A TWO-STAGE COMPRESSED-DOMAIN FRAMEWORK FOR UAV PRESENCE DETECTION AND MULTI-CLASS CLASSIFICATION FROM CONTROLLER SIGNALS UNDER RF OVERLAP CONDITIONS

KHUNG HAI TẦNG TRONG MIỀN ĐO NÉN CHO PHÁT HIỆN HIỆN DIỆN VÀ PHÂN LOẠI ĐA LỚP UAV TỪ TÍN HIỆU BỘ ĐIỀU KHIỂN DƯỚI ĐIỀU KIỆN CHỒNG LẤN RF

Nguyen Dang Hieu[1],[2], Nguyen Le Cuong[1], Tran Vu Kien[1], Nguyen Cong Thom[2]

[1] Electric Power University, [2] Academy of Security Engineering and Technology

Received: March 01, 2026 Accepted: May 7, 2026

Abstract:

In passive unmanned aerial vehicle (UAV) surveillance, identifying the UAV type becomes meaningful only after the system first determines whether the observed radio-frequency (RF) segment contains UAV-related activity. However, most existing studies still formulate UAV detection and classification as a unified recognition process and often rely on full RF signals or information-rich time-frequency representations. In this paper, we propose Direct Compressed-Sensing Network (Direct-CSNet), a two-stage inference framework operating directly in the compressed-domain for UAV presence detection and multi-class classification from controller signals under RF overlap conditions. In the proposed framework, Stage 1 performs binary UAV presence detection from compressed measurements, whereas Stage 2 performs multi-class UAV classification only for samples accepted by Stage 1. This conditional inference mechanism enables a lightweight pipeline that is suitable for deployment-oriented RF surveillance. Experimental results on the CardRF dataset show that Stage 1 achieves 95.56% accuracy, 0.9520 F1-score, and 0.9902 AUROC, while Stage 2 achieves 83.77% accuracy and 0.8415 macro-F1. At the full-system level, the proposed framework achieves 91.66% accuracy and 0.8602 macro-F1, while forwarding only 46.55% of the input samples to Stage 2. In addition, the average latency is 0.1099 ms per sample, and the throughput reaches 9102.6 samples/s. These results indicate that conditional inference in the compressed-domain is a practical and efficient solution for RF-based UAV surveillance.

Keywords: UAV detection; RF signal classification; compressed sensing; conditional inference; edge computing.

Tóm tắt:

Trong giám sát phương tiện bay không người lái (UAV) thụ động, việc xác định loại UAV chỉ có ý nghĩa khi hệ thống trước hết phát hiện được liệu đoạn tín hiệu vô tuyến (RF) quan sát có chứa hoạt động liên quan đến UAV hay không. Tuy nhiên, phần lớn các nghiên cứu hiện nay vẫn xem phát hiện và phân loại UAV như một quy trình suy luận thống nhất, đồng thời chủ yếu khai thác tín hiệu đầy đủ hoặc các biểu diễn giàu thông tin trong miền thời gian-tần số. Trong bài báo này, chúng tôi đề xuất Direct-CSNet (Direct Compressed-Sensing Network), một khung suy luận hai tầng hoạt động trực tiếp trên miền đo nén cho bài toán phát hiện hiện diện và phân loại đa lớp UAV từ tín hiệu của bộ điều khiển trong điều kiện chồng lấn RF. Cụ thể, Tầng 1 thực hiện phát hiện hiện diện UAV từ các phép đo nén, trong khi Tầng 2 chỉ thực hiện phân loại đa lớp đối với các mẫu được Tầng 1 chấp nhận. Cơ chế suy luận có điều kiện này cho phép xây dựng một quy trình gọn nhẹ, định hướng triển khai trên thiết bị biên. Kết quả thực nghiệm trên tập dữ liệu CardRF cho thấy Tầng 1 đạt độ chính xác 95,56%, F1-score 0,9520 và AUROC 0,9902; trong khi Tầng 2 đạt độ chính xác 83,77% và macro-F1 0,8415. Ở cấp độ toàn hệ thống, phương pháp đề xuất đạt độ chính xác 91,66% và macro-F1 0,8602, đồng thời chỉ chuyển tiếp 46,55% số mẫu đầu vào sang Tầng 2. Ngoài ra, hệ thống đạt độ trễ trung bình 0,1099 ms/mẫu và thông lượng 9102,6 mẫu/giây. Những kết quả này cho thấy suy luận có điều kiện trong miền đo nén là một giải pháp khả thi và hiệu quả cho giám sát UAV dựa trên tín hiệu RF.

Từ khóa: Phát hiện UAV; phân loại tín hiệu RF; lấy mẫu nén; tính toán biên.

1. INTRODUCTION

RF signal-based UAV surveillance is a passive and relatively low-cost approach, especially in scenarios where camera- or radar-based surveillance is limited. In practice, however, UAV surveillance is not merely a single-step classification problem. The system must first determine whether an observed RF segment contains UAV-related controller activity. Only then does recognizing the UAV type become meaningful. This makes the problem a conditional inference task, in which presence detection and classification are performed sequentially but serve different roles.

Based on the CardRF benchmark [1], previous studies [2]-[5] have shown that UAV RF signals can be distinguished from WiFi and Bluetooth signals in mixed environments. Nevertheless, most current methods still focus on directly optimizing the final recognition result, without clearly separating the presence screening step and the subsequent classification step within a unified pipeline.

Another limitation is that many high-performing methods [2]-[5] rely on either full RF signals or transformed representations such as Wavelet, HHT/WPT, or time-frequency features. These representations can increase computational cost and system complexity. In this context, direct inference from compressed measurements has emerged as a practical way to reduce acquisition and processing cost without requiring signal reconstruction. The feasibility of UAV detection and classification from sub-Nyquist compressed sensing was demonstrated in [8]. More recent benchmarks such as RFUAV and CageDroneRF [9], [10] also reflect the growing interest in broader and more reproducible evaluation settings for RF-based UAV surveillance.

In this paper, we propose Direct-CSNet, a two-stage compressed sensing inference framework for UAV presence detection and multi-class classification from controller RF signals under RF overlapping conditions. In the proposed architecture, Stage 1 acts as a binary UAV presence detector (present or absent), while Stage 2 performs multi-class classification only on samples detected as present by Stage 1. Both stages share a common compressed in-phase and quadrature (I/Q) representation, creating a unified and lightweight inference pipeline. The main contribution of this work lies in the design of the two-stage inference framework and the deployment-oriented evaluation methodology, rather than proposing a new architecture.

The main contributions of this paper are as follows:

1. We model controller-signal-based UAV surveillance as a two-stage inference pipeline in which UAV presence detection precedes multi-class classification.

2. We propose Direct-CSNet, a unified compressed-domain framework in which both stages share the same I/Q representation of compressed measurements.

3. We provide a stage-level and system-level evaluation of recognition performance, robustness under RF overlap, and deployment-related behavior, including latency,

throughput, forwarding rate, and end-to-end error propagation.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 details the proposed Direct-CSNet framework. Section 4 reports and discusses the experimental results. Section 5 concludes the paper.

2. RELATED WORK AND RESEARCH GAP

2.1. RF-Based UAV detection and identification

RF-based UAV detection and classification have been extensively studied with the emergence of public datasets and machine learning/deep learning methods. Among them, CardRF [1] is a standard benchmark, providing real-world RF recordings from UAVs, controllers, Bluetooth, and WiFi devices under both line-of-sight (LOS) and non-line-of-sight (NLOS) conditions.

Exploiting this dataset, previous studies have proposed diverse approaches, including hierarchical learning with HHT/WPT features [2], Wavelet processing in noisy environments [3], end-to-end deep learning [4], and STFT models combined with Transformers [5]. These studies show that RF sensing can effectively support UAV detection and identification. However, the majority utilize methods that analyze the full RF signal or information-rich time-frequency representations, rather than performing direct inference through compressed sensing. Recently, robustness, scalability, and comprehensive evaluation setups have received special attention, as demonstrated by interference-resistant oriented evaluations and the introduction of new benchmarks such as RFUAV or CageDroneRF [6], [9]-[12].

2.2. Direct RF inference from compressed sensing

Alongside full RF signal processing methods, direct inference using compressed sensing has emerged as a practical solution to minimize costs and signal processing requirements. By operating directly on compressed measurements, this approach can reduce data acquisition, storage, and transmission requirements without reconstructing the original signal. Study [8] demonstrated the feasibility of UAV detection and classification from sub-Nyquist RF measurements using a multi-stage deep learning pipeline. On a broader scope, hierarchical RF recognition models have also been applied in other problems, such as spatio-temporal models or bispectrum analysis [13]. Nevertheless, there remains a lack of studies constructing a unified compressed sensing inference framework that integrates both tasks: UAV presence detection and classification from controller signals under overlapping RF conditions.

2.3. Research gap and positioning of this work

Based on the literature review, three primary gaps are identified as the motivation for this study. First, many current RF-based UAV methods [2]-[5] still rely on processing the full RF signal or analyzing signals in various domains, imposing significant computational pressure for deployment on practical systems. Second, although direct RF inference using

compressed sensing has been proven feasible [8], the development of a unified framework integrating both detection and classification remains under-explored. Third, studies based on CardRF [1]-[5] have not clearly modeled the controller signal surveillance process as a conditional inference problem, where the presence screening step and multi-class classification operate within the same compressed sampled signal representation under RF overlapping conditions.

To exploit the advantages of compressed sensing in the UAV detection and classification problem, this paper proposes a two-stage inference framework, Direct-CSNet, with the objective not to create a new architectural model, but to focus on optimizing the UAV surveillance problem toward practical deployability.

3. THE PROPOSED TWO-STAGE DIRECT-CSNET FRAMEWORK

3.1. Problem formulation

Assume r[n] is the observed I/Q RF signal segment, with n = 0, 1,…, N-1. After vectorization, the compressed measurement vector y is generated by the sensing operator, as defined in Eq. (1):

$$y = \Phi r + v \tag{1}$$

In Eq. (1), $\mathbf{r} \in C^N$ denotes the vectorized RF I/Q segment, $y \in C^M$ denotes the compressed measurement vector, $\Phi \in C^{M\times N}$ is the compressed sensing operator, and $v \in C^M$ represents measurement or receiver noise. The measurement rate is defined in Eq. (2), where N is the input length and M is the number of retained compressed measurements. A smaller value of ρ corresponds to stronger compression.

$$\rho = \frac{M}{N} \tag{2}$$

From y, the system performs two sequential inference tasks. At Stage 1, the system produces a binary decision $\hat{d}$, as shown in Eq. (3):

$$\hat{d} \in \{0,1\} \tag{3}$$

Here, $\hat{d} = 1$ indicates UAV-related activity and $\hat{d} = 0$ indicates NO UAV background.

At Stage 2, only when $\hat{d} = 1$ does the system produce a multi-class UAV label $\hat{c}$ as defined in Eq. (4):

$$\hat{c} \in \{1,2,\ldots,C\} \tag{4}$$

Accordingly, the final system output can be summarized by Eq. (5):

$$\hat{y} = \begin{cases} \text{NO UAV}, & \hat{d} = 0 \\ \hat{c}, & \hat{d} = 1 \end{cases} \tag{5}$$

The system of equations from (1) to (5) models the RF surveillance logic compared to single-step classification systems. In practical surveillance, UAV type identification is meaningful only after the input segment has passed the presence detection stage.

Figure 1 illustrates the overall architecture of the two-stage inference framework, and Table 1 summarizes the core mathematical notations used throughout the model.

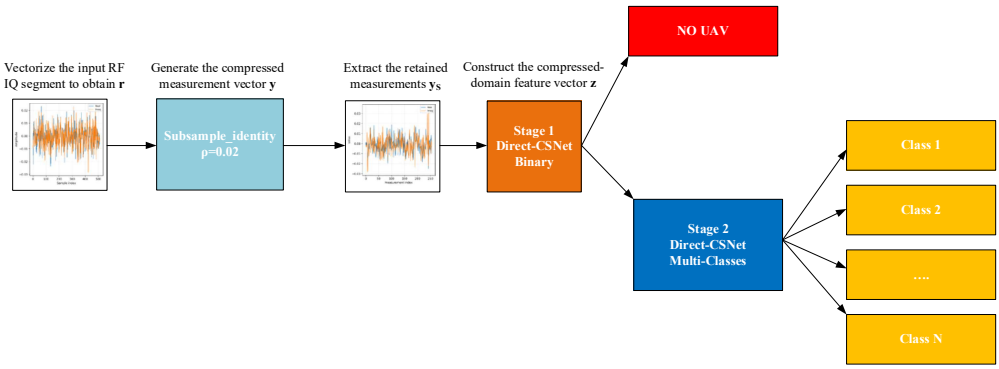

Figure 1. Two-stage compressed-domain framework for UAV presence detection and multi-class classification using Direct-CSNet

Table 1. Main symbols used in the two-stage model

Symbol	Meaning
r[n]	Input RF IQ segment
Ö	Compressed sensing operator
y	Compressed measurement vector
M	Number of measurements after compression
N	Input signal length
ρ	Measurement rate
S	Index set of retained measurements
z	Compressed-domain input feature
$\hat{d}$	Stage-1 UAV presence decision
τ	Stage-1 decision threshold
$\hat{c}$	Stage-2 multi-class UAV label
C	Number of UAV classes
T_{avg}	Average inference cost
T1	Stage-1 inference cost
T2	Stage-2 inference cost
p+	Forwarding probability from Stage 1 to Stage 2

3.2. Compressed-domain feature representation

To maintain a unified inference process for both stages, the proposed framework shares a common feature representation using compressed sensing. Let S be the index set of the retained compressed measurements, and yS be the vector containing the corresponding complex measurements. The input feature vector z is constructed by concatenating the real and imaginary parts of yS, defined according to equation (6):

$$z = \begin{bmatrix} \Re(y_S) \\ \Im(y_S) \end{bmatrix} = \left[\Re(y_S)^\top, \Im(y_S)^\top \right]^\top \qquad (6)$$

Equation (6) preserves the I/Q signal structure directly after compression, while avoiding explicit signal reconstruction and time-frequency transformations. Consequently, the same feature vector can be shared for both presence detection and multi-class classification tasks.

In this work, Direct-CSNet is instantiated as a lightweight multilayer perceptron (MLP). The network is not used to reconstruct the original RF waveform. Instead, it learns a task-oriented nonlinear mapping from the compressed-domain feature vector z to the target decision space. Let $h^{(\ell)}$ denote the hidden representation at layer ℓ. The hidden-layer transformation is written in Eq. (7):

$$h^{(\ell)} = \phi\left(W^{(\ell)}h^{(\ell-1)} + b^{(\ell)}\right), \qquad h^{(0)} = z \quad (7)$$

In Eq. (7), $W^{(\ell)}$ and $b^{(\ell)}$ are the trainable weights and biases of layer ℓ, and $\phi(\cdot)$ denotes the nonlinear activation.

3.3. Two-Stage inference with Direct-CSNet

3.3.1. Stage 1: UAV presence detection

At Stage 1, Direct-CSNet takes the compressed-domain feature vector z as input and outputs the probability p1 predicting whether the current segment contains UAV-related control activity, defined according to equation (8):

$$p_1 = \sigma\left(f_{\theta_1}(z)\right) \quad (8)$$

In Eq. (8), $f_{\theta_1}(\cdot)$ denotes the Stage 1 network with parameters θ_1, and $\sigma(\cdot)$ is the sigmoid activation.

The binary decision rule is given by Eq. (9), where τ is the Stage 1 decision threshold.

$$\hat{d} = \begin{cases} 1, & p_1 \geq \tau \\ 0, & p_1 < \tau \end{cases} \quad (9)$$

The Stage 1 model is trained using the binary cross-entropy loss in Eq. (10).

$$\mathcal{L}_{BCE} = -\frac{1}{N_b}\sum_{i=1}^{N_b}\left[d_i \log p_i + (1-d_i)\log(1-p_i)\right] \quad (10)$$

In Eq. (10), N_b is the mini-batch size, d_i is the ground-truth binary label, and p_i is the predicted probability for sample i. Under this design, Stage 1 acts as an inference gate that suppresses NO UAV background signals before invoking the multi-class branch.

3.3.2. Stage 2: Multi-Class UAV classification

Samples predicted as UAV-present by Stage 1 are forwarded to Stage 2 for multi-class classification. Here, the second neural network predicts the probability distribution p_2 over C UAV classes, according to equation (11):

$$p_2 = \text{soft}\max\left(f_{\theta_2}(z)\right) \quad (11)$$

In Eq. (11), $f_{\theta_2}(\cdot)$ denotes the Stage 2 model. The final class label is determined by Eq. (12):

$$\hat{c} = \arg\max_{k\in\{1,\ldots,C\}} p_{2,k} \quad (12)$$

Stage 2 is trained using the multi-class cross-entropy loss in Eq. (13):

$$\mathcal{L}_{CE} = -\frac{1}{N_m}\sum_{i=1}^{N_m}\log p_{i,c_i} \quad (13)$$

In Eq. (13), N_m is the Stage 2 mini-batch size, c_i is the ground-truth UAV class of sample i, and p_{i,c_i} is the predicted probability assigned to that class.

3.3.3. Conditional inference and average inference cost

Because Stage 2 is only conditionally activated when samples pass Stage 1, the average inference cost per input does not equal the sum of the costs of both stages. Let T_1 and T_2 denote the inference cost of Stage 1 and Stage 2, respectively, and let $\mathbf{p}_+$ denote the forwarding probability from Stage 1 to Stage 2. The average inference cost is defined in Eq. (14):

$$T_{avg} = T_1 + p_+ T_2 \quad (14)$$

The forwarding probability is estimated empirically by Eq. (15):

$$p_+ \approx \frac{N_{forward}}{N_{total}} \quad (15)$$

In Eq. (15), $N_{forward}$ is the number of samples passed to Stage 2 and N_{total} is the total number of evaluated samples. Equations (14) and (15) clarify the central advantage of the proposed framework: the more expensive multi-class branch is applied only to samples that have already been judged likely to contain UAV-related activity.

3.4. Algorithmic summary of the proposed framework

Algorithm 1 summarizes the end-to-end inference process of the proposed model. Starting from an input I/Q RF signal segment, the system constructs a feature vector using shared compressed sensing. Stage 1 then uses it to detect the presence of a UAV; the multi-class classification branch at Stage 2 activates when the sample passes the binary screening gate. In this way, compressed-domain preprocessing is shared across both stages, while conditional branching is introduced only after Stage 1.

Algorithm 1. Two-stage compressed-domain inference using Direct-CSNet

Input: RF IQ segment r[n], compressed sensing operator Φ, retained-index set S, decision threshold τ

Output: Final label $\hat{y}$

1. Vectorize the input RF IQ segment to obtain r.

2. Generate the compressed measurement vector according to Eq. (1): $y = \Phi r + v$

3. Extract the retained measurements y_S

4. Construct the compressed-domain feature vector according to Eq. (6):

$$z = \left[\Re(y_S)^\top, \Im(y_S)^\top\right]^\top$$

5. Compute the Stage-1 UAV-presence probability according to Eq. (8):

6. Apply the Stage-1 decision rule in Eq. (9). If $p < \tau$, assign $\hat{y} = \text{NO UAV}$, and stop

7. Otherwise, compute the Stage-2 class-probability vector according to Eq. (11):

$$p_2 = \text{soft max}\left(f_{\theta_2}(z)\right)$$

8. Determine the UAV class label according to Eq. (12):

$$\hat{c} = \arg \max_{k \in \{1 \ldots C\}} p_{2,k}$$

9. Assign the final output $\hat{y} = \hat{c}$, and return $\hat{y}$

4. EXPERIMENTAL RESULTS

4.1. Experimental setup

The experiments are conducted on the CardRF dataset under a surveillance scenario focusing on controller signals. Stage 1 performs binary presence detection, distinguishing UAV-related samples from WiFi/Bluetooth background signals, using 73,780 training samples and 31,620 test samples; whereas Stage 2 executes 6-class UAV classification on 33,915 training samples and 14,535 test samples. The data split follows the split_exp field in the CardRF manifest, retaining LOS samples and excluding NLOS samples from the main protocol. The overlapping scenarios (Clean, Moderate, Heavy) are synthesized after this splitting step to evaluate performance at each target SNR level (-15dB to 20dB, in 5dB steps), ensuring complete isolation between the training and testing sets. The overlap protocol is intended to provide a controlled evaluation of interference robustness rather than session-

wise generalization. Representative UAV-controller, Wifi, Bluetooth, and overlapped signals in the Stage-1 protocol are illustrated in Figure 2.

The system is deployed on a unified pipeline with an input data size of 65,536 samples. First, the *subsample_identity* compression matrix is applied with a nominal compression ratio of $\rho = 0.02$. After compression, a subset of 512 complex measurements is retained for inference. The input feature is further structured by concatenating the real and imaginary parts of these measurements, yielding a real-valued representation with a dimension of 1024. Both stages employ a lightweight MLP architecture (Hidden dimension: 256, Dropout: 0.1). The model is optimized using the AdamW algorithm ($\mathrm{LR} = 10^{-3}$, $\mathrm{WD} = 10^{-4}$, Batch size: 1024). Stage 1 is trained for 20 epochs with the *BCEWithLogits* loss function, while Stage 2 is trained for 25 epochs using the *CrossEntropy* loss combined with a class weighting technique. Performance is quantified through Accuracy, Precision, Recall, F1-score, and AUROC metrics for Stage 1; along with Macro/ Weighted metrics for Stage 2 and the overall system. In addition, practical deployment parameters such as latency and throughput are analyzed in detail.

Table 2. Overall component-level results

Component	Stage 1	Stage 2
Task	Binary detection	6-class classification
Test samples	31620	14535
Classes	2	6
Accuracy	95.56%	83.77%
Precision / Macro-Precision	0.946	0.845
Recall / Macro-Recall	0.958	0.8418
F1 / Macro-F1	0.952	0.8415
Weighted-F1	-	0.8376
AUROC	0.9902	-
Loss	0.1575	-

4.2. Main recognition results

Table 2 presents the experimental results for both stages. Stage 1 performs effectively as a UAV presence detector, achieving an accuracy of 95.56%, an F1-score of 0.952, and an AUROC of 0.9902 on 31,620 binary test samples. Stage 2, which undertakes the six-class classification task, achieves an accuracy of 83.77% and a macro-F1 of 0.8415

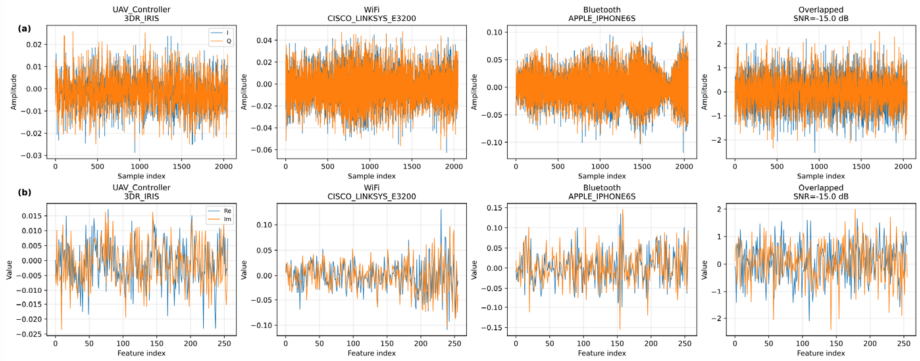

Figure 2. Representative UAV-controller, WiFi, Bluetooth, and overlapped signals (a) raw RF IQ, (b) compressed-domain features

on 14,535 positive samples. Analysis of the confusion matrix (Figure 3) and the impact of overlapping levels (Figure 4) reveals that Stage 1 maintains high robustness from clean to moderate overlap conditions, with noticeable degradation only under heavy overlap at very low SNR levels. In contrast, Stage 2 is more sensitive to noise due to the signal similarity among controllers from the same manufacturer.

Classification details (Table 3, Figure 3) indicate that errors are primarily concentrated in the DJI INSPIRE and DJI M600 pair, whereas the 3DR IRIS, BEEBEERUN, and DJI PHANTOM series are distinctly separated; demonstrating that the core challenge lies in the fine-grained classification step rather than presence screening. This pattern suggests that controllers from the same manufacturer may exhibit similar RF characteristics. Under heavy overlap and low-SNR conditions, the compressed-domain representation preserves coarse discriminative cues but may lose some finer class-specific details, which mainly affects Stage 2.

Table 3. Per-class performance of Stage 2

Class	Support	Precision	Recall	F1-score
3DR IRIS	2550	0.9409	0.9682	0.9544
BEEBEERUN	1785	0.9122	0.9199	0.916
DJI INSPIRE	2550	0.6653	0.7773	0.7169
DJI M600	2550	0.7698	0.6663	0.7143
DJI MAVICPRO	2550	0.898	0.8008	0.8466
DJI PHANTOM	2550	0.8834	0.9184	0.9006

Regarding the end-to-end system performance (Table 4, Table 5; Figure 5, Figure 6), the model records 91.66% accuracy and a weighted-F1 of 0.9149 across all 7 output classes. Notably,

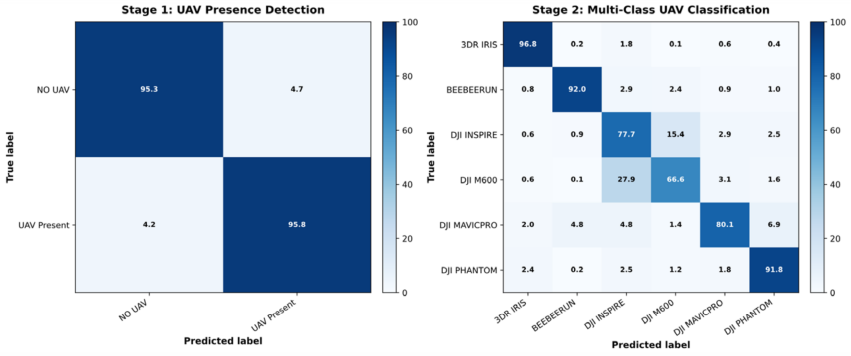

Figure 3. Component-level confusion matrices of Stage 1 and Stage 2

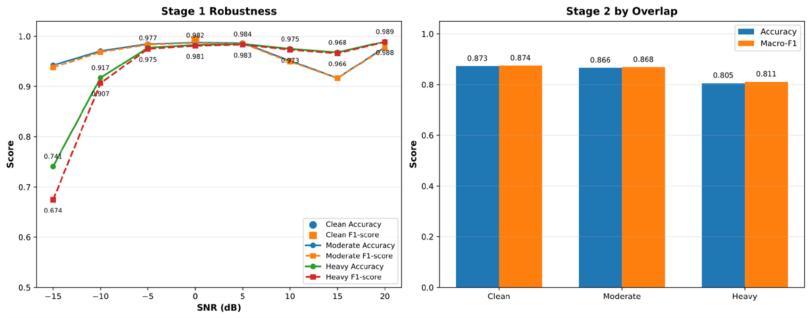

Figure 4. Component-level performance by overlap condition

the system forwards only 46.55% of the input samples to Stage 2, with the global error source primarily stemming from internal confusion within Stage 2 rather than false rejections at Stage 1. This result indicates that the proposed framework preserves recognition capability while implementing a practically meaningful conditional inference mechanism.

4.3. Comparative evaluation and deployability

To objectively evaluate the value of the hierarchical architecture, Table 6 compares the proposed framework against a one-stage

Table 4. Overall end-to-end performance of the two-stage system

Metric	Value
Test samples	31620
Output classes	7
Accuracy	91.66%
Macro-Precision	0.8755
Macro-Recall	0.8487
Macro-F1	0.8602
Weighted-F1	0.9149

Table 5. Per-class end-to-end performance of the two-stage system

Class	Support	Precision	Recall	F1-score
NO UAV	17085	0.9655	1	0.9825
3DR IRIS	2550	0.955	0.9325	0.9437
BEEBEERUN	1785	0.9212	0.9171	0.9191
DJI INSPIRE	2550	0.6918	0.7624	0.7254
DJI M600	2550	0.7868	0.6569	0.716
DJI MAVICPRO	2550	0.9082	0.7718	0.8344
DJI PHANTOM	2550	0.8997	0.9004	0.9

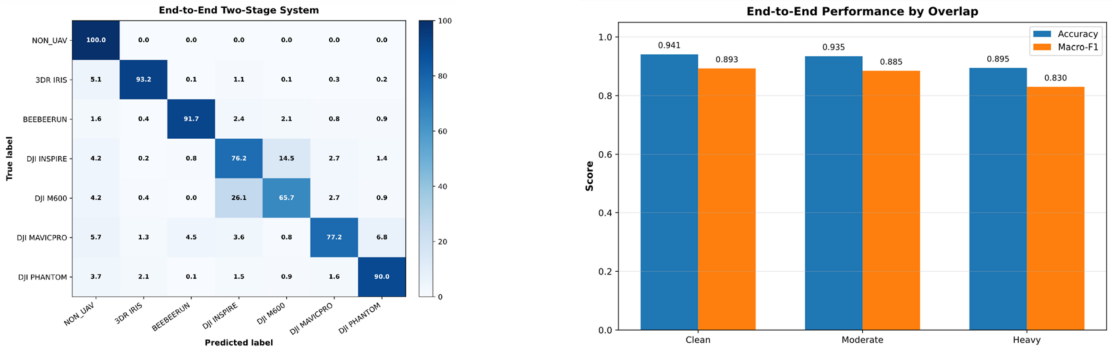

Figure 5. End-to-end confusion matrix and overlap-wise performance of the proposed two-stage system

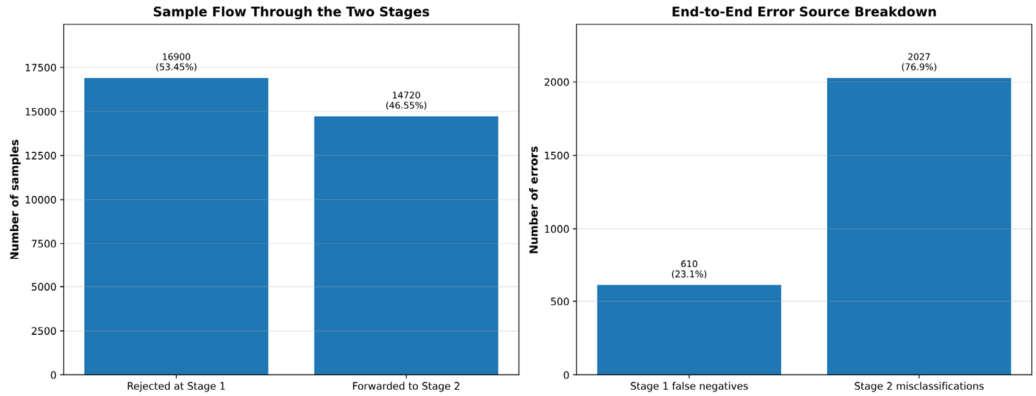

Figure 6. System-level behavior of the proposed two-stage framework

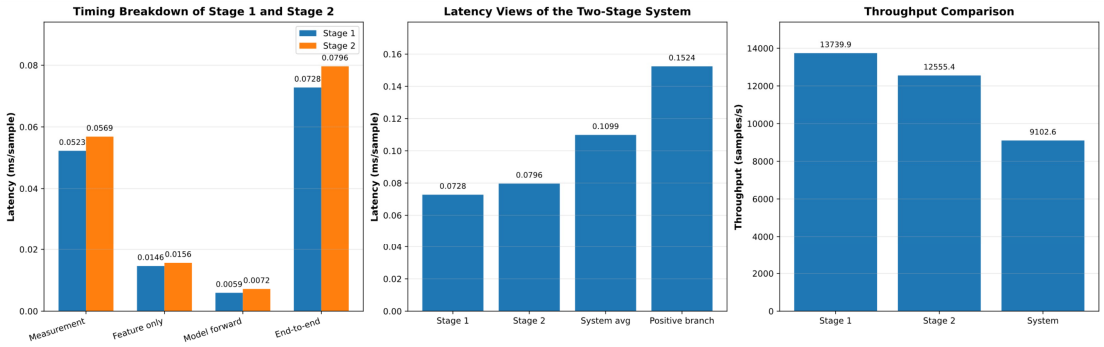

Figure 7. Computational efficiency and deployability comparison

baseline model. The baseline model is configured fairly by utilizing the same parameters (the subsample_identity compression operator, raw_ri features) and MLP network architecture, but it performs direct 7-class classification (No UAV + 6 UAV classes). The results reveal a clear trade-off: the baseline holds an advantage in speed (latency of 0.0751 ms/sample, throughput of 13,311.3 samples/s), whereas the two-stage framework holds an advantage in end-to-end recognition quality (achieving an accuracy of 91.66% versus 89.48%, and a weighted-F1 of 0.9149 versus 0.8955).

Practical deployability is illustrated in detail in Figure 7, with the individual latencies of Stage 1 and Stage 2 being 0.0728 and 0.0796 ms/sample, respectively. Thanks to the analyzed conditional computation mechanism (only activating Stage 2 for positive samples), the average system-wide latency is maintained at 0.1099 ms/sample, with an effective throughput reaching 9,102.6 samples/s. This result shows that the proposed two-stage framework improves recognition quality while maintaining a practical average inference cost.

4.4. Sensitivity and ablation studies

To evaluate the robustness and parameter selection in the proposed model, we conduct a sensitivity analysis and an ablation study.

Table 6. Comparison between a matched one-stage compressed-domain baseline and the proposed two-stage framework

System	One-stage compressed-domain baseline	Proposed two-stage framework
Label space	NO UAV + 6 UAV classes	Stage 1 binary + Stage 2 six-class UAV
Accuracy	89.48%	91.66%
Macro-F1	0.8455	0.8602
Weighted-F1	0.8955	0.9149
Avg. latency (ms/sample)	0.0751	0.1099
Throughput (samples/s)	13,311.3	9,102.6

First, tuning the decision threshold τ of Stage 1 (Figure 8) in the range from 0.3 to 0.7 shows that the system maintains stability, with the end-to-end macro-F1 and average latency remaining nearly unchanged (fluctuating only within the ranges of 0.8579-0.8592 and

0.1087-0.1094 ms/sample, respectively). As τ increases, the forwarding rate decreases and false negative errors at Stage 1 slightly increase, but this effect is largely compensated by a reduction in errors at Stage 2. Therefore, $\tau = 0.5$ is established as the default balanced operating point.

Second, regarding the compression ratio, analysis on the reduced dataset is summarized in Table 7. Unlike the main protocol, this reduced ablation does not cap the retained measurements at 512.

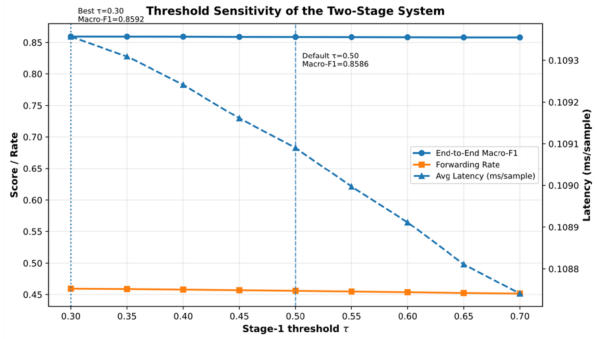

Figure 8. Threshold sensitivity and conditional inference tradeoff of the proposed two-stage framework

The results show that ρ = 0.02 provides the best trade-off among the tested configurations. This compression level not only achieves an accuracy of 79.95% and a macro-F1 of 0.8019 but also optimizes latency and throughput. Deeper compression ($\rho = 0.01$) causes information loss, whereas $\rho > 0.02$ inflates the feature representation and increases processing costs without providing noticeable improvement in recognition performance.

Table 8. Hidden-dimension ablation on the reduced Stage-2 overlap multi-class subset (ρ=0.02)

Hidden dim	Accuracy	Macro-F1
128	0.794	0.7979
256	0.7995	0.8019
512	0.8023	0.8077
768	0.798	0.8025
1024	0.799	0.8052

Finally, the network capacity (hidden dimension) analysis reveals: although increasing the hidden dimension from 256 to 512 enhances local performance at each

Table 7. Measurement-rate ablation in the Stage-2 overlap multi-class setting

ρ	Operator shape	Feature dim	Accuracy	Macro-F1	End-to-end latency (ms/sample)	Throughput (samples/s)
0.01	655 × 65536	1310	0.7893	0.794	0.0709	14103.2
0.02	1311 × 65536	2622	0.7995	0.8019	0.0654	15288.7
0.03	1966 × 65536	3932	0.7835	0.7909	0.0699	14306.1
0.04	2621 × 65536	5242	0.7913	0.795	0.0768	13023.6
0.05	3277 × 65536	6554	0.7915	0.7959	0.0791	12637.1

discrete stage (Table 8), it degrades the overall system quality when subjected to full-protocol validation (Table 9). Specifically, the end-to-end accuracy drops from 91.66% to 91.39%, and the macro-F1 decreases to 0.8591. The main reason is that the larger feature space makes the screening gate at Stage 1 more stringent (the forwarding rate drops to only 44.88%), thereby increasing early rejection errors. Therefore, a hidden dimension of 256 is retained as the default configuration because it provides a better balance between end-to-end recognition performance and deployability.

Table 9. Full-protocol validation of hidden dimensions 256 and 512

Configuration	hidden = 256	hidden = 512
Stage 1 Accuracy	95.56%	95.84%
Stage 1 F1	0.952	0.9542
Stage 2 Accuracy	83.77%	84.00%
Stage 2 Macro-F1	0.8415	0.8433
End-to-end Accuracy	91.66%	91.39%
End-to-end Macro-F1	0.8602	0.8591
Forwarding rate	46.55%	44.88%

4.5. Comparison with prior CardRF-based studies and discussion

Direct comparison with existing studies on the CardRF dataset (Table 10) should be interpreted as a comparison of design trade-offs rather than as a strict head-to-head benchmark. The reason is that previous methods [2]-[5] all analyze the full RF signal or time-frequency representations (such as Wavelet, HHT/WPT, or STFT-Transformer). Therefore, the objective of this work is not to achieve the highest absolute accuracy, but to explore a different operating regime based on compressed-domain processing and conditional inference.

While models utilizing raw features can achieve accuracies from 97% to 98.9%, they generally rely on full RF signals or information-rich time-frequency representations, which require heavier front-end processing. In contrast, the proposed Direct-CSNet framework operates directly on compressed measurements and activates the multi-class branch only when needed. This design reduces representation size and avoids more expensive front-end transformations, although it does not aim to achieve the highest absolute accuracy. By forwarding only 46.55% of the data to the classification branch, the system maintains a relatively low average latency of 0.1099 ms/sample. Therefore, the proposed method should be interpreted as a deployment-oriented alternative that offers a different trade-off between recognition quality and practical deployability, rather than as a direct replacement for information-rich full-signal pipelines [2]-[5].

The experimental results also point out the limitations of Direct-CSNet: the primary end-to-end error source does not originate from the detection stage but lies in the ability to delineate closely matched class boundaries in the multi-class classification branch (especially the DJI INSPIRE and M600 pair). This indicates that the compressed-domain representation, while effective for coarse

classification, still faces limitations when processing highly similar controller signals under heavy noise conditions. Although the ablation studies are performed on a reduced dataset to serve the design justification process, the results still support the main claim of the paper: The conditional two-stage compressed-domain framework is a promising solution for RF- based UAV surveillance under practical deployment constraints.

Table 10. Indicative comparison with prior CardRF-based studies from the perspective of performance and deployability.

Studies	Representation/ input	Task	Reported result	Reported latency	Deployment implication
[2]	HHT/WPT; UAV, controller, WiFi/Bluetooth	Hierarchical learning for detection and identification	Semi-supervised detection followed by three-level hierarchical identification; the paper does not report a single end-to-end metric directly comparable to the present two-stage system	N/A	A foundational CardRF baseline based on information-rich features rather than compressed-domain inference
[3]	Wavelet scattering / wavelet features	RF-based UAV signal classification under wireless interference	Accuracy 98.9% at 10 dB SNR (WST + SqueezeNet)	N/A	High performance, but dependent on time-frequency feature extraction
[4]	Full RF signal; end-to-end deep learning	One-stage end-to-end RF detection/ identification framework	Reported detection metrics: Accuracy 97.53%; Precision 98.06%; Sensitivity 98.00%; F1 98.00%	0.37 ms	A strong detection baseline, but it neither uses compressed measurements nor conditional invocation
[5]	STFT + hybrid transformer	Coarse-grained signal identification on CARDRF	Accuracy 97.57% under Gaussian noise; 98.03% under co-channel/co-frequency interference	Not clearly reported	Very high accuracy, but relies on STFT and a heavier model family
Proposed method	Direct compressed-domain; raw_ri + Direct-CSNet	Two-stage framework: presence detection + multi-class classification	Stage 1: Accuracy 95.56%; F1 0.9520; AUROC 0.9902; Stage 2: Accuracy 83.77%; Macro-F1 0.8415; Overall system: Accuracy 91.66%; Macro-F1 0.8602	Stage 1: 0.0728 ms/sample; Stage 2: 0.0796 ms/sample; Overall system: 0.1099 ms/ sample	Emphasizes lightweight, near-real-time inference and suitability for edge deployment

5. CONCLUSION

This paper proposed Direct-CSNet, a two-stage inference framework operating directly in the compressed-domain for UAV presence detection and multi-class classification under RF overlap conditions. By combining a lightweight binary screening stage with a conditional multi-class classifier on a shared compressed-domain representation, the proposed framework achieves favorable end-to-end recognition performance while maintaining a practical inference cost.

The main contribution of this work is not the introduction of a new backbone model, but the design of a practical compressed-domain conditional inference framework for RF-based UAV surveillance. The results support the feasibility of maintaining useful recognition performance under practical constraints on sensing resources, feature dimensionality, and inference latency. A remaining limitation is the reduced separability between closely matched classes, especially DJI INSPIRE and DJI M600, under heavy overlap and low-SNR conditions.

REFERENCES

[1]. O. O. Medaiyese, M. Ezuma, A. P. Lauf, and A. A. Adeniran, "Cardinal RF (CardRF): An Outdoor UAV/UAS/Drone RF Signals with Bluetooth and WiFi Signals Dataset", IEEE DataPort, 2022. https://dx.doi.org/10.21227/1xp7-ge95.

[2]. O. O. Medaiyese, M. Ezuma, A. P. Lauf, and A. A. Adeniran, "Hierarchical Learning Framework for UAV Detection and Identification", IEEE Journal of Radio Frequency Identification, vol. 6, pp. 176-188, 2022. https://doi.org/10.1109/JRFID.2022.3157653.

[3]. 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, Art. no. 101569, 2022. https://doi.org/10.1016/j.pmcj.2022.101569.

[4]. S. S. Alam, A. Chakma, M. H. Rahman, R. B. Mofidul, M. M. Alam, I. B. K. Y. Utama, and Y. M. Jang, "RF-Enabled Deep-Learning-Assisted Drone Detection and Identification: An End-to-End Approach", Sensors, vol. 23, no. 9, Art. no. 4202, 2023. https://doi.org/10.3390/s23094202.

[5]. X. Yan, B. Han, Z. Su, and J. Hao, "SignalFormer: Hybrid Transformer for Automatic Drone Identification Based on Drone RF Signals", Sensors, vol. 23, no. 22, Art. no. 9098, 2023. https://doi.org/10.3390/s23229098 .

[6]. H. Elyousseph and M. Altamimi, "Robustness of Deep-Learning-Based RF UAV Detectors", Sensors, vol. 24, no. 22, Art. no. 7339, 2024. https://doi.org/10.3390/s24227339.

[7]. A. AlKhonaini, T. Sheltami, A. Mahmoud, and M. Imam, "UAV Detection Using Reinforcement Learning", Sensors, vol. 24, no. 6, Art. no. 1870, 2024. https://doi.org/10.3390/s24061870.

[8]. Y. Mo, J. Huang, and G. Qian, "Deep Learning Approach to UAV Detection and Classification by Using Compressively Sensed RF Signal", Sensors, vol. 22, no. 8, Art. no. 3072, 2022. https://doi.org/10.3390/s22083072.

[9]. R. Shi, X. Yu, S. Wang, Y. Zhang, L. Xu, P. Pan, and C. Ma, "RFUAV: A Benchmark Dataset for Unmanned Aerial Vehicle Detection and Identification", arXiv preprint arXiv:2503.09033, 2025. https://doi.org/10.48550/arXiv.2503.09033.

[10]. M. Rostami, A. Faysal, H. Xia, H. Kasasbeh, Z. Gao, and H. Wang, "CageDroneRF: A Large-Scale RF Benchmark and Toolkit for Drone Perception", arXiv preprint arXiv:2601.03302, 2026. https://doi.org/10.48550/arXiv.2601.03302.

[11]. X. Yan, T. Fu, H. Lin, F. Xuan, Y. Huang, Y. Cao, H. Hu, and P. Liu, "UAV Detection and Tracking in Urban Environments Using Passive Sensors: A Survey", Applied Sciences, vol. 13, no. 20, Art. no. 11320, 2023. https://doi.org/10.3390/app132011320.

[12]. 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, no. 5, Art. no. 879, 2024. https://doi.org/10.3390/rs16050879.

[13]. H. Bai, S. Li, Y. Jia, and B. Xiao, "Radio Signal Recognition Using Two-Stage Spatiotemporal Network with Bispectral Analysis", Sensors, vol. 25, no. 17, Art. no. 5449, 2025. https://doi.org/10.3390/s25175449.