drones

MDPI

Article

PeirceTriadicNet: A Peircean Triadic Three-Branch Fusion Network for Non-Cooperative Signal Recognition in UAV Communications

Cheng Chang *, Tingting Lu, Zixin Li, Xin Gu, Zhe Li and Haipeng Chen

Research and Development Center, China Academy of Launch Vehicle Technology, Beijing 100076, China
* Correspondence: loeibx@163.com

Highlights

What are the main findings?

- In UAV communications, signals from unknown emitters are hard to recognize because they are weak, have unknown parameters, and may be jammed. This work shows that combining three complementary information sources—the received signal waveform, physical source parameters, and prior knowledge about modulation families—lets a single network recognize non-cooperative signals far more accurately than using the waveform alone.
- A controlled comparison shows that this gain comes from the way the three information sources are combined (a simple fixed three-way fusion) rather than from a more complex learned weighting mechanism.

What are the implications of the main findings?

- The proposed framework offers a practical recipe for spectrum awareness in UAV communications: it remains accurate at low signal-to-noise ratio and does not require target-domain data or additional annotation, leaving room for lightweight on-board deployment.
- The same three-way fusion idea can be extended to other low-altitude security tasks, such as drone detection and emitter identification in emergency-response and anti-drone scenarios.

Abstract

Non-cooperative signal recognition is essential for spectrum awareness and interference management in unmanned aerial vehicle (UAV) communications, where unknown emitters, jamming threats, and low signal-to-noise ratio (SNR) make modulation classification and parameter estimation particularly challenging. Existing approaches often treat modulation classification, parameter estimation, and semantic reasoning as isolated tasks, making it difficult to exploit the complementarity of cross-domain information. This paper proposes PeirceTriadicNet (PTN), a Peircean triadic-relation network for this problem. The Representamen branch processes observable signal representations (IQ waveforms, time–frequency maps, and power spectral densities), the Object branch embeds physical source parameters (modulation type, symbol rate, carrier-frequency offset, roll-off, bandwidth, and SNR), and the Interpretant branch incorporates modulation-family and SNR priors for semantic reasoning. The final model, PTN-ResIcon, fuses the three branch embeddings with fixed equal weights ($\alpha_R = \alpha_O = \alpha_I = 1/3$). On RML-S-Full-Std, a UAV-oriented synthetic dataset that re-generates the RML2016.10a modulation set under representative UAV link conditions (Rician fading, Doppler shift, and barrage, partial-band, follower, and deceptive jamming), with 220,000 IQ frames of 11 modulation formats over −20 to +18 dB,

Check for updates

Academic Editor: Emmanouel T. Michailidis

Received: 12 October 2026
Revised: 7 November 2026
Accepted: 11 November 2026
Published: 13 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

per-frame physical-parameter and semantic annotations, and a released open-data generation script—the covered PSK/QAM OFDM video downlinks and FSK/CPM frequency-hopping control links mirror the signal families observed in real commercial drone RF links—PTN-ResIcon achieves an overall modulation-recognition accuracy of 80.67%, which is 17.6 percentage points higher than the Representamen branch alone and 24.0, 20.1, and 20.2 percentage points higher than CNN2, CNN-LSTM, and ResNet1D, respectively; the +18 dB accuracy is 91.50%. On the real DroneRFa corpus (USRP-acquired recordings of four DJI devices at three flight distances), the complete triadic configuration reaches 72.50% device-identification accuracy, demonstrating that the framework transfers from synthetic to genuine UAV signals. An SNR-biased learnable-attention variant, PTN-ResIcon-Learned, dynamically adjusts the triadic weights from the oracle SNR and branch confidences; it reaches 76.86%, showing that fixed equal-weight triadic fusion itself provides the main gain and that overly aggressive SNR gating slightly degrades generalization. In the UAV-swarm scenario of RML-S-Full-Std, PTN-ResIcon maintains an accuracy of 81.44%, demonstrating its applicability to UAV communication environments and to spectrum-awareness tasks in emergency-response missions.

Keywords: automatic modulation recognition (AMR); UAV communications; UAV swarm; Peircean semiotics; triadic-relation modeling; parameter estimation

1. Introduction

Automatic modulation classification (AMC) and signal parameter estimation are core tasks in non-cooperative signal recognition, spectrum monitoring, and cognitive radio [1,2]. Traditional methods can be grouped into two categories: statistical-estimation approaches offer strong interpretability but suffer from threshold effects at low SNR; deep-learning approaches achieve high accuracy when data are abundant but are less interpretable [2].

Unmanned aerial vehicle (UAV) communications have become a critical component of modern wireless networks, enabling surveillance, relay, and swarm missions. In emergency-response applications such as disaster relief and search-and-rescue, UAV swarms are often deployed to restore situational awareness and communications in congested or contested spectrum, where spectrum awareness is indispensable: UAV links face intentional jamming and spoofing and must identify non-cooperative emitters without prior synchronization. Because UAV platforms are constrained in size, weight, and power, and because UAV-to-ground and UAV-to-UAV links frequently encounter low SNR, Doppler shift, and fast-fading conditions, recognition algorithms deployed in this context must remain accurate at low SNR while keeping computational cost manageable. These requirements directly motivate robust, multi-source fusion methods for automatic modulation classification (AMC) and parameter estimation.

In complex non-cooperative scenarios, single-feature or single-task methods face three common challenges. First, low SNR severely corrupts constellation details and time–frequency structure. Second, parameters such as roll-off factor and occupied bandwidth lack strong periodic structure and are difficult to estimate. Third, conventional AMC outputs only a modulation label and cannot directly support risk assessment or jamming decisions.

Peirce's semiotics describes every sign process through three interrelated elements: the *Representamen*, the form that the sign takes; the *Object*, what the sign stands for; and the *Interpretant*, the effect or meaning that the sign produces in an interpreter. The triadic sign relation—Representamen–Object–Interpretant—provides a conceptual framework for unifying these tasks [3]. In this analogy, the Representamen corresponds to observable signal representations (IQ waveforms, time–frequency maps, power spectra); the Object

corresponds to physical parameters that generate the signal (modulation type, symbol rate, carrier-frequency offset, roll-off, bandwidth, SNR); and the Interpretant corresponds to semantic interpretations (platform type, risk level, jamming strategy). This paper emphasizes that the mapping is analogical rather than identical; the concrete correspondence is given in Table 1. The value of the analogy lies in clarifying how different information sources complement one another: when waveform details are corrupted by noise, spectral shape and prior knowledge can constrain inference, and vice versa.

Table 1. Peircean triadic analogy for signal recognition.

Peirce Element	Signal-Domain Analogy	Learned Branch	Concrete Form
Representamen	Observable signal representation	IconResNet + IndexNet	IQ waveform; STFT magnitude and PSD
Object	Physical source parameters	ObjectEncoder	Modulation type, symbol rate, CFO, roll-off, bandwidth, SNR
Interpretant	Semantic interpretation	SymbolNet + task heads	Platform type, risk level, jamming strategy, confidence

This paper proposes PeirceTriadicNet (PTN), a Peircean triadic-relation network. PTN instantiates the triad with three branches—a Representamen branch that fuses Icon and Index encoders, an Object branch that encodes physical source parameters, and an Interpretant branch based on SymbolNet—and fuses their embeddings. The final model, PTN-ResIcon, uses fixed equal-weight fusion; a learnable SNR-biased attention variant is studied as an ablation. The main contributions are summarized as follows:

1. The Peircean triad is formalized as a strict computational analogy for signal recognition in UAV communications, with an explicit three-branch architecture (Representamen, Object, Interpretant) and a clear mapping table relating each Peircean element to a concrete signal-domain information source.
2. PTN is proposed to fuse the three Peircean branches, supporting modulation classification, semantic tasks, and parameter regression for non-cooperative signal recognition; an SNR-biased learnable attention mechanism is introduced as an interpretability ablation.
3. Systematic evaluations are conducted on RML-S-Full-Std, including single-branch baselines, fixed-fusion and learned-attention ablations, parameter-regression analysis, and SNR-robustness experiments, together with scenario-wise validation covering a UAV-swarm scenario (a simulated multi-UAV spectrum-sharing environment in which PTN-ResIcon maintains 81.44% accuracy) and a held-out test set.
4. The related work is extended and updated to cover deep-learning AMC, attention and multi-modal fusion, multi-task learning, interpretability, cognitive radio, and Peircean semiotics.

2. Related Work

2.1. Deep-Learning AMC

Since the release of the RadioML datasets, deep-learning-based AMC has been studied extensively [1]. O'Shea et al. demonstrated that convolutional networks can classify modulations directly from IQ data [1]. Subsequent works explored CNN-LSTM architectures [4], residual-network (ResNet) variants [5,6], attention mechanisms [7–9], vision transformers [10], and Transformer–CNN hybrid networks [11]. Graph neural networks have also been used to model structural relationships among modulation samples [12]. Surveys by Jdid et al. [2], Sanna et al. [5], and Rehman et al. [13] summarize recent progress and challenges.

2.2. Multi-Modal and Multi-Task Fusion

Multi-modal fusion in AMC combines IQ data, constellation diagrams, and time–frequency representations [14–16]. Feature-level and decision-level fusion strategies have been compared for modulation classification [16]. Multi-task learning has been applied to joint modulation and wireless-signal classification [17]. However, most existing methods do not explicitly model which view should be relied upon under which SNR condition.

2.3. Interpretability and Semiotics

Interpretable artificial intelligence remains an active topic in signal processing. Semiotic perspectives have been used to explain intermediate representations of deep networks in vision and language [18,19]. The Peircean triadic framework has been applied to cognitive systems and communication theory [19]. To the best of the authors' knowledge, a direct mapping of the Peircean triad to signal recognition via SNR-aware fusion has not been reported.

2.4. Cognitive Radio and Parameter Estimation

Cognitive radio emphasizes the perception–action cycle and adaptive spectrum use [20,21]. Regression estimation of parameters such as carrier-frequency offset (CFO), symbol rate, and bandwidth has been studied, whereas roll-off estimation has received less attention. In the context of UAV communications, spectrum sensing and non-cooperative emitter identification have also attracted increasing attention, driven by security and spectrum-sharing requirements in UAV networks and swarms.

2.5. Gap Summary

The above review highlights three persistent limitations of existing research. First, in terms of *information coverage*, most AMC methods operate on a single view (IQ waveforms or time–frequency images) and treat parameter estimation and semantic reasoning as isolated tasks; even multi-modal fusion studies [14–16] combine signal views but do not explicitly incorporate physical source parameters or semantic priors, so cross-domain complementarity remains underexploited. Second, in terms of *reliability*, existing models are largely SNR-agnostic: they do not encode an explicit mechanism for deciding when to trust waveform details (reliable at high SNR) and when to fall back on prior knowledge (necessary at low SNR), and none of the reviewed methods provides a quantitative study of how recognition degrades as SNR estimation becomes noisy, as required for UAV links with fast-fading conditions. Third, in terms of *data*, publicly available benchmarks that simultaneously provide modulation labels, physical parameters, and semantic interpretations are scarce, and standard datasets such as RML2016.10a offer modulation labels only; this forces most studies to evaluate AMC in isolation and leaves the value of fused multilayer annotations unmeasured. These gaps motivate a triadic framework that explicitly links features, parameters, and semantics, an SNR-aware fusion mechanism, and a jointly annotated benchmark for UAV-oriented signal recognition.

3. System Model and Triadic Modeling

3.1. System-Level Model and Information Provenance

Figure 1 presents the system-level model of non-cooperative signal recognition in UAV communications, from the RF environment to the recognition outputs. A non-cooperative emitter—a UAV, its ground control link, or an unidentified transmitter—radiates in the presence of a jamming source; a monitoring receiver without synchronization or prior knowledge captures the waveform, and the RF front-end produces the complex baseband

frame $\mathbf{x} \in \mathbb{C}^{128}$. Three information sources are then formed—observable representations, physical source parameters, and semantic priors—and drive the Representamen, Object, and Interpretant branches, whose fused representation feeds the shared task heads for signal-parameter recognition and intent inference.

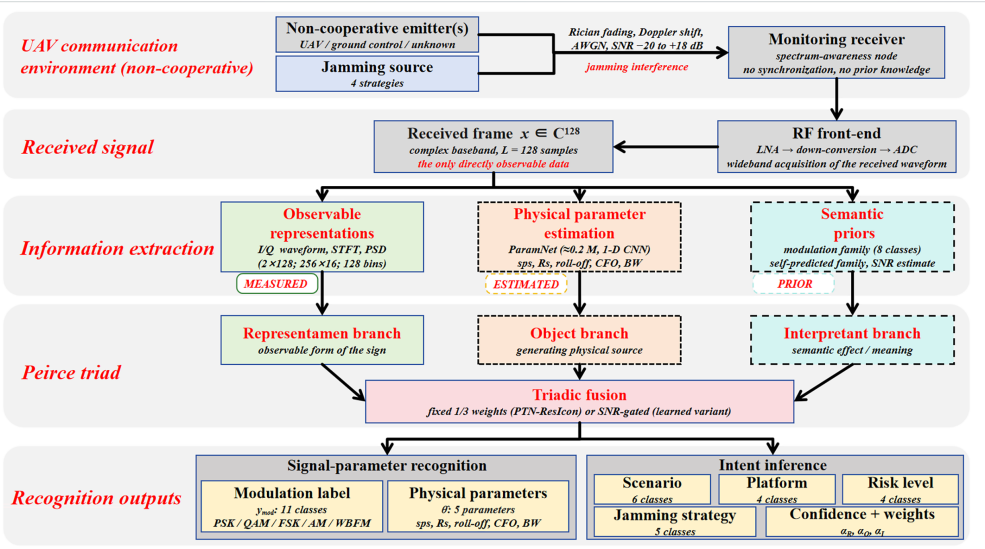


Figure 1. System-level model of non-cooperative signal recognition in UAV communications. From the received frame $\mathbf{x} \in \mathbb{C}^{128}$, three information sources—observable representations (solid borders), estimated physical parameters (dashed borders), and semantic/channel priors (dotted borders)—feed the three Peircean branches, whose fused output produces the recognition outputs.

The three sources differ in provenance, which is central to non-cooperative operation. Observable representations are computed deterministically from $\mathbf{x}$ and require no annotation (solid borders). Physical parameters are estimated from $\mathbf{x}$ by the regression network (dashed borders). Semantic priors cannot be measured: the modulation-family index comes from the classifier's coarse prediction, and the SNR value is treated as a channel-state estimate (dotted borders). Section 6.11 evaluates how these estimates affect recognition accuracy and reports the annotated setting as an upper bound. Figure 2 details the branch internals summarized here.

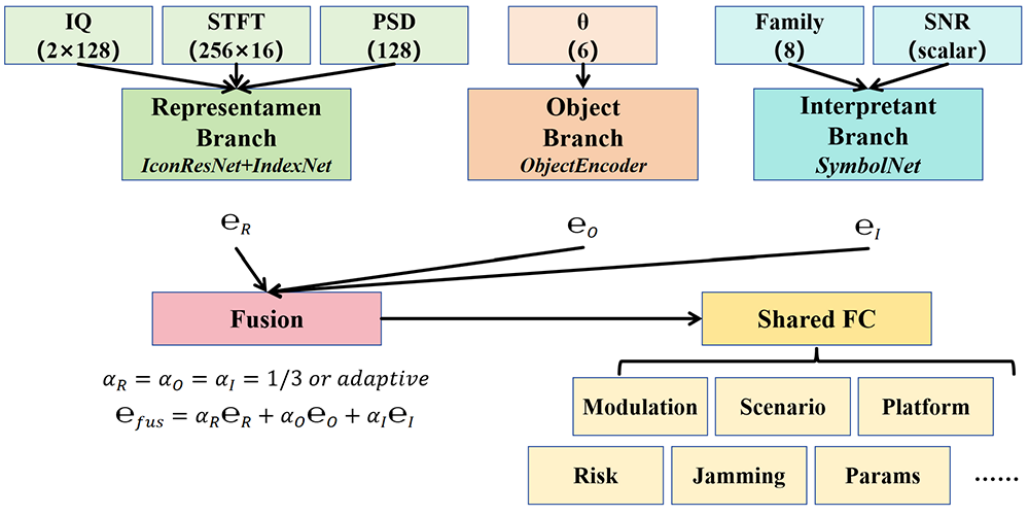


Figure 2. PeirceTriadicNet (PTN-ResIcon) architecture with three Peircean branches and shared task heads.

3.2. Peircean Triadic Analogy

Peirce defined the sign process as a triadic relation: Representamen → Object → Interpretant. Table 1 lists the analogy of this triad to signal recognition. The mapping is analogical: the Peircean elements are realized as neural branches, with Representamen instantiated by the union of Icon and Index signal representations, Object by the physical-parameter embedding, and Interpretant by the Symbol branch and task heads.

3.3. Input–Output Mapping to the Peircean Triad

To make the computational role of each Peircean element concrete, we state the inputs, the triadic decomposition, and the outputs of PTN.

Inputs. As summarized by Stage 3 of Figure 1, the raw input to the network is a received complex baseband frame $\mathbf{x} \in \mathbb{C}^L$ ($L = 128$) together with three groups of side information: (1) *observable signal representations* derived from $\mathbf{x}$ by standard DSP, i.e., the IQ waveform, the short-time Fourier transform (STFT) magnitude, and the power spectral density (PSD); (2) *estimated physical parameters* (samples per symbol, symbol rate, CFO, roll-off, bandwidth, and SNR), which are taken during training from the annotated dataset and at test time are estimated from $\mathbf{x}$ by a lightweight parameter-estimation network (Section 6.11); (3) *semantic priors*, namely the modulation-family index—at test time derived from the network's own coarse prediction (Section 6.11)—and an SNR scalar treated as a channel-state estimate. The estimation procedure, as well as an end-to-end evaluation in which no ground-truth modulation information, no oracle physical parameters, and no target-derived semantic labels are available, are given in Section 6.11.

Triadic decomposition. These inputs are mapped to Peirce's triad as follows:

- *Representamen*—the observable form of the sign—is instantiated by the Icon encoder (IQ waveform) and the Index encoder (STFT magnitude and PSD). Together, they form the Representamen branch and encode what is directly "perceivable" from the signal.
- *Object*—what the sign stands for—is instantiated by the ObjectEncoder, which takes the physical source parameters (modulation type, symbol rate, CFO, roll-off, bandwidth, SNR) as input. They are the material cause that generates the received waveform.
- *Interpretant*—the effect produced in the receiver—is instantiated by the SymbolNet (modulation-family and SNR priors) and the subsequent task heads. It corresponds to the semantic/decisional consequence of recognizing the signal: platform type, risk level, jamming strategy, and confidence.

The three neural branches therefore realize the Peircean triad in a computational analogy: the Representamen branch covers observable signal forms, the Object branch covers the generating physical parameters, and the Interpretant branch covers semantic reasoning and decisions. The SNR-biased fusion combines the three branch embeddings.

Outputs. The network outputs a tuple

$$(y_{\text{mod}}, y_{\text{sce}}, y_{\text{plat}}, y_{\text{risk}}, y_{\text{jam}}, \hat{\boldsymbol{\theta}}, \boldsymbol{\alpha}), \tag{1}$$

where y_{mod} is the estimated modulation label, y_{sce}, y_{plat}, y_{risk}, and y_{jam} are semantic/intent labels (scenario, platform, risk, jamming strategy), $\hat{\boldsymbol{\theta}}$ is the vector of estimated physical parameters, and $\boldsymbol{\alpha}$ is the SNR-biased triadic attention weight vector. Thus, PTN performs both *signal-parameter recognition* (modulation and physical parameters) and *intent inference* (scenario/platform/risk/jamming) within one triadic framework.

3.4. Problem Formulation

Let the received complex baseband frame be $\mathbf{x} \in \mathbb{C}^L$, with $L = 128$. The goal is to learn the mapping

$$f : \mathbf{x} \mapsto \left(y_{\text{mod}}, y_{\text{sce}}, y_{\text{plat}}, y_{\text{risk}}, y_{\text{jam}}, \hat{\boldsymbol{\theta}}, \boldsymbol{\alpha}\right). \tag{2}$$

where $y_{\cdot}$ are discrete labels, $\hat{\boldsymbol{\theta}}$ is a continuous parameter-estimation vector, and $\boldsymbol{\alpha} = (\alpha_R, \alpha_O, \alpha_I)$ are the learned triadic attention weights.

3.5. Justification for Multi-Branch Fusion

The following proposition serves as a sanity check that fusion is, under ideal conditions, no worse than a single branch; it is not intended as a deep theoretical result.

Proposition 1 (**Justification for multi-branch fusion**). *Let $P_e(B_j)$ denote the Bayes-optimal classification error using only branch $B_j \in \{Representamen, Object, Interpretant\}$. Then, the Bayes-optimal error of the fused classifier satisfies*

$$P_e^{Bayes}(B_R, B_O, B_I) \leq \min_{j \in \{R,O,I\}} P_e(B_j). \tag{3}$$

Proof. Let $j^* = \arg\min_j P_e(B_j)$. A fusion rule that ignores the other branches and uses only B_{j^*} achieves error $P_e(B_{j^*})$. The Bayes-optimal fusion rule minimizes the error over all measurable fusion rules, so its error cannot exceed that of the single-branch rule. □

This proposition guarantees that, given a sufficiently rich hypothesis space and adequate optimization, the fusion network is in principle no worse than the best single branch. Actual performance depends on optimization, data, and architecture.

3.6. Attention as an SNR-Gating Interpretant

Let s denote SNR. Ideally, the attention for branch B_j should increase with the conditional mutual information $I(Y; B_j \mid S = s)$. Using $H(Y|B_j, S = s) = H(Y|S = s) - I(Y; B_j|S = s)$, a temperature-scaled softmax over the negative conditional entropy yields

$$\alpha_j(s) \propto \exp\left[\beta\, I(Y; B_j \mid S = s)\right]. \tag{4}$$

At high SNR, Representamen details are informative, so α_R increases; at low SNR, waveform details are submerged in noise, and the network relies more on the Object parameters and prior knowledge from the Interpretant branch. In practice, PTN approximates this gating with a multilayer perceptron (MLP) that takes normalized SNR and per-branch confidences as inputs.

3.7. Conceptual Versus Algorithmic Novelty

It is useful to separate the conceptual contribution of the Peircean formulation from the algorithmic operations that instantiate it. Conceptually, the triad prescribes *which* information sources should be combined—the observable waveform, the generating physical parameters, and semantic priors—and *why* they are complementary: as one source degrades (e.g., waveform details at low SNR), the others still constrain recognition, and the fixed equal-weight fusion imposes a robust inductive bias rather than trusting a learned weighting that may overfit. Algorithmically, the corresponding operations are the multi-view Representamen encoder (a residual 1D-CNN on IQ plus 2D/1D-CNNs on STFT and PSD), the Object encoder with its parameter-regression head, the SymbolNet Interpretant embedding, the fixed equal-weight or SNR-gated fusion, and the shared multi-task heads. The standard building blocks (residual convolutions, 2D convolutions, MLPs, and softmax attention) are used as generic components; the novelty lies in how the triadic decomposition organizes

these components and in the equal-weight fusion rule derived from the Peircean argument, not in any single module.

4. PTN Network Architecture

The full PTN-ResIcon architecture is illustrated in Figure 2. Whereas Figure 1 specifies the system-level information flow and the provenance of each source, Figure 2 details the internal architecture inside the three branches and the shared task heads.

4.1. Three Peircean Branches

- *Representamen branch.* This branch aggregates the Icon and Index signal representations. IconResNet is a ResNet-style 1D-CNN operating on two-channel IQ waveforms; IndexNet is a 2D-CNN operating on STFT magnitude and a 1D-CNN operating on PSD. Their features are concatenated and projected to a common embedding space, yielding the Representamen embedding $\mathbf{e}_R$.
- extitObject branch (ObjectEncoder). A fully connected encoder that takes normalized physical parameters (samples per symbol, symbol rate, CFO, roll-off, bandwidth, and SNR) and maps them to an Object embedding $\mathbf{e}_O$. During training, the input parameters are perturbed by Gaussian noise to prevent label leakage. At test time, when ground-truth annotations are not available in a non-cooperative setting, the branch receives parameters estimated from the received IQ frame by the estimator described in Section 6.11; the idealized setting with ground-truth parameters is additionally reported there as an upper bound.
- *Interpretant branch (SymbolNet).* An embedding network that encodes modulation-family indices and an SNR scalar into the Interpretant embedding $\mathbf{e}_I$.

4.2. SNR-Biased Learnable Attention

For the final model PTN-ResIcon, the three branch embeddings are fused with fixed equal weights:

$$\boldsymbol{\alpha} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \tag{5}$$

$$\mathbf{e}_{\text{fus}} = \alpha_R \mathbf{e}_R + \alpha_O \mathbf{e}_O + \alpha_I \mathbf{e}_I. \tag{6}$$

This fixed fusion avoids over-reliance on an oracle SNR while preserving triadic complementarity.

To answer whether SNR-dynamic weighting can improve performance, we also train a learnable-attention variant, **PTN-ResIcon-Learned**, in which the triadic weights are conditioned on the instantaneous SNR. For each branch $j \in \{R, O, I\}$, let $\mathbf{z}_j$ be the classification logits produced by that branch. The branch confidence is defined as

$$c_j = \max_k \text{softmax}(\mathbf{z}_j / T)_k, \tag{7}$$

where T is a temperature parameter. The attention MLP takes $[s/20, c_R, c_O, c_I]$ as input, where s is the oracle SNR in dB. To prevent the attention from averaging over branches at high SNR and diluting the Representamen branch, an SNR-biased prior is added to the attention logits:

$$\boldsymbol{\ell}_{\text{prior}}(s) = [s/20,\ 0,\ -s/20], \tag{8}$$

where the three entries correspond to Representamen, Object, and Interpretant, respectively. The prior favors Interpretant at low SNR and Representamen at high SNR, while keeping Object neutral. A learnable scalar γ controls its strength:

$$\boldsymbol{\alpha} = \text{softmax}\Big(\text{MLP}([s/20, c_R, c_O, c_I]) + \gamma\, \boldsymbol{\ell}_{\text{prior}}(s)\Big). \tag{9}$$

4.3. Task Heads and Loss Function

The fused embedding $\mathbf{e}_{\text{fus}}$ first passes through two shared fully connected layers (ReLU, Dropout) to obtain a high-level, task-agnostic representation, which is then distributed to task-specific heads:

- *Classification heads*: modulation recognition (11 classes), scenario recognition (6 classes), platform type (4 classes), risk level (4 classes), and jamming strategy (5 classes);
- *Parameter-regression head*: estimates the five continuous parameters—samples per symbol, symbol rate in Hz, roll-off factor, CFO in Hz, and bandwidth in Hz.

Training all tasks jointly raises two issues:

1. **Gradient-scale mismatch.** The regression MSE is typically much larger in magnitude than the classification cross-entropy losses; without re-weighting, the shared representation would be dominated by the regression objective.
2. **Class imbalance across semantic tasks.** Some tasks have few samples per class in each mini-batch, so their heads may underfit.

The following measures address these issues:

- *Loss weighting:* the parameter-regression loss is multiplied by $\lambda = 0.2$ so its magnitude is comparable to the classification losses;
- *Independent task heads:* each classification task and the regression task uses its own fully connected output layer, allowing task-specific fine-tuning;
- *End-to-end joint training:* the shared layers and all task heads are updated simultaneously without freezing any branch, so the three-branch embedding trades off among classification and regression objectives.

The total loss is

$$\mathcal{L} = \mathcal{L}_{\text{mod}} + \mathcal{L}_{\text{sce}} + \mathcal{L}_{\text{plat}} + \mathcal{L}_{\text{risk}} + \mathcal{L}_{\text{jam}} + \lambda \mathcal{L}_{\text{params}}, \tag{10}$$

where $\lambda = 0.2$, each $\mathcal{L}_{\cdot}$ denotes the cross-entropy loss of the corresponding classification task, and $\mathcal{L}_{\text{params}}$ is the mean-squared-error loss for parameter regression.

4.4. Complexity

Table 2 lists the parameter counts of the main PTN-ResIcon modules. The total parameter count is 5.69 M, of which the Representamen branch accounts for 4.95 M (IconResNet 4.77 M plus IndexNet 0.18 M), the ObjectEncoder for 0.06 M, the Interpretant branch (SymbolNet) for 6.2 K, and the fusion MLP plus task heads for 0.68 M. Compared with the single-stream IconResNet (4.77 M), PTN-ResIcon increases the parameter count by about 19.2%, mainly because the shared task heads operate on the 256-dimensional fused embedding.

The average per-batch inference time for a batch size of 256 is about **2.82 ms** for IconResNet and **5.97 ms** for PTN-ResIcon, i.e., a latency increase of about **112%** (measured on an NVIDIA GeForce RTX 5070 Ti). At the per-sample level, PTN-ResIcon requires **113.19 M multiply–accumulate operations (MACs)** (226.38 M FLOPs) and runs at **1.92 ms per frame** on the same GPU, which corresponds to a throughput of about 520 frames per second and leaves a large margin for real-time spectrum awareness on a UAV platform (typical uplink control frames are produced at most every few milliseconds). In the current implementation, the three branches execute sequentially; because IndexNet, SymbolNet, and ObjectEncoder are much smaller than IconResNet, data-level parallelism is expected to reduce the extra latency to 20%–30% of IconResNet. Memory-wise, training with batch size 256 requires storing intermediate activations for all three branches simultaneously, so peak GPU memory is about 40%–50% higher than that of IconResNet.

Table 2. Parameter counts of PTN-ResIcon modules.

Module	Parameters	Fraction
IconResNet	4.77 M	83.8%
IndexNet	0.18 M	3.1%
ObjectEncoder	0.06 M	1.0%
SymbolNet	6.2 K	0.1%
Fusion + task heads	0.68 M	12.0%
Total	**5.69 M**	**100%**

5. Dataset and Experimental Setup

5.1. Datasets

All experiments use synthetic extensions of the RML2016.10a framework [1], the de facto public benchmark for deep-learning AMC. The dataset names are unified in Table 3. RML-S-Full-Std is a UAV-oriented re-synthesis of RML2016.10a: the 11 modulation formats and the −20 to +18 dB SNR grid follow the original benchmark, while each frame additionally carries object-layer physical parameters and Interpretant semantic labels (scenario, platform, risk, jamming strategy) that are not present in the original dataset. The dataset is released as open data (see Data Availability) with the full manifest, split indices, and generation script so that every number reported in this paper is reproducible. In addition, a held-out test set, RML-S-Full-Std-test, contains 11,000 frames at +18 dB generated under the same distribution but never used for training or validation; it is used in Section 6.5 to assess generalization to a fresh sample of the same distribution, and it is released together with the main datasets (see Data Availability). The +18 dB level is chosen as the high-SNR reference of the benchmark so that the generalization assessment is not confounded by SNR degradation, whose effects are analyzed separately in Sections 6.1 and 6.12.

Table 3. Dataset configurations.

Name	Samples	Modulations	SNR Range	Purpose
RML-S-Full-Std	220,000	11	−20 to +18 dB	Main modulation classification and regression
RML-S-Full-Std-test	11,000	11	+18 dB	Held-out generalization (Section 6.5)
RML-S-Full-Scene	300,000	15	−20 to +18 dB	Scenario distortion and transfer (described only in this paper)

RML-S-Full-Std contains the 11 modulations of RML2016.10a: BPSK, QPSK, 8PSK, QAM16, QAM64, PAM4, CPFSK, GFSK, AM-DSB, AM-SSB, and WBFM. Table 4 summarizes their roles in UAV communications.

Table 4. Modulation formats of RML-S-Full-Std and their relevance to UAV communications.

Modulation	Role in UAV Communications
BPSK	Robust low-rate telemetry and control; preamble and synchronization
QPSK	Digital control links and downlink signaling
8PSK	Bandwidth-efficient control and data links
QAM16	OFDM-based video downlinks and high-rate data links
QAM64	OFDM-based video downlinks with higher spectral efficiency
PAM4	Short-range baseband data and sensor links
CPFSK	Frequency-hopping control links of drone remote controllers
GFSK	Bluetooth-class and low-cost control links
AM-DSB	Analog voice and relay in low-altitude operations
AM-SSB	Analog voice with reduced occupied bandwidth
WBFM	Analog voice broadcast and relay

This selection is retained for two reasons. First, it constitutes the most widely used public AMC benchmark, which keeps our results directly comparable with the large body of published work [1,2,4,5,11]. Second, it covers the signal families most relevant to UAV links: PSK/QAM formats underlie OFDM-based video downlinks and digital data links, FSK/CPFSK/GFSK formats typify the frequency-hopping control links of commercial drone remote controllers, and AM/WBFM cover analog voice and relay scenarios in low-altitude operations. Indeed, the same families—frequency-hopping FSK control signals and OFDM-based video signals in the 915 MHz, 2.4 GHz, and 5.8 GHz ISM bands—are observed in real drone RF datasets [22]. Each sample contains an IQ waveform (length 128), STFT (256-point FFT, hop 8), PSD (128 points), object-layer parameters (samples per symbol, symbol rate in Hz, roll-off, CFO in Hz, bandwidth in Hz), and Interpretant labels (scenario, platform type, risk level, jamming strategy). Example parameter distributions under different scenarios are shown in Figure 3; the clusters confirm that scenario context induces distinct parameter configurations that the Object branch can exploit.

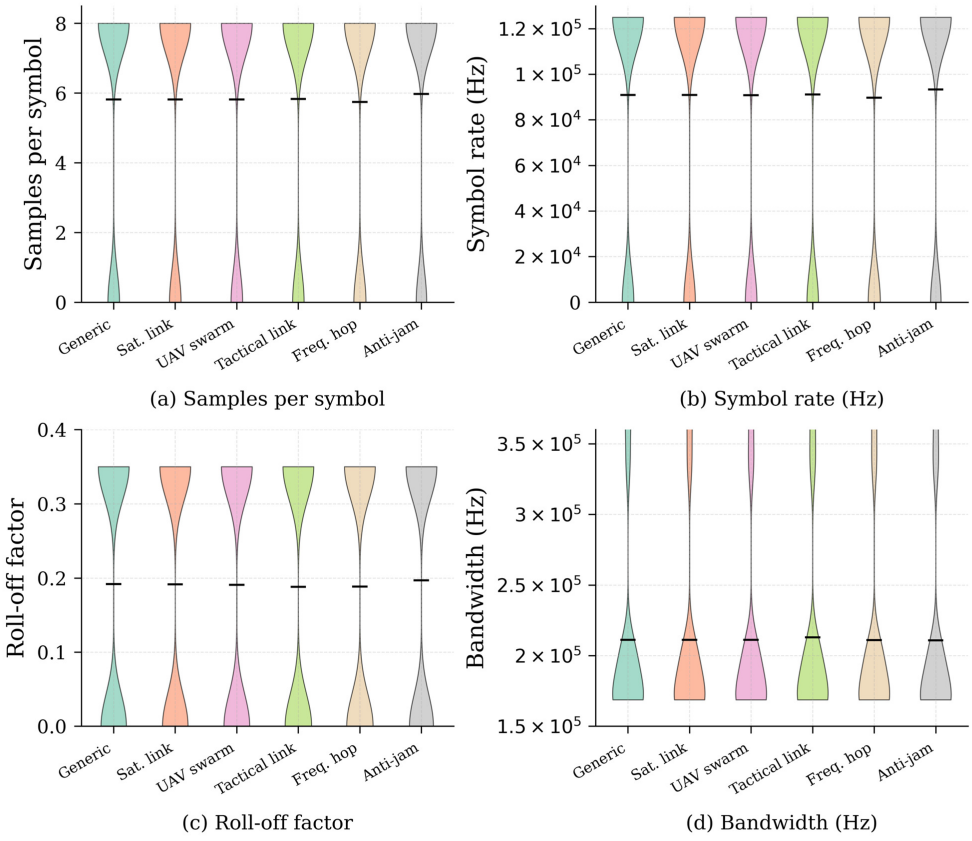


Figure 3. Example parameter distributions under different scenarios in RML-S-Full-Std: (**a**) samples per symbol; (**b**) symbol rate; (**c**) roll-off factor; (**d**) bandwidth.

UAV channel assumptions. Because the primary target of this work is UAV communications, the synthetic channel model deliberately reflects representative UAV link conditions. The baseline channel follows the RML2016.10a model (AWGN over a flat-fading Rayleigh channel) so that results remain comparable with the AMC literature; on top of it, the scenario layer applies UAV-specific distortions: Rician fading with K-factor 5 dB and Doppler shift drawn uniformly from ± 500 Hz to emulate platform mobility for UAV-to-ground links; co-channel emitters with random arrival times in the uav_swarm scenario to model multi-UAV spectrum sharing; and barrage, partial-band, follower, and deceptive jamming to model intentional interference in contested spectrum. These assumptions are consistent

with the channel conditions reported for real drone RF links [22] and are documented in the released generation script.

5.2. Training, Validation, and Test Splitting

All datasets are stratified by (modulation, SNR) pairs into 70%/15%/15% training/validation/test splits, ensuring that every modulation–SNR combination is represented in all three folds. The random seed is set to 42.

5.3. Reproducibility

Models are trained with the Adam optimizer (learning rate 10^{-3}, weight decay 10^{-4}), a cosine-annealing learning-rate schedule over 50 epochs, and a batch size of 256. Cross-entropy loss is used for classification tasks. The software environment is Python 3.13.5 and PyTorch 2.12.0+cu130. Code, data manifests, and figures are available in the repository; this includes the IQ-based parameter estimator (ParamNet) and the channel-variant generation pipeline used in Sections 6.11 and 6.12, so that every number reported in this paper is reproducible. The DroneRFa transfer experiments reproduction code, is provided as Supplementary Materials (File S1).

5.4. Baselines and Ablations

- Single-branch baselines: Representamen (IconResNet + IndexNet), Object (ObjectEncoder), and Interpretant (SymbolNet).
- Widely cited comparison baselines: CNN2 [1], CNN-LSTM [4], ResNet1D [6].
- Fusion ablations: PTN-ResIcon (fixed equal weights $\alpha_R = \alpha_O = \alpha_I = 1/3$, final model); PTN-ResIcon-Learned (SNR-biased learnable attention).
- Parameter-regression baseline: per-modulation mean predictor.

6. Experimental Results

6.1. Modulation Recognition on RML-S-Full-Std

Table 5 reports the overall and +18 dB accuracies. PTN-ResIcon achieves an overall accuracy of 80.67%, which is 17.6 percentage points higher than the Representamen branch alone and 24.0, 20.1, and 20.2 percentage points higher than CNN2, CNN-LSTM, and ResNet1D, respectively. The gain is concentrated in the low-to-moderate SNR range, where the Object and Interpretant priors compensate for corrupted waveform details.

Table 5. Modulation-recognition accuracy (%) on RML-S-Full-Std.

Method	Overall Accuracy	+18 dB	Type
Representamen	63.02	95.63	Single branch
Object	36.28	36.02	Single branch
Interpretant	54.29	53.57	Single branch
CNN2 [1]	56.72	89.35	Comparison baseline
CNN-LSTM [4]	60.60	92.30	Comparison baseline
ResNet1D [6]	60.45	97.91	Comparison baseline
PTN-ResIcon-Learned	76.86	82.64	Ablation
PTN-ResIcon	**80.67**	**91.50**	**Proposed**

At +18 dB, ResNet1D reaches 97.91% and CNN-LSTM reaches 92.30%, indicating that the dataset is highly separable at high SNR. PTN-ResIcon achieves 91.50% at +18 dB, which is lower than ResNet1D because the fixed equal-weight fusion dilutes the strong Representamen branch when waveform details are already reliable; this motivates future work on sharper high-SNR gating. The overall gain over CNN2, CNN-LSTM, and ResNet1D

is 24.0, 20.1, and 20.2 percentage points, respectively, and is concentrated in the low-to-moderate SNR range.

Table 6 reports accuracies at selected SNR points. PTN-ResIcon retains 64.67% at −20 dB, whereas CNN2, CNN-LSTM, and ResNet1D are all below 11%. It is important to attribute this low-SNR behavior correctly. At −20 dB, the Representamen branch alone—which operates only on the received IQ waveform, exactly like the baselines—achieves 10.1%, i.e., the same level as CNN2 (9.7%), CNN-LSTM (10.6%), and ResNet1D (9.9%); the architecture itself does not extract low-SNR information from the waveform that the baselines miss. The 64.67% figure instead comes from the Object and Interpretant branches, which consume side information (physical parameters and modulation-family priors) that is nearly SNR-insensitive: they provide floors of about 36% and 54% across the whole SNR range, and their fusion with the Representamen branch produces the large low-SNR gain. This is precisely the intended triadic behavior—recognition is driven by auxiliary knowledge when the waveform is uninformative—but it also means that the reported low-SNR numbers assume this side information is available. Section 6.11 evaluates the fully autonomous setting and shows how the low-SNR advantage is then affected. The complete accuracy-versus-SNR curves of all models are shown in Figure 4.

Table 6. Accuracy (%) at selected SNR points on RML-S-Full-Std.

Method	−20 dB	0 dB	+18 dB
CNN2	9.66	72.16	89.35
CNN-LSTM	10.57	77.37	92.30
ResNet1D	9.90	78.86	97.91
Representamen	10.14	82.52	95.63
Object	38.29	35.21	36.02
Interpretant	55.74	54.00	53.57
PTN-ResIcon-Learned	64.49	80.66	82.64
PTN-ResIcon	**64.67**	**84.19**	**91.50**

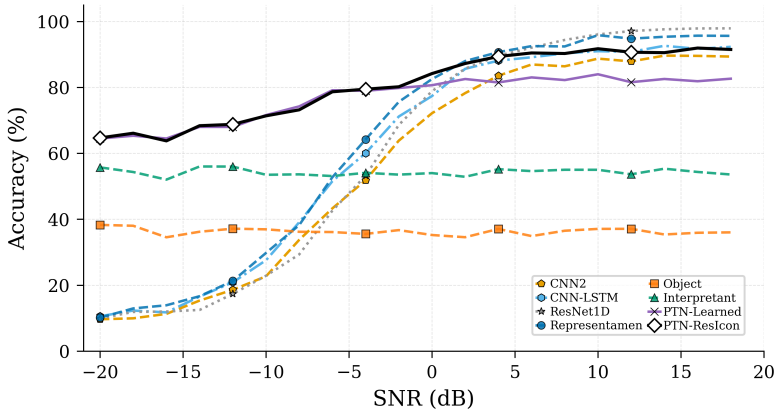


Figure 4. Modulation-recognition accuracy versus SNR on RML-S-Full-Std.

6.2. Multi-Task Learning on RML-S-Full-Std

Table 7 reports test accuracies for the five semantic tasks on RML-S-Full-Std. PTN-ResIcon outperforms the single-branch baselines on modulation recognition while maintaining competitive performance on scenario, platform, risk, and jamming-strategy tasks. The Interpretant branch achieves the highest single-branch modulation accuracy (54.29%) because the modulation-family prior is strongly correlated with modulation type, but it performs poorly on semantic tasks that depend on waveform details. The Object branch

provides strong parameter estimates and modest classification performance. PTN-ResIcon achieves a balanced trade-off among the multiple objectives.

Table 7. Multi-task test accuracy (%) on RML-S-Full-Std.

Method	Modulation	Scenario	Platform	Risk	Jamming Strategy
Representamen	63.02	22.14	31.89	43.10	31.76
Object	36.33	16.61	32.98	38.46	32.98
Interpretant	54.29	16.58	32.98	38.46	32.98
PTN-ResIcon-Learned	76.86	22.40	32.98	43.41	32.98
PTN-ResIcon	**80.67**	**22.02**	**32.98**	**43.42**	**32.98**

To evaluate the semantic tasks more rigorously, Table 8 reports macro-F1, balanced accuracy, and the majority-class baseline in addition to raw accuracy. On the scenario task, PTN-ResIcon reaches 22.02% accuracy with a macro-F1 of 0.2113 and balanced accuracy of 0.2203, all above the majority-class baseline (17.13%), indicating a small but real signal. On platform and jamming tasks, however, both the accuracy (32.98%) and the majority-class baseline coincide (32.98%) and macro-F1 is only 0.0992, showing that these two auxiliary tasks are dominated by the majority class in the current corpus and do not provide discriminative supervision yet; we therefore treat them as exploratory auxiliary tasks rather than as primary evaluation targets, and we emphasize modulation recognition, parameter regression, and scenario awareness as the main results. The risk task shows the strongest relative signal (43.42% accuracy, macro-F1 0.3044, balanced accuracy 0.3974 vs. a majority baseline of 38.46%). These auxiliary-task numbers should be interpreted with the class-imbalance caveat in mind; enriching the semantic annotation balance is part of the dataset release's future work. The same multi-task results are visualized as a grouped bar chart in Figure 5.

Table 8. Rigorous semantic-task metrics of PTN-ResIcon on RML-S-Full-Std.

Task	Accuracy (%)	Macro-F1	Balanced Acc.	Majority-Class (%)
Scenario	22.02	0.2113	0.2203	17.13
Platform	32.98	0.0992	0.2000	32.98
Risk	43.42	0.3044	0.3974	38.46
Jamming strategy	32.98	0.0992	0.2000	32.98

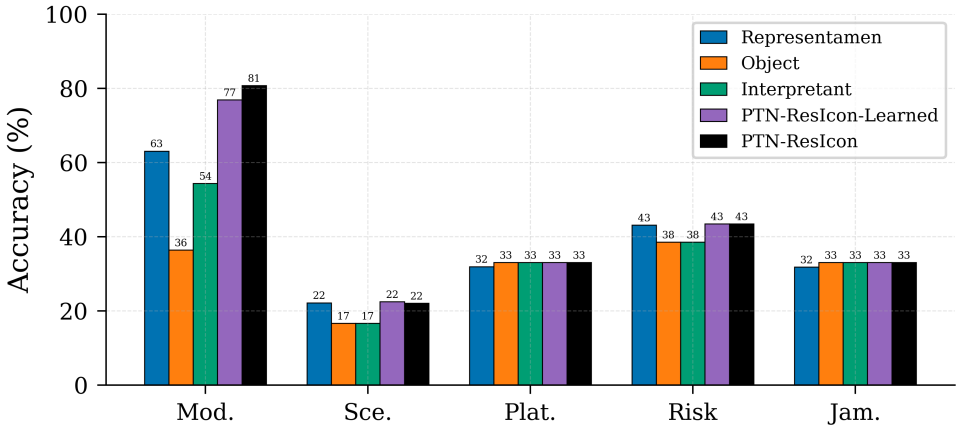


Figure 5. Multi-task accuracy comparison on RML-S-Full-Std.

6.3. Attention Analysis

Figure 6 shows the average attention weights learned by PTN-ResIcon-Learned on the test set. At low SNR, the Interpretant branch dominates because the Representamen branch is unreliable; as SNR increases, the Representamen weight rises while Interpretant falls; the Object branch contributes a smaller but steady share. This trend is consistent with the SNR-gating interpretation. However, the final model PTN-ResIcon uses fixed equal weights $(1/3, 1/3, 1/3)$; its overall accuracy (80.67%) is higher than that of PTN-ResIcon-Learned (76.86%). This indicates that the triadic decomposition provides a strong prior, while the learned attention in this variant is too aggressive and slightly degrades generalization.

The underperformance of the SNR-dynamic variant can be attributed to two factors. First, the attention MLP is trained on the oracle SNR and branch confidences from the training set; at test time, small distribution shifts in SNR or confidence can push the weights away from the robust fixed average. Second, the prior in Equation (8) already encodes the intended SNR gating, but the learnable component overfits to training-set confidence patterns, especially for modulations that are easily confused at intermediate SNR. Consequently, fixing the weights at 1/3 yields a more stable and more accurate classifier across the full SNR range.

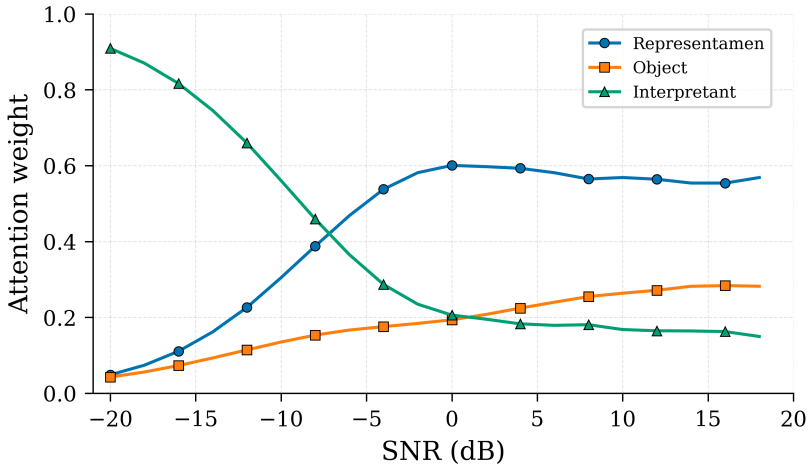


Figure 6. SNR-biased triadic attention: learned average weights of PTN-ResIcon-Learned on the test set.

More broadly, this result highlights a bias–variance trade-off in triadic fusion. Fixed equal weights impose a strong inductive bias: they force the model to rely on the complementary structure of the Peircean branches rather than on data-driven reliability estimates. In the low-SNR regime of RML-S-Full-Std, where training samples per (modulation, SNR) cell are limited and branch confidences are noisy, this strong bias stabilizes the decision boundary. Dynamic weighting offers lower bias, but the additional MLP parameters increase variance unless the reliability estimator is well-calibrated or heavily regularized. In other words, equal-weight fusion is not inherently superior to SNR-adaptive fusion; rather, it is the more robust choice when the dynamic mechanism has insufficient reliable supervision to learn a generalizable weighting policy.

Figure 7 shows the confusion matrix of PTN-ResIcon on RML-S-Full-Std. The main confusions occur among similar modulations: BPSK–QPSK–8PSK, QAM16–QAM64, and CPFSK–GFSK. Modulations with distinctive waveform or spectral features, such as PAM4, AM-DSB, AM-SSB, and WBFM, achieve high recall.

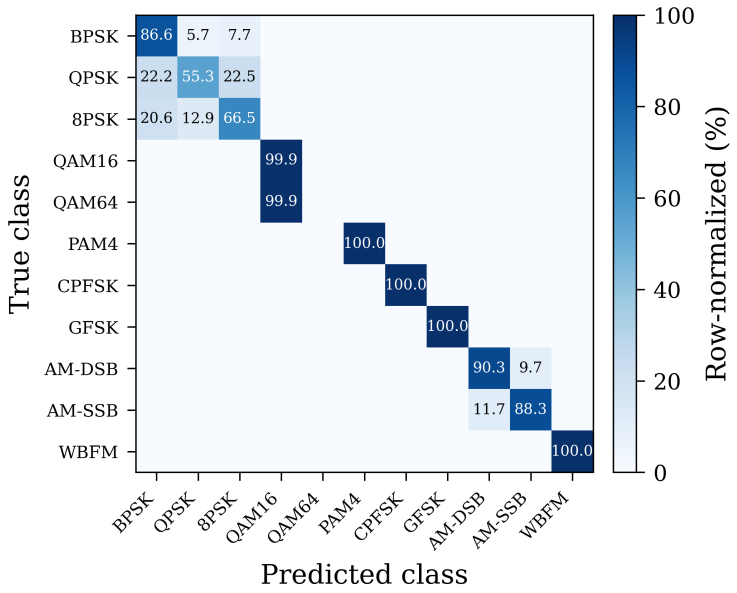


Figure 7. Confusion matrix of PTN-ResIcon on RML-S-Full-Std.

6.4. Scenario-Dependent Performance

To examine how operating context affects recognition, Table 9 reports the modulation-recognition accuracy of each model for every scenario in RML-S-Full-Std. The single-branch Object and Interpretant models provide SNR-insensitive floors of about 36% and 54%, respectively. PTN-ResIcon and PTN-ResIcon-Learned substantially exceed these floors, with PTN-ResIcon reaching 80.67% overall. The largest drop occurs in cognitive_frequency_hopping (78.18% for PTN-ResIcon), where rapid carrier switching disrupts the Representamen branch; the Object and Interpretant branches still preserve a reasonable floor. For the uav_swarm scenario, PTN-ResIcon reaches 81.44%, the second-highest accuracy among the six scenarios, confirming that triadic fusion remains effective when multiple UAV emitters share the spectrum and produce mixed signal sources. Notably, the satellite_link (81.64%) and anti_jam_evasive (81.05%) scenarios—which mirror the coordination with satellite networks and the secure-information-exchange requirements emphasized in emergency-response UAV communications—are handled with almost no loss relative to the generic scenario (80.93%), indicating that the triadic model preserves recognition accuracy precisely in the contested-spectrum conditions typical of emergency deployments. These scenario-dependent trends are visualized in Figure 8.

Table 9. Scenario-dependent modulation-recognition accuracy (%) on RML-S-Full-Std.

Scenario	Representamen	Object	Interpretant	PTN-ResIcon-Learned	PTN-ResIcon
generic	63.33	36.34	54.35	77.09	80.93
satellite_link	63.28	36.29	54.29	77.61	81.64
uav_swarm	62.94	36.44	54.30	77.25	81.44
tactical_data_link	62.86	36.76	54.27	76.75	80.73
cognitive_frequency_hopping	62.43	36.62	53.70	74.50	78.18
anti_jam_evasive	63.32	36.37	54.37	77.16	81.05
Overall	**63.02**	**36.28**	**54.29**	**76.86**	**80.67**

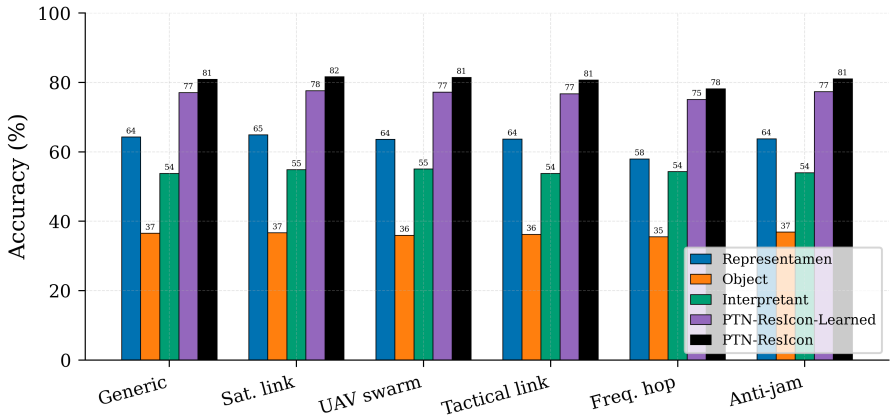


Figure 8. Scenario-dependent modulation-recognition accuracy across single-branch and fused models.

6.5. Generalization to Held-Out Test Set

A separate held-out test set, RML-S-Full-Std-test, contains 11,000 samples at +18 dB generated under the same distribution but held out from training and validation. Table 10 compares the held-out accuracies of all models. PTN-ResIcon achieves 90.76%, which is 4.67 percentage points below the Representamen-only baseline (95.43%) but far above the Object (36.63%) and Interpretant (54.55%) floors. The gap between PTN-ResIcon and Representamen at +18 dB is consistent with Table 6: the fixed equal-weight fusion dilutes the strong waveform branch when waveform details are already reliable. Figure 9 summarizes the held-out accuracies as a bar chart.

Table 10. Held-out test-set accuracy (%) at +18 dB.

Method	Accuracy
Representamen	95.43
Object	36.63
Interpretant	54.55
PTN-ResIcon-Learned	82.21
PTN-ResIcon	**90.76**

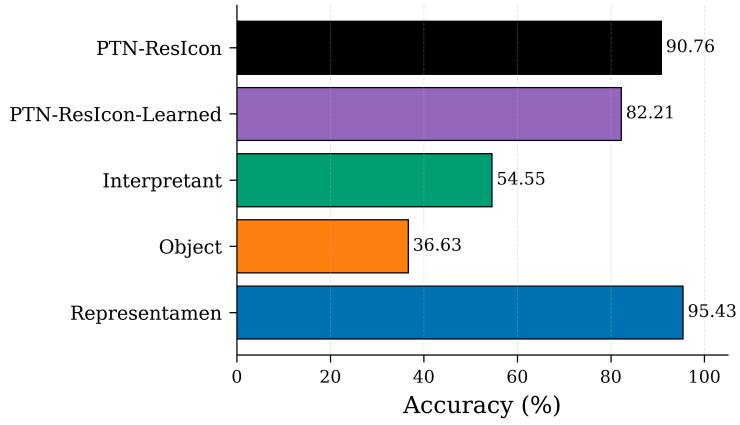


Figure 9. Held-out test-set accuracy comparison at +18 dB.

6.6. Out-of-Distribution Scenario Distortion

To test generalization beyond the training distribution, we evaluate PTN-ResIcon on RML-S-Full-Scene, a scenario-distortion corpus generated with the same simulator but

with stronger scenario layers (15 modulations, six scenarios, 300,000 frames). We consider two aspects. First, for the 11 modulations seen during training, accuracy drops from 80.67% on the in-distribution test set to 54.94% on the distorted corpus, with the largest degradation in the cognitive_frequency_hopping scenario (47.59%) and the smallest in the tactical_data_link scenario (58.75%); the model therefore degrades gracefully rather than collapsing, but the drop confirms that strong scenario distortions remain a challenge. Second, for the four modulations never seen during training (GMSK, OOK, FSK, OFDM), the model assigns an average maximum softmax confidence of 0.85, i.e., it is overconfident on unseen classes. This is an expected limitation of closed-set softmax classifiers and motivates future work on open-set recognition or calibrated uncertainty for unknown emitters; we report it here to give a complete picture of the model's behavior in the contested spectrum.

6.7. Transfer to Real Drone RF Signals with Full Triadic Supervision

To validate the framework on genuine UAV signals with per-frame physical annotations, we use **DroneRFa** [22], a large-scale corpus of software-defined-radio recordings of drone–controller communication (NI USRP-2955, dual channel, 100 MS/s) covering the three ISM bands (915 MHz, 2.4 GHz, 5.8 GHz) and 24 device classes plus background. Every segment carries the triadic annotation scheme of the dataset: device type (T), flight distance (D: near 20–40 m, medium 40–80 m, far 80–150 m), and segment code (S encoding the initial and switched ISM band). We select four commercial devices—DJI Phantom 3 (*T0001*), DJI Phantom 4 Pro (*T0010*), DJI MATRICE 200 (*T0011*), and DJI MATRICE 100 (*T0100*)—each recorded at the three distances, i.e., twelve ".mat" segments of 1.4×10^8 dual-channel complex samples each. Following the same preprocessing as for RML-S-Full-Std (activity-window selection, decimation to 1 MS/s, 128-point frame slicing with an energy gate, per-frame IQ/STFT/PSD extraction), we obtain 240,000 frames. We emphasize that the two corpora address *different recognition tasks at different levels*: RML-S-Full-Std targets *signal-level modulation recognition* (an intrinsic property of the emitted waveform, 11 classes), whereas DroneRFa targets *emitter (device) identification* (the identity of the transmitting platform, four classes), which must be inferred from the same RF observables. Evaluating the triadic framework on both tasks demonstrates that it is not tied to a single recognition granularity but generalizes across signal-level and device-level recognition.

Because DroneRFa provides per-frame physical parameters and semantic labels, the *full* triadic model can be exercised on real data. The Object branch receives the true physical parameters—normalized flight distance, segment code, and the per-frame received power $P_{\mathrm{dB}} = 10 \log_{10} \langle |x|^2 \rangle$, which is *measured directly* from each raw frame before normalization (it is *not* derived from the distance label by any propagation formula; the corpus provides no per-frame SNR annotations, so this directly measurable received power is the honest physical quantity). As a sanity check, the measured power follows the expected distance-decay trend for the DJI Phantom 4 Pro recordings: near/medium/far mean powers of about $-57/-60/-64$ dB, i.e., drops of roughly 3–4 dB per distance step that are consistent with free-space attenuation over 20–40 m, 40–80 m, and 80–150 m; no SNR value is mapped from distance. The Interpretant branch receives a semantic prior that is aligned with the target hierarchy: the four devices are grouped into two coarse application classes (aerial-imaging platforms: Phantom 3 and Phantom 4 Pro; industrial platforms: MATRICE 200 and MATRICE 100), mirroring the modulation-family prior used on RML-S-Full-Std; the shared fusion is fine-tuned to predict the fine-grained device type. Table 11 reports the four configurations (70/15/15 split, fine-tuned from the RML pre-trained PTN-ResIcon). Each branch contributes an independent, quantifiable gain: the Representamen branch alone reaches 33.69% (close to the 25% random-guess level, reflecting the synthetic-to-real domain gap); adding the Interpretant coarse-class prior raises accuracy to 64.86% (+31.2 points),

adding the Object branch raises it to 47.86% (+14.2 points), and the *full triadic combination is the best configuration* (72.50%, +38.8 points over the Representamen-only baseline). The complete triadic model thus outperforms every partial combination, and the gains are complementary: the Interpretant provides the semantic hierarchy while the Object provides genuine physical parameters of the recording, together with the Representamen features they drive recognition on real signals. This is, to the best of our knowledge, the first time the full three-branch triadic model is exercised on genuine UAV signals with per-frame physical annotations.

Table 11. Device-identification accuracy (%) on the DroneRFa real-drone corpus for the triadic-ablation configurations (70/15/15 split, fine-tuned from the RML pre-trained PTN-ResIcon). Random guess is 25%.

Configuration	Object Branch	Interpretant Branch	Test Acc. (%)
Representamen only	–	–	33.69
+Object	Yes	–	47.86
+Interpretant	–	Yes	64.86
Full triadic	**Yes**	**Yes**	**72.50**

6.8. SNR Robustness

PTN-ResIcon and PTN-ResIcon-Learned use the oracle SNR as an input to the Interpretant branch. To test robustness, Gaussian noise with standard deviations of 0, 1, 2, 3, and 5 dB is added to the oracle SNR at inference. Table 12 shows that both fused models are essentially unaffected (PTN-ResIcon stays at 80.67%–80.75%; PTN-ResIcon-Learned stays at 76.76%–76.92%). The single-branch models do not consume oracle SNR, so their accuracies are unchanged by this perturbation. The result confirms that moderate SNR-estimation errors have limited impact on the fusion models. Figure 10 plots the same robustness trend, confirming that PTN-ResIcon remains stable even under 5 dB SNR-estimation noise.

Table 12. Accuracy (%) under noisy SNR estimates on RML-S-Full-Std.

Method/Noise Std. (dB)	0	1	2	3	5
Representamen	63.02	63.02	63.02	63.02	63.02
Object	36.28	36.28	36.28	36.28	36.28
Interpretant	54.29	54.29	54.29	54.29	54.29
PTN-ResIcon-Learned	76.86	76.85	76.86	76.92	76.76
PTN-ResIcon	**80.67**	**80.68**	**80.75**	**80.69**	**80.71**

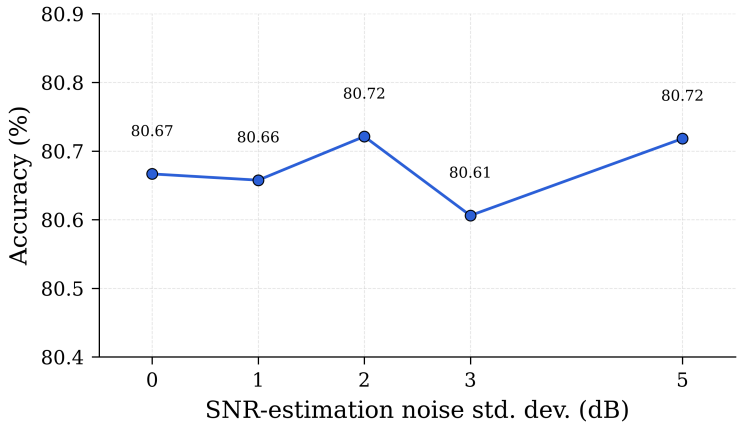


Figure 10. Robustness of PTN-ResIcon to SNR-estimation errors.

6.9. Parameter Regression

Table 13 reports the mean absolute error (MAE) for the five object-level parameters for the Object branch, PTN-ResIcon-Learned, PTN-ResIcon, and a per-modulation mean predictor (the class-conditional mean of each parameter, which is the strongest non-parametric baseline when a parameter is nearly constant within a modulation). Because sps, symbol_rate_hz, and roll-off are nearly fixed per modulation in the current dataset, the per-modulation mean baseline achieves near-zero MAE for those parameters and is therefore a very strong reference; the meaningful comparison is on CFO, where the per-modulation mean predictor is far worse (3635.9 Hz) than both the Object branch (343.3 Hz) and PTN-ResIcon (596.0 Hz), confirming that CFO carries within-modulation variability that only learned regression captures. The Object branch alone achieves the lowest CFO MAE because it is trained exclusively for parameter regression. PTN-ResIcon trades some CFO accuracy (596.0 Hz) for large gains in modulation recognition, while still keeping all normalized MAEs below 0.015. Figure 11 visualizes the regression MAE across the five parameters.

Table 13. Parameter-regression MAE on RML-S-Full-Std.

Method	Sps	Symbol_RATE (kHz)	Roll-Off	CFO (Hz)	Bandwidth (kHz)
Object branch	0.050	0.774	0.003	343.3	1.184
Per-modulation mean	0.000	0.001	0.000	3635.9	0.000
PTN-ResIcon-Learned	0.051	0.804	0.003	575.0	1.379
PTN-ResIcon	**0.038**	**0.600**	**0.002**	**596.0**	**0.887**

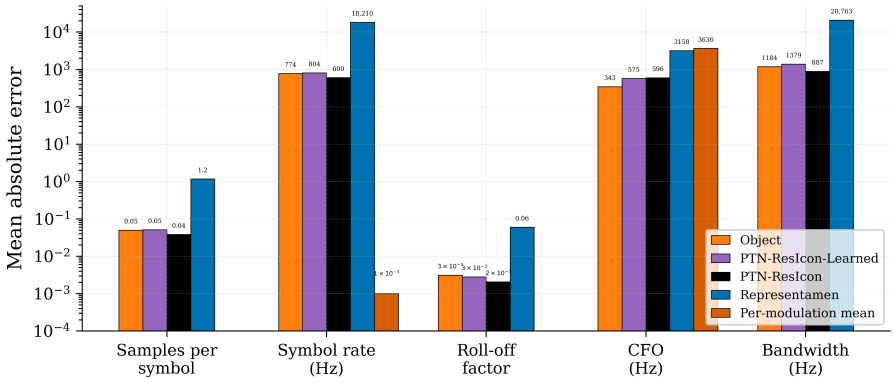


Figure 11. Parameter-regression MAE comparison across single-branch and fused models.

6.10. Ablation Summary

Table 14 and Figure 12 summarize the ablation and baseline results. The comparison baselines (CNN2, CNN-LSTM, ResNet1D) and the single Peircean branches verify the complementary contribution of the Representamen, Object, and Interpretant branches. Three additional controlled variants further localize the source of the triadic gain. (i) **PTN-ResIcon-NoFamily** removes the modulation-family prior from the Interpretant branch, which is the information most strongly correlated with the target modulation label; its accuracy drops from 80.67% to 70.28%, quantifying the contribution of the family prior while showing that the remaining fusion still clearly outperforms the best single branch (63.02%). (ii) **ConcatMLP-Fusion** replaces the triadic architecture by a conventional multi-modal baseline that simply concatenates waveform, time–frequency, and spectral features with the physical parameters and feeds them to an MLP; it reaches 70.91%, i.e., about ten points below PTN-ResIcon, demonstrating that the gain is not explained by mere feature

concatenation. (iii) **AttnFusion-Triadic** replaces the fixed equal weights by a learned self-attention over the three branch embeddings; it reaches 81.37%, statistically comparable to the fixed equal-weight fusion and notably better at −20 dB (65.88% vs. 64.67%). Taken together, these ablations show that the performance advantage comes jointly from the multi-source information and from the triadic decomposition, while the SNR-gated learned attention of PTN-ResIcon-Learned (76.86%) is the only variant that hurts accuracy, confirming that aggressive SNR conditioning degrades generalization. The Object branch alone achieves 36.28% modulation accuracy, showing that physical parameters carry useful but incomplete information; the Interpretant branch alone achieves 54.29%, reflecting the strength of the modulation-family prior. The large gap between the best single branch (Representamen, 63.02%) and PTN-ResIcon (80.67%) demonstrates the gain of triadic fusion.

Table 14. Ablation and baseline comparison summary on RML-S-Full-Std.

Configuration	Overall Accuracy (%)	−20 dB (%)	+18 dB (%)
CNN2	56.72	9.66	89.35
CNN-LSTM	60.60	10.57	92.30
ResNet1D	60.45	9.90	97.91
Representamen	63.02	10.14	95.63
Object	36.28	38.29	36.02
Interpretant	54.29	55.74	53.57
PTN-ResIcon-NoFamily	70.28	38.77	90.83
ConcatMLP-Fusion	70.91	37.92	91.44
AttnFusion-Triadic	81.37	65.88	90.39
PTN-ResIcon-Learned	76.86	64.49	82.64
PTN-ResIcon	**80.67**	**64.67**	**91.50**

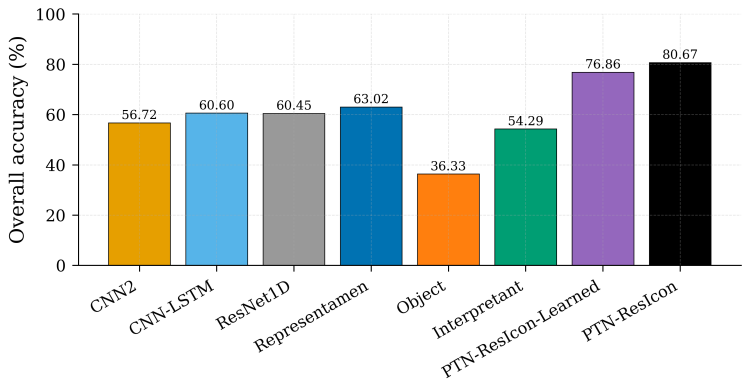


Figure 12. Overall modulation-recognition accuracy on RML-S-Full-Std: comparison of representative baselines (CNN2, CNN-LSTM, ResNet1D), single Peircean branches (Representamen, Object, Interpretant), and proposed fusion variants (PTN-ResIcon and PTN-ResIcon-Learned).

6.11. End-to-End Evaluation with Independently Estimated Side Information

The experiments above assume that the Object and Interpretant branches receive side information (physical parameters and the modulation-family prior) from the dataset annotations. In a non-cooperative setting, such annotations are unavailable, so this subsection evaluates the fully autonomous pipeline in which every input is derived only from the received IQ frame. Three sources are estimated: (i) the physical parameters are predicted by a lightweight 1D-CNN estimator (ParamNet, about 0.2 M parameters) trained on the same training split to regress the five object parameters from the IQ waveform; (ii) the modulation-family prior is derived from the argmax of the signal-only Representamen classifier (the best single branch); (iii) the SNR scalar is treated as a channel-state estimate, as

is standard in the AMC literature. Table 15 compares five inference configurations ordered by the amount of side information available: configurations A and B use ground-truth or encoder-refined parameters with the ground-truth family prior (the settings reported in the earlier sections); C uses IQ-estimated parameters with the ground-truth family; E uses ground-truth parameters with the predicted family; and D is the fully autonomous configuration with IQ-estimated parameters and the predicted family.

Table 15. Modulation-recognition accuracy of PTN-ResIcon and the robustness-adapted PTN-ResIcon-Est under different side-information configurations at inference (RML-S-Full-Std test set).

Model	Configuration	Parameters	Family	Overall (%)	−20 dB (%)
PTN-ResIcon	A/B: oracle or refined	Ground truth	Ground truth	80.67	64.67
PTN-ResIcon	C: estimated parameters	IQ-estimated	Ground truth	70.49	55.74
PTN-ResIcon	E: predicted family	Ground truth	Predicted	65.19	27.29
PTN-ResIcon	D: fully autonomous	IQ-estimated	Predicted	54.93	10.27
PTN-ResIcon-Est	A/B: oracle or refined	Ground truth	Ground truth	81.84	65.00
PTN-ResIcon-Est	C: estimated parameters	IQ-estimated	Ground truth	61.73	11.00
PTN-ResIcon-Est	E: predicted family	Ground truth	Predicted	69.68	37.00
PTN-ResIcon-Est	**D: fully autonomous**	**IQ-estimated**	**Predicted**	**61.53**	**10.90**

ParamNet achieves a test-set MAE of 0.718 (sps), 11.22 kHz (symbol rate), 0.051 (roll-off), 2.96 kHz (CFO), and 20.93 kHz (bandwidth), with the error dominated by low-SNR frames. Estimated parameters cost about 10 points (C vs. B, 80.67% → 70.49%) and the predicted family costs about 15 points (E, 80.67% → 65.19%). The fully autonomous configuration D of the main model retains 54.93% overall—71.2% at 0 dB, 82.1% at 10 dB, and 83.3% at +18 dB—but drops to 10.3% at −20 dB, i.e., to the signal-only level. Because the main model is trained with clean side information, we additionally train a robustness-adapted variant, PTN-ResIcon-Est, with the same architecture and data but with training-time corruption that matches the inference-time side-information errors. With probability 0.5, the Object branch receives the IQ-estimated parameters of ParamNet and the Interpretant branch receives the predicted family instead of the ground-truth ones. The adaptation barely affects the idealized setting (A/B: 81.84% vs. 80.67%) but substantially improves the autonomous configuration: D reaches 61.53% overall and 83.0/91.6/90.9% at 0/10/+18 dB, essentially matching the idealized values at moderate SNR, and the autonomous confidence becomes well-calibrated (ECE 0.013 vs. 0.258). The −20 dB accuracy of the adapted variant remains at 10.9%: at such low SNR, the side information itself cannot be estimated reliably. This confirms the attribution in Section 6.1: the low-SNR advantage originates from side information that cannot be estimated reliably at −20 dB; the idealized numbers of the earlier sections should be read as upper bounds.

6.12. Robustness Under Additional Channel Conditions and Open-Set Behavior

Section 6.6 already reported graceful degradation under scenario-domain distortion (54.94% for known modulations). Using the same standard-mode generation pipeline as RML-S-Full-Std, we synthesize five test sets that isolate a single stress condition: (i) Doppler shifts up to ±1 kHz (nominal ±500 Hz), (ii) Rician fading with $K = 0$ dB and (iii) $K = 10$ dB, (iv) heavy in-band jamming (jammer-to-signal ratio +3 dB), and (v) extreme SNRs of −25 and −30 dB, below the training range. All conditions are evaluated on the existing PTN-ResIcon weights without retraining.

Table 16 shows that, with oracle side information (configuration B), the model degrades gracefully: Doppler and Rician variations cost 1–6 points; heavy jamming costs about 8 points; even 7 points below the nominal SNR range (−30 dB), the accuracy stays at 63.8% because the Object and Interpretant branches are largely channel- and SNR-insensitive. In

the fully autonomous configuration D of the adapted variant, the same channel robustness holds at moderate SNR, but heavy jamming and extreme SNR still collapse accuracy (22.8% and 10.0%), since the IQ-based parameter and family estimates themselves fail under these conditions.

Table 16. Known-modulation accuracy under additional channel conditions. Configuration B denotes the encoder-refined (oracle side information) setting of the main model, and configuration D denotes the fully autonomous setting of the robustness-adapted variant PTN-ResIcon-Est (Section 6.11).

Condition	B Overall (%)	D Overall (%)	B at 0 dB (%)	B at +18 dB (%)
Baseline (nominal test set)	80.67	61.53	84.19	91.50
Doppler ±1 kHz	80.30	61.39	84.9	90.7
Rician $K = 0$ dB	75.02	33.50	78.8	81.8
Rician $K = 10$ dB	79.66	60.25	83.5	90.7
Heavy jamming (+3 dB INR)	73.00	22.77	75.2	79.4
Extreme SNR −25/−30 dB	63.82	9.97	—	—

We further address open-set behavior, confidence calibration, and false alarms, which matter for non-cooperative operations where unseen emitters must be recognized as unknown. Table 17 summarizes the results. The in-distribution confidence of configuration B is excellently calibrated (ECE 0.007, per-SNR below 0.012 for all tested SNR levels); configuration D of the main model is markedly miscalibrated (ECE 0.258), whereas the robustness-adapted variant is well-calibrated in the autonomous setting (ECE 0.013). However, both configurations are overconfident on unseen classes (GMSK, OOK, FSK, OFDM): the mean max-softmax confidence on unknowns is 0.85, higher than on known classes (0.81); the open-set AUROC is close to chance (0.49 for max-softmax and 0.38 for energy in configuration B); and the FAR at 90% true-positive rate is 1.0. A confidence threshold therefore cannot reliably separate known from unknown signals. This honest limitation of closed-set supervision motivates the calibrated-uncertainty and open-set-recognition directions listed as future work.

Table 17. Open-set detection and calibration metrics on the standard test set (known, 33,000 frames) and four unseen modulation classes (GMSK, OOK, FSK, OFDM; 80,000 frames). B: main model with encoder-refined (oracle) side information; D: main model in the fully autonomous setting; D-Est: robustness-adapted variant PTN-ResIcon-Est in the fully autonomous setting (Section 6.11).

Metric	B	D	D-Est
ECE (overall)	0.007	0.258	0.013
AUROC, max-softmax score	0.494	0.574	0.451
AUROC, energy score	0.381	0.590	0.400
FAR at 90% TPR	1.000	1.000	1.000
TPR at 10% FAR	0.115	0.069	0.093
Mean max-softmax on known/unknown	0.814/0.850	0.808/0.857	0.624/0.624

7. Conclusions

This paper proposes PeirceTriadicNet, a Peircean triadic-relation network for non-cooperative signal recognition. PTN-ResIcon instantiates the Representamen, Object, and Interpretant branches and fuses them with fixed equal weights. On RML-S-Full-Std, it achieves 80.67% overall modulation-recognition accuracy (+24.0, +20.1, and +20.2 points over CNN2, CNN-LSTM, and ResNet1D), and controlled ablations localize the gain: the modulation-family prior contributes about 10 points, the triadic decomposition about 10 points over plain concatenation, and attention-based fusion (81.37%) is comparable to the fixed equal-weight rule. The model generalizes to a held-out test set (90.76%), stays stable across operational scenarios (81.44% in the UAV-swarm scenario), and degrades

gracefully under scenario distortion (54.94%). An end-to-end evaluation with side information estimated from the received IQ frame shows that the fully autonomous pipeline, using the robustness-adapted variant PTN-ResIcon-Est, retains 61.53% overall accuracy (83–92% at moderate SNR), with the idealized numbers serving as upper bounds; open-set analysis reveals overconfidence on unseen classes (AUROC close to chance), which we report as an honest limitation and a target for future work. On the real DroneRFa corpus, the complete triadic configuration reaches 72.50% device-identification accuracy, confirming transfer to genuine UAV signals. We emphasize that the demonstrated results are modulation classification and device identification; downstream functions such as anti-jamming decision support, spectrum adaptation, and emergency-communication management are proposed applications that are not demonstrated here. Future work includes calibrated uncertainty and open-set recognition for unseen emitters, as well as further validation on larger real-world corpora.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/drones10110853/s1, File S1: DroneRFa transfer experiments.

Author Contributions: Conceptualization, C.C. and T.L.; methodology, C.C. and Z.L. (Zixin Li); software, C.C.; validation, C.C., T.L. and Z.L. (Zixin Li); formal analysis, C.C.; investigation, X.G.; resources, C.C.; data curation, Z.L. (Zhe Li); writing—original draft preparation, C.C.; writing—review and editing, T.L.; visualization, C.C.; supervision, H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets RML-S-Full-Std, RML-S-Full-Std-test, and RML-S-Full-Scene are openly available in Figshare at https://doi.org/10.6084/m9.figshare.33706927.

Acknowledgments: The authors would like to thank the Cyber Electromagnetic Technology Laboratory of the Research and Development Center for providing the research facilities and experimental environment that made this work possible.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AMC	Automatic Modulation Classification
AMR	Automatic Modulation Recognition
AUROC	Area Under the Receiver Operating Characteristic Curve
CFO	Carrier-Frequency Offset
ECE	Expected Calibration Error
FAR	False Alarm Rate
IQ	In-phase/Quadrature
MAE	Mean Absolute Error
MLP	Multilayer Perceptron
OFDM	Orthogonal Frequency-Division Multiplexing
PSD	Power Spectral Density
PTN	PeirceTriadicNet
ResNet	Residual Network
SNR	Signal-to-Noise Ratio
STFT	Short-Time Fourier Transform
TPR	True-Positive Rate

UAV	Unmanned Aerial Vehicle

References

1. O'Shea, T.J.; Roy, T.; Clancy, T.C. Over-the-Air Deep Learning Based Radio Signal Classification. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 168–179.
2. Jdid, B.; Hassan, K.; Dayoub, I.; Jouide, M.; Jadane, M. Machine Learning Based Automatic Modulation Recognition for Wireless Communications: A Comprehensive Survey. *IEEE Access* **2021**, *9*, 54009–54032.
3. Peirce, C.S. *Collected Papers of Charles Sanders Peirce*; Hartshorne, C., Weiss, P., Eds.; Harvard University Press: Cambridge, MA, USA, 1931–1958.
4. Nwakanma, C.; Kim, J.; Lim, D.-W.; Hassan, N.U.; Sun, C. CNN-LSTM Based Automatic Modulation Classification for Beyond 5G Networks. In *Proceedings of the IEEE Vehicular Technology Conference (VTC)*; IEEE: New York, NY, USA, 2021; pp. 1–5.
5. Sanna, M.; Cianca, E.; De Sanctis, M. Automatic Modulation Classification for 5G and Beyond Using Deep Learning: A Survey. *IEEE Access* **2022**, *10*, 10330–10346.
6. Qi, P.; Zhou, X.; Zheng, S.; Li, Z. Automatic Modulation Classification Based on Deep Residual Networks with Multimodal Information. *IEEE Trans. Cogn. Commun. Netw.* **2021**, *7*, 21–33.
7. Wei, S.; Qu, Q.; Zeng, X.; Li, M.; Zhang, J.; Yang, Y. Self-Attention Bi-LSTM Networks for Radar Signal Modulation Recognition. *IEEE Trans. Microw. Theory Tech.* **2021**, *69*, 4556–4567.
8. Lin, S.; Zeng, Y.; Gong, Y. Learning of Time-Frequency Attention Mechanism for Automatic Modulation Recognition. *IEEE Wirel. Commun. Lett.* **2022**, *11*, 939–943.
9. Ansari, S.; Alnajjar, K.A.; Majzoub, S.; Ballal, T.; Al-Naffouri, T.Y. Attention-Enhanced Hybrid Automatic Modulation Classification for Advanced Wireless Communication Systems: A Deep Learning-Transformer Framework. *IEEE Access* **2025**, *13*, 3891–3906.
10. Chen, B.; Rao, B.; Zou, X.; Wang, W. Emerging Versatile Context Representation on Modulation Recognition with Vision Transformer. In *Proceedings of the IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*; IEEE: New York, NY, USA, 2024; pp. 1–7.
11. Zhang, W.; Xue, K.; Yao, A.; Sun, Y. CTRNet: An Automatic Modulation Recognition Based on Transformer-CNN Neural Network. *Electronics* **2024**, *13*, 3408.
12. Ke, Y.; Zhang, W.; Zhang, Y.; Zhao, H.; Fei, Z. GIGNet: A Graph-in-Graph Neural Network for Automatic Modulation Recognition. *IEEE Trans. Veh. Technol.* **2025**, *74*, 10058–10062.
13. Rehman, M.D.U.; Rehman, M.H.H.; Butt, M.F.U.; Hassan, S.A.; Jung, H. Deep Learning Based AMC: A Survey of Advances and Future Prospects. *Comput. Netw.* **2023**, *232*, 109936.
14. Zhang, M.; Wei, G.; Tang, P.; Ding, G.; Wang, H. Open Set Modulation Recognition for Robust Channel Based on Multimodal Information Fusion. In *Proceedings of the IEEE/CIC International Conference on Communications in China (ICCC)*; IEEE: New York, NY, USA, 2024; pp. 1686–1691.
15. Zhang, W.; Xue, K.; Yao, A.; Zhang, H. Automatic Modulation Recognition Based on Multimodal Information Processing: A New Approach and Application. *Electronics* **2024**, *13*, 4568.
16. Wang, M.; Fan, Y.; Fang, S.; Wang, Y. Automatic Modulation Classification with Multi-Modal Data Fusion Located at the Feature Layer. In *Proceedings of the International Conference on Intelligent Control, Measurement and Signal Processing (ICMSP)*; IEEE: New York, NY, USA, 2024; pp. 624–629.
17. Jagannath, A.; Jagannath, J. Multi-Task Learning Approach for Modulation and Wireless Signal Classification for 5G and Beyond: Edge Deployment via Model Compression. *Phys. Commun.* **2022**, *54*, 101793.
18. Chandler, D. *Semiotics: The Basics*, 3rd ed.; Routledge: London, UK, 2017.
19. Deacon, T.W. *The Symbolic Species: The Co-Evolution of Language and the Brain*; W. W. Norton: New York, NY, USA, 1997.
20. Gao, F.; Zhang, K. Enhanced Multi-Parameter Cognitive Architecture for Future Wireless Communications. *IEEE Commun. Mag.* **2015**, *53*, 86–92.
21. Haykin, S. Cognitive Radio: Brain-Empowered Wireless Communications. *IEEE J. Sel. Areas Commun.* **2005**, *23*, 201–220.
22. Yu, N.; Mao, S.; Zhou, C.; Sun, G.; Shi, Z.; Chen, J. DroneRFa: A Large-Scale Dataset of Drone Radio Frequency Signals for Detecting Low-Altitude Drones. *J. Electron. Inf. Technol.* **2024**, *46*, 1147–1156.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.