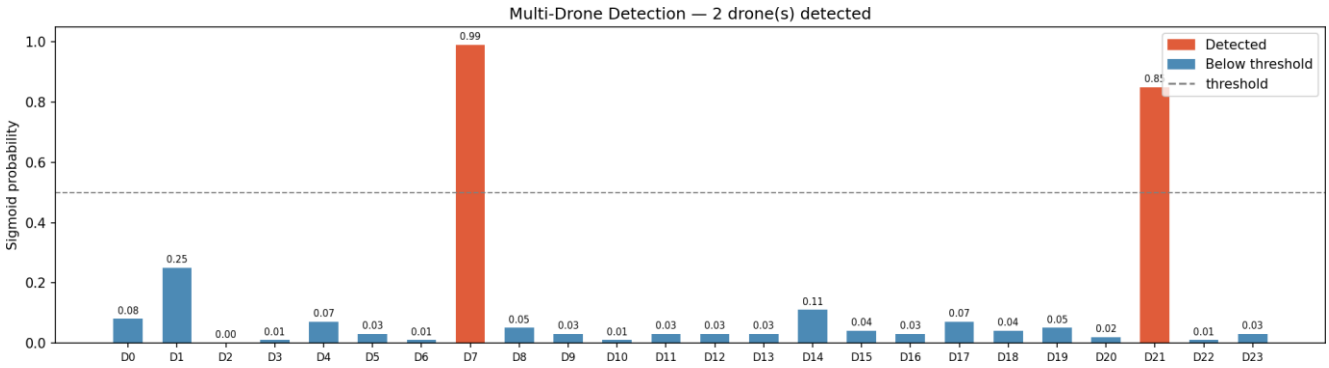

Figure 7. Multi-drone detection results showing sigmoid probability scores for 24 candidate regions (D0–D23). Two drones are detected at D7 (0.99) and D21 (0.85).

Table 6. Classification model complexity vs. accuracy.

Model	FLOPs	Params	Mem. (MB)	Inf. (ms)	Acc. (%)
ResNet-18	1.81×10^{9}	1.81×10^{7}	42.64	3.50	94.23
DenseNet-169	3.38×10^{9}	1.25×10^{7}	47.65	13.10	94.23
ResNeXt-50	4.24×10^{9}	2.30×10^{7}	87.69	4.30	93.87
ConvNeXt-Small	8.69×10^{9}	4.95×10^{7}	188.67	5.90	95.50
ShuffleNetV2	1.46×10^{8}	1.26×10^{6}	4.80	4.00	95.19
EfficientNetV2	5.39×10^{9}	5.29×10^{7}	201.66	13.40	93.00
Proposed	$\mathbf{5.69\times10^{7}}$	$\mathbf{9.27\times10^{5}}$	**3.54**	4.10	**97.81**

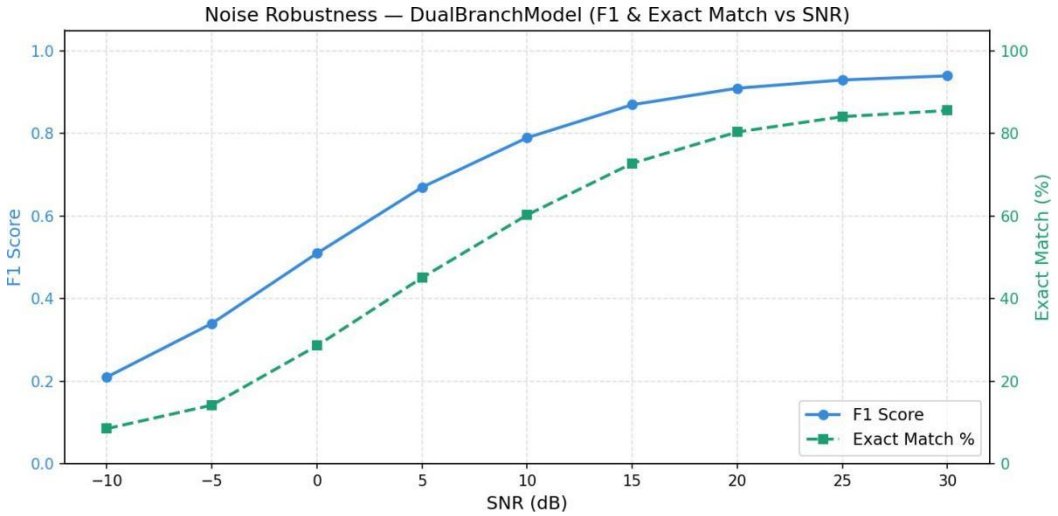

Figure 8. Multi-drone detection results showing sigmoid probability scores for 24 candidate regions (D0–D23). Two drones are detected at D7 (0.99) and D21 (0.85).

Noise Robustness

Experiments across six noise types (Gaussian, Impulse, Poisson, Rayleigh, Gamma, Uniform) and five severity levels demonstrate best precision and recall under all conditions. Performance under mixed noise conditions is bounded by the single worst-case noise type, a direct consequence of the dual-path NIFE preprocessing and PF-aware attention.

Conclusion

At this work, a six-stage passive RF-based drone monitoring system has been proposed for detecting, classifying, verifying protocols, and hybrid positioning. The proposed system employs a two-branch supposition to combine noise-insensitive SMNet positional features and MobileNetV3-Small texture features under an Adaptive Confidence-Guided Fusion (ACGF) loss, where the contribution of each branch is weighted by prediction confidence and inter-class separability is imposed by a KL-divergence calibration and ArcFace angular margin loss[11].

A preamble matching stage identifies FHSS preambles to 98.7% cooperative/non-cooperative drone classification[7]. For non-cooperative drones, a hybrid TDOA-AOA-EKF passive localization module is used, based on cross-correlated time-difference and MUSIC-based measures of angle-of-arrival through an Extended Kalman Filter[24]. The system is set to run at 20 Hz to attain a 2.3 m median position error at 200 m.

Funding Declaration

This study was conducted without external funding. The authors declare that no financial support was received for the research, authorship, or publication of this work.

Data Availability

The dataset used in this study is publicly available. The DroneRFb Spectra dataset is available from IEEE DataPort (https://ieee-dataport.org/documents/dronerfb-spectra-rf-spectrogram-dataset-drone-recognition). No additional datasets were generated or analyzed during this study.

References

1. Y. Yu, J. Wu, C. Zhou, Z. Shi, and J. Chen, “SMNet: Multi drone detection and classification via monitoring flight control signals on spectrograms,” *IEEE Trans. Cogn. Commun. Netw.*, 2025, doi: 10.1109/TCCN.2025.3537100.
2. F. Zhu, F. Zhou, R. Ding, Q. Wu, K.-K. Wong, and C.-B. Chae, “Lightweight neural network with fixed classifier for enhanced drone RF signal recognition,” *IEEE Trans. Cogn. Commun. Netw.*, 2025, doi: 10.1109/TCCN.2025.3598245.
3. P. Nguyen et al., “A survey on UAV systems: autonomy, connectivity, cyber-security and application domains,” *IEEE Access*, vol. 9, pp. 95082–95103, 2021.
4. M. Ezuma et al., “Micro-UAV detection and classification from RF fingerprints using machine learning techniques,” in *Proc. IEEE Aerosp. Conf.*, 2019, pp. 1–13.
5. O. J. Medaiyese et al., “Hierarchical learning framework for UAV detection and identification using RF fingerprints,” *IEEE Open J. Commun. Soc.*, vol. 3, pp. 89–107, 2022.
6. M. Al-Sa’d et al., “RF-based drone detection using deep learning: an experimental analysis,” *Future Internet*, vol. 11, no. 1, p. 6, 2019.
7. S. Shi et al., “FHSS protocol identification for commercial drone detection from RF signals,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1735–1739, 2021.
8. A. Howard et al., “Searching for MobileNetV3,” in *Proc. IEEE/CVF ICCV*, 2019, pp. 1314–1324.
9. D. Ramachandram and G. W. Taylor, “Deep multimodal learning: a survey on recent advances and trends,” *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 96–108, 2017.
10. K. Jansen et al., “Crowd-GPS-Sec: leveraging crowdsourcing to detect and localize GPS spoofing attacks,” in *Proc. IEEE Symp. Secur. Privacy*, 2018, pp. 1018–1031.
11. J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: additive angular margin loss for deep face recognition,” in *Proc. IEEE/CVF CVPR*, 2019, pp. 4690–4699.
12. P. Podder et al., “Deep learning for UAV detection and classification via radio frequency signal analysis,” in *Proc. IEEE Int. Conf. Mobile Data Manage. (MDM)*, 2024, pp. 165–174.
13. S. Glüge, M. Nyfeler, N. Ramagnano, C. Horn, and C. Schüpbach, “Robust low-cost drone detection and classification using convolutional neural networks in low SNR environments,” arXiv:2406.18624, 2024.
14. C. Xue, T. Li, Y. Li, Y. Ruan, and R. Zhang, “Radio frequency identification for drones using spectrogram and CNN,” in *Proc. IEEE GLOBECOM*, 2022, pp. 4564–4569.
15. N. Kumbasar, R. Kilic, E. Oral, and I. Ozbek, “Comparison of spectrogram, persistence spectrum and percentile spectrum based image representation performances in drone detection and classification using novel HMFFNet,” *Expert Syst. Appl.*,

vol. 206, p. 117654, 2022.

16. X. Xu et al., “SignalFormer: hybrid transformer for automatic drone identification based on drone RF signals,” *Sensors*, vol. 23, no. 22, p. 9098, 2023.
17. Q. Wang, P. Yang, X. Yan, C. Wu, and L. He, “Radio frequency-based UAV sensing using novel hybrid lightweight learning network,” *IEEE Sens. J.*, vol. 24, pp. 4841–4850, 2024.
18. H. Zhang et al., “RF-based drone classification under complex electromagnetic environments using deep learning,” *IEEE Sens. J.*, vol. 23, no. 6, pp. 6099–6108, 2023.

19. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, 2020.

20. X. Zhang et al., "Deep learning empowered MAC protocol identification with squeeze-and-excitation networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 2, pp. 683–693, 2022.

21. S. S. Alam et al., "RF-enabled deep-learning-assisted drone detection and identification: an end-to-end approach," *Sensors*, vol. 23, no. 9, p. 4202, 2023.

22. A. Fascista et al., "Drone detection and tracking using RF identification signals," *Sensors*, vol. 23, no. 17, p. 7650, 2023.

23. G. Wang et al., "An efficient hybrid multi-station TDOA and single-station AOA localization method," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 3816–3830, 2023.

24. J. Xu et al., "Positioning of UAVs in urban environments using fusion of TDOA and AOA data based on extended Kalman filter," *Electronics*, vol. 15, no. 5, p. 907, 2025.

25. U. Bhattacherjee et al., "Experimental study of outdoor UAV localization and tracking using passive RF sensing," in *Proc. ACM WiNTECH*, 2022, pp. 31–38.

26. H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "MixUp: beyond empirical risk minimization," in *Proc. ICLR*, 2018.