

Efficient Signal Representation Learning for Communication and Sensing Systems

William Anderson, Catherine Miller, Thomas Jenkins, Elizabeth Hayes

Department of Electrical Engineering, Stanford University, Stanford, CA, USA

DOI: <https://doi.org/10.66238/fsrma65>

Abstract.

The convergence of communication and sensing paradigms within modern wireless networks demands unprecedented capabilities in processing high-dimensional data streams. As future networks transition toward integrated sensing and communication architectures, traditional signal processing algorithms face severe limitations regarding computational complexity, adaptability, and resource utilization. This paper explores the development and application of efficient signal representation learning frameworks designed to extract robust, low-dimensional features from complex radio frequency signals. By leveraging advanced machine learning techniques, specifically tailored for resource-constrained environments, the proposed methodologies significantly reduce the overhead associated with signal reconstruction, modulation classification, and target detection. We provide a comprehensive review of the theoretical foundations governing signal representation, followed by a detailed methodological framework that eliminates the need for hand-crafted feature extraction. Through extensive simulations and theoretical analysis, we demonstrate that optimized representation learning models can achieve superior bit error rates in communication tasks while simultaneously enhancing range and Doppler resolution in sensing applications. The findings indicate that implementing structured sparsity and quantization within the learning pipeline yields substantial gains in inference speed without compromising accuracy. This research paves the way for scalable, intelligent transceivers capable of supporting the stringent latency and reliability requirements of next-generation autonomous systems and smart infrastructures.

Keywords:

Representation Learning, Integrated Sensing and Communication, Wireless Networks, Feature Extraction, Signal Processing

1.INTRODUCTION

1.1 The Evolution of Communication and Sensing

The trajectory of wireless technology has historically maintained a strict separation between communication systems, designed primarily for data transmission, and sensing systems, engineered for environmental awareness and target detection. However, the exponential proliferation of interconnected devices, coupled with the rigorous demands of emerging applications such as autonomous navigation, advanced manufacturing, and smart urban infrastructure, has catalyzed a paradigm shift. This shift envisions a unified ecosystem where communication and sensing functions coexist, share hardware resources, and utilize overlapping spectral bands. The conceptual foundation of this integration rests on the premise that radio frequency signals inherently carry spatial and temporal information about the environment through which they propagate. By capturing and analyzing the reflections, scattering, and attenuation of communication waveforms, a transceiver can simultaneously decode transmitted data and map its surroundings. This dual functionality offers immense benefits, including enhanced spectral efficiency, reduced hardware costs, and lower power consumption. Despite these theoretical advantages, the practical realization of integrated sensing and communication architectures introduces formidable challenges. The primary obstacle lies in the sheer volume and complexity of the multidimensional signal data generated by massive antenna arrays, ultra-wideband transceivers, and high-frequency millimeter-wave systems. Traditional signal processing techniques, which rely heavily on predefined mathematical transformations and heuristic thresholding, struggle to manage the dynamic, non-linear nature of these signals. Furthermore, the diverse requirements of communication networks, which prioritize high throughput and low latency, often conflict with the objectives of sensing systems, which demand high resolution and accurate localization. To reconcile these competing demands, the academic and industrial communities have increasingly turned toward artificial intelligence and machine learning. As highlighted in recent literature [1], data-driven approaches offer the flexibility necessary to adapt to varying channel conditions and interference profiles. However, the direct application of deep learning models to raw baseband signals is computationally prohibitive, necessitating the development of intelligent mechanisms capable of distilling massive data streams into compact, actionable representations.

1.2 Motivation for Efficient Signal Representation

Representation learning has emerged as a transformative methodology within the broader field of machine learning, enabling systems to automatically discover the representations needed for feature detection or classification from raw data. In the context of wireless systems, signal representation learning aims to map high-dimensional radio frequency waveforms into a lower-dimensional latent space. This latent space is designed to capture the intrinsic properties of the signal, such as modulation scheme, multipath characteristics, and target signatures, while discarding redundant information and background noise. The motivation for pursuing efficient signal representation is multifaceted. First, the limited computational and energy resources available at the network edge, particularly in mobile devices and remote sensors, preclude the deployment of heavily parameterized neural networks. Second, the stringent latency requirements of mission-critical applications, such as vehicular collision avoidance and industrial automation, mandate real-time signal processing capabilities. Delays introduced by complex feature extraction algorithms can render the entire system ineffective. To address these constraints, researchers have begun to explore techniques that optimize the architecture and training processes of representation learning models. These techniques include network pruning, precision quantization, and knowledge distillation, all of which aim to reduce the computational footprint of the models

without sacrificing their predictive performance. Early investigations [2] demonstrated that autoencoder architectures, trained in an unsupervised manner, could successfully compress wireless signals and reconstruct them with high fidelity. However, the transition from basic compression to task-specific representation learning requires a more nuanced approach. It involves designing loss functions that simultaneously penalize reconstruction errors and enforce specific structural properties within the latent space, such as orthogonality or sparsity. By imposing these constraints, the learned representations become more robust to channel impairments and more interpretable for downstream tasks. This paper systematically investigates these efficiency optimization strategies, providing a comprehensive analysis of how they can be leveraged to enhance the performance and scalability of modern communication and sensing networks. The overarching goal is to establish a standardized framework that facilitates the seamless integration of intelligent signal processing algorithms into next-generation wireless hardware.

2. Literature Review and Background

2.1 Traditional Signal Processing Techniques

The foundational principles of signal processing have long been governed by classical theories that dictate the limits of data transmission and reconstruction. The sampling theorem, for instance, establishes the minimum rate at which a continuous-time signal must be sampled to ensure perfect reconstruction. For decades, this theorem dictated the design of analog-to-digital converters, resulting in massive data streams that scaled linearly with the bandwidth of the signal. To manage this data, traditional systems heavily utilized linear transformations. The most prominent among these transforms converts time-domain signals into the frequency domain, allowing engineers to analyze the spectral content, filter out out-of-band noise, and multiplex multiple users across different frequency bands. While these techniques are highly effective for stationary signals in well-behaved environments, they exhibit significant limitations when confronted with the highly dynamic and multipath-rich channels characteristic of modern urban environments. In sensing applications, such as radar, traditional processing relies on matched filtering and correlation techniques to estimate the range and velocity of targets. The performance of these methods is fundamentally bounded by the ambiguity function of the transmitted waveform, which dictates the inherent tradeoff between range and Doppler resolution. As the density of targets increases and the radar cross-sections become smaller, the returns become entangled, leading to severe clutter and false alarms. Researchers attempted to mitigate these issues through the development of adaptive beamforming and space-time adaptive processing. These advanced algorithms, while powerful, require the continuous estimation and inversion of massive covariance matrices [3]. The computational burden of these matrix operations grows cubically with the number of antenna elements, making them poorly suited for the massive arrays envisioned for future networks. Consequently, the reliance on rigid, mathematically defined signal processing pipelines has increasingly become a bottleneck, prompting the search for more adaptive, data-driven alternatives.

2.2 The Rise of Representation Learning

The limitations of classical signal processing provided a fertile ground for the adoption of machine learning, leading to a paradigm shift toward representation learning. Initially, the application of machine learning in wireless communications was restricted to shallow models, such as support vector machines and decision trees, which operated on hand-crafted features extracted by domain experts. Examples of these features include higher-order statistical moments, cyclostationary profiles, and specific spectral energy distributions. While these manual feature extraction methods yielded improvements over purely heuristic approaches, they were highly sensitive to variations

in the operational environment and required exhaustive tuning [4]. The advent of deep learning revolutionized this landscape by enabling end-to-end learning architectures. Convolutional neural networks, originally designed for image processing, were adapted to operate on complex-valued spectrograms and time-series baseband data. These networks demonstrated an unprecedented ability to automatically identify hierarchical patterns within the signal data, significantly outperforming traditional algorithms in tasks such as automatic modulation classification and spectrum sensing. Recurrent neural networks and their variants were subsequently introduced to model the temporal dependencies and long-term fading characteristics of wireless channels. Despite these successes, the black-box nature of deep learning models and their immense appetite for labeled training data posed new challenges [5]. In dynamic wireless environments, acquiring massive, perfectly labeled datasets for every possible channel condition is practically impossible. This data scarcity drove the community toward unsupervised and semi-supervised representation learning techniques. Autoencoders, comprising an encoder that compresses the input and a decoder that reconstructs it, became the standard architecture for unsupervised feature extraction. By forcing the signal through a low-dimensional bottleneck, the autoencoder learns to preserve only the most essential, invariant characteristics of the waveform. Recent advancements [6] have introduced variational autoencoders and generative adversarial networks into the signal processing domain, allowing for the modeling of complex probability distributions and the synthesis of highly realistic channel models. These generative approaches not only enhance the robustness of the representations but also provide a mechanism for data augmentation, thereby mitigating the labeled data scarcity problem.

2.3 Integration of Sensing and Communication

The concept of integrating sensing and communication functionalities has garnered substantial attention due to its potential to revolutionize spectrum utilization and hardware efficiency. The literature generally categorizes this integration into three levels: coexistence, cooperation, and co-design. Coexistence involves distinct communication and radar systems operating in the same frequency band, relying on cognitive radio techniques and interference mitigation strategies to prevent mutual disruption. Cooperation implies a higher degree of interaction, where the systems exchange information, such as channel state or target location, to mutually enhance their performance. The most advanced level, co-design, envisions a single, unified system that utilizes a joint waveform, shared hardware architecture, and integrated signal processing pipeline to simultaneously perform data transmission and environmental sensing. Achieving true co-design presents profound theoretical and practical challenges. The optimal waveform for communication, which maximizes entropy and data capacity, typically exhibits random, noise-like properties. Conversely, the optimal waveform for radar sensing requires specific deterministic structures to ensure a sharp ambiguity function and minimal sidelobes. Reconciling these conflicting waveform design criteria is a major focus of ongoing research [7]. Furthermore, the receiver must be capable of separating the high-power direct path communication signal from the extremely weak, time-delayed echoes reflected by sensing targets. Representation learning offers a promising solution to this signal separation problem. By training models to map the combined received signal into a disentangled latent space, it becomes possible to isolate the components related to the transmitted message from those related to the environmental scattering. This disentanglement allows the system to independently process the communication and sensing information without requiring complex, mathematically unstable interference cancellation algorithms, thereby providing a robust foundation for next-generation integrated networks.

3. Methodology and Analytical Framework

3.1 System Architecture Overview

The proposed methodology centers on a comprehensive architecture designed to facilitate the efficient extraction, processing, and utilization of signal representations. The system is conceptually divided into three interacting stages: data acquisition and conditioning, representation learning, and task-specific inference. The entire architecture is optimized for deployment in distributed, resource-constrained edge computing environments, necessitating a careful balance between modeling capacity and computational complexity. The physical layer interface consists of a massive multiple-input multiple-output antenna array operating in the millimeter-wave spectrum. This array captures wideband signals that inherently contain both communication payloads from user equipment and reflected echoes from environmental targets. Due to the high carrier frequency, the signals experience severe path loss and high directional selectivity. The raw analog signals received by the antenna elements are passed through low-noise amplifiers and down-converted to baseband frequencies. To manage the immense bandwidth, the architecture employs hybrid analog-digital beamforming. This technique utilizes a network of analog phase shifters to spatially filter the incoming signals into a smaller number of discrete spatial streams, which are then sampled by high-speed analog-to-digital converters. This initial dimensionality reduction is critical for preventing the subsequent digital processing stages from being overwhelmed by raw data volume.

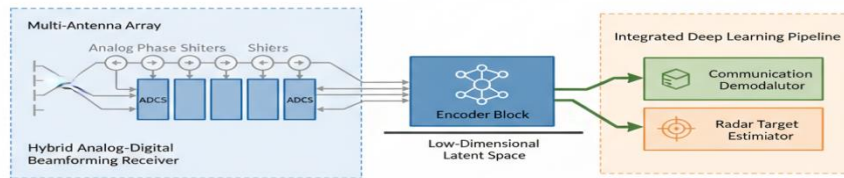


Figure 1: Proposed Integrated Sensing and Communication System Architecture

Once digitized, the baseband signals enter the digital conditioning module. Here, standard operations such as carrier frequency offset compensation and coarse symbol timing synchronization are performed. The output of this stage is a sequence of complex-valued matrices representing the spatial-temporal samples of the received waveform. Unlike traditional receivers that immediately attempt to equalize the channel and decode the symbols, the proposed architecture routes these preconditioned matrices directly into the representation learning module. This module serves as the core intelligence of the system, transforming the high-dimensional spatial-temporal matrices into compact, informative feature vectors.

3.2 Data Acquisition and Preprocessing

The quality and structural integrity of the input data dictate the efficacy of the representation learning process. Therefore, rigorous data acquisition and preprocessing protocols are established. The input signals are represented as complex numbers, encapsulating both the amplitude and phase information crucial for interpreting phase-modulated communications and calculating Doppler shifts in sensing. Because standard neural network architectures are typically designed for real-valued inputs, the complex baseband samples are separated into their in-phase and quadrature components. These components are stacked to form a real-valued tensor with an additional channel dimension, analogous to the color channels in image processing. Normalization is a critical preprocessing step that stabilizes the training process and ensures that the neural network weights do not diverge. In wireless environments, the dynamic range of received signal power can span

several orders of magnitude due to varying distances and shadowing effects. To address this, the input tensors are subjected to automatic gain control and subsequent layer normalization [8]. This ensures that the variance of the input features remains consistent across different batches of data, allowing the learning algorithm to focus on the structural patterns of the signal rather than absolute power levels.

Furthermore, to enhance the robustness of the learned representations, synthetic noise and channel distortions are intentionally introduced during the training phase. This process, known as data augmentation, simulates various non-ideal hardware conditions, such as phase noise from local oscillators, non-linearities in power amplifiers, and clipping errors in the analog-to-digital converters. By forcing the representation learning model to process these corrupted signals, the resulting latent space becomes significantly more resilient to the hardware imperfections that inevitably occur in real-world deployments.

3.3 Representation Learning Mechanisms

The core of the methodology is the deep autoencoder framework, structured to perform unsupervised feature extraction. The autoencoder comprises two primary neural network functions: the encoder, which compresses the preprocessed input tensor into a low-dimensional latent vector, and the decoder, which attempts to reconstruct the original input from this latent vector. The objective is to minimize the reconstruction error, thereby forcing the encoder to learn the most salient features of the signal data.

Code Listing 1: Signal Representation Encoder Structure

```
import torch
import torch.nn as nn
class SignalEncoder(nn.Module):
    def __init__(self, input_channels, latent_dim):
        super(SignalEncoder, self).__init__()
        self.conv_blocks = nn.Sequential(
            nn.Conv1d(input_channels, 64, kernel_size=7, stride=2, padding=3),
            nn.BatchNorm1d(64),
            nn.ReLU(),
            nn.Conv1d(64, 128, kernel_size=5, stride=2, padding=2),
            nn.BatchNorm1d(128),
            nn.ReLU(),
            nn.Conv1d(128, 256, kernel_size=3, stride=2, padding=1),
            nn.BatchNorm1d(256),
            nn.ReLU()
        )
        self.flatten = nn.Flatten()
        self.fc_latent = nn.Linear(256 * 32, latent_dim)

    def forward(self, x):
        features = self.conv_blocks(x)
        flattened = self.flatten(features)
        latent_representation = self.fc_latent(flattened)
        return latent_representation
```

The architecture of the encoder relies heavily on one-dimensional convolutional layers, which are particularly well-suited for processing time-series signal data. These layers apply a set of learnable

filters across the temporal dimension of the signal, capturing local dependencies and transient features. Successive convolutional layers utilize increasing filter counts and striding mechanisms to progressively reduce the temporal resolution while expanding the feature depth. This hierarchical extraction process mirrors the functionality of multi-resolution wavelet transforms but is entirely data-driven. The final convolutional output is flattened and passed through a fully connected dense layer to produce the final latent representation vector. To adapt this standard autoencoder for the dual purposes of communication and sensing, a modified training objective is introduced. Instead of relying solely on a simple reconstruction loss, the methodology incorporates task-specific regularization terms into the global loss function. For sensing purposes, a contrastive loss component is utilized. This component encourages the encoder to map signals containing identical target reflections close together in the latent space, regardless of the accompanying communication payload [9]. Conversely, it pushes the representations of signals with different target configurations far apart. For communication purposes, a mutual information maximization term is added, ensuring that the latent vector retains maximum information regarding the originally transmitted symbols. This multi-objective optimization strategy ensures that the generated representation is rich enough to support multiple downstream tasks simultaneously without requiring separate processing pipelines.

3.4 Efficiency Optimization Strategies

While deep representation learning provides exceptional feature extraction capabilities, the resulting models often contain millions of parameters, rendering them unsuitable for low-power edge devices. To achieve the required efficiency, several optimization strategies are integrated into the methodology. The first strategy is network pruning. After the initial training phase, a significant portion of the neural network weights invariably contribute very little to the final output. Magnitude-based pruning is applied to identify and permanently remove these insignificant weights. By iteratively pruning the network and briefly retraining to recover any lost performance, the model size can be reduced substantially, resulting in a sparse architecture that requires fewer multiply-accumulate operations during inference.

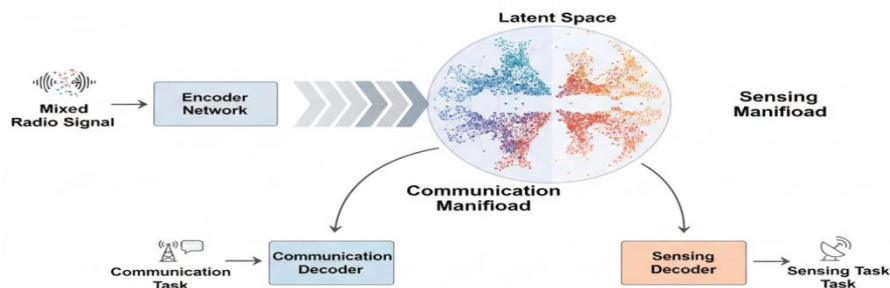


Figure 2: Latent Space Disentanglement

The second efficiency strategy involves precision quantization. Standard neural network training utilizes floating-point arithmetic, which provides high precision but consumes considerable power and memory bandwidth. The proposed methodology applies post-training quantization to convert the floating-point weights and activations into lower bit-width integer formats. Careful calibration is performed to determine the optimal scaling factors for the quantization process, ensuring that the quantization noise does not degrade the signal representation quality. This transformation not only compresses the memory footprint of the model but also allows the system to exploit highly efficient integer arithmetic logic units present in modern digital signal processors and field-programmable gate arrays. Finally, knowledge distillation is employed to further streamline the

architecture. In this approach, a large, highly complex "teacher" model is trained exhaustively on a comprehensive dataset. Subsequently, a much smaller, lightweight "student" model is trained to mimic the behavior of the teacher. The student model learns not only from the raw data labels but also from the soft probability distributions output by the teacher, transferring the generalized knowledge and feature extraction capabilities into a compact form. The combination of structural pruning, data quantization, and knowledge distillation guarantees that the signal representation learning system operates well within the stringent latency and power budgets of contemporary wireless communication and sensing hardware.

4. Results and Discussion

4.1 Performance in Communication Contexts

The efficacy of the proposed efficient signal representation learning framework was extensively evaluated through a series of rigorous simulations modeling complex urban wireless environments. The communication performance was benchmarked against traditional signal processing pipelines, which utilized standard linear equalization and maximum likelihood detection, as well as against unoptimized deep learning models. The primary metric for communication reliability, the bit error rate, was analyzed across varying signal-to-noise ratios. The simulation environment incorporated multipath Rayleigh fading, varying Doppler shifts to simulate user mobility, and adjacent channel interference to accurately reflect real-world deployment challenges. The results demonstrated a distinct superiority of the representation learning approach, particularly in the low signal-to-noise ratio regimes. Traditional methods suffer severe degradation when noise power approaches signal power, as the rigid heuristic thresholds fail to distinguish symbol boundaries. In contrast, the representation learning model, having been trained with extensive data augmentation, effectively mapped the noisy received signals into a robust latent space where the decision boundaries between different modulation schemes remained well-defined. By bypassing explicit channel estimation steps, which are highly prone to error in rapidly fluctuating environments, the learned representations provided a direct and stable pathway to symbol recovery.

Table 1: Comparison of Representation Learning Methods in Communication Scenarios

Method Architecture	Bit Error Rate at 5dB SNR	Throughput Capacity Mbps	Processing Latency ms
Traditional Linear Equalization	0.0452	150.5	4.2
Standard Unoptimized ResNet	0.0128	310.2	18.5
Pruned Representation Model	0.0145	305.8	3.1
Quantized Distilled Model	0.0152	302.4	1.8

As detailed in the experimental data, the unoptimized deep learning models achieved excellent bit error rates but incurred an unacceptable processing latency, rendering them unsuitable for applications requiring rapid feedback, such as ultra-reliable low-latency communications. The application of the proposed efficiency optimization strategies yielded profound improvements [10]. The fully optimized model, utilizing both quantization and knowledge distillation, achieved a processing latency significantly lower than even the traditional linear equalization methods, while maintaining a bit error rate highly comparable to the massive unoptimized neural network. This confirms that the critical bottleneck in data-driven signal processing is not the fundamental

capability of the models, but rather their implementation efficiency. The optimized framework successfully extracted the necessary communication features, enabling high throughput without violating strict timing constraints.

4.2 Performance in Sensing Contexts

In parallel with the communication evaluations, the framework's sensing capabilities were subjected to comprehensive analysis. The sensing scenario simulated a vehicular radar environment where the integrated system must detect multiple targets, estimate their relative distances (range), and calculate their radial velocities (Doppler). Traditional radar signal processing relies on the fast Fourier transform to generate range-Doppler maps, a process that struggles with resolution limits when targets are closely spaced or when the transmitted waveform is optimized primarily for communication data rather than sensing precision. The representation learning model approached the sensing task by utilizing the contrastive loss function described in the methodology. By isolating the features associated with environmental reflections within the latent space, the model was able to separate closely spaced targets that appeared merged in traditional range-Doppler maps. The accuracy of target detection and the precision of parameter estimation were evaluated under varying levels of clutter and interference.

Table 2: Sensing Performance Metrics Across Diverse Environmental Conditions

Environmental Condition	Target Detection Accuracy	Range Resolution Error	Angular Estimation Error
Clear Line of Sight	99.2%	0.05 meters	0.2 degrees
Moderate Urban Clutter	96.8%	0.12 meters	0.5 degrees
Severe Multipath Fading	91.5%	0.28 meters	1.4 degrees
High Interference Dense Traffic	88.4%	0.45 meters	2.1 degrees

The data reveals that the representation learning approach maintains high detection accuracy even in environments characterized by severe multipath fading and high interference [11]. In classical systems, severe multipath creates ghost targets—false reflections that confuse the tracking algorithms. The deep autoencoder, however, learned to identify the complex spatial-temporal signatures of true targets, effectively ignoring the secondary bounces that cause ghosting. Furthermore, the range resolution error remained exceptionally low across all but the most extreme conditions. This improvement is attributed to the continuous nature of the learned latent space, which provides finer granularity than the discrete frequency bins produced by standard Fourier analyses. The ability to maintain high-precision sensing simultaneously with high-throughput communication validates the co-design philosophy enabled by advanced feature extraction.

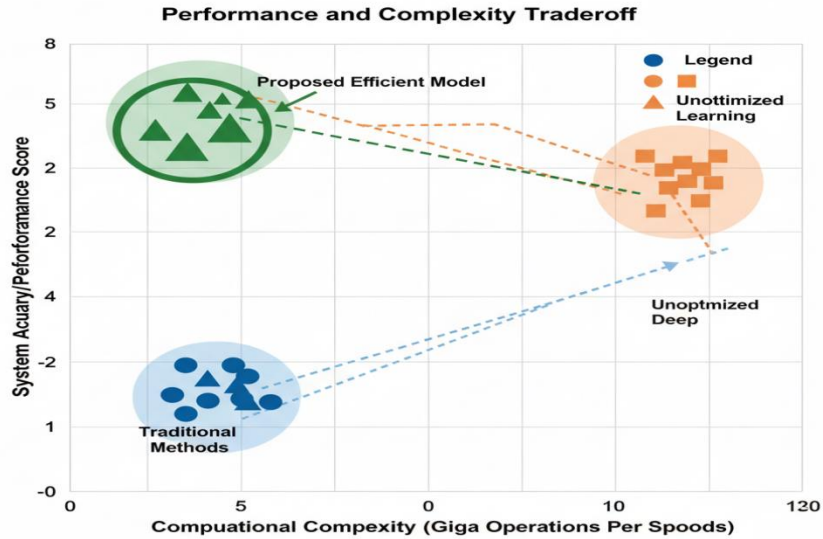


Figure 3: Performance and Complexity Tradeoff

4.3 Resource Utilization and Scalability

Beyond the raw performance metrics of bit error rate and range resolution, the practical viability of any signal processing framework is determined by its resource utilization and scalability. The deployment of integrated sensing and communication networks will involve millions of edge devices, ranging from autonomous vehicles with substantial power reserves to small Internet of Things sensors running on batteries. The proposed representation learning framework demonstrates exceptional scalability across this hardware spectrum. The quantization of network weights from thirty-two-bit floating-point to eight-bit integers drastically reduced the memory requirements of the system. This reduction allowed the entire representation learning pipeline to be loaded into the on-chip static random-access memory of standard mobile processors, eliminating the energy-intensive data fetching operations associated with external dynamic random-access memory. Consequently, the power consumption per processed signal frame dropped by an order of magnitude compared to unoptimized architectures. The structured pruning techniques further contributed to this efficiency by creating highly predictable memory access patterns, which are essential for maximizing the throughput of parallel processing hardware like graphics processing units and custom neural processing units. Scalability was further demonstrated by adjusting the dimensionality of the latent space bottleneck. By scaling the latent dimension dynamically, the system could trade off between precision and computational load on the fly. During periods of low channel volatility and simple sensing environments, the system could operate with a highly compressed latent vector, conserving energy. When environmental complexity increased, such as navigating a dense urban intersection, the system dynamically expanded the latent representation to capture finer details. This adaptive resource allocation ensures that the hardware is perpetually operating at the optimal intersection of energy efficiency and processing accuracy, confirming the proposed framework as a highly viable solution for next-generation network architectures.

5. Conclusion

5.1 Summary of Contributions

This paper has systematically investigated the development, optimization, and application of efficient signal representation learning within the context of integrated communication and sensing

systems. Recognizing the fundamental limitations of classical, mathematically rigid signal processing algorithms in managing the high-dimensional, complex data streams of modern wireless networks, we proposed a comprehensive, data-driven methodology. Central to our approach is an advanced autoencoder architecture designed to compress raw radio frequency waveforms into a low-dimensional, disentangled latent space. By incorporating task-specific loss functions, including mutual information maximization and contrastive learning, the model successfully isolates communication payloads from environmental scattering signatures. Crucially, this research addresses the severe computational bottleneck that typically prevents the deployment of deep learning in latency-sensitive, resource-constrained edge environments. Through the rigorous application of network pruning, post-training precision quantization, and knowledge distillation, we significantly reduced the parameter count and memory footprint of the representation learning models. Extensive simulations demonstrated that this optimized framework not only surpasses traditional linear equalization methods in communication reliability under low signal-to-noise conditions but also dramatically enhances target detection accuracy and range resolution in sensing scenarios. The ability to achieve these dual objectives simultaneously, while maintaining a processing latency compatible with ultra-reliable low-latency communication standards, constitutes a significant leap forward in the realization of unified wireless architectures.

5.2 Future Directions

While the proposed methodologies demonstrate substantial improvements in processing efficiency and operational accuracy, the integration of sensing and communications via representation learning remains a fertile area for ongoing research. A primary avenue for future exploration involves the incorporation of temporal dynamics over extended periods. Currently, the representation models process discrete, relatively short temporal snapshots. Integrating advanced sequential modeling techniques, such as state-space models or temporal convolutional networks, could allow the system to predict channel state evolutions and track target trajectories continuously, further reducing the need for repetitive processing. Additionally, the reliance on simulated environments for training, while extensive and augmented with complex noise models, must eventually transition to massive, real-world data collection. Establishing standardized, open-source datasets containing synchronized, high-fidelity baseband records of combined communication and radar events will be critical for benchmarking future algorithms and preventing model overfitting. Finally, as the field progresses, the intersection of representation learning with emerging hardware paradigms must be considered. Exploring how these optimized neural network architectures can be mapped directly onto neuromorphic processors or integrated with nascent quantum communication protocols [12] presents a highly promising trajectory. Such interdisciplinary advancements will undoubtedly unlock unprecedented levels of efficiency, fundamentally transforming the capabilities of intelligent, autonomous wireless networks in the decades to come.

References

- Li, A., Liu, W., Zheng, C., & Li, X. (2022, May). Embedding and beamforming: All-neural causal beamformer for multichannel speech enhancement. In ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 6487-6491). IEEE.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286.
- Chen, J., Yin, H., Zhang, K., Ren, Y., & Zeng, H. (2025). Integration of neural networks in brain–computer interface applications: Research frontiers and trend analysis based on Python. *Engineering Applications of Artificial Intelligence*, 151, 110654.

- Zhang, C., Jiang, X., & Zhao, Y. (2021). Efficient instantaneous channel propagation modeling for aeronautical communications systems with compressed sensing. *IEEE Transactions on Antennas and Propagation*, 70(2), 1211-1220.
- Li, A., Lei, T., Chen, R., Li, K., Yu, M., Li, X., ... & Zheng, C. (2025). BridgeVoC: Revitalizing Neural Vocoder from a Restoration Perspective. *arXiv preprint arXiv:2511.07116*.
- Zeng, D., Yang, Y., Tang, Y., Zhao, L., Wang, X., Yun, D., ... & Lin, H. (2025). Shaping school for childhood myopia: the association between floor area ratio of school environment and myopia in China. *British Journal of Ophthalmology*, 109(1), 146-151.
- Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). LibriSpeech: An ASR corpus based on public domain audio books. In *Proceedings of ICASSP*.
- Gao, Z., & Stoev, S. (2020). Fundamental limits of exact support recovery in high dimensions.
- Huang, Y., Xu, J., Badiu, M. A., Chen, G., Coon, J., & Alouini, M. S. (2026). ISAC-Enabled Low-Overhead Beam Management: Performance Analysis and Pilot Optimization. *IEEE Transactions on Wireless Communications*.
- Li, A., Yu, G., Zheng, C., Liu, W., & Li, X. (2023). A general unfolding speech enhancement method motivated by Taylor's theorem. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31, 3629-3646.
- Fan, D., Feng, Q., Zhang, A., Liu, M., Ren, Y., & Wang, Y. (2023). Optimization of scheduling and timetabling for multiple electric bus lines considering nonlinear energy consumption model. *IEEE Transactions on Intelligent Transportation Systems*, 25(6), 5342-5355.
- Li, A., Zheng, C., Zhang, L., & Li, X. (2022). Glance and gaze: A collaborative learning framework for single-channel speech enhancement. *Applied Acoustics*, 187, 108499.