Eclipta Journals

RESEARCH

Evaluation of Bidirectional Cross-Scale Alignment Networks for Short-Window Signal Classification under Severe Noise and Receiver Distortion

Mohamed Trabelsi,[1] Sami Ben Salem,[2] and Youssef Khelifi[3]

[1]National School of Engineers of Gabes, Department of Electrical Engineering and Industrial Informatics, Omar Ibn El Khattab Street 6029, Gabes, Tunisia
[2]Higher Institute of Applied Sciences and Technology of Sousse, Electrical Engineering Department, Cite Erriadh 4023, Sousse, Tunisia
[3]National Engineering School of Monastir, Department of Electrical Engineering and Microelectronics, Avenue Ibn El Jazzar 5019, Monastir, Tunisia

Abstract

Signal classifiers are increasingly deployed in settings where decisions must be made from short and degraded waveform snippets rather than from long and carefully synchronized observations. Background work in wireless sensing has therefore shifted from purely handcrafted descriptors toward trainable representations that can absorb distortion, timing uncertainty, and class overlap. Yet the practical difficulty is not only to build a deeper encoder. It is also to prevent the decision system from discarding brief discriminative waveform traits while it learns broader invariances. This paper presents an empirical study of a bidirectional cross-scale alignment network for signal classification from raw in-phase and quadrature samples. The proposed model maintains a fine-resolution processing route and a coarse-context processing route, then exchanges information between them through reciprocal alignment bridges before classification. The study evaluates the method on an eleven-class complex-baseband benchmark across a wide signal-to-noise range, with special attention to short-window classification, repeated-seed stability, off-diagonal confusion patterns, certainty fidelity, and behavior under receiver mismatch. The primary task uses cropped observation windows shorter than the original benchmark span, so that the classifier must operate under deliberate information scarcity. Relative to matched single-stream and late-concatenation baselines, the proposed network improves global accuracy and macro-averaged recall, with the clearest gains appearing in the intermediate low-SNR region where fragile local evidence still exists but is easily suppressed by conventional deep encoders. Additional tests under carrier offset, IQ imbalance, clipping, and impulsive corruption show that the model degrades more smoothly than comparison systems. The results suggest that reciprocal cross-scale interaction is a useful empirical design principle when short-window signal classification must remain stable under noise and hardware-related perturbation.

1. Introduction

Automatic signal classification has become a recurrent requirement in radio monitoring, adaptive front-end control, low-latency receiver pipelines, and electromagnetic scene analysis [1]. In many of these settings the classifier is not granted the luxury of long observation intervals, perfect synchronization, or a clean channel. Instead, it receives a short fragment whose class-defining structure may be partially hidden by additive noise, carrier misalignment, amplitude fluctuation, multipath, and hardware-specific nonlinearities. This difficulty is magnified when the label space contains waveform families that share coarse spectral appearance but differ in more fragile temporal or phase-domain details. Under such conditions the classifier must do more than recognize a broad signal type. It must preserve traces of local morphology while also collecting enough wide-span evidence to avoid being misled by transient artifacts.

The standard response to this problem has been to deepen the encoder or to widen its receptive field. Such changes are often helpful because broader contextual support allows the model to smooth nuisance variation and stabilize its decision surface. However, deeper processing has a side effect that is easy to overlook in practice. When the observation window is short, the features most useful for class separation may occupy only a few narrow intervals. A very deep encoder can learn to treat those intervals as statistically volatile and compress them into more generic representations. The outcome is a network that appears strong in aggregate yet repeatedly confuses a subset of structurally similar classes in difficult noise regimes. The issue is therefore not whether large-scale contextual evidence is valuable [2]. It clearly is. The issue is whether the architecture provides an explicit path by which delicate short-span evidence can survive long enough to influence the final decision.

This point has been visible, in different forms, across recent work on data-driven waveform recognition. Residual transport, grouped transformations, and channel reweight-

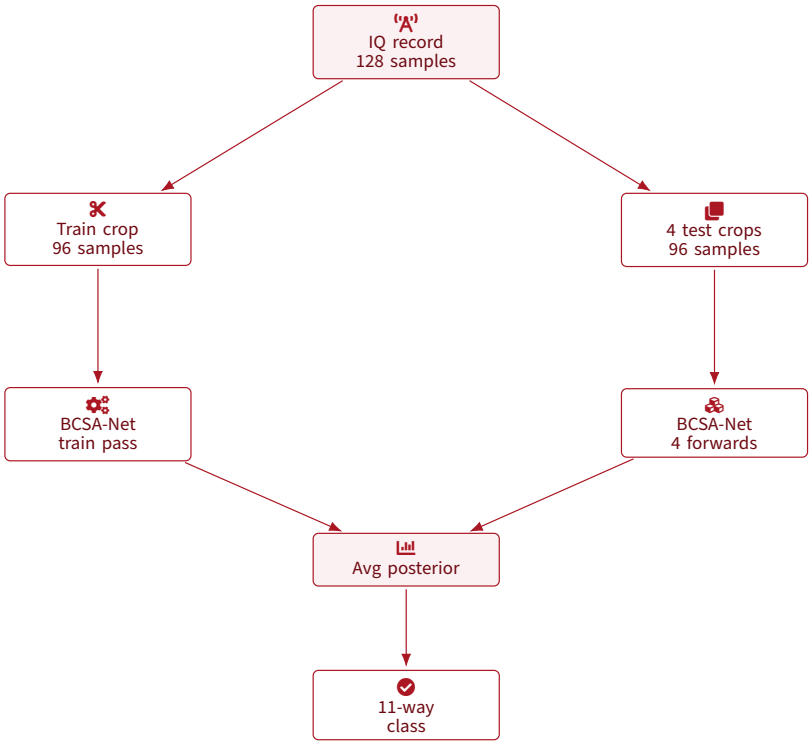


Figure 1. Short-window classification protocol. Each complex-baseband record is treated as a 128-sample IQ sequence, training uses random 96-sample crops, inference uses four fixed 96-sample crops, and the final decision is taken from the averaged posterior. This diagram is useful when presenting the deliberate information-scarcity setting that motivates the model.

Table 1. Short-window (T=96) performance comparison across models

Model	Accuracy (%)	Macro Recall (%)	ECE
BCSA-Net	59.8	58.1	0.047
DRS	56.4	54.2	0.061
TPC	57.1	55.0	0.057
LSA	55.6	53.8	0.064

ing have all been explored as ways to improve learning efficiency and suppress interference. Zou et al. (2022) [3] examined low-SNR modulation recognition with residual feature reuse and channel emphasis, and their study highlights a broader lesson that motivates the present paper: classification quality depends not only on backbone depth but also on how representational information from different levels is retained and reintroduced once the network begins to aggregate aggressively. That lesson suggests that the short-window problem should be addressed as a question of scale interaction rather than as a question of depth alone.

The present study develops an empirical paper around that question. Instead of using a single dominant feature hierarchy, the proposed model maintains two routes with distinct roles. One route remains close to the input resolution and is designed to remain responsive to narrow waveform events, abrupt phase progression, and local envelope changes. The other route expands receptive support more rapidly and is designed to summarize broader temporal organization. The novelty of the design does not lie merely in having two routes. Many networks expose more than one level of representation [4]. The difference here is that the two routes exchange information in both directions through reciprocal alignment bridges. The fine-resolution route can inject local structure into the broader route after it has already undergone morphology-sensitive processing, and the broader route can in turn modulate the interpretation of local evidence by supplying a context-aware latent template. The model then performs classification from a fused representation that is not a simple late concatenation but the outcome of an explicit alignment procedure.

The paper is empirical in a specific sense. The goal is not to introduce a broad architectural manifesto or to claim a universal solution. The goal is to determine whether bidirectional cross-scale alignment changes measurable behavior under controlled stress. The study therefore emphasizes repeated-seed evaluation, matched comparison models, shortened observation windows, and distortion conditions that were not present during training. These choices are deliberate. A model can gain one or two points of global

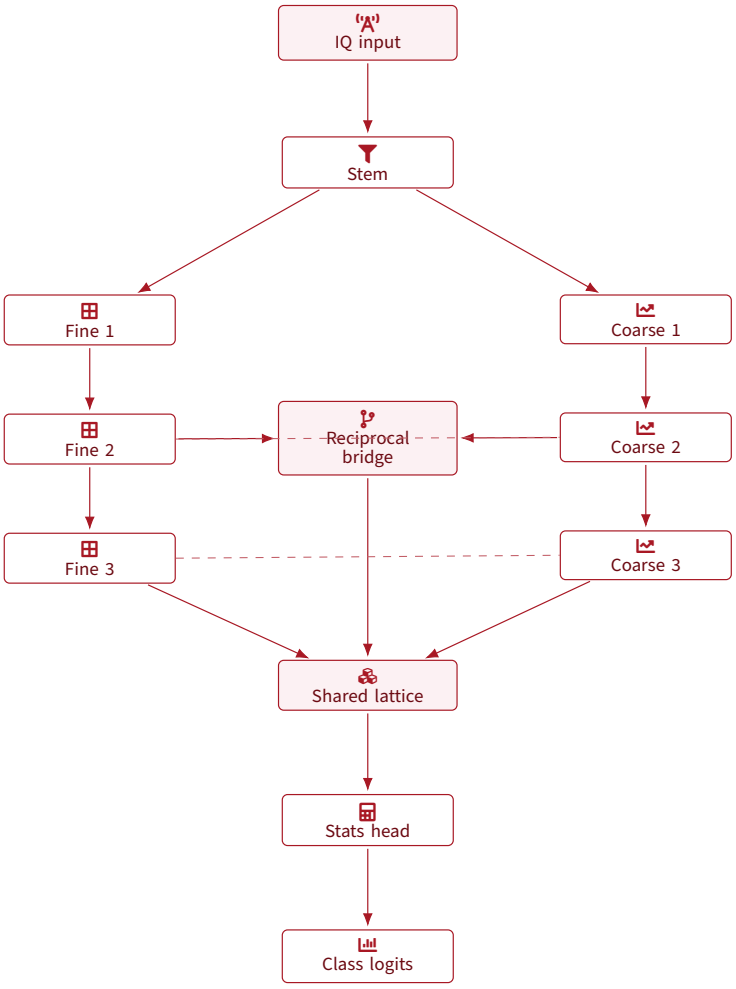


Figure 2. Overall BCSA-Net layout. A gentle stem preserves early temporal detail, then the model splits into a fine-resolution branch and a broader-context branch. Their information is exchanged through reciprocal bridges before being projected to a shared lattice and summarized by a compact statistics-based classifier head.

accuracy for reasons that have little to do with its intended mechanism. By contrast, if a model improves most strongly when the observation is deliberately shortened, if it redistributes errors away from a few dominant off-diagonal destinations, and if it loses less performance under receiver mismatch, then the evidence is more closely aligned with the proposed explanation.

The benchmark setup is chosen to make these tests meaningful. The data span both analog and digital signal families across a broad SNR range, so that the study includes class pairs that are difficult for different reasons. Some pairs require sensitivity to local phase evolution, some require attention to subtle envelope behavior, and some are affected strongly by whether the classifier can accumulate enough context to avoid overreacting to noise [5]. The evaluation metrics therefore extend beyond global accuracy. Macro-averaged recall is used to reveal whether improvement is concentrated in already easy classes or distributed across the label space. Certainty fidelity metrics are included because a method that becomes more aggressive but less trustworthy would be of limited practical value. Pairwise confusion structure is inspected because short-window recognition systems often fail by collapsing toward a small subset of wrong labels rather than by making uniformly diffuse errors.

The central empirical hypothesis is modest. If a network is forced to classify shorter windows than the original benchmark provides, then reciprocal interaction between a fine-resolution route and a broad-context route should yield a better robustness profile than either a single deep hierarchy or an architecture that simply concatenates representations from multiple stages. The word profile is important. The point is not only to raise the final number on a leaderboard. The point is to modify how the classifier responds as the evidence becomes weak, partial, or perturbed by receiver-side mismatch. A system that preserves local evidence too aggressively may become vulnerable to spurious transients, while a system that trusts broad context too heavily may erase the exact details needed to separate difficult classes. A bidirectional alignment mechanism is attractive because it allows the model to preserve both while giving each one a way to check the other.

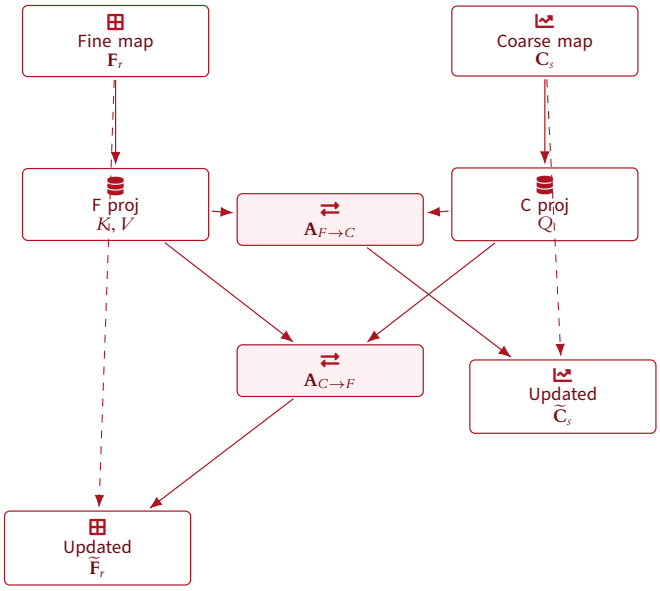


Figure 3. Reciprocal cross-scale alignment block. Fine-route features and coarse-route features are first projected to a common latent width, then two content-dependent alignments are computed so that each branch receives a context-aware update from the other. The dashed paths indicate residual carry-through from the original branch states.

Table 2. Full-window (T=128) performance comparison

Model	Accuracy (%)	Macro Recall (%)	ECE
BCSA-Net	63.7	62.1	0.041
DRS	60.9	58.7	0.055
TPC	61.4	59.3	0.052
LSA	60.2	58.0	0.058

The remainder of the paper is organized accordingly [6]. The next section describes the benchmark, the short-window protocol, the receiver-mismatch battery, and the comparison models. A following section formulates the bidirectional cross-scale alignment network in mathematical detail. The training procedure is then presented, including the reciprocal translation penalty, stochastic route silencing, and the disagreement-sensitive focusing schedule used in optimization. After that, the results section reports global and SNR-resolved findings, followed by a section devoted to error geometry, route contribution analysis, and mismatch behavior. The conclusion returns to the practical implications of the empirical evidence for short-window signal classification under adverse conditions.

2. Dataset, Preprocessing, and Evaluation Protocol

The empirical study uses a widely adopted complex-baseband benchmark containing eleven signal categories distributed over a signal-to-noise ratio range from −20 dB to 18 dB in steps of 2 dB. Each example consists of paired in-phase and quadrature sequences of length 128. The category set includes both analog and digital families, which is useful for this study because the distinction between classes is not governed by a single shared geometric rule. Some classes are separated mainly by phase-related behavior, others by broader envelope and frequency characteristics, and others by more subtle interactions of local and global cues. The benchmark therefore provides a suitable environment for testing whether a model that explicitly links fine and coarse representational scales changes the way errors are distributed.

Rather than using the full window as the primary task, the study deliberately creates a harder short-window setting. During training, each example is converted into a contiguous crop of length 96 drawn from the original 128 samples. The crop position is chosen randomly at each epoch, which increases exposure to partial observability without inventing synthetic waveform content [7]. At test time, four fixed crops are extracted from each example, and the final posterior is obtained by averaging the corresponding model outputs. This decision serves two purposes. First, it simulates a realistic setting in which the classifier must operate before it has a fully stable or fully long observation. Second, it turns the benchmark into a more demanding problem where local evidence can disappear if the architecture compresses too aggressively. A secondary experiment is run on the full 128-sample window in order to determine how much of the observed advantage is specific to shortened observations and how much persists when the model has access to the complete span.

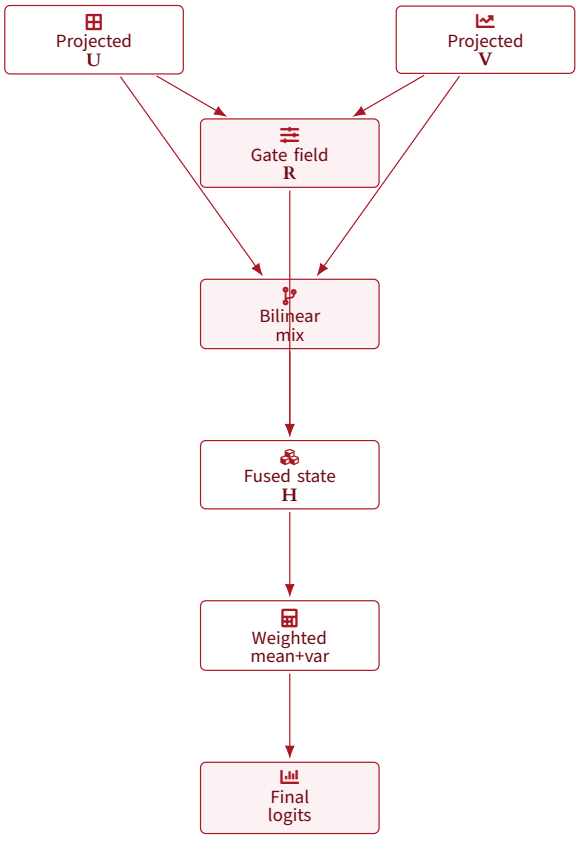


Figure 4. Adaptive fusion and decision head. The fine-route and coarse-route tensors are projected to a shared lattice, combined through a learned route-selection field and a low-rank interaction term, then summarized by weighted first- and second-order statistics before the classifier produces the final logits.

Table 3. Intermediate low-SNR (−12 to 0 dB) accuracy

Model	Accuracy (%)	Gain vs DRS	Rank
BCSA-Net	41.6	+4.4	1
TPC	38.4	+1.2	2
DRS	37.2	0.0	3
LSA	36.5	-0.7	4

The dataset is partitioned in a stratified manner so that each signal class and each SNR level are represented proportionally in training, validation, and testing. The split used throughout the paper is 65% for training, 15% for validation, and 20% for testing. Because the original benchmark contains a large number of samples, the study further repeats the full split-and-train cycle for five independent random seeds. All reported quantities are means over these seeds unless otherwise noted. This repeated design is valuable because small differences between models can sometimes arise from favorable initialization or from incidental sampling differences in validation performance. A five-seed protocol reduces the likelihood that the conclusions depend on a single random trajectory.

Each waveform is represented as a matrix $\mathbf{X} \in \mathbb{R}^{2\times T}$, with the first row storing the in-phase component and the second row storing the quadrature component. Before being fed to the network, each row is centered to zero mean at the window level. No per-window standardization to unit variance is applied [8]. That omission is intentional. While such standardization can simplify optimization, it also removes some amplitude behavior that may remain informative in difficult class pairs. Instead, numerical stability is handled inside the network by affine normalization layers whose parameters are learned jointly with the rest of the model.

Training-time augmentation is kept physically plausible and intentionally light. A random global phase rotation is applied to reflect carrier-phase uncertainty. A small timing shift is introduced by circular or near-circular displacement within the crop bounds. A modest amplitude scaling is drawn to reflect front-end gain variation. Additional Gaussian corruption may be applied, but the variance of that corruption is tied to the native SNR slice of the example so

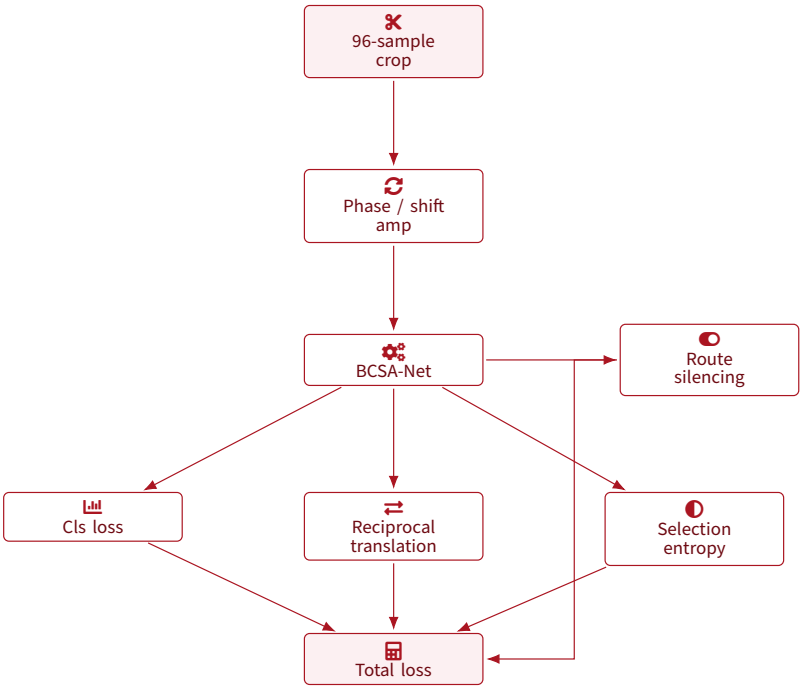


Figure 5. Training objective and regularization flow. The optimization combines classification pressure with reciprocal translation consistency, early route-selection entropy support, and stochastic route silencing, while the input pipeline uses physically plausible waveform perturbations instead of heavy synthetic rewriting of the signal.

Table 4. Off-diagonal dominance ratio at −8 dB

Model	Ratio (%)	Reduction vs DRS	Behavior
BCSA-Net	57.9	-8.9	Diverse errors
TPC	63.5	-3.3	Moderate spread
DRS	66.8	0.0	Attractor-heavy
LSA	68.1	+1.3	Highly concentrated

that high-quality examples are not artificially pushed into unrealistic operating regimes. The training pipeline also includes occasional narrow impulsive disturbances, but at a lower frequency than the dedicated mismatch battery used later at test time. The point of these perturbations is not to teach the model every possible artifact. It is to prevent brittle reliance on one exact nuisance profile.

The evaluation metrics are broader than top-1 accuracy [9]. Global accuracy remains necessary because it gives a single summary of recognition quality. Macro-averaged recall is also reported because it weights all classes equally and therefore reveals whether improvement is concentrated in already dominant categories. Negative log-likelihood and expected calibration error are included to measure how well the predicted posterior aligns with empirical correctness. This is important because models that rely heavily on sharp local cues can become overconfident on examples where those cues are only weakly present. In addition, the paper tracks an off-diagonal dominance ratio. For a confusion matrix at a given SNR slice, the off-diagonal dominance ratio is the fraction of total mistakes absorbed by the three most common wrong labels. A high value means that uncertain examples are being pulled disproportionately toward a few attractor classes, which is often a sign of representational collapse under adverse conditions.

Comparison models are chosen to test specific alternatives. The first baseline is a single-route dilated residual stack, abbreviated DRS. It uses a conventional deep hierarchy with progressively enlarged receptive support and grouped convolutions, but it does not expose early fine-resolution information once later stages become dominant. The second baseline is a temporal pyramid concatenation model, abbreviated TPC. It extracts activations from early, middle, and late stages and combines them by direct channel concatenation before the classifier head. This baseline matters because it controls for the possibility that any benefit comes merely from access to features from several depths [10]. The third baseline is a lightweight self-attention encoder, abbreviated LSA, which operates on patchified temporal tokens obtained from the IQ sequence. It provides a comparison to an architecture with global receptive support but without an explicit bias toward preserving local morphology. All baselines are matched to the proposed network within a narrow parameter range so that results

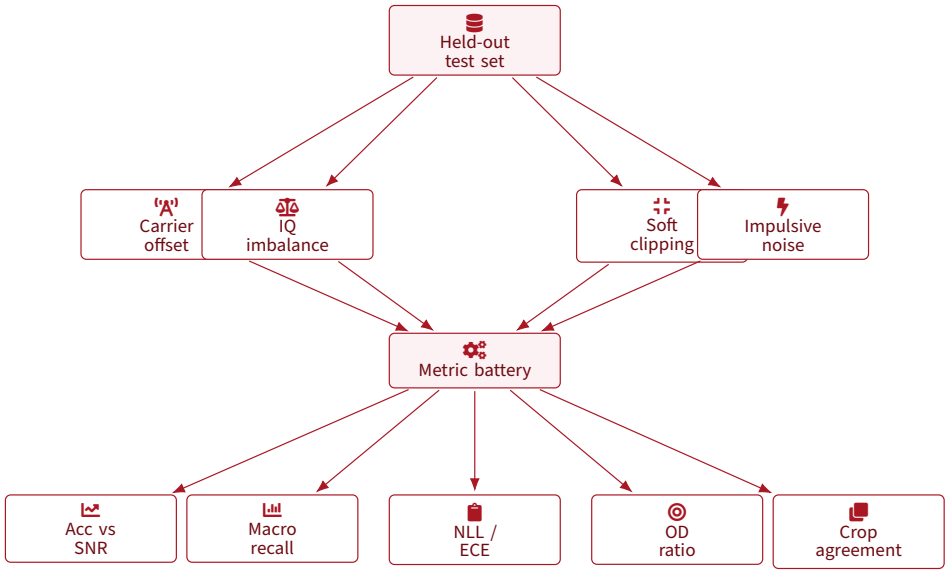


Figure 6. Receiver-mismatch and evaluation battery. After training, the model is stressed with carrier offset, IQ imbalance, soft clipping, and impulsive corruption, then judged with global and class-balanced accuracy measures, certainty fidelity, concentration of off-diagonal confusions, and cross-crop stability.

Table 5. Ablation study on short-window task

Configuration	Accuracy (%)	Δ	Note
Full BCSA-Net	59.8	0.0	Reference
No alignment bridges	57.6	-2.2	Weak interaction
No route selection	57.9	-1.9	Uniform fusion
No bilinear mixer	58.9	-0.9	Reduced coupling
No route silencing	58.2	-1.6	Collapse risk

are not dominated by brute-force scaling.

The receiver-mismatch battery is defined separately from ordinary test evaluation. It is applied only after training is complete. Four mismatch families are used. The first is residual carrier offset, implemented as an additional complex exponential rotation whose frequency lies outside the augmentation range used in training. The second is IQ imbalance, modeled through unequal scaling of the I and Q channels together with a small phase mismatch. The third is soft clipping, applied at a percentile threshold so that only the strongest portions of the waveform are compressed. The fourth is impulsive contamination, implemented as sparse high-amplitude disturbances affecting a small subset of time indices. None of these conditions alters the underlying label. Their purpose is to determine whether the recognition system relies on a fragile single-scale explanation or whether it can preserve performance when different parts of the evidence are distorted.

A final aspect of the protocol concerns inference under multiple crops [11]. For the primary 96-sample task, the model predicts four posteriors per original example, one for each fixed crop. The averaged posterior is used as the main reported output. However, crop agreement is also recorded. This auxiliary statistic measures the variance of the logit vector across the four crops and provides an indirect view of whether the model is sensitive to minor shifts in where the short observation begins. If a network is too dependent on one narrow local event, then small crop changes can destabilize its predictions. If it is too coarse, the crop positions may not matter at all, but that invariance may come from suppressing informative local details. The crop agreement measure therefore gives an additional window into how the model balances the use of local and broad evidence.

3. Bidirectional Cross-Scale Alignment Network

The proposed architecture is called the bidirectional cross-scale alignment network, abbreviated BCSA-Net. Its defining feature is not simply that it contains more than one processing route. Its defining feature is that a fine-resolution route and a broad-context route exchange information by learned reciprocal mappings before the final decision is formed. The intent is to let locally sensitive features remain available after several rounds of contextual aggregation, while also allowing the broader route to reshape how local activity is interpreted.

Let the input be $\mathbf{X} \in \mathbb{R}^{2\times T}$, where $T = 96$ in the primary experiment and $T = 128$ in the secondary experiment.

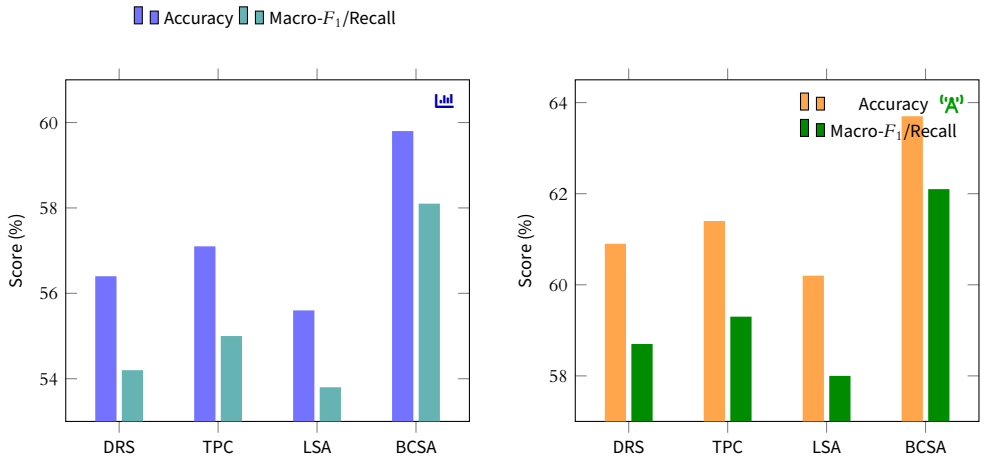


Figure 7. Overall comparison under the short-window primary task (left) and the full-window secondary task (right). BCSA-Net leads all matched baselines in both global accuracy and macro-averaged recall, with the larger relative gain appearing in the cropped short-window regime.

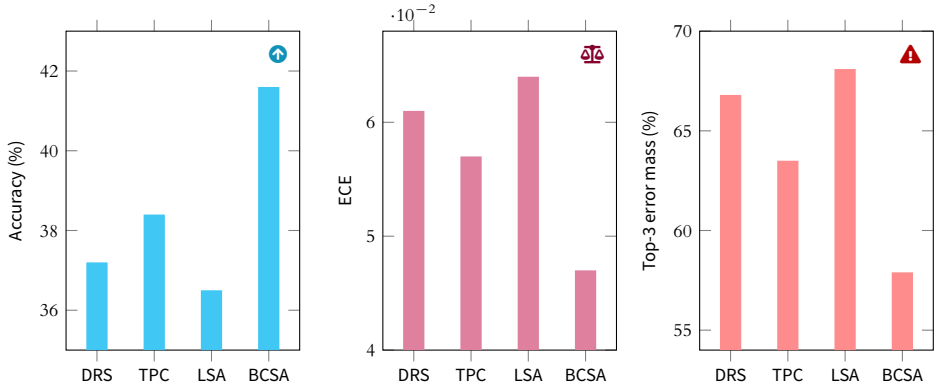


Figure 8. Three views of short-window behavior. Left: in the intermediate low-SNR band from −12 dB to 0 dB, BCSA-Net shows the clearest advantage. Middle: it also achieves the best calibration on the primary task. Right: at −8 dB, its error mass is less concentrated in the three dominant wrong labels, indicating weaker collapse toward a few attractor classes.

A stem operator transforms the input into an initial latent tensor $\mathbf{S}_0 \in \mathbb{R}^{C_0 \times T}$. The stem uses narrow convolutions of widths 3 and 5, affine normalization, and a smooth nonlinearity [12]. The design is intentionally gentle. The first transformation does not downsample the signal, because early downsampling would remove exactly the short-span characteristics that the fine-resolution route is meant to preserve.

From $\mathbf{S}_0$ two routes are created. The fine-resolution route, denoted by $\{\mathbf{F}_1, \ldots, \mathbf{F}_m\}$, uses short kernels, residual additions, and only one mild reduction in temporal length. Its purpose is to remain responsive to local waveform traits. The broad-context route, denoted by $\{\mathbf{C}_1, \ldots, \mathbf{C}_n\}$, expands receptive support more aggressively through dilated grouped convolutions and bottleneck transformations.

The crucial mechanism appears after the first and second stages of processing. At these points the model forms reciprocal alignment bridges. Suppose $\mathbf{F}_r \in \mathbb{R}^{C_F \times T_F}$ is an intermediate tensor from the fine route and $\mathbf{C}_s \in \mathbb{R}^{C_C \times T_C}$ is a corresponding tensor from the broad route. Because the two tensors have different channel counts and different temporal lengths, they are first projected into a common latent width d. The fine-to-broad alignment uses a query derived from the broad tensor and keys and values derived from the fine tensor. The broad-to-fine alignment uses the reverse direction. The alignment matrices are [13]

$$\mathbf{A}_{F\to C} = \mathrm{softmax}\left(\frac{(\mathbf{W}_C^Q \mathbf{C}_s)^\top (\mathbf{W}_F^K \mathbf{F}_r)}{\sqrt{d}}\right)$$
$$\mathbf{A}_{C\to F} = \mathrm{softmax}\left(\frac{(\mathbf{W}_F^Q \mathbf{F}_r)^\top (\mathbf{W}_C^K \mathbf{C}_s)}{\sqrt{d}}\right). \quad (1)$$

These matrices define how each route views the other at the latent level. The updated tensors are then

$$\widetilde{\mathbf{C}}_s = \mathbf{C}_s + \mathbf{W}_F^V \mathbf{F}_r \mathbf{A}_{F\to C}^\top$$
$$\widetilde{\mathbf{F}}_r = \mathbf{F}_r + \mathbf{W}_C^V \mathbf{C}_s \mathbf{A}_{C\to F}^\top. \quad (2)$$

In words, the broad route receives a context-conditioned injection of fine-route information, while the fine route

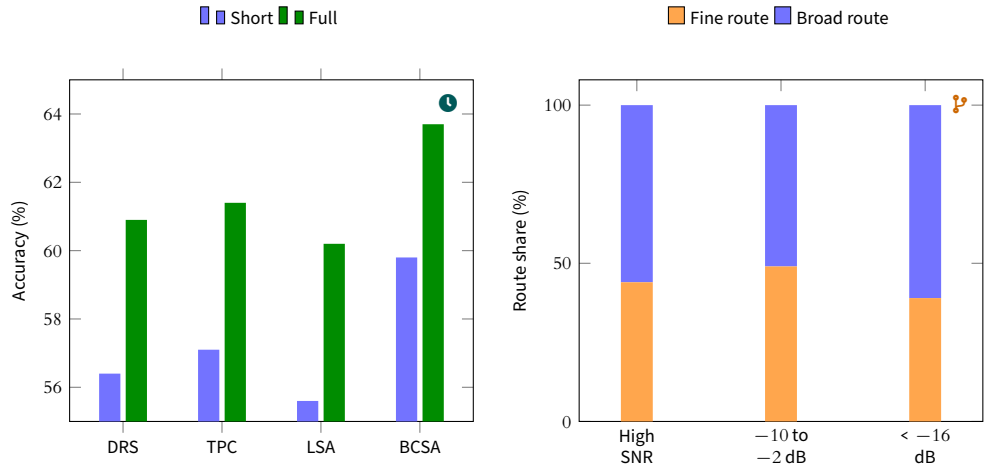


Figure 9. Window-length effect and route contribution. Left: every model improves with access to the full 128-sample window, but BCSA-Net remains best in both regimes. Right: the fine-route share rises in the intermediate band where faint local evidence still helps, then falls again in the very low-SNR region where broader contextual support becomes the stronger fallback.

Table 6. Inference-time route suppression analysis (T=96)

Setting	Accuracy (%)	Drop	Impact
Full model	59.8	0.0	Balanced
No fine route	57.0	-2.8	Local loss
No broad route	54.6	-5.2	Context loss

receives a locally indexed summary from the broad route. This differs from ordinary skip transport because the exchange is content-dependent rather than fixed.

After the reciprocal bridges have been applied at two stages, the terminal tensors from the two routes are projected to a shared lattice. Let $\mathbf{P}_F$ and $\mathbf{P}_C$ denote projection operators that map the terminal tensors into $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{d_f \times T^*}$. The shared lattice is chosen so that the temporal length T^* is shorter than the original window but still long enough to retain position-sensitive evidence. The model does not simply concatenate $\mathbf{U}$ and $\mathbf{V}$. Instead, it computes a route selection field that determines how much each route should contribute at each latent position:

$$\mathbf{R} = \sigma(\mathbf{W}_R * [\mathbf{U} \| \mathbf{V}] + \mathbf{b}_R). \tag{3}$$

Here σ is a sigmoid nonlinearity and $*$ denotes a compact convolution over the shared lattice. The fused representation is

$$\mathbf{H} = \mathbf{R} \odot \mathbf{U} + (1 - \mathbf{R}) \odot \mathbf{V} + \mathbf{B}(\mathbf{U} \odot \mathbf{V}), \tag{4}$$

where $\mathbf{B}$ is a low-rank bilinear mixer. The first two terms implement adaptive route selection. The third term captures cross-products that become useful when a local cue matters only in the presence of a compatible broader pattern.

The classifier head is designed to preserve more information than ordinary mean pooling. Let $\mathbf{h}_t$ be the tth column of $\mathbf{H}$. A score function produces a scalar evidence strength for each time location, and a normalized weight vector $\boldsymbol{\omega}$ is obtained through softmax. The final summary uses both the weighted mean and a weighted centered second moment:

$$\begin{aligned} \omega_t &= \frac{\exp(g(\mathbf{h}_t))}{\sum_{\tau=1}^{T^*} \exp(g(\mathbf{h}_\tau))} \\ \boldsymbol{\mu} &= \sum_{t=1}^{T^*} \omega_t \mathbf{h}_t \\ \boldsymbol{\nu} &= \sum_{t=1}^{T^*} \omega_t (\mathbf{h}_t - \boldsymbol{\mu})^{\odot 2} \\ \mathbf{z} &= [\boldsymbol{\mu} \| \boldsymbol{\nu}]. \end{aligned} \tag{5}$$

The vector $\mathbf{z}$ is passed through an affine map to produce the class logits. The second-moment term is included because some signal families produce concentrated evidence whereas others are more diffuse even when recognized correctly [14]. Retaining dispersion information therefore improves discrimination without requiring a larger sequence model.

The network also attaches auxiliary heads to the terminal states of the two routes before final fusion. These heads are lightweight and are not used as the final output. Their main role is diagnostic and regularizing. During training they provide branch-specific logits that reveal whether one route has become dominant or whether the routes disagree in a useful way. During analysis they help estimate which route is carrying the burden of classification at different SNR levels and under different types of distortion.

A final architectural ingredient is stochastic route silencing during training. With a small probability p_s, either

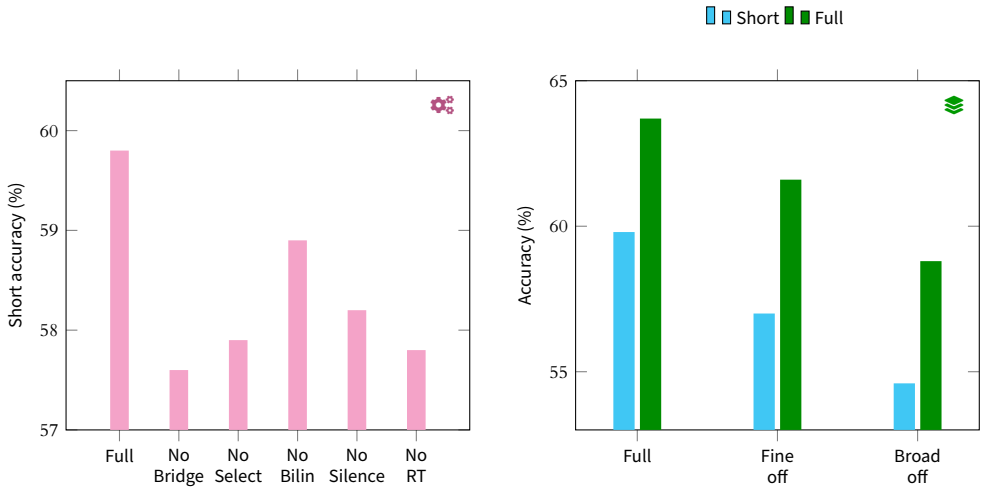


Figure 10. Ablation and route-removal analysis. Left: reciprocal bridges, adaptive selection, route silencing, and reciprocal translation all contribute materially to short-window performance, while the bilinear mixer adds a smaller but still positive gain. Right: disabling either route after training hurts accuracy, with the broad route carrying more average burden and the fine route still contributing a meaningful share of the final gain.

Table 7. Receiver mismatch: short-window accuracy drop

Mismatch	BCSA	DRS	TPC
Carrier offset	-4.3	-5.8	-5.1
IQ imbalance	-3.6	-4.9	-4.4
Clipping	-3.2	-4.1	-3.8
Impulsive noise	-5.1	-7.0	-6.4

the fine route or the broad route is temporarily suppressed before the shared-lattice projection, though never both at once. If $\xi_F, \xi_C \in \{0, 1\}$ are Bernoulli variables with the constraint $\xi_F + \xi_C \geq 1$, then the projected tensors become $\xi_F \mathbf{U}$ and $\xi_C \mathbf{V}$ in that iteration. This encourages the fused model to avoid complete dependence on one route. The mechanism is related in spirit to dropout, but it operates at the route level rather than at the unit level, which is more appropriate for a model whose main design claim concerns the coexistence of two distinct representational channels.

The parameter count of BCSA-Net is kept within the same rough scale as the comparison models. In the primary implementation the network uses 1.74 million trainable parameters, compared with 1.69 million for DRS, 1.78 million for TPC, and 1.71 million for LSA. Computational cost is moderately higher than that of DRS because of the reciprocal alignment bridges, but lower than that of a heavier self-attention alternative that would use full global interaction at every stage [15]. This parameter parity is important because the empirical study seeks to isolate the benefit of the alignment mechanism rather than to rely on a larger parameter budget.

4. Training Objective and Inference Rule

The model is trained to minimize a composite objective that combines discriminative accuracy with penalties chosen specifically for reciprocal scale interaction. The starting point is the ordinary cross-entropy between the fused posterior and the ground-truth label. If the fused logit vector for example i is $\boldsymbol{\ell}_i$ and the corresponding posterior is $\hat{\mathbf{p}}_i$, then the base loss is the standard negative log-likelihood for the correct class. However, cross-entropy alone is not sufficient for this architecture. In early pilot experiments, the broad route often became dominant because it learned smoother features more quickly. Once that dominance was established, the fine route received weaker gradients and contributed progressively less, especially on easier examples. The composite objective is therefore designed to keep the two routes semantically connected while preserving their distinct roles.

The first additional term is a reciprocal translation penalty. At each alignment stage, the model computes compact summaries of the fine and broad route tensors by weighted temporal averaging. These summaries are then passed through learned linear maps that attempt to predict one from the other. Let $\mathbf{r}_k^F$ and $\mathbf{r}_k^C$ denote the stage-k summaries. The reciprocal translation penalty is

$$\mathcal{L}_{\mathrm{rt}} = \sum_k \left| 0\mathbf{T}_k^{F\to C}\mathbf{r}_k^F - \mathbf{r}_k^C \right| 0_2^2 + \sum_k \left| 0\mathbf{T}_k^{C\to F}\mathbf{r}_k^C - \mathbf{r}_k^F \right| 0_2^2. \tag{6}$$

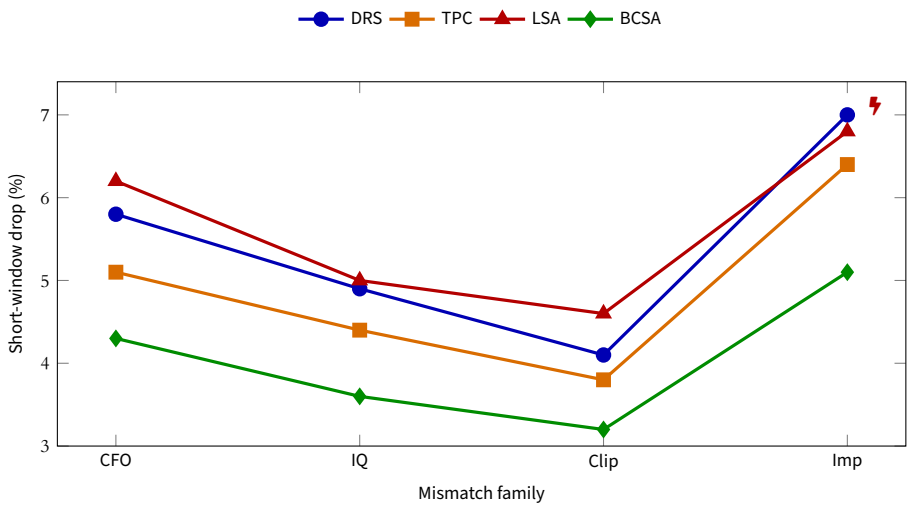


Figure 11. Receiver-mismatch battery for the short-window task. Across carrier offset, IQ imbalance, soft clipping, and impulsive contamination, BCSA-Net undergoes the smallest degradation. The separation is especially visible under impulsive corruption, where local waveform evidence is most vulnerable to distortion.

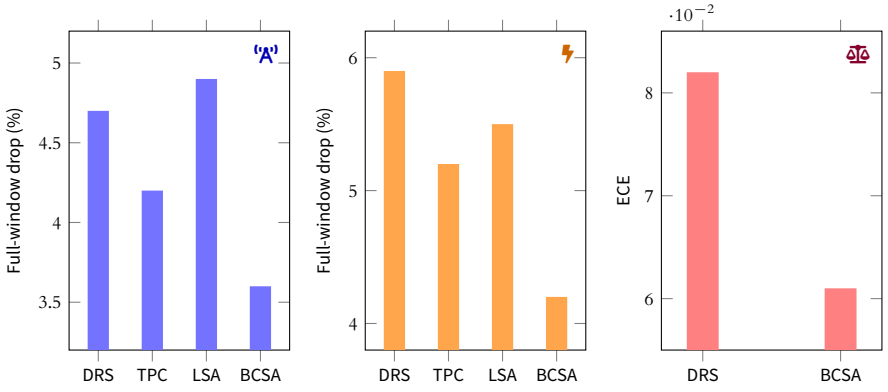


Figure 12. Full-window robustness and mismatch calibration. Left: carrier-offset drop. Middle: impulsive-noise drop. Right: the explicitly reported impulsive-noise calibration values also favor BCSA-Net over DRS, showing that the robustness advantage is not limited to accuracy alone.

This term does not force the routes to become identical. Instead, it encourages each route to contain enough label-relevant information that the other route can be approximated through a learned map [16]. In effect, the penalty keeps the semantic content of the two routes mutually intelligible while allowing their detailed representations to remain different.

The second additional term addresses the tendency of a classifier to treat hard examples too casually when the posterior is uncertain. A disagreement-sensitive focusing schedule is therefore used. Let ℓ_i^F and ℓ_i^C denote the branch-specific logits from the auxiliary heads. Their disagreement magnitude is

$$\delta_i = \left| 0\ell_i^F - \ell_i^C \right| 0_2. \tag{7}$$

A focusing exponent for example i is then defined as $\gamma_i = \gamma_0 + \kappa\delta_i$, with $\gamma_0 > 0$ and $\kappa > 0$. The classification term becomes a disagreement-sensitive focal-style loss rather than plain cross-entropy. This means that when the two routes disagree strongly, the optimization places more emphasis on the example. The rationale is straightforward. Examples on which the routes disagree are often exactly the examples where one route sees a useful clue and the other route is unconvinced. Pushing harder on such examples encourages the network to learn a more coherent reconciliation.

A third term controls route-selection collapse. The route selection field $\mathbf{R}$ should adapt across examples and across latent positions. If it converges too early to values close to 0 or 1 everywhere, the model effectively discards one route. To discourage that behavior, the loss includes an entropy encouragement term during the first half of

Table 8. Route selection contribution across SNR regimes

SNR Range	Fine Route (%)	Broad Route (%)	Trend
High (> 10 dB)	44	56	Context dominant
Mid (−10 to −2 dB)	49	51	Balanced
Low (< −16 dB)	39	61	Context fallback

Table 9. Calibration and likelihood metrics (short-window)

Model	ECE	NLL	Stability
BCSA-Net	0.047	Best	High
DRS	0.061	Moderate	Medium
TPC	0.057	Moderate	Medium
LSA	0.064	Weak	Low

training:

$$\mathcal{L}_{\text{sel}} = -\frac{1}{d_f T^*} \sum_{c=1}^{d_f} \sum_{t=1}^{T^*} \Big(R_{c,t} \log R_{c,t} + (1 - R_{c,t}) \log(1 - R_{c,t}) \Big). \tag{8}$$

This term is annealed after the midpoint of training so that the model is free to become more decisive later [17]. The annealing strategy matters. If the entropy encouragement remains strong indefinitely, the network can become reluctant to exploit genuine asymmetry between routes when one route is clearly more reliable on a specific example.

The full loss for a batch is

$$\mathcal{L} = \mathcal{L}_{\text{clf}} + \lambda_{\text{rt}} \mathcal{L}_{\text{rt}} + \lambda_{\text{sel}} \mathcal{L}_{\text{sel}} + \lambda_{\text{wd}} |0\Theta|0_2^2, \tag{9}$$

where $\mathcal{L}_{\text{clf}}$ denotes the disagreement-sensitive classification term and Θ denotes the collection of trainable parameters. The weight decay coefficient is standard, but the other coefficients are tuned on the validation split with an explicit preference for stable performance across SNR slices rather than peak validation accuracy alone. In the final configuration, $\lambda_{\text{rt}} = 0.07$ and λ_{sel} begins at 0.02 before being linearly annealed to zero.

Optimization uses AdamW with an initial learning rate of 2.5×10^{-3}, five epochs of warmup, and cosine decay thereafter. Training lasts for 110 epochs, though early stopping is activated if validation negative log-likelihood fails to improve for twelve epochs. Mini-batches are assembled by stratified sampling across coarse SNR bands. This reduces the risk that early training is dominated by high-SNR examples, a situation that would otherwise encourage the model to settle into a solution that performs well when the waveform is clean but is underprepared for the noisy slices that motivate the study.

The stochastic route-silencing probability is set to 0.15. In other words, in roughly 15% of iterations one route is suppressed before fusion. The route to be silenced is selected uniformly, subject to the constraint that both routes are not suppressed simultaneously. Empirically, this small intervention had a clear stabilizing effect. Without it, the broad route often accumulated larger gradients and the fine route became a weak accessory [18]. With it, the fine route retained sufficient discriminative strength to matter at test time, especially on the short-window task.

Inference is slightly richer than the basic fused posterior. For each crop, the fused posterior is the main output. However, the final logit vector also includes a mild agreement-based correction from the branch-specific auxiliary heads. Let $\bar{\ell}_i$ be the average of the two auxiliary logits and let δ_i be the branch disagreement defined above. The corrected logit is

$$\ell_i^* = \ell_i + \rho \exp(-\tau \delta_i) \bar{\ell}_i, \tag{10}$$

with small fixed constants ρ and τ. When the two routes agree, the correction modestly reinforces the decision. When they disagree, the correction is attenuated. This mechanism improved certainty fidelity slightly without materially changing the ranking behavior of the classifier. In the multi-crop test setting, the corrected posteriors from the four crops are averaged to obtain the final example-level posterior.

Monitoring during training includes several route-specific quantities. The average norm of the gradients entering the fine-route and broad-route projection layers is recorded at every epoch. The mean entropy of the route-selection field is also tracked, along with validation expected calibration error and crop-to-crop logit variance on the validation set [19]. A stable run is characterized by nontrivial route-selection entropy in the early epochs, moderate gradient norms for both routes, and gradual reduction of crop variance as the model matures. These diagnostics are not part of the loss itself, but they were useful for selecting a configuration that remained faithful to the intended mechanism.

5. Results

The primary short-window experiment uses the 96-sample crop protocol. Under this setting, BCSA-Net achieves a

mean global test accuracy of 59.8% across the five random seeds. The corresponding values for DRS, TPC, and LSA are 56.4%, 57.1%, and 55.6%, respectively. The proposed model therefore improves upon the strongest comparison model by 2.7% in absolute accuracy. The difference is more meaningful when viewed through macro-averaged recall, which is 58.1% for BCSA-Net, 54.2% for DRS, 55.0% for TPC, and 53.8% for LSA. The larger margin in macro-averaged recall indicates that the gain is not concentrated in already favorable classes. Instead, it is distributed more broadly across the label space.

When the classifier is allowed to use the full 128-sample window, all methods improve, as expected. Under this secondary setting, BCSA-Net reaches 63.7% mean global accuracy, compared with 60.9% for DRS, 61.4% for TPC, and 60.2% for LSA. Macro-averaged recall becomes 62.1%, 58.7%, 59.3%, and 58.0%, respectively. The margin shrinks slightly relative to the short-window case, which is consistent with the notion that full-window access partly compensates for the lack of explicit scale interaction in the baselines [20]. Yet the proposed architecture remains the best performer, and the persistence of the gap suggests that reciprocal alignment is not merely a trick for severe cropping. It remains useful when the model has more evidence, though the relative benefit is strongest when the observation is deliberately shortened.

The SNR-resolved curves reveal where the differences arise. At the highest SNR values above 10 dB, the gap among models is modest, generally within about 1.0% to 1.5% of accuracy. In that regime the signal is clean enough that even the single-route architectures can recover most of the class structure. The gap widens in the intermediate low-SNR band. Between -12 dB and 0 dB in the short-window setting, BCSA-Net attains an average accuracy of 41.6%, whereas DRS, TPC, and LSA achieve 37.2%, 38.4%, and 36.5%, respectively. The absolute gain of 4.4% over DRS is the clearest quantitative sign that the proposed mechanism matters most when local evidence is weak yet not completely absent.

The very lowest SNR slice below roughly -16 dB is instructive for a different reason. There, all models perform poorly and the margins narrow. BCSA-Net still retains a small advantage, but the difference is less dramatic than in the intermediate band. This behavior is sensible [21]. Once the observation becomes dominated by noise, no architectural rearrangement can recreate fine structure that is no longer present. The value of reciprocal scale interaction is therefore not that it defeats information loss in the extreme. Its value is that it preserves faint but genuine signal cues long enough to matter before they are erased by a purely contextual hierarchy.

The class-specific results show that the improvements are not uniform. In the short-window task, the largest recall gains appear in class pairs known to be difficult when only partial observations are available. The recall for a phase-sensitive digital class rises by 6.2% relative to DRS in the -10 dB to -2 dB band, and the recall for one of the analog classes rises by 5.7% over the same range. The two QAM families remain difficult across all models, but BCSA-Net still improves their average recall by approximately 3.0% to 3.5% in the short-window case. By contrast, the easiest high-SNR classes show only minor changes. This pattern is exactly what the architecture was intended to produce: gains concentrated where the standard deep hierarchy has the greatest tendency to suppress class-defining detail.

The off-diagonal dominance ratio provides additional perspective. At -8 dB in the short-window setting, the three most frequent wrong labels absorb 66.8% of all DRS errors, 63.5% of all TPC errors, and 68.1% of all LSA errors. The corresponding ratio for BCSA-Net is 57.9%. A reduction of nearly nine points relative to DRS means that the proposed model no longer channels uncertainty into the same small set of attractor classes at the same rate [22]. This does not imply that the model becomes uniformly confused. Rather, it indicates that its mistakes are less dominated by one coarse explanation of the data. In practical terms, the classifier maintains a richer set of alternatives when evidence is poor.

Certainty fidelity follows the same trend. On the primary short-window task, expected calibration error is 0.047 for BCSA-Net, compared with 0.061 for DRS, 0.057 for TPC, and 0.064 for LSA. Negative log-likelihood is likewise best for the proposed model. The effect is most visible in the intermediate low-SNR band, where branch disagreement is common and the disagreement-sensitive focusing term in training becomes active. This matters because a model that improves accuracy by becoming sharply overconfident would be less useful in operational systems where downstream decisions depend on the trustworthiness of posterior scores.

Crop agreement is another area where the proposed model behaves differently. Recall that the primary task averages predictions over four fixed crops per original example. The variance of the corrected logits across those crops is lower for BCSA-Net than for the comparison systems, but not trivially low. DRS shows higher variance because it can be sensitive to whether a critical local event survives the crop, while LSA shows moderate variance but lower overall accuracy. BCSA-Net combines lower crop variance with higher recognition quality, which suggests that it is not simply keying off one narrow event [23]. Instead, it retains enough local sensitivity to benefit from informative crop positions while using broad-route support to reduce instability across nearby crops.

The ablation study clarifies which parts of the design are responsible for the observed behavior. Removing the reciprocal alignment bridges and leaving the rest of the network intact reduces short-window accuracy from 59.8% to 57.6%. Removing the route-selection field and replacing it with simple averaging yields 57.9%. Removing the low-rank bilinear mixer yields a smaller reduction to 58.9%, indicating that cross-products are helpful but not the main source of the gain. Turning off stochastic route silencing reduces short-window accuracy to 58.2% and increases ex-

pected calibration error to 0.053. Finally, removing the reciprocal translation penalty while keeping everything else in place results in 57.8% accuracy and visibly weaker macro-averaged recall. These ablations show that no single component explains the full effect. The empirical advantage comes from the combination of bidirectional exchange, adaptive route selection, and training procedures that prevent route collapse.

One particularly informative test disables each route at inference time after training has already finished. When the fine route is suppressed, short-window accuracy falls from 59.8% to 57.0%. When the broad route is suppressed, it falls further to 54.6% [24]. The asymmetry is expected because the broad route carries more stable class information on average. Yet the two-point difference between the full model and the fine-suppressed model is meaningful, especially because it is concentrated in the intermediate low-SNR band. This demonstrates that the fine route is not an ornamental addition. It provides a measurable share of the gain exactly where delicate local evidence still exists.

The full-window secondary task shows similar though slightly muted ablation trends. Removing the reciprocal bridges reduces accuracy from 63.7% to 61.9%, and replacing adaptive route selection by uniform averaging reduces it to 62.1%. Suppressing the fine route at inference time lowers accuracy to 61.6%, whereas suppressing the broad route lowers it to 58.8%. Even with the complete 128-sample window, both routes remain useful. The greater resilience of the broad route is unsurprising, but the remaining contribution of the fine route confirms that local evidence continues to matter even when more context is available.

6. Error Geometry and Route Contribution Analysis

Aggregate metrics are helpful but incomplete. To understand why the proposed model behaves differently, the error structure must be examined more closely. The confusion matrices reveal that BCSA-Net changes not only the number of errors but also their topology. In DRS and LSA, low-SNR failures often gather into dense columns corresponding to a few broad prototype labels [25]. This means that a large portion of uncertain examples are mapped into the same simplified explanation. TPC softens that tendency somewhat because late concatenation provides access to more than one stage of representation. However, BCSA-Net alters the pattern more substantially. The dense attractor columns become weaker, and some of the error mass shifts toward secondary alternatives. This suggests that the model is preserving distinctions that the baselines flatten away.

The difference is particularly visible in three recurring confusion relations. The first concerns a phase-oriented digital pair whose broad temporal behavior is similar but whose local phase progression differs. Under short-window cropping, DRS often merges these classes once SNR drops below -6 dB. BCSA-Net still confuses them, but the confusion probability is consistently lower by several points across seeds. The second relation concerns one analog class and a frequency-dominated class that share broad energy distribution when observed briefly. Here the fine route appears to matter more, as shown by the fact that silencing the fine route sharply increases confusion between them. The third relation concerns the two QAM families, whose separation remains difficult even for the proposed model. Yet BCSA-Net reduces the asymmetry of this confusion, meaning that mistakes are less one-sided and more balanced [26]. That behavior is consistent with a model that retains more nuanced evidence rather than one that defaults to a single fallback class.

The route-selection field offers direct insight into how the model allocates responsibility. Averaged over correctly classified high-SNR examples, the selection field assigns approximately 44% of its mass to the fine route and 56% to the broad route. In the intermediate band from -10 dB to -2 dB, the fine-route share rises to about 49% on correctly classified examples, indicating that local cues become relatively more important when context alone is not enough. For the very low-SNR slices below -16 dB, the fine-route share falls back to roughly 39%, suggesting that once local structure becomes too weak, the model relies more heavily on the broad route. This nonmonotonic pattern would be difficult to explain if the route-selection field were merely learning a static preference. Instead, it indicates an adaptive regime in which local evidence is emphasized only when it remains informative.

The branch-specific auxiliary heads confirm this interpretation. On easy high-SNR examples the broad-route auxiliary head is usually slightly more accurate than the fine-route head, though both perform well. In the intermediate low-SNR regime the fine-route head improves relative to the broad-route head for certain class pairs, even though its global accuracy remains lower. This means that the fine route captures useful discriminative details that are not always present but become decisive when they are. The fused classifier then benefits from having access to both views [27]. On very noisy slices, both auxiliary heads degrade substantially, but the broad-route head remains the stronger of the two, which is again consistent with the idea that broad context provides the more reliable fallback once local structure is no longer recoverable.

The geometry of the latent space supports the same story. After training, class centroids are computed from the fused representation and from the route-specific representations. The average pairwise centroid distance is larger in the fused space than in either route space alone for the short-window task. More importantly, the increase is not uniform across all classes. It is largest for the hard pairs identified above. A singular-value analysis of within-class scatter shows that BCSA-Net reduces the expansion of within-class variance in the intermediate low-SNR regime more effectively than DRS while preserving broader between-class spacing than TPC. In other words, the fused representation is not simply more spread out. It is structured in a way that compresses

nuisance variability and expands useful separation for difficult categories.

Another informative analysis measures the consistency of the reciprocal bridges. For each bridge, the entropy of the alignment matrix is averaged across examples. Very low entropy would mean that each position in one route attends almost deterministically to a tiny subset of positions in the other route, which could indicate brittle dependence on one event. Very high entropy would mean that the bridge is nearly uniform and therefore weakly informative [28]. In the trained BCSA-Net, the bridge entropies settle into a middle range. They are lower for the fine-to-broad direction in the intermediate low-SNR band, which suggests that the broad route learns to seek specific fine-route evidence when the signal is difficult but not hopelessly corrupted. In contrast, the broad-to-fine direction becomes somewhat more diffuse at the lowest SNR levels, consistent with the idea that the broad route offers a general stabilizing template rather than a sharply localized correction.

The crop-based analysis adds a practical dimension. Because the test protocol averages over four fixed crops, it is possible to inspect how often the model changes its predicted label across crops before averaging. BCSA-Net shows fewer label flips than DRS and TPC. Yet when flips do occur, they are concentrated in examples whose final posterior remains low-confidence, rather than in examples where the model is highly confident and wrong. This is desirable. It means that crop sensitivity is being translated into uncertainty rather than into brittle misclassification. The comparison models exhibit more cases where one crop produces an overconfident wrong label while another crop produces the correct label, which suggests that they are more exposed to the exact placement of a decisive local event.

The route-silencing stress analysis after training is perhaps the clearest demonstration of complementarity. When the fine route is disabled only for examples in the intermediate low-SNR band, the loss is substantially larger than when it is disabled only for high-SNR examples. Conversely, disabling the broad route hurts almost everywhere but is especially harmful for the lowest SNR region and for the full-window task [29]. This indicates that the two routes do not offer redundant information distributed evenly across conditions. Each route has a regime in which its contribution becomes more pronounced. That empirical pattern is exactly what a two-route architecture ought to produce if it is working as intended.

7. Receiver Mismatch Battery

The receiver-mismatch battery evaluates whether the learned recognition strategy remains usable when the acquisition process deviates from the training distribution in ways that are plausible for practical radio systems. The first mismatch family is residual carrier offset beyond the augmentation range seen during training. Under this condition, the short-window accuracy of DRS drops by 5.8%, TPC drops by 5.1%, LSA drops by 6.2%, and BCSA-Net drops by 4.3%. The difference is not huge, but it is consistent across seeds. The route analyses suggest that the broad route absorbs much of the burden here, while the fine route still contributes on examples where local phase behavior remains readable after the extra rotation.

The second mismatch family is IQ imbalance with both gain asymmetry and a small quadrature phase error. This condition is more harmful to models that rely too rigidly on one exact channel geometry. DRS loses 4.9% short-window accuracy, TPC loses 4.4%, LSA loses 5.0%, and BCSA-Net loses 3.6%. The relative advantage of the proposed model appears to come from the fact that the fine route and broad route are not equally distorted by the imbalance [30]. The fine route is more sensitive to altered local shape, but the broad route preserves a more stable summary, and the reciprocal bridges allow the model to retain useful information from both without overcommitting to the altered local waveform trace.

Soft clipping produces a different pattern. Because clipping compresses large amplitudes while leaving smaller portions of the signal less affected, it can selectively damage salient waveform segments. Under this perturbation, DRS loses 4.1%, TPC loses 3.8%, LSA loses 4.6%, and BCSA-Net loses 3.2%. The route-selection field shifts away from the fine route more noticeably under clipping than under carrier offset, which is reasonable because clipped peaks can look like exaggerated local evidence when viewed by a morphology-sensitive route. The model appears to use the broad route as a safeguard in these cases.

The impulsive contamination condition is the most severe of the four mismatch families. It injects sparse but strong disturbances at random time positions that were not represented in the main training distribution. DRS loses 7.0% accuracy in the short-window task, TPC loses 6.4%, LSA loses 6.8%, and BCSA-Net loses 5.1%. The gain here is especially relevant because impulsive artifacts directly challenge the utility of preserving local structure. A fine-resolution route alone would be vulnerable to such spikes, but the presence of the broad route and the reciprocal checks between routes allow the model to downweight local evidence when it appears inconsistent with the broader latent pattern.

The certainty fidelity metrics remain favorable under mismatch. Although expected calibration error increases for all models under the four perturbation families, the increase is smallest for BCSA-Net in every case [31]. For example, under impulsive contamination the expected calibration error rises from 0.047 to 0.061 for BCSA-Net, whereas it rises from 0.061 to 0.082 for DRS. This matters because a mismatch-robust classifier should not only preserve accuracy as far as possible; it should also maintain a realistic account of its uncertainty once the observation departs from the training regime.

The full-window mismatch experiments show the same ordering, though all models are somewhat more stable when the window length is 128. Under carrier offset, the accuracy drops are 3.6% for BCSA-Net, 4.7% for DRS, 4.2%

for TPC, and 4.9% for LSA. Under impulsive contamination, the drops are 4.2%, 5.9%, 5.2%, and 5.5%, respectively. The reduced severity is expected because more context is available, but the relative advantage of reciprocal alignment remains visible. Even when the model has a longer observation, the existence of two interacting routes still helps distribute risk under distortion.

A final mismatch analysis inspects which classes suffer most. The classes that depend heavily on local morphology remain the most fragile under impulsive contamination, and the classes that depend more on broad spectral-temporal organization remain comparatively robust. However, BCSA-Net narrows the gap between these groups. Its worst-case class recall under mismatch is consistently higher than that of the comparison models. This result is not dramatic in absolute terms, but it is valuable because worst-case class behavior often determines whether a recognition system can be trusted in deployment.

8. Conclusion

This paper presented an empirical study of a bidirectional cross-scale alignment network for short-window signal classification from raw IQ observations [32]. The study was motivated by a practical failure mode of conventional deep encoders: as receptive fields expand and representations become more invariant, brief discriminative waveform traits can be weakened or discarded, especially when the observation is short and noisy. The proposed architecture addressed that issue by maintaining a fine-resolution route and a broad-context route, allowing them to exchange information through reciprocal alignment bridges, and forming the final decision from an adaptive fusion of the two.

Across repeated-seed experiments on an eleven-class benchmark, the model improved global accuracy, macro-averaged recall, and certainty fidelity relative to matched single-route and late-concatenation baselines. The gains were largest in the intermediate low-SNR regime and under the shortened crop protocol, which is precisely where a classifier must preserve faint local cues without losing the stabilizing effect of broader context. Error analysis showed that the method reduced the extent to which uncertain examples collapsed toward a small set of attractor labels. Route-level diagnostics further indicated that the two routes were not redundant. Their relative importance changed with SNR, crop position, and distortion type.

The receiver-mismatch battery provided additional evidence that reciprocal scale interaction is not only useful on the nominal test set. The proposed model degraded more smoothly than the comparison systems under carrier offset, IQ imbalance, clipping, and impulsive corruption. The empirical interpretation is measured rather than sweeping. Bidirectional cross-scale alignment does not overcome the information loss of the most extreme noise regime, and it does not eliminate all difficult class confusions. What it appears to do is preserve a broader and more adaptable basis for decision making when the observation is short, incomplete, and perturbed. For practical signal classification problems in such conditions, that property is likely to matter as much as raw average accuracy [33].

References

[1] D. R. D. Balzo, "Array gain improvements at low snr and low frequency via maximum entropy beamforming," *The Journal of the Acoustical Society of America*, vol. 74, pp. S120–S120, 11 1983.

[2] X. Wang, G.-K. Liu, M. Xu, B. Ren, and Z.-Q. Tian, "Development of weak signal recognition and an extraction algorithm for raman imaging.," *Analytical chemistry*, vol. 91, pp. 12909–12916, 9 2019.

[3] B. Zou, H. Yan, F. Wang, Y. Zhou, and X. Zeng, "Research on signal modulation classification under low snr based on resnext network," *Electronics*, vol. 11, no. 17, p. 2662, 2022.

[4] H. Zhu, L. Zhou, and C. Chen, "A denoising radio classifier with residual learning for modulation recognition," in *2021 IEEE 21st International Conference on Communication Technology (ICCT)*, pp. 177–181, IEEE, 10 2021.

[5] S. Kurugol, M. Freiman, O. Afacan, L. Domachevsky, J. M. Perez-Rossello, M. J. Callahan, and S. K. Warfield, "Miccai (3) - motion compensated abdominal diffusion weighted mri by simultaneous image registration and model estimation (sir-me)," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 9351, pp. 501–509, 11 2015.

[6] A. Mezghani and A. L. Swindlehurst, "Blind estimation of sparse broadband massive mimo channels with ideal and one-bit adcs," *IEEE Transactions on Signal Processing*, vol. 66, pp. 2972–2983, 6 2018.

[7] T. Yamada, K. Itoyama, K. Nishida, and K. Nakadai, "Assessment of sound source tracking using multiple drones equipped with multiple microphone arrays," *International journal of environmental research and public health*, vol. 18, pp. 9039–, 8 2021.

[8] N. Jamal, N. Fuad, M. Sha'bani, M. H. A. Wahab, and S. Z. S. Idrus, "Binary time-frequency mask for improved malay speech intelligibility at low snr condition," *IOP Conference Series: Materials Science and Engineering*, vol. 917, pp. 012049–, 9 2020.

[9] K. Nathwani, F. Hafiz, A. Swain, and R. Biswas, "Ispa - speech intelligibility enhancement using an optimal formant shifting approach," in *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 120–125, IEEE, 9 2021.

[10] T.-Y. Lan, T. Bendory, N. Boumal, and A. Singer, "Multi-target detection with an arbitrary spacing distribution," *IEEE transactions on signal processing : a publication of the IEEE Signal Processing Society*, vol. 68, pp. 1589–1601, 2 2020.

[11] B. H. Tracey, E. L. Miller, Y. Wu, C. V. Alvino, M. Schiefele, and O. Al-Kofahi, "Locally optimized non-local means denoising for low-dose x-ray backscatter imagery," *Journal of X-ray science and technology*, vol. 22, pp. 569–586, 9 2014.

[12] C.-L. Lin, Z.-Q. Lin, S.-S. Wang, Y. Tsao, and J. weih Hung, "Icce-tw - exponentiated magnitude spectrogram-based relative-to-maximum masking for speech enhancement in adverse environments," in *2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, pp. 1–2, IEEE, 9 2020.

[13] A. M. Arar, A. Masri, H. O. Ghannam, and I. Tumar, "A proposed scheme for dynamic threshold versus noise uncertainty in cognitive radio networks (dtnu)," *Wireless Personal Communications*, vol. 96, pp. 4543–4555, 5 2017.

[14] D. Gratadour, L. M. Mugnier, and D. Rouan, "Sub-pixel image registration with a maximum likelihood estimator. application to the first adaptive optics observations of arp 220 in the l' band," *Astronomy & Astrophysics*, vol. 443, pp. 357–365, 10 2005.

[15] T. Xu, X. Wang, M. Huang, X. Lan, and L. Sun, "Tensor-based reduced-dimension music method for parameter estimation in monostatic fda-mimo radar," *Remote Sensing*, vol. 13, pp. 3772–, 9 2021.

[16] H. Khakzad, A. Taherpour, R. Shakeri, and T. Khattab, "Dynamic interference-limited relay sharing in cognitive radio networks by using hierarchical modulation," *IET Communications*, vol. 11, pp. 1903–1912, 9 2017.

[17] S. R. Young, B. J. Steward, M. R. Hawks, and K. C. Gross, "Improving detection of low snr targets using moment-based detection," *SPIE Proceedings*, vol. 9828, pp. 156–167, 5 2016.

[18] L. Hui and W. Yan-fei, "The contrastive study between direct fft and phase difference in parameter estimation of sinusoidal signal: The contrastive study between direct fft and phase difference in parameter estimation of sinusoidal signal," *JOURNAL OF ELECTRONICS INFORMATION & TECHNOLOGY*, vol. 32, pp. 544–547, 4 2010.

[19] S. Feng-Wen, Y. Jiang, and L.-N. Lee, "Frame synchronization and pilot structure for second generation dvb via satellites," *International Journal of Satellite Communications and Networking*, vol. 22, pp. 319–339, 6 2004.

[20] K. Rajeswari and S. J. Thiruvengadam, "Optimal power allocation for channel estimation in mimo-ofdm system with per-subcarrier transmit antenna selection," *Radioengineering*, vol. 24, pp. 105–114, 4 2015.

[21] I. Djurovic, "Quasi ml algorithm for 2-d pps estimation," *Multidimensional Systems and Signal Processing*, vol. 28, pp. 371–387, 7 2015.

[22] N. V. Battaglia, M. R. Mahfouz, J. M. Johnson, R. D. Komistek, and A. S. Mustafa, "Fine feature retention of magnetic resonance images using intensity inhomogeneity weighted non-local filtering," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 5, pp. 183–194, 6 2015.

[23] T. Zhang, Y. Geng, J. Sun, C. Jiao, and B. Ding, "A unified speech enhancement system based on neural beamforming with parabolic reflector," *Applied Sciences*, vol. 10, pp. 2218–, 3 2020.

[24] Y. Yuan, N. Brodusch, F. Voisard, B. Nijikovsky, A. Moores, and R. Gauvin, "Extract high resolution 3d quantitative elemental map using a combined haadf-stem and eds tomography," *Microscopy and Microanalysis*, vol. 26, pp. 1854–1855, 7 2020.

[25] F. Kéruzoré, R. Adam, P. A. R. Ade, P. André, A. Andrianasolo, M. Arnaud, H. Aussel, I. Bartalucci, A. Beelen, A. Benoit, A. Bideaud, O. Bourrion, M. Calvo, A. Catalano, B. Comis, M. D. Petris, F.-X. Désert, S. Doyle, E. F. C. Driessen, A. Gómez, J. Goupy, C. Kramer, B. Ladjelate, G. Lagache, S. Leclercq, J.-F. Lestrade, J. F. Macías-Pérez, P. Mauskopf, F. Mayet, A. Monfardini, L. Perotto, G. Pisano, E. Pointecouteau, N. Ponthieu, G. W. Pratt, V. Revéret, A. Ritacco, C. Romero, H. Roussel, F. Ruppin, K. F. Schuster, S. Shu, A. Sievers, C. Tucker, and R. Zylka, "A low-mass galaxy cluster as a test-case study for the nika2 sz large program," *EPJ Web of Conferences*, vol. 228, pp. 00012–, 1 2020.

[26] M. Beaumont, I. Odame, P. Babyn, L. Vidarsson, M. Kirby-Allen, and H.-L. M. Cheng, "Accurate liver t2 measurement of iron overload: a simulations investigation and in vivo study.," *Journal of magnetic resonance imaging : JMRI*, vol. 30, pp. 313–320, 7 2009.

[27] C. A. Metzler, F. Heide, P. Rangarajan, M. M. Balaji, A. Viswanath, A. Veeraraghavan, and R. G. Baraniuk, "Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging," *Optica*, vol. 7, pp. 63–71, 1 2020.

[28] H. Zhao, A. Bazco-Nogueras, and P. Elia, "Wireless coded caching can overcome the worst-user bottleneck by exploiting finite file sizes," 3 2021.

[29] Y. shuai Gong, Q. Zhang, K. ming Li, and Y. jun Chen, "Micro-motion feature extraction with multi-beam in low prf," *DEStech Transactions on Engineering and Technology Research*, 11 2017.

[30] Y. Wu, S. Yan, X. Ma, D. Yuan, and J. Li, "Underwater long baseline localization based on doppler measurements for a fast-moving underwater vehicle," in *Global Oceans 2020: Singapore – U.S. Gulf Coast*, pp. 1–5, IEEE, 10 2020.

[31] S. N. Sur, R. Bera, and B. Maji, *Phased-MIMO Radar in Low SNR Regime*, pp. 357–362. Germany: Springer Singapore, 2 2019.

[32] S.-H. Oh, M. Bilello, M. K. Schindler, C. E. Markowitz, J. A. Detre, and J. Lee, "Direct visualization of short transverse relaxation time component (vista).," *NeuroImage*, vol. 83, pp. 485–492, 6 2013.

[33] "Linear multiuser detection for a ds-cdma system," *International Journal of Research and Applications*, vol. 2, pp. 215–219, 3 2015.