electronics

MDPI

Article

Neural-Assisted Detector with Low False-Alarm Rate for Dense Multi-Target Sensing in OTFS-ISAC Systems

Imran Younas, Zhenzhen Gao * and Kejia Ma

Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China; imran.younas@stu.xjtu.edu.cn (I.Y.); 312215@stu.xjtu.edu.cn (K.M.)
* Correspondence: zhenzhengao@xjtu.edu.cn

Abstract

Dense multi-target sensing in orthogonal time–frequency space (OTFS)-based integrated sensing and communication (ISAC) is challenged by sub-resolution target spacing, unknown target cardinality, and inter-target radar-cross-section (RCS) variation. This paper develops a neural-assisted detector (NAD) that combines convolutional-neural-network candidate generation with continuous delay–Doppler refinement, physics-based validation, active splitting, and explicit model-order control. The CNN generates candidates from the nominal delay–Doppler grid while predicting fractional offsets, so that the final target positions are continuous-valued rather than confined to grid centers. Isolated candidates are refined individually, whereas closely spaced candidates are jointly optimized through alternating position and complex-gain updates, with active splitting used to resolve dense target pairs. For unknown target number, NAD-Hybrid combines conservative candidate generation with information-criterion-based weak-target rescue and backward pruning to control false alarms under RCS mismatch. On a frozen 8000-frame blind test covering two scenes and four RCS standard deviations, NAD-Hybrid achieves a macro-averaged F_1 score of 0.9338 and reduces false alarms relative to the unknown-order OMPFR-BIC baseline, at the cost of lower detection probability under severe RCS mismatch. The proposed detector therefore provides a conservative operating point that prioritizes low false alarms and reliable target-count estimation. The sensing framework is further demonstrated to coexist with an index-modulated OTFS communication link using a shared frame.

Keywords: orthogonal time–frequency space; integrated sensing and communications; dense targets; radar cross section; model-order selection; bayesian information criterion; low false alarm; B5G/6G

Check for updates

Academic Editor: Yaochun Shen

Received: 14 September 2026
Revised: 17 October 2026
Accepted: 21 October 2026
Published: 26 October 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.

1. Introduction

Integrated sensing and communications (ISAC) require a common physical-layer waveform to support data transmission and environmental perception [1]. In high-mobility B5G/6G scenarios, time-varying channels and pronounced Doppler spreads challenge conventional time–frequency processing [2]. Orthogonal time–frequency space (OTFS) modulation represents propagation paths in the delay–Doppler (DD) domain and therefore provides a natural interface between communication-channel estimation and radar estimation of target range and radial velocity [3,4]. Furthermore, by transforming the rapidly time-varying fading channel into a quasi-static and sparse representation, OTFS significantly simplifies transceiver design while preserving the fine-grained geometric features necessary for high-resolution sensing [5]. OTFS therefore provides a suitable

waveform framework for high-mobility terrestrial and non-terrestrial communication [6,7] and sensing applications. In this work, we focus specifically on short-to medium-range OTFS-ISAC scenarios in which multiple targets or scattering centers may occupy neighboring delay–Doppler resolution cells. This setting is representative of applications such as traffic monitoring, convoy or formation monitoring, unauthorized UAV identification in the low-altitude economy [8], and access-point-assisted sensing, where reliable weak-target detection and low false-alarm operation are important.

The sensing difficulty is not simply the presence of multiple targets. Vehicles, unmanned aerial vehicles (UAVs), or scattering centers can have delay and Doppler separations below the nominal resolution. Their mainlobes and sidelobes then overlap, producing a distorted broad peak instead of distinct local maxima [9,10]. A detector that repeatedly fits and subtracts the strongest peak can leave structured residual sidelobes, which may be mistaken for new targets and can corrupt later estimates [11]. These residuals are deterministic and signal-dependent rather than stochastic noise. Traditional constant false alarm rate detectors and the fixed-grid sparse recovery methods therefore cannot reliably distinguish genuine weak targets from such structured sidelobe artifacts [12,13]. The problem becomes more severe when nearby targets have dissimilar radar cross sections (RCSs). Although their propagation losses are similar, differences in size, material, aspect angle, polarization, and scattering-center interference can yield echo powers that differ by several or even tens of decibels [14]. A weak target may consequently be masked by the sidelobes or the subtraction residual of a strong neighbor [15].

Early OTFS sensing studies established maximum-likelihood radar estimation and analyzed estimation limits [3]. MIMO-OTFS extensions introduced coarse detection followed by continuous refinement and can separate targets when at least one of the elements including delay, Doppler, or angle remains resolvable [16], and recent cell-free massive MIMO-OTFS architectures have extended this application to distributed multi-target detection [17]. Continuous-domain sparse approaches reduce off-grid error: Liu et al. formulated two-dimensional atomic-norm recovery followed by two-dimensional MUSIC [18]; Zegrar et al. exploited the physical sidelobe structure of rectangular OTFS pulses for super-resolution range–velocity profiling [19]; Zhang et al. combined sparse reconstruction preprocessing with weighted subspace fitting [20]. Orthogonal matching pursuit with fractional refinement (OMPFR) recovers a sparse support and then refines fractional delay and Doppler without an excessively fine dictionary [21]. For unknown target numbers, sequential generalized-likelihood-ratio testing has also been applied to OTFS sensing [22]. Deep learning has recently shown strong potential for communication-focused tasks such as plug-and-play channel estimation and symbol detection [23]; for sensing, deep learning is integrated to execute model-order selection in low-complexity decoupled estimation [24]. However, this decoupled framework depends on the sequential detection and elimination, thereby neglecting the combined issues of dense target overlap, unknown target quantity, and mismatched RCS discrepancies.

These studies demonstrate the potential of OTFS for multi-target and super-resolution sensing, yet jointly handling targets that are dense in both DD dimensions, unknown target count, and strong inter-target echo imbalance remains insufficiently explored. Array-based angular separation does not address a SISO or unresolved-angle setting; sparse continuous refinement improves localization but may still require a model order; and aggressive support expansion can improve detection probability at the cost of false targets. The latter cost is operationally important because false detections can initiate spurious tracks, enlarge multi-frame association trees, trigger unnecessary beam revisits, and consume sensing and communication resources.

To address these challenges, this paper develops a neural-assisted dense multi-target detector for scenes containing a densely overlapping target pair coexisting with additional dispersed targets. A lightweight convolutional neural network predicts fractional delay–Doppler offsets from the nominal grid, thereby removing the fixed-grid localization constraint. Generated candidates are categorized into isolated and closely-spaced groups based on their parameter-space distances. Alternating optimization of target locations and complex gains mitigates parameter-coupling interference for closely-spaced targets. Following iterative candidate reconstruction, two post-processing model-order control strategies are devised for unknown-target-number scenarios. One directly adopts the Bayesian information criterion (BIC) for target-count estimation as an internal reference variant named Neural Assisted Dectector (NAD)-BIC; the other, NAD-Hybrid, implements a hybrid information-criterion workflow and serves as the primary practical solution for low-false-alarm operation.

1. A CNN-assisted dense multi-target detection framework is presented to mitigate discrete-grid errors. The CNN predicts continuous fractional delay–Doppler offsets from the nominal grid, and candidates are jointly optimized via alternating location-gain updates. Active splitting resolves pairwise overlapping targets within one resolution bin, while severe cases with dense multi-target clusters are left for future research.
2. A hybrid information-criterion model-order control scheme (NAD-Hybrid) is proposed for target-count estimation without a ground-truth prior. Compared with an internal raw-BIC baseline (NAD-BIC), the hybrid scheme suppresses over-splitting artifacts under heavy sidelobe interference and RCS mismatch, achieving a favorable high-precision, low-false-alarm operating point at the accepted cost of reduced detection probability under severe mismatch.
3. A rigorous frozen blind-test evaluation protocol is established with separate validation and blind-test datasets to prevent data leakage. Comprehensive metrics benchmark the scheme against OTFS-ISAC baselines, and a randomized-cardinality full-orientation stress test validates robustness. Complementary experiments further demonstrate sensing-communication coexistence under a power-fair index-modulated OTFS shared-frame architecture.

The rest of the paper is organized as follows. Section 2 gives the OTFS sensing and RCS models. Section 3 presents the detector and model-order controllers. Section 4 defines the scenarios, baselines, and metrics. Section 5 reports the results and limitations, and Section 6 concludes the paper.

2. System Model and Problem Formulation

2.1. OTFS-ISAC Echo Model

Consider a colocated OTFS-ISAC transmitter and sensing receiver observing K moving targets, indexed by $k = 1, \ldots, K$. Multipath reflections of order higher than the dominant target returns are neglected. Let $\mathbf{x} \in \mathbb{C}^{MN}$ denote the known vectorized transmitted DD-domain frame over the $M \times N$ delay–Doppler grid, and let $\mathbf{\Psi}(\tau_k, \nu_k) \in \mathbb{C}^{MN \times MN}$ denote the equivalent OTFS response operator of target k, which imposes the delay–Doppler shift (τ_k, ν_k) on the DD-domain frame through the OTFS modulation/demodulation chain. The vectorized DD-domain sensing observation $\mathbf{y} \in \mathbb{C}^{MN}$ is

$$\mathbf{y} = \sum_{k=1}^{K} h_k \, \mathbf{\Psi}(\tau_k, \nu_k) \, \mathbf{x} + \mathbf{w} = \mathbf{A}(\boldsymbol{\theta}) \, \mathbf{h} + \mathbf{w}, \tag{1}$$

where h_k is the complex reflection coefficient of the kth target; $\boldsymbol{\theta} = \{(\tau_k, \nu_k)\}_{k=1}^{K}$ collects the delay–Doppler pairs of all K targets; $\mathbf{A}(\boldsymbol{\theta}) = [\boldsymbol{\Psi}(\tau_1, \nu_1)\mathbf{x}, \ldots, \boldsymbol{\Psi}(\tau_K, \nu_K)\mathbf{x}] \in \mathbb{C}^{MN \times K}$ is the sensing dictionary whose kth column is the noise-free response of the target k; $\mathbf{h} = [h_1, \ldots, h_K]^{\mathrm{T}} \in \mathbb{C}^K$ stacks the reflection coefficients; and $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_{MN})$ is an additive white Gaussian noise with variance σ_w^2, where $\mathbf{I}_{MN}$ is the identity matrix $MN \times MN$.

For carrier wavelength λ_c and propagation speed c, the delay and Doppler of target k map to its range R_k and radial velocity v_k as

$$R_k = \frac{c\,\tau_k}{2}, \qquad v_k = \frac{\lambda_c\,\nu_k}{2}. \tag{2}$$

with subcarrier spacing Δf and symbol duration T, the continuous delay–Doppler parameters can be decomposed as

$$\tau_k = \frac{\ell_k + \delta_{\tau,k}}{M\,\Delta f}, \qquad \nu_k = \frac{q_k + \delta_{\nu,k}}{N\,T}, \tag{3}$$

where ℓ_k and q_k are the integer delay and Doppler grid indices, and $\delta_{\tau,k}, \delta_{\nu,k} \in [-0.5, 0.5)$ are the corresponding fractional offsets. Consequently, $\Delta\tau_{\mathrm{res}} = 1/(M\Delta f)$ and $\Delta\nu_{\mathrm{res}} = 1/(NT)$ denote the delay and Doppler resolution cells used in Section 2.2.

2.2. Dense Targets and Structured Residuals

Targets i and j are defined as a dense pair if both their normalized delay and Doppler separations are smaller than one resolution-cell dimension.

$$\frac{|\tau_i - \tau_j|}{\Delta\tau_{\mathrm{res}}} < 1, \qquad \frac{|\nu_i - \nu_j|}{\Delta\nu_{\mathrm{res}}} < 1. \tag{4}$$

their responses are then non-orthogonal. If a single-target estimate $(\widehat{\tau}, \widehat{\nu}, \hat{h})$ is fitted to the composite observation formed by closely-spaced targets, subtraction yields

$$\mathbf{r} = \mathbf{y} - \widehat{h}\boldsymbol{\Psi}(\widehat{\tau}, \widehat{\nu})\mathbf{x}, \tag{5}$$

where $\widehat{\tau}$ and $\widehat{\nu}$ are delay–Doppler estimates obtained under the single-target assumption, and scalar $\hat{h}$ denotes the conditional least-squares estimate of the equivalent complex reflection coefficient given $(\widehat{\tau}, \widehat{\nu})$. This residual generally contains deterministic sidelobe-like structure rather than white noise. This structure motivates explicit target splitting and physical validation instead of blindly selecting the largest residual sample, as illustrated in Figure 1.

Figure 1a shows the echo-amplitude distributions of two closely-spaced targets in the delay–Doppler domain, plotted by the blue dashed line and red dashed line, respectively. The black solid line represents their superposed echo-amplitude distribution. It can be observed that when two targets are closely separated in the delay–Doppler domain, their mainlobes overlap significantly. The superposed amplitude no longer exhibits two well-separated peaks, but forms a distorted broad peak. Consequently, target detection based merely on local peaks tends to misinterpret the superposed amplitude contributed by multiple dense targets as a single target. Under the single-target assumption, conventional iterative detection methods generally take the composite-peak position as the target location and perform parameter fitting. As illustrated in Figure 1b, although the single-target fitting result can roughly match the overall trend of the composite echo, it cannot accurately characterize the echo structure arising from two coupled dense targets. Therefore, after subtracting this single-target response from the observation, prominent modeling errors remain in the residual signal. Figure 1c further depicts the residual-amplitude distribution

after single-target response subtraction. One can see that the residual is not Gaussian white noise, but deterministic sidelobe structures with non-negligible amplitude appear on both sides of the targets. Such residual sidelobes may be misidentified as candidate targets in subsequent iterative detection, yielding false detections. Meanwhile, residual energy contaminates parameter estimation of nearby targets and induces error accumulation.

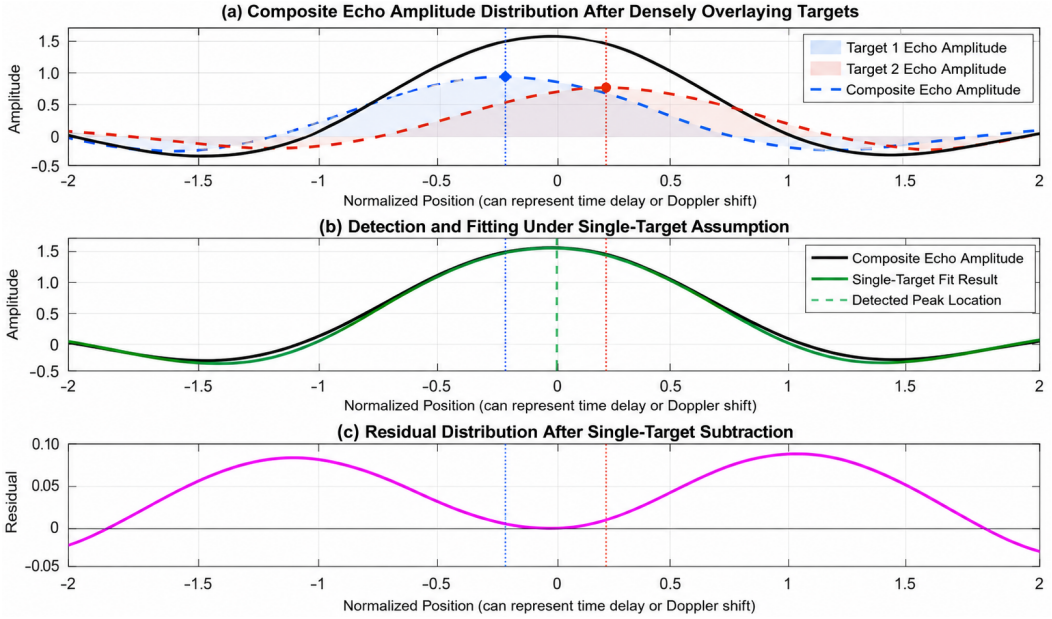


Figure 1. Dense–target superposition and structured residuals.

2.3. Inter-Target RCS Mismatch

Besides the geometrical closeness in the delay–Doppler domain, realistic dense-target scenarios also suffer from large amplitude imbalance among targets due to RCS differences. To emulate the realistic inter-target scattering, we adopted an independent log-normal RCS perturbation model, commonly used for random reflection-strength variations [25]. According to the monostatic radar equation, the following proportional relations hold

$$P_{r,k} \propto \frac{\sigma_k}{R_k^4}, \qquad |h_k^p| \propto \frac{\sqrt{\sigma_k}}{R_k^2}, \tag{6}$$

where σ_k is the target RCS and $P_{r,k}$ is the received echo power at (monostatic) sensing receiver due to target k. Thus, close physical positions imply similar range loss, but not necessarily similar echo strength. In our robustness experiments, the original ranges, velocities, target counts, and dense geometry are preserved. Only an independent log-normal RCS perturbation is applied:

$$x_k \sim \mathcal{N}(0, s_{\mathrm{RCS}}^2), \qquad a_k = 10^{x_k/20}, \tag{7}$$

where x_k represents zero-mean Gaussian RCS fluctuation in the decibel domain. The perturbed amplitude before normalization reads

$$|h_k^p| = |h_k| a_k, \tag{8}$$

and the final normalized amplitude is given by

$$|h_k^p|' = \frac{|h_k^p|}{\sqrt{\frac{1}{K}\sum_{j=1}^{K}\left(|h_j^{p(0)}|a_j\right)^2}}. \tag{9}$$

Here, $h_j^{p(0)}$ denotes the complex reflection coefficient of the jth target before applying the independent RCS perturbation, i.e., $h_j^{p(0)} = h_j$, and a_j is the corresponding RCS perturbation factor defined in (7). This normalization ensures that the average target return power remains comparable across different RCS-perturbation settings, while the relative echo strengths among targets within each frame are diversified. Here, s_{RCS} stands for the standard deviation of RCS fluctuations in dB. Setting $s_{\mathrm{RCS}} = 0$ corresponds to the case without additional inter-target mismatch; increasing the value of s_{RCS} gradually aggravate the amplitude imbalance between strong and weak targets. It should be emphasized that this is an inter-target independent mismatch model, rather than a temporally correlated Swerling fluctuation model.

3. Neural-Assisted Dense Multi-Target Detection

3.1. Overall Processing Chain

As shown in Figure 2, the proposed receiver takes the sensing echo and the known delay–Doppler domain transmit frame as inputs, and outputs a variable-cardinality target set $\widehat{\mathcal{T}} = \{(\widehat{\tau}_k, \widehat{\nu}_k, \widehat{h}_k)\}$. It initializes the residual $\boldsymbol{r}^{(0)} = \mathbf{y}$ and enters an iterative loop of candidate extraction, validation and splitting, joint refinement, and global residual reconstruction. In each round, the CNN scans the current residual; if the peak existence probability falls below the validity threshold, the loop ends with the accumulated target set. Otherwise, the peak index and its fractional offsets form a continuous candidate coordinate, routed by its normalized distance to the accepted targets: an isolated candidate enters active splitting, where a single-target hypothesis is tested against symmetric two-target hypotheses to select single-point or dual-point Adam refinement, whereas a neighboring candidate undergoes the amplitude–distance plausibility check and is either admitted into the closest group or recorded as a rejected sidelobe position before joint group refinement (Section 3.4). Each round closes with a global least-squares gain re-estimation and a residual refresh $\boldsymbol{r} \leftarrow \mathbf{y} - \widehat{\mathbf{y}}$, so that the next CNN input contains only the unexplained echo. The loop terminates when the peak probability falls below the existence-probability threshold of 0.01 in the frozen configuration; a strongest-to-weakest amplitude-ratio stopping rule also exists in the implementation but is disabled in the frozen configuration (ratio threshold ∞, Table 3). After termination, a final screening discards targets violating the physical range–gain relation. Thus, the CNN governs candidate generation, yet the final target number is settled jointly by the validation, splitting, and model-order mechanisms of Sections 3.4 and 3.5.

3.2. Lightweight CNN Candidate Extractor

The residual is reshaped into a two-channel $M \times N$ tensor containing its real and imaginary parts. The network has four convolutional blocks. A 3×3 circularly padded convolution maps two inputs to 32 channels; a second 3×3 layer uses dilation two and maps to 64 channels; a third circular convolution returns to 32 channels; and a final convolution produces target-existence logits and two fractional-offset maps. Batch normalization and ReLU are used in the hidden layers, with dropout after the first two blocks. Sigmoid maps the existence output to $\hat{p}_{\ell,q} \in [0,1]$, and a scaled hyperbolic tangent constrains $\hat{\delta}_{\tau,\ell,q}, \hat{\delta}_{\nu,\ell,q} \in [-0.5, 0.5]$. The complete network is depicted in Figure 3. The originally deployed checkpoint accepted five input maps, but three of these maps were identically zero for every inference sample. Folding the corresponding zero-map weights into the first convolution produces an equivalent two-channel checkpoint without retraining. Over the 8000-frame blind test, the two implementations produced identical TP, FP, FN, and target-count decisions, with range- and velocity-RMSE differences of only 3.24×10^{-8} m and 4.80×10^{-7} m/s, respectively. Two thresholds govern this decision. On the first scan of

a frame, the peak probability is compared with the empty-scene gate of 0.66; if it falls below this gate, the frame is declared target-free and the detector returns an empty target set. On every scan, the iterative loop terminates when the peak falls below the existence-probability threshold of 0.01, returning the targets accepted up to that point. Otherwise, the integer grid index of the peak is combined with the two fractional offsets read from the same cell to form the continuous candidate coordinate, which initializes the least-squares gain estimate and the refinement stage of Section 3.3. Consistent with this division of labor, during training, the network receives the original received signal rather than stripped residuals, so that the iterative reconstruction loop is not unfolded through the optimizer; the network therefore learns candidate extraction directly from superposed multi-target echoes. The training objective combines the probability-map loss and the fractional-offset regression losses:

$$\mathcal{L} = \mathcal{L}_{\text{prob}} + \lambda_{\text{off}}(\mathcal{L}_\tau + \mathcal{L}_\nu), \tag{10}$$

where λ_{off} weights the offset losses so that the total loss is not dominated by the probability-map term. Because the vast majority of DD-grid cells contain no target, the positive–negative sample ratio is extremely imbalanced and standard binary cross-entropy would be dominated by the negative samples; the probability-map loss therefore adopts the focal loss, whose modulating factor concentrates the optimization on the few hard-to-classify positive cells, augmented by a ReLU-based false-alarm penalty:

$$\begin{aligned} \mathcal{L}_{\text{prob}} = & -\frac{1}{MN}\sum_{\ell,q}\Big[\alpha\Big(1-\hat{p}_{\ell,q}\Big)^{\gamma} p^{\text{gt}}_{\ell,q}\log\hat{p}_{\ell,q} \\ & + (1-\alpha)\,\hat{p}^{\gamma}_{\ell,q}\Big(1-p^{\text{gt}}_{\ell,q}\Big)\log\Big(1-\hat{p}_{\ell,q}\Big)\Big] \\ & + \frac{\beta}{MN}\sum_{\ell,q}\max\Big(0,\ \hat{p}_{\ell,q}-p^{\text{gt}}_{\ell,q}\Big), \end{aligned} \tag{11}$$

where $\hat{p}_{\ell,q}$ is the predicted existence probability at cell (ℓ, q); $p^{\text{gt}}_{\ell,q} \in \{0,1\}$ is the corresponding ground-truth existence label, set to 1 when the integer grid indices of a true target coincide with (ℓ, q); α is the positive–negative sample balancing coefficient; γ is the focusing parameter; and $\beta \geq 0$ is the false-alarm penalty coefficient. When an occupied cell is already correctly classified with high confidence, $\hat{p}_{\ell,q}$ approaches 1, the modulating factor $(1-\hat{p}_{\ell,q})^{\gamma}$ approaches 0, and the cell's contribution to the loss is strongly suppressed; conversely, for hard-to-classify cells the loss weight is retained. The third term penalizes the network only when its prediction exceeds the ground truth: since $\hat{p}_{\ell,q} \leq 1 = p^{\text{gt}}_{\ell,q}$ at occupied cells, it constrains only over-confident predictions at target-free cells and thereby suppresses false alarms.

The offset losses are meaningful only at cells containing a target. With the mask $m_{\ell,q} \in \{0,1\}$ equal to 1 at occupied cells, they are computed as mask-normalized mean-squared errors,

$$\mathcal{L}_\tau = \frac{\sum_{\ell,q} m_{\ell,q}\Big(\hat{\delta}_{\tau,\ell,q}-\delta_{\tau,\ell,q}\Big)^2}{\sum_{\ell,q} m_{\ell,q}+\varepsilon}, \qquad \mathcal{L}_\nu = \frac{\sum_{\ell,q} m_{\ell,q}\Big(\hat{\delta}_{\nu,\ell,q}-\delta_{\nu,\ell,q}\Big)^2}{\sum_{\ell,q} m_{\ell,q}+\varepsilon}, \tag{12}$$

where $\delta_{\tau,\ell,q}$ and $\delta_{\nu,\ell,q}$ are the true fractional offsets of the target assigned to cell (ℓ, q), and $\varepsilon > 0$ is a small constant preventing division by zero; no constraint is imposed on the offset outputs at target-free cells. The scalars α, γ, λ_{off}, and β are fixed on the independent validation set.

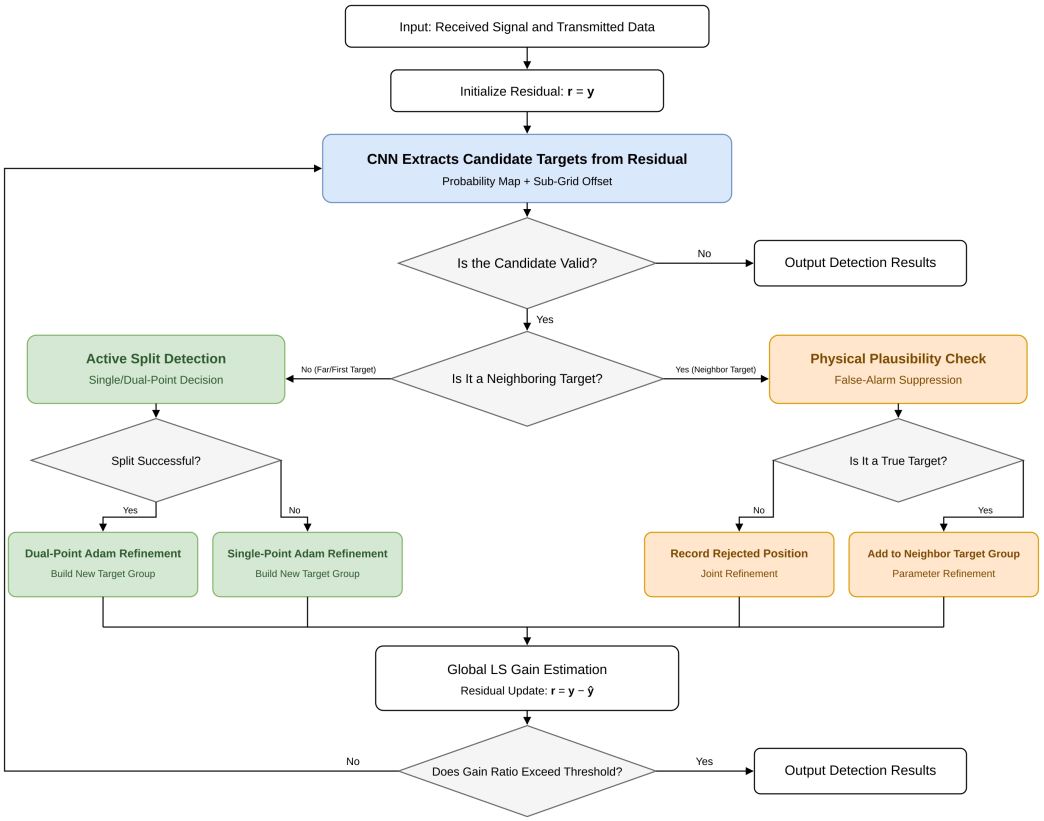


Figure 2. Dense multi-target detection workflow of the NAD.

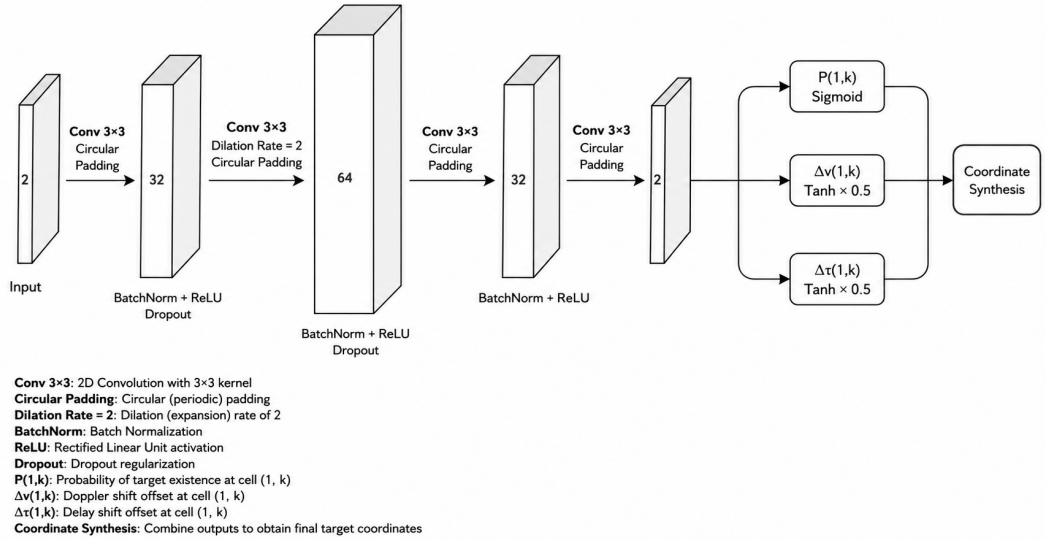


Figure 3. Architecture of the lightweight CNN candidate extractor and its coordinate synthesis output stage.

3.3. Decoupled Position–Gain Refinement

Let K_c denote the number of currently accepted targets and construct $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{C}^{MN \times K_c}$ as in Equation (1). Because closely spaced targets render the columns of $\mathbf{A}(\boldsymbol{\theta})$ ill-conditioned, complex gains are updated by regularized least squares:

$$\widehat{\mathbf{h}} = \left(\mathbf{A}^{\mathrm{H}}\mathbf{A} + \lambda_h \mathbf{I}\right)^{-1} \mathbf{A}^{\mathrm{H}}\mathbf{y}, \tag{13}$$

where $\lambda_h > 0$ is a regularization parameter that stabilizes the inversion of $\mathbf{A}^{\mathrm{H}}\mathbf{A}$ and limits noise amplification at low SNR, and $\mathbf{I}$ is the $K_c \times K_c$ identity matrix. To maintain numerical stability without introducing excessive bias, λ_h is adaptively adjusted based on target proximity: a base value of 1×10^{-4} is employed generally, but it is dynamically increased up to a maximum of 0.025 when the normalized delay–Doppler distance between jointly optimized targets falls below 0.8 resolution bins (as detailed in Table 3). With gains fixed, Adam minimizes the reconstruction mean squared error (MSE) over continuous positions. The algorithm alternates several position steps with an LS gain update. For a dense group, all positions are refined jointly against the received signal $\mathbf{y}$ with the reconstructed echoes of targets outside the group removed, using

$$\mathcal{L}_{\text{joint}} = \frac{1}{MN}\left\|\mathbf{y} - \mathbf{A}(\boldsymbol{\theta})\widehat{\mathbf{h}}\right\|_2^2 + \lambda_{\text{sep}} \sum_{i<j} \left[\rho - d_{ij}\right]_+^2, \tag{14}$$

where d_{ij} is the normalized DD distance in resolution cells, $[z]_+ = \max(0, z)$, and ρ is a repulsion radius that discourages two estimates from collapsing to the same location, while λ_{sep} is the separation penalty weight. The decoupled position–gain refinement framework follows the workflow of Figure 4. Both branches share a unified Adam foundation and the position–gain alternation described above, yet adopt separate configurations. The single-point branch serves isolated and successfully split candidates: the CNN candidate coordinates initialize the trainable position parameters, the loss is the MSE between the current residual and the reconstructed echo, and the closed-form gain refresh (13) is invoked every fixed number of Adam steps. The joint branch serves densely clustered groups: all group coordinates are initialized simultaneously, the objective is (14), and a larger iteration budget is allocated to resolve the additional degrees of freedom. Both branches terminate in a common estimate-output and residual-update node, which returns control to the global reconstruction step of Section 3.1. While the conceptual block in Figure 4 may be broadly referred to as a gain update, the rigorous mathematical implementation strictly follows the regularized least-squares (LS) formulation in (13), rather than a heuristic multiplicative scaling.

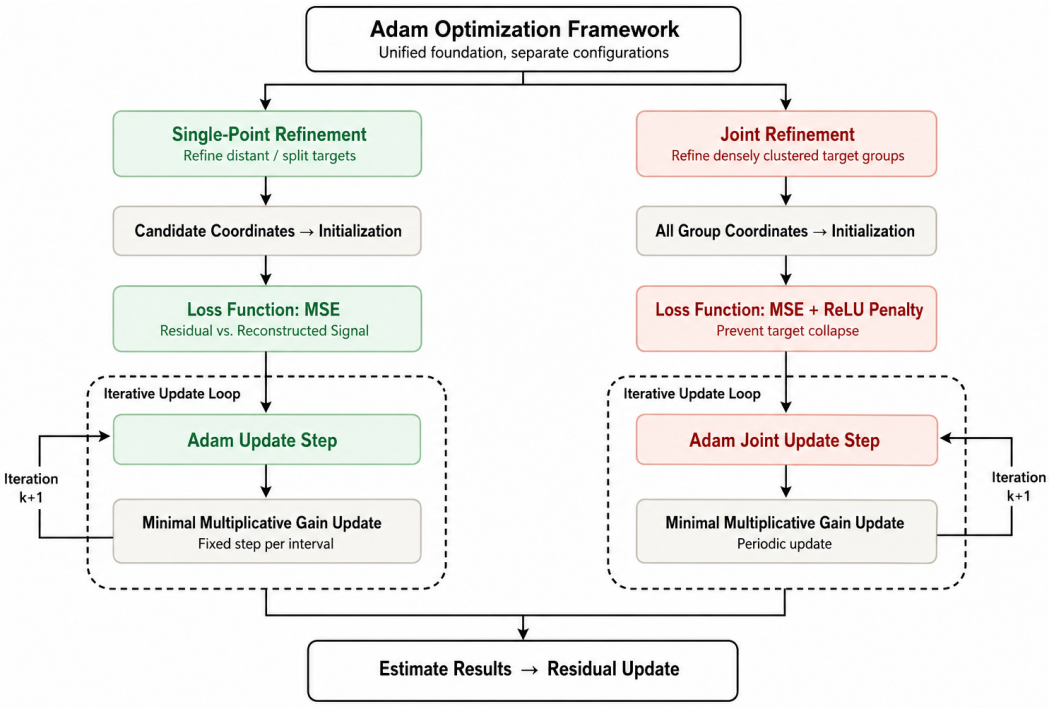


Figure 4. Position–gain refinement framework based on decoupled position–gain optimization.

3.4. Near-Target Validation and Active Splitting

A new candidate is classified as "far" when its minimum normalized distance from the accepted targets exceeds d_{close}. The first candidate and far candidates are subjected to active splitting, whereas near candidates require additional physical validation because they may correspond either to weak targets or to structured sidelobes. For a near candidate, the fitted amplitude is compared with a distance-dependent threshold referenced to the strongest target in the closest group. If the candidate is rejected, its location is recorded and the neighboring group is jointly refined to suppress the sidelobe energy; if a subsequent rejection occurs at the same location, the candidate is forcibly absorbed into the group to clear the residual energy, preventing an infinite processing loop. Conversely, a near candidate whose amplitude exceeds the threshold is admitted into the closest group for joint parameter refinement, while a far candidate that passes active splitting forms a new target cluster. All branches then converge to the global update and residual refresh step, which returns the refreshed residual to the iterative loop of Section 3.1. The far/near target classification and physical validation procedure is illustrated in Figure 5.

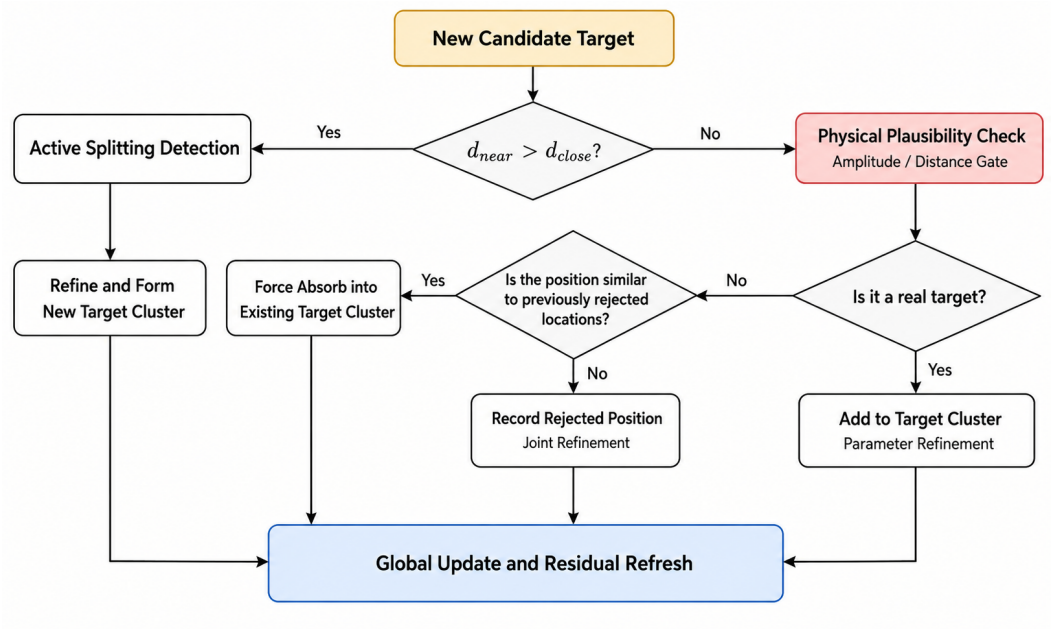


Figure 5. Far/Near target classification and physical validation flowchart.

Active splitting determines whether a CNN-generated candidate represents a composite peak. Let $\mathcal{E}_1$ denote the optimized single-target mean squared error (MSE), which serves as the splitting baseline. Four symmetric two-target seeds are generated along the delay axis, Doppler axis, and the two diagonal directions. These four directions (delay, Doppler, and two diagonals) are selected to provide comprehensive angular coverage of the 2D delay–Doppler plane. Because dense targets are separated by less than one resolution bin, these symmetric seeds serve as robust initializations that place the starting guesses well within the convergence basin of the subsequent joint Adam refinement, ensuring accurate final localization regardless of the pair's actual relative angle. For each seed the gains are solved by least squares (LS) at fixed positions, and the seed yielding the smallest MSE is retained, with its error denoted by $\mathcal{E}_2$. Splitting is accepted only if the relative reduction in MSE satisfies

$$\frac{\mathcal{E}_1 - \mathcal{E}_2}{\mathcal{E}_1} > \eta_{\text{split}}, \tag{15}$$

in which case the retained seed enters joint dual-point Adam refinement. As shown in Figure 6, the refined pair then passes two sequential physical gates: if the dual-point

spacing is too close, or if the amplitude difference falls outside the admissible range, the split is discarded and the candidate reverts to the single-point estimate; only when both gates are passed is the pair confirmed as a genuine dense dual-target and forwarded to group formation and the global residual update. Thus, the current splitter implements a one-target versus-two-target hypothesis test: it can resolve a dense pair in the presence of additional dispersed targets, but the present implementation does not establish reliable separation of three or more targets within the same resolution neighborhood. While the $\mathcal{E}_D$ and $\mathcal{E}_S$ represent the generalized distance and amplitude-ratio thresholds in Figure 6, corresponding to d_{close} and the near-candidate gates detailed in the text.

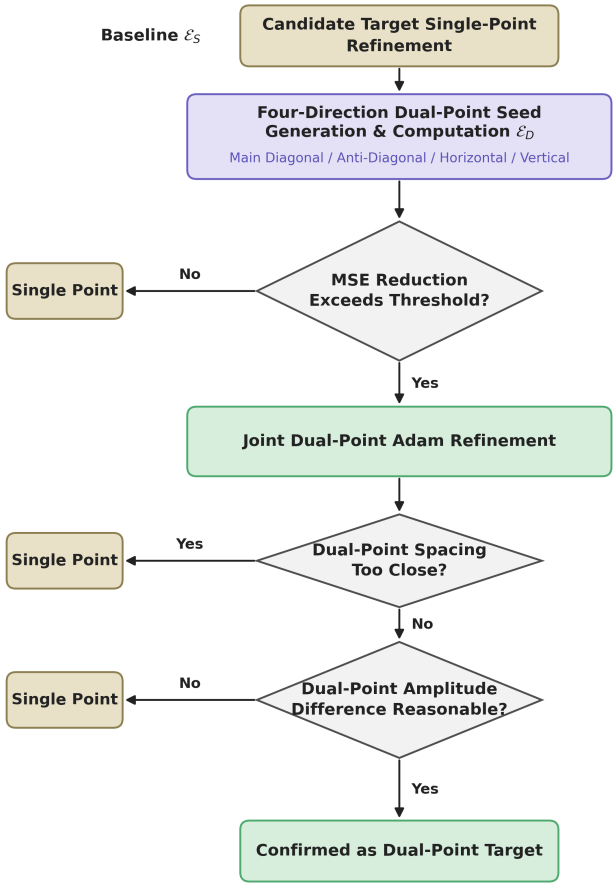


Figure 6. Active splitting hypothesis test and sequential physical gates for dual-target confirmation.

3.5. Unknown Model Order: BIC and Hybrid Control

The base detector (NAD) uses conservative amplitude, relative-MSE, and stopping thresholds. This produces few false alarms when targets have comparable RCS, but fixed thresholds can miss weak targets and some original stopping logic depends on a known count or count upper bound. To handle this problem, an enhanced scheme based on NAD that directly adopts the Bayesian information criterion for target-count estimation is proposed, which is denoted as NAD-BIC. NAD-BIC evaluates a candidate order K using

$$\mathrm{BIC}(K) = N_o \ln\left(\frac{\mathrm{SSE}_K}{N_o}\right) + q_K \ln N_o, \tag{16}$$

where N_o is the number of effective real-valued observations, SSE_K is residual squared error, and q_K is the number of free parameters. A target is added only when the fit improvement outweighs the complexity penalty. This recovers more weak targets but may over-split structured residuals.

NAD-Hybrid has three stages: (i) retain high-confidence targets using the conservative NAD pipeline; (ii) rescue a rejected weak candidate when its BIC improvement exceeds a frozen margin; and (iii) remove candidates through backward BIC pruning. The procedure is summarized in Table 1. The chosen validation configuration is `r20_p0`, i.e., a rescue margin of 20 and no additional pruning margin beyond the BIC comparison. During inference, the detector operates without prior knowledge of the true target cardinality K, whereas the training and evaluation datasets encompass scenarios containing one to five targets.

Table 1. Steps of the proposed NAD-Hybrid unknown-order controller.

Step	Operation
1	Generate and refine conservative candidates using NAD.
2	While a rejected candidate provides sufficient BIC improvement, rescue it and jointly refine positions and LS gains.
3	Tentatively remove every accepted candidate and recompute BIC. Permanently remove the candidate yielding the largest valid BIC decrease.
4	Repeat Step 3 until no removal decreases BIC; return the remaining target set.

4. Simulation Setup and Evaluation Protocol

MATLAB (release R2025b) generates OTFS-ISAC frames and echoes, while PyTorch (v2.8.0) implements the CNN and inference pipeline. Tables 2 and 3 list the waveform and algorithm parameters, respectively. Sensing SNR is defined as the ratio of the normalized average target-return power to the complex Gaussian noise power per delay–Doppler sample, measured before receiver processing; it therefore excludes any coherent integration or matched-filter processing gain. It is distinct from the communication SNR used in Section 5.4.

Table 2. OTFS-ISAC simulation parameters.

Parameter	Value
Carrier frequency	64 GHz
Number of subcarriers M	64
Number of OTFS symbols N	32
Subcarrier spacing Δf	120 kHz
Maximum delay tap	20
Range resolution	19.53 m
Maximum Doppler tap	2
Velocity resolution	8.789 m/s
Modulation	BPSK + QPSK index modulation
Sensing-SNR range	10–18 dB

All detector thresholds were frozen before the final blind test with seed offset 200,000. Sensitivity analyses were conducted only on the independent validation data with seed offset 100,000. In particular, the fission initialization offset δ_{seed} and the relative-MSE activation threshold η_{split} are distinct parameters. To clarify threshold selection, all decision thresholds were frozen on the independent validation split prior to the blind test. A sensitivity scan of the relative-MSE threshold $\eta_{\text{split}} \in \{0.06, 0.09, 0.12\}$ yields F_1 scores of 0.9069, 0.9101, and 0.9114, with false-alarm rates of 0.1083, 0.0750, and 0.0667 per frame; the frozen value $\eta_{\text{split}} = 0.09$ was retained to avoid post-hoc retuning. One-factor-at-a-time validation tests for the grouping distance d_{close} (1.4 and 2.2 bins) and near-candidate gates (0.35 and 0.55 bins) alter the F_1 score by at most 0.0029 and the false-alarm rate by at most 0.0167 per frame. Furthermore, enabling the forced-absorption rule triggers 245 times

across 120 validation frames, but only one trigger (0.41%) corresponds to a true dense weak target. It increases runtime from 1.448 to 2.169 s/frame without reducing false alarms, justifying its exclusion.

Table 3. CNN training and detector hyperparameters.

Parameter	Value Used	Provenance
CNN candidate extractor (training)		
Input maps	2	architecture; real/imaginary maps
Dropout, blocks 1–2	0.2	training design
Loss coefficients ($\alpha, \gamma, \beta, \lambda_{\text{off}}$)	0.9, 2, 0.5, 10	development split
Optimizer	Adam	training design
Learning rate	$1 \times 10^{-4} \rightarrow 1 \times 10^{-7}$, cosine	training design
Weight decay	1×10^{-4}	training design
Max. epochs; patience	100; 20	training design
Batch; train/validation	32; 80/20	fixed split
Training seed	20,260,807	predefined
Detection, refinement, and order control (inference)		
Existence threshold	0.01	development split
Grouping (d_{close}; $K_{\max}$; attempts)	1.8 bins; 8; 40	development split
Empty-scene threshold	0.66	seed-offset 50,000 development only
Single refinement	15 steps; LR 0.02; early stop 3	development split
Joint refinement	30 steps; LR 0.01; early stop 8	development split
Gain refresh	every 2 Adam steps	algorithmic design
Gain regularization λ_h	1×10^{-4} base; max 0.025	development split
Separation (ρ; λ_{sep})	0.15 bin; 0.03	development split
Near-candidate gates A/B	0.45 bin, 0.50; 1.8 bins, 0.28	development split
Fission seed offset δ_{seed}	$\pm$0.15 bin, four directions	initialization; validation scan
Relative-MSE threshold η_{split}	0.09	development split; validation scan
Split-pair gates	0.25 bin; amp. ratio 0.50 or 0.25	development split
Stopping gain ratio	∞	frozen implementation

The detector parameters in Table 3 are fixed using the normalized delay–Doppler resolution as the reference scale; thus, geometric quantities such as grouping distance, splitting offsets, and separation radius are expressed in resolution cells. The empty-scene gate is tuned on a dedicated development split, while the remaining detection, refinement, and splitting parameters remain fixed. For unknown-order control, nine NAD-Hybrid configurations are evaluated on an independent validation set, using the criterion specified in Section 4.1. The frozen parameters are applied without condition-specific retuning across the evaluated SNR range (10–18 dB), target counts, and RCS variations (0–10 dB). Substantially different conditions, such as altered delay–Doppler resolution, noise characteristics, clutter, or hardware impairments, may require offline recalibration on a representative development set; the resolution-normalized formulation facilitates such recalibration.

4.1. Scenes and Data Split

Three original scenes are used when K is known and $s_{\text{RCS}} = 0$ dB. Scene 1 contains one dense pair and one separated target. Scene 2 contains one dense pair and three separated targets with a wider nominal dynamic range. Scene 3 contains three fully separated targets. In the code and later experiments, `dense3` denotes Scene 1 and `mixed5` denotes Scene 2; it does not mean that all three or all five targets are mutually dense.

Unknown-order experiments are separated into independent validation and final blind testing (Table 4). Validation uses a random-seed offset of 100,000. Nine hybrid rescue/pruning configurations are compared with NAD and NAD-BIC; the frozen configuration maximizes overall F_1 subject to at most 0.12 false alarms/frame, with ties resolved by lower false alarms and then runtime. Blind testing uses a previously unseen seed offset of 200,000. Each scene–RCS dataset has 1000 frames: 200 at each SNR of 10, 12, 14, 16, and

18 dB. Paired differences use 5000 stratified bootstrap replications for 95% confidence intervals and 5000 paired-label permutations for significance tests. A separate 2100-frame stress test randomizes $K \in \{0, \ldots, 6\}$, sub-resolution offset orientation (Uniform$[0, 2\pi)$), offset magnitude (0.3–0.9 bins), and SNR over five values; the $K = 0$ frames explicitly test the empty-scene gate. In training, $K = 1$–5 frames with log-normal RCS coefficients are used; the dense-pair separation is ≈0.45 bins in both scenes, and same-cell collisions account for 32.74% of training frames.

Table 4. Independent validation and final blind-test protocol.

Stage	Scenes	RCS SD (dB)	Frames	Purpose
Validation	`dense3`, `mixed5`	0, 6, 10	6×100	Select/freeze hybrid parameters
Blind test	`dense3`, `mixed5`	0, 3, 6, 10	8×1000	Compare three internal variants

To quantify the train–inference distribution shift introduced by iterative residual cancellation, we evaluate the candidate generator on the original received mixture and on residuals produced using both oracle and mismatched previous-stage cancellations. As shown in Table 5, the candidate hit probability decreases progressively with successive cancellation. On oracle residuals, the hit probability decreases from 1.000 at the original input to 0.858, 0.538, and 0.390 at stages 1–3, respectively. With mismatched cancellation, the corresponding values are 0.895, 0.407, and 0.223. These results confirm a residual-domain distribution shift between the training input and later inference stages. Residual-augmented fine-tuning did not yield a material end-to-end improvement; therefore, we disclose this limitation rather than claiming that the distribution shift has been eliminated.

Table 5. Stage-wise diagnostic of CNN candidate extraction under original and residual-domain inputs.

Residual	Stage	Frames	Hit Prob.	Mean Peak Prob.	Median Nearest Dist.
Original	0	600	1.000	0.931	0.049
Mismatched	1	600	0.895	0.789	0.125
Oracle	1	600	0.858	0.735	0.094
Mismatched	2	600	0.407	0.697	3.459
Oracle	2	600	0.538	0.552	0.132
Mismatched	3	300	0.223	0.656	4.732
Oracle	3	300	0.390	0.528	3.829

4.2. Metrics

Predictions and truths are assigned one-to-one using the Hungarian algorithm after range and velocity errors are normalized by their resolutions. Only assignments within a one-resolution-cell gate count as correct. For frame i, let TP_i, FP_i, and FN_i denote matched targets, unmatched predictions, and unmatched truths. With $TP = \sum_i TP_i$, $FP = \sum_i FP_i$, $FN = \sum_i FN_i$, and N_f frames,

$$P_D = \frac{TP}{TP + FN} \qquad \text{Precision} = \frac{TP}{TP + FP} \tag{17}$$

$$F_1 = \frac{2TP}{2TP + FP + FN} \qquad \overline{N}_{\text{FA}} = \frac{1}{N_f} \sum_{i=1}^{N_f} FP_i \tag{18}$$

$$\text{Acc}_{\text{count}} = \frac{1}{N_f} \sum_{i=1}^{N_f} \mathbb{1}\left(\hat{K}_i = K_i\right) \tag{19}$$

To aggregate performance across the C evaluation conditions (scenes and RCS settings), a macro-averaged F_1 score is computed as $\bar{F}_1 = \frac{1}{C}\sum_{c=1}^{C} F_1^{(c)}$, where $F_1^{(c)}$ is the condition-level F_1 score calculated via (18) for condition c. For the successful matching set $\mathcal{M}$,

$$\mathrm{RMSE}_R = \sqrt{\frac{1}{TP}\sum_{(i,j)\in\mathcal{M}}(\widehat{R}_{ij} - R_{ij})^2} \quad \mathrm{RMSE}_v = \sqrt{\frac{1}{TP}\sum_{(i,j)\in\mathcal{M}}(\widehat{v}_{ij} - v_{ij})^2} \tag{20}$$

In Equation (19), K_i denotes the true target counts of frame i, and for each matched pair $(i,j) \in \mathcal{M}$, $\widehat{R}_{ij}$ and $\widehat{v}_{ij}$ are the range and radial velocity obtained by mapping the refined delay–Doppler estimate of the j-th matched prediction in frame i through Equation (2). Accordingly, P_D measures missed detections; precision and false alarms/frame quantify over-detection; F_1 summarizes their trade-off; exact-count accuracy evaluates frame-level model order; and conditional RMSE measures localization only after a successful match. The latter must therefore be interpreted together with the detection metrics: a detector that retains only easy strong targets can obtain a deceptively small RMSE. In addition, exact-frame accuracy, used in the known-order comparison, is the fraction of frames in which every true target is matched within the one-resolution-cell gate and no unmatched estimate remains.

4.3. External Baselines

The following baselines are independently implemented and run on identical frames with identical matching and scoring. The suffixes "Oracle" and "BIC" specify model-order control in this study; they are not names used in the cited original papers.

- **SIC+PIC-Oracle:** Successive interference cancellation estimates, reconstructs, and removes the current strongest echo; parallel interference cancellation then reconstructs other components from the previous iteration and updates each target. The classical SIC and PIC principles follow [26,27], adapted here to the OTFS echo model. True K is supplied at inference, so this is a known-order upper-bound configuration.
- **OMPFR-Oracle:** OMPFR-Oracle independently reproduces the support recovery and fractional delay/Doppler refinement mechanism of [21]. True K is supplied, providing a known-order sparse-recovery upper bound.
- **OMPFR-BIC:** OMPFR-BIC retains the OMPFR support and continuous-refinement mechanism but does not receive true K; BIC is added in this study to compare candidate orders. This constructed practical version is the direct unknown-order external baseline for NAD-Hybrid.

OMPFR-original applies the published OMPFR-style candidate generator with its original stopping rule. OMPFR-BIC replaces that stopping rule by BIC order selection. OMPFR-Hybrid uses the same OMPFR candidate source but passes the candidates through the same practical Hybrid order controller used by the proposed method. This four-method comparison separates the contribution of candidate generation from that of order control. All four methods are evaluated on identical blind frames under the same one-resolution-cell matching rule.

The two-dimensional atomic-norm ADMM approach in [18] was not assigned a numerical result. At $M = 64, N = 32$, a direct implementation requires repeated semidefinite projections involving a 2049-order Hermitian matrix. Without author code and adequate high-performance resources, a reduced proxy would risk misrepresenting the published method.

5. Results and Discussion

5.1. Known Target Number and No Additional RCS Mismatch

Table 6 reports a strictly known-order comparison. Both OMPFR-Oracle and SIC+PIC-Oracle receive the true target count K and are scored with the same one-resolution-cell assignment. In dense3, SIC+PIC-Oracle yields slightly lower conditional RMSE, whereas OMPFR-Oracle gives higher exact-frame accuracy. In mixed5, OMPFR-Oracle is better in both exact-frame accuracy and conditional RMSE. These oracle-K results are not directly ranked against the proposed unknown-order detector; the latter is evaluated in the blind comparisons below.

Table 6. Unified known-order comparison at RCS SD = 0 dB, averaged over SNR = 10, 12, 14, 16, and 18 dB. Exact-frame accuracy means that every true target is correctly matched and no unmatched estimate remains.

Scene	Known-K Method	Range RMSE (m)	Velocity RMSE (m/s)	Exact-Frame Accuracy
dense3	OMPFR–Oracle	0.934	0.441	0.995
dense3	SIC+PIC–Oracle	0.871	0.395	0.958
mixed5	OMPFR–Oracle	0.816	0.418	0.997
mixed5	SIC+PIC–Oracle	0.873	0.429	0.900

Runtime also favors the neural-assisted pipeline in the reported implementation: Scene 1 decreases from 977 to 281 s (71.2%), and Scene 3 decreases from 445 to 198 s (55.5%). These are implementation-level measurements rather than hardware-independent complexity claims.

5.2. Internal Unknown-Order Comparison

Table 7 expands all eight blind-test conditions. With no additional RCS mismatch, the conservative NAD already has very high F_1 and NAD-BIC rescue can introduce false alarms. At moderate mismatch (3–6 dB), NAD-Hybrid yields the highest F_1 in both scenes. At 10 dB, the more aggressive NAD-BIC attains slightly higher F_1, but its false-alarm rate is approximately 1.9–2.8 times that of NAD-Hybrid. Hence the NAD-Hybrid version is preferable when precision, low false alarms, and a stable target set matter; NAD-BIC alone can be preferable when severe mismatch makes detection probability the dominant objective.

Table 7. Blind-test F_1 and false alarms/frame for the three internal variants.

Scene	RCS SD	NAD $\bar{F}_1$	NAD-BIC $\bar{F}_1$	NAD-Hybrid $\bar{F}_1$	NAD-BIC FA/Frame	NAD-Hybrid FA/Frame
dense3	0	1.0000	0.9640	0.9846	0.224	0.094
dense3	3	0.9664	0.9516	0.9689	0.207	0.101
dense3	6	0.9033	0.9210	0.9307	0.154	0.085
dense3	10	0.8166	0.8654	0.8641	0.132	0.068
mixed5	0	0.9931	0.9685	0.9838	0.323	0.142
mixed5	3	0.9552	0.9591	0.9750	0.320	0.119
mixed5	6	0.8680	0.9228	0.9285	0.190	0.082
mixed5	10	0.7669	0.8381	0.8348	0.131	0.046

Paired statistics on the same 8000 frames are reported in Table 8. Relative to NAD-BIC, NAD-Hybrid reduces false alarms by 0.118/frame (56.2%) and improves exact-count accuracy by 7.45 percentage points, while losing 0.70 percentage points in P_D. Both $\bar{F}_1$ improvements are statistically significant; the 95% intervals do not cross zero and paired-label permutation tests give $p = 0.0002$ (reported as $p < 0.001$). The $\bar{F}_1$ gain over NAD-BIC is +0.998 pp; the small p-value reflects the 8000-frame sample rather than a large effect size.

Table 8. Macro-averaged paired blind-test differences (NAD-Hybrid minus comparator).

Comparator	ΔP_D (pp)	ΔF_1 (pp), 95% CI	ΔFA/Frame	ΔCount Accuracy (pp)	$\bar{F}_1$ Permutation p
NAD	+5.918	+2.512 [2.324, 2.700]	+0.0879	+7.500	<0.001
NAD-BIC	−0.698	+0.998 [0.875, 1.121]	−0.1180	+7.450	<0.001

5.3. External Baseline Comparison

Table 9 shows that NAD-Hybrid exceeds the unknown-order OMPFR-BIC in five of eight scene–RCS conditions: dense3 at 0, 3, and 6 dB, and mixed5 at 0 and 3 dB. OMPFR-BIC becomes superior for dense3 at 10 dB and mixed5 at 6–10 dB because aggressive support expansion recovers more weak targets. Across all eight datasets, macro F_1 is 0.9338 for NAD-Hybrid and 0.9180 for OMPFR-BIC. Oracle methods receive true K and must not be interpreted as fair deployment comparisons.

Table 9. External-baseline F_1 by scene and RCS condition. An asterisk denotes true-K input.

Scene	RCS SD	NAD-Hybrid	OMPFR–BIC	OMPFR–Oracle *	SIC+PIC–Oracle *
dense3	0	0.9846	0.8933	0.9983	0.9850
dense3	3	0.9689	0.8967	0.9997	0.9560
dense3	6	0.9307	0.8971	0.9827	0.9277
dense3	10	0.8641	0.8838	0.9353	0.8940
mixed5	0	0.9838	0.9547	0.9994	0.9798
mixed5	3	0.9750	0.9540	0.9996	0.9786
mixed5	6	0.9285	0.9523	0.9884	0.9504
mixed5	10	0.8348	0.9120	0.9070	0.8754

Across all unknown-order methods, the proposed Hybrid detector (summarized in Table 10) achieves the highest precision (0.9743), the highest $\bar{F}_1$ score (0.9338), the lowest false-alarm rate (0.0921 per frame), and the highest exact-count accuracy (0.6664). OMPFR-Hybrid and OMPFR-BIC obtain higher detection probabilities (0.9787 and 0.9825) and lower conditional RMSEs, but they generate 0.4980 and 0.5746 false alarms per frame, respectively. Thus, the proposed method should be described as a conservative low-false-alarm operating point rather than as uniformly superior in every metric.

Table 10. Macro-averaged blind-test comparison of four unknown-order methods over 8000 frames. R-RMSE and V-RMSE are the conditional range (m) and velocity (m/s) RMSEs; s/fr. denotes seconds per frame.

Method	Prec.	P_D	$\bar{F}_1$	FA/fr.	Count	R-RMSE	V-RMSE	s/fr.
OMPFR-BIC	0.8631	0.9825	0.9180	0.5746	0.4818	0.794	0.359	1.825
OMPFR-Hybrid	0.8810	0.9787	0.9266	0.4980	0.5355	0.790	0.361	2.648
OMPFR-original	0.6369	0.9813	0.7685	2.1409	0.2442	0.769	0.349	3.594
Proposed-Hybrid	0.9743	0.9022	0.9338	0.0921	0.6664	1.098	0.528	1.295

Conditional localization in Table 11 favors OMPFR-BIC for successfully matched targets: it has lower range and velocity RMSE in seven of eight conditions. The exceptions are mixed5/10 dB for range and dense3/0 dB for velocity. This is a real advantage of fractional sparse refinement. It does not, however, imply a better complete detector because OMPFR-BIC also outputs substantially more false targets.

Table 11. Conditional range/velocity RMSE for unknown-K methods. Each entry is range RMSE (m)/velocity RMSE (m/s).

Scene	RCS SD	NAD-Hybrid	OMPFR-BIC
dense3	0	0.671/0.307	0.646/0.311
dense3	3	0.947/0.435	0.648/0.311
dense3	6	1.323/0.620	0.699/0.342
dense3	10	1.453/0.695	1.315/0.589
mixed5	0	0.949/0.511	0.616/0.272
mixed5	3	1.036/0.540	0.601/0.278
mixed5	6	1.267/0.568	0.648/0.284
mixed5	10	1.136/0.548	1.175/0.485

The low-false-alarm distinction is explicit in Table 12: NAD-Hybrid produces fewer false alarms in all eight conditions. A paired bootstrap with 10,000 frame-level resamples within each of the eight blind datasets, macro-averaged across datasets (the same aggregation as the displayed performance tables), gives a NAD-Hybrid minus OMPFR-BIC precision difference of +0.1112 (95% CI [+0.1077, +0.1146]), a P_D difference of −0.0803 [−0.0833, −0.0772], a $\bar{F}_1$ difference of +0.0158 [+0.0131, +0.0184], an FA/frame difference of −0.4825 [−0.4980, −0.4670], and an exact-count difference of +0.1846 [+0.1706, +0.1990]. The $\bar{F}_1$ interval center therefore coincides exactly with the difference of the $\bar{F}_1$ values in Table 10 (0.9338−0.9180). Thus, the method exchanges detection probability for fewer false alarms and more reliable frame-level cardinality.

Table 12. Direct unknown-K comparison of F_1 and false alarms/frame.

Scene	RCS SD	NAD-Hybrid $\bar{F}_1$	OMPFR–BIC $\bar{F}_1$	NAD-Hybrid FA/Frame	OMPFR–BIC FA/Frame
dense3	0	0.9846	0.8933	0.094	0.717
dense3	3	0.9689	0.8967	0.101	0.691
dense3	6	0.9307	0.8971	0.085	0.655
dense3	10	0.8641	0.8838	0.068	0.607
mixed5	0	0.9838	0.9547	0.142	0.474
mixed5	3	0.9750	0.9540	0.119	0.482
mixed5	6	0.9285	0.9523	0.082	0.467
mixed5	10	0.8348	0.9120	0.046	0.504

This operating point is relevant to persistent airborne or UAV OTFS-ISAC sensing. A frame-level detection set may feed track initiation, multi-frame association, cooperative fusion, and beam-revisit scheduling. When a large aircraft and a small UAV occupy adjacent DD cells with substantial RCS mismatch, each sidelobe-induced false candidate can initiate a spurious tentative track, enlarge the association problem, and trigger an unnecessary revisit. In tasks where intermittent misses can be repaired by subsequent frames or multi-sensor fusion but false-track cost is high, high precision and a conservative target set have direct system value.

Table 13 extends the evaluation to randomized geometries and cardinalities. At RCS SD 0/6/10 dB the $\bar{F}_1$ scores are 0.9244/0.8975/0.8168, and the empty-scene correct-rejection probabilities are 1.000/0.970/0.970. A development-tuned 2-D CA-CFAR reference ($P_{FA} = 10^{-6}$) yields fewer false alarms (0.0235/frame) but far lower P_D (0.5404) and count accuracy (0.0130), confirming that the proposed detector retains high target recovery at a low-false-alarm operating point.

Table 13. Variable-cardinality stress test (2100 frames) and CA-CFAR reference.

Condition/Method	Frames	Prec.	P_D	$\bar{F}_1$	FA/fr.	Count
Proposed (RCS 0 dB)	700	0.9089	0.9405	0.9244	0.2829	0.7000
Proposed (RCS 6 dB)	700	0.9179	0.8781	0.8975	0.2357	0.6729
Proposed (RCS 10 dB)	700	0.9129	0.7390	0.8168	0.2114	0.4986
CA-CFAR ($P_{FA} = 10^{-6}$)	2100	0.9874	0.5404	0.6980	0.0235	0.0130

5.4. Communication Feasibility

To leverage the spectral and energy efficiency advantages of index modulation in the delay–Doppler domain, the communication experiment reuses the $M = 64$, $N = 32$ shared index-modulated OTFS frame. The center data region activates one of every four delay–Doppler positions; two index bits and one BPSK bit are conveyed per group. The outer region uses four-choice indexing and QPSK. The revised communication experiment enforces a fixed average transmit-frame energy over the complete 64×32 grid. One pilot and 430 active payload symbols are used. The pilot-to-data power ratio is 10 dB, corresponding to normalized data and pilot amplitudes of 2.157 and 6.822, respectively. The corrected frame accounting comprises 2048 physical bins: 430 active payload symbols, 1 pilot, and 369 pilot/guard reserved bins (including the pilot). Index-modulation inactive positions carry index information and are not additional physical cells; the payload consists of 778 index bits and 676 symbol bits, totaling 1454 bits per frame. The mean, 95th-percentile, and maximum PAPR values of the transmitted frame are 9.077 dB, 10.243 dB, and 12.256 dB, respectively. The communication receiver is a remote downlink terminal, distinct from the colocated sensing branch. It observes an independently generated four-path, integer-delay/integer-Doppler cyclic OTFS channel. For each frame, the four path delays are uniformly drawn from 0–20 delay bins and the Doppler indices from -2 to 2 bins, while the complex Gaussian path gains are normalized to unit total power. The channel is therefore redrawn independently for every frame. The receiver estimates the one-way communication channel from the embedded pilot and applies two-dimensional FFT linear-MMSE equalization followed by groupwise index/BPSK/QPSK decisions. RCS perturbations are applied only to the target-dependent sensing echoes and are not added to this independent communication link.

Figure 7 presents the overall uncoded BER versus data-link SNR for perfect CSI and embedded-pilot CSI. For perfect CSI, the BER at 6, 8, and 10 dB is 5.42×10^{-3}, 7.37×10^{-4}, and 3.78×10^{-5}, respectively. For embedded-pilot CSI, the BER at 10, 12, and 14 dB is 2.63×10^{-2}, 6.52×10^{-3}, and approximately 9.10×10^{-4}, respectively. As expected, the BER decreases as the data-link SNR increases. The experiment demonstrates basic uncoded communication feasibility using an embedded pilot; channel coding and iterative/message-passing reception are outside the scope of this dense-sensing study.

5.5. Limitations and Reproducibility Boundary

The evidence supports a robust detection–false-alarm trade-off under unknown K and inter-target RCS mismatch, not superiority in every metric. OMPFR-BIC is more sensitive under severe mismatch and gives better conditional localization in most conditions; oracle methods benefit from true-K input. The current active splitter is limited to one dense pair, training/testing covers mainly one to five targets, all results are simulated, and clutter, hardware impairments, and measured echoes remain untested. Because the study employs a normalized baseband model to evaluate algorithmic trade-offs, no end-to-end hardware link budget or specific RF front-end is claimed. The OMPFR baselines reproduce the algorithmic mechanism described in [21] rather than line-by-line author code. Compared with OMPFR-BIC, NAD-Hybrid provides a more conservative low-false-alarm operating

point and more stable target-count estimation, but this comes at the cost of detection probability and conditional localization accuracy under severe RCS mismatch. Therefore, the proposed method is intended for operating conditions where reliable low-false-alarm detection and unknown-order control are prioritized over maximum detection sensitivity.

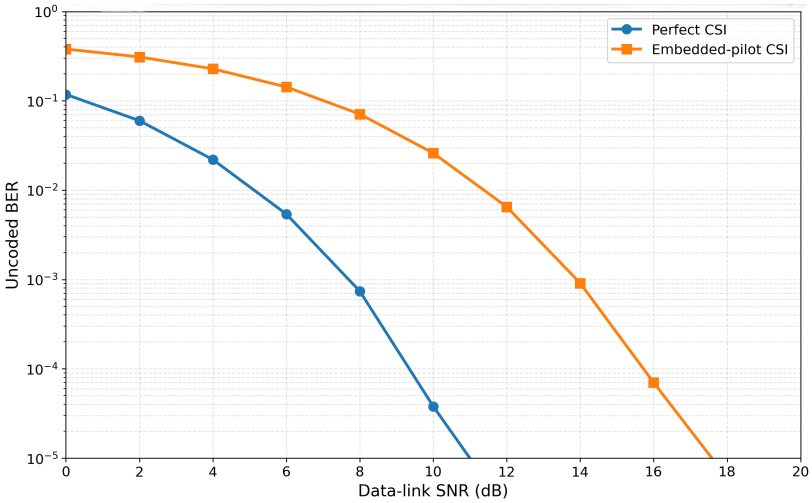


Figure 7. Uncoded BER of the shared IM-OTFS frame (10 dB pilot-to-data ratio).

6. Conclusions

This paper presented a hybrid OTFS sensing detector for dense targets with unknown cardinality. The frozen blind test shows that the method primarily improves precision, false-alarm control, and exact-count reliability rather than recall under every condition. Compared with OMPFR-BIC, the proposed detector increases macro-averaged F_1 by 0.0158 and reduces false alarms by 0.4825 per frame, while sacrificing 0.0803 in detection probability. Random-K full-orientation experiments further show graceful degradation as RCS variation increases, with precision remaining above 0.90 but recall decreasing at 10 dB RCS standard deviation. The method is therefore most suitable for applications in which false tracks, unnecessary confirmation actions, or overestimated target sets are costly. Limitations include residual-distribution shift after repeated cancellation, reduced weak-target recall under severe RCS mismatch, and the absence of an end-to-end hardware link budget. Future work will investigate residual-aware candidate learning, adaptive operating-point selection, and measured-channel validation.

Author Contributions: Conceptualization, Z.G.; methodology, I.Y., K.M. and Z.G.; software, I.Y. and K.M.; validation, I.Y. and K.M.; formal analysis, I.Y.; investigation, K.M.; data curation, K.M.; writing—original draft preparation, I.Y.; writing—review and editing, I.Y. and Z.G.; supervision, Z.G.; project administration, Z.G.; funding acquisition, Z.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by Open Research Fund of CEMEE under Grant 2023K0201, and the Project of Natural Science Foundation of China under Grant 62071367.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The simulation data, source code, frozen evaluation protocols, and per-frame results supporting this study are available from the corresponding author upon reasonable request.

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

Abbreviations

The following abbreviations are used in this manuscript:

NAD	neural-assisted detector
NAD-BIC	neural-assisted detector with BIC order control
NAD-Hybrid	neural-assisted detector with hybrid information-criterion order control
BIC	bayesian information criterion
BER	bit-error rate
CNN	convolutional neural network
DD	delay–Doppler
FA	false alarm
IM	index modulation
ISAC	integrated sensing and communications
OMPFR	orthogonal matching pursuit with fractional refinement
OTFS	orthogonal time frequency space
PIC	parallel interference cancellation
RCS	radar cross section
RMSE	root mean squared error
SIC	successive interference cancellation
SNR	signal to noise ratio
ADMM	alternating direction method of multipliers

References

1. Liu, F.; Cui, Y.; Masouros, C.; Xu, J.; Han, T.X.; Eldar, Y.C.; Buzzi, S. Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 1728–1767. [CrossRef]
2. Wei, X.; Yuan, W.; Zhang, K.; Liu, F. OTFS-assisted ISAC system: Delay Doppler channel estimation and SDR-based implementation. *IEEE Trans. Mob. Comput.* **2025**, *24*, 11865–11878. [CrossRef]
3. Gaudio, L.; Kobayashi, M.; Caire, G.; Colavolpe, G. On the Effectiveness of OTFS for Joint Radar Parameter Estimation and Communication. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 5951–5965. [CrossRef]
4. Yuan, W.; Zhou, L.; Dehkordi, S.K.; Li, S.; Fan, P.; Caire, G.; Poor, H.V. From OTFS to DD-ISAC: Integrating sensing and communications in the delay Doppler domain. *IEEE Wirel. Commun.* **2024**, *31*, 152–160. [CrossRef]
5. Chu, T.M.C.; Zepernick, H.J.; Hüük, A.; Westerhagen, A.; Granbom, B. OTFS modulation for non-terrestrial networks: Concepts, applications, benefits, and challenges. In *Proceedings of the 2023 16th International Conference on Signal Processing and Communication System (ICSPCS)*; IEEE: New York, NY, USA, 2023; pp. 1–10. [CrossRef]
6. Buzzi, S.; Caire, G.; Colavolpe, G.; D'Andrea, C.; Foggi, T.; Piemontese, A.; Ugolini, A. LEO satellite diversity in 6G non-terrestrial networks: OFDM vs. OTFS. *IEEE Commun. Lett.* **2023**, *27*, 3013–3017. [CrossRef]
7. Shi, J.; Li, Z.; Hu, J.; Tie, Z.; Li, S.; Liang, W.; Ding, Z. OTFS enabled LEO satellite communications: A promising solution to severe Doppler effects. *IEEE Netw.* **2023**, *38*, 203–209. [CrossRef]
8. Xiao, Y.; Du, Q.; Zhao, Z.; Li, B.; Tang, X.; Zhang, S.; Song, H. RF-based identification framework against unauthorized UAV networking in low-altitude economy. *IEEE Trans. Netw. Sci. Eng.* **2026**, *13*, 6538–6555. [CrossRef]
9. Chong, R.; Mohammadi, M.; Ngo, H.Q.; Cotton, S.L.; Matthaiou, M. Multiple target detection in OTFS-ISAC. In *Proceedings of the ICC 2025-IEEE International Conference on Communications*; IEEE: New York, NY, USA, 2025; pp. 6735–6740. [CrossRef]
10. Sruti, S.; Kumar, A.A.; Giridhar, K. Joint association and localization of multiple targets in an OTFS-Based multistatic ISAC system. *IEEE Trans. Aerosp. Electron. Syst.* **2025**, *61*, 13459–13473. [CrossRef]
11. Cui, K.; Zhao, J.; Jia, W.; Jin, W.; He, F.; Hu, H.; Zhang, F. Multi-Target Sensing with Interference Analysis for the Multi-UAV-Enabled ISAC Network. *Electronics* **2025**, *14*, 3984. [CrossRef]
12. Kang, S.w.; Jang, M.h.; Lee, S. Autoencoder-based target detection in automotive MIMO FMCW radar system. *Sensors* **2022**, *22*, 5552. [CrossRef] [PubMed]
13. Li, P.; Li, M.; Liu, R.; Liu, Q.; Swindlehurst, A.L. Low range-Doppler sidelobe ISAC waveform design: A low-complexity approach. *IEEE Trans. Veh. Technol.* **2025**, *74*, 16544–16549. [CrossRef]
14. Elster, L.; Holder, M.F.; Rapp, M. A dataset for radar scattering characteristics of vehicles under real-world driving conditions: Major findings for sensor simulation. *IEEE Sens. J.* **2023**, *23*, 4873–4882. [CrossRef]
15. Hussain, K.; Ali, E.M.; Hussain, W.; Raza, A.; Elkamchouchi, D.H. Robust OTFS-ISAC for vehicular-to-base station end-to-end sensing and communication. *Electronics* **2025**, *14*, 4340. [CrossRef]

16. Gaudio, L.; Kobayashi, M.; Caire, G.; Colavolpe, G. Joint Radar Target Detection and Parameter Estimation with MIMO OTFS. In Proceedings of the 2020 IEEE Radar Conference (RadarConf20), Florence, Italy, 21–25 September 2020; pp. 1–6. [CrossRef]
17. Singh, S.; Nakkeeran, A.; Singh, P.; Sharma, E.; Bapat, J. Target detection for OTFS-aided cell-free MIMO ISAC system. *IEEE Trans. Veh. Technol.* **2025**, *74*, 11568–11573. [CrossRef]
18. Liu, S.; Zhang, H.; Li, L.; Gong, Z.; Huang, Y.; Yuan, J. Super-Resolution Delay–Doppler Estimation for OTFS-Based Automotive Radar. *Signal Process.* **2024**, *224*, 109596. [CrossRef]
19. Zegrar, S.A.E.; Haif, H.; Arslan, H. OTFS-Based ISAC for Super-Resolution Range–Velocity Profile. *IEEE Trans. Commun.* **2024**, *72*, 3934–3946. [CrossRef]
20. Zhang, Z.; Shang, X.; Xiao, Y. Target Parameter Estimation for OTFS-Integrated Radar and Communications Based on Sparse Reconstruction Preprocessing. *Front. Inf. Technol. Electron. Eng.* **2024**, *25*, 742–754. [CrossRef]
21. Zacharia, O.; Devi, M.V. Fractional Delay and Doppler Estimation for OTFS Based ISAC Systems. In Proceedings of the 2023 IEEE Wireless Communications and Networking Conference (WCNC), Glasgow, UK, 26–29 March 2023; pp. 2274–2279. [CrossRef]
22. Zhang, K.; Li, Z.; Yuan, W.; Cai, Y.; Gao, F. Radar Sensing via OTFS Signaling. *China Commun.* **2023**, *20*, 34–45. [CrossRef]
23. Zhang, X.; Ni, Z.; Yuan, W.; Zhang, J.A.; Quek, T.Q. Deep learning-based OTFS channel estimation and symbol detection with plug-and-play framework. Deep learning-based OTFS channel estimation and symbol detection with plug-and-play framework. *arXiv* **2025**, arXiv:2503.11102. [CrossRef]
24. Marchese, M.; Keskin, M.F.; Savazzi, P.; Wymeersch, H. Exploiting Structural Sparsity and Delay-Doppler Decoupling for Low-Complexity OTFS-ISAC Receivers. *arXiv* **2025**, arXiv:2504.20659. [CrossRef]
25. Zhang, Y.; Zhang, J.; Gong, H.; Hu, X.; Zhang, J.; Xing, H.; Luo, S.; Xiong, Y.; Yu, L.; Yuan, Z.; et al. A unified RCS modeling of typical targets for 3GPP ISAC channel standardization and experimental analysis. *IEEE J. Sel. Areas Commun.* **2025**, *44*, 702–716. [CrossRef]
26. Patel, P.R.; Holtzman, J.M. Analysis of a Simple Successive Interference Cancellation Scheme in a DS/CDMA System. *IEEE J. Sel. Areas Commun.* **1994**, *12*, 796–807. [CrossRef]
27. Divsalar, D.; Simon, M.K.; Raphaeli, D. Improved Parallel Interference Cancellation for CDMA. *IEEE Trans. Commun.* **1998**, *46*, 258–268. [CrossRef]

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.