A Survey on Modulation Recognition Model Segmentation Inference Acceleration Multi-node Cooperation Signal Processing and Distributed Computing in Communication Systems

Abstract

This survey paper provides a comprehensive analysis of advanced computational methodologies integrated within communication systems, focusing on modulation recognition, model segmentation, inference acceleration, multi-node cooperation, signal processing, and distributed computing. These domains are pivotal for addressing computational challenges and enhancing the efficiency of modern networks. Modulation recognition is crucial for reliable signal processing and network optimization, especially with the proliferation of diverse modulation types. The survey highlights the importance of efficient spectrum management, particularly in radar and communication systems, and the integration of sensing technologies in next-generation networks. It explores neural network models for compressing and accelerating computations, addressing complexities in optical networks, and the application of deep learning in wireless communication. The review also examines wireless spectrum monitoring, cybersecurity challenges, and advancements in near-field communications. By systematically organizing these topics, the survey underscores their collective role in improving communication systems' efficiency and reliability. Future research directions include optimizing neural architectures, refining hybrid models, and enhancing distributed systems' robustness. These efforts aim to advance communication systems, emphasizing the transformative potential of innovative methodologies in modern networks.

1 Introduction

1.1 Scope and Significance

This survey investigates the integration of advanced computational methodologies in communication systems, focusing on modulation recognition, model segmentation, inference acceleration, multi-node cooperation, signal processing, and distributed computing. These areas are pivotal in addressing computational challenges and enhancing the efficiency of modern communication networks. Modulation recognition is crucial for effective communication, facilitating reliable signal processing and network optimization. Automatic Modulation Recognition (AMR) plays a vital role in classifying modulation modes of radio signals, essential for cognitive radio applications and wireless communication [1]. The complexity of accurately recognizing modulation methods has escalated due to rapid advancements in wireless techniques and the diversity of modulation types [2].

The importance of efficient spectrum management and utilization in the evolving landscape of wireless networks is also emphasized. The coexistence of radar and communication systems, which share overlapping bandwidths, necessitates advanced strategies to support high-quality wireless communications and reliable sensing capabilities [3]. Furthermore, integrating sensing technologies with communication systems, as seen in Integrated Sensing and Communication (ISAC) systems,

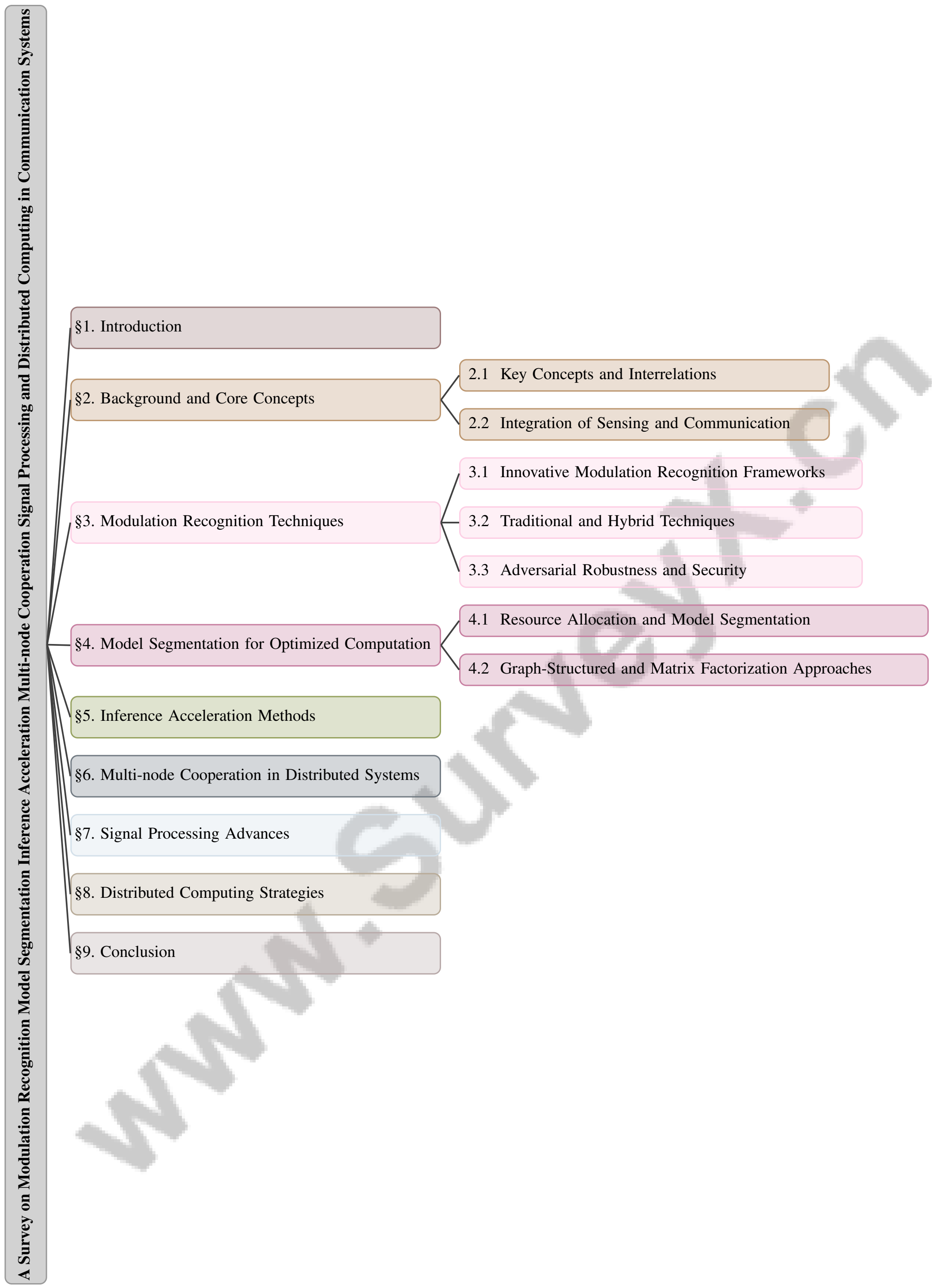


Figure 1: chapter structure

enhances accuracy and continuity in Internet of Everything (IoE) applications [4]. This integration is recognized as a hallmark of next-generation cellular networks, improving spectrum utilization and hardware resource efficiency.

In exploring neural networks (NNs) within communication systems, this survey examines methods for compressing and accelerating these models to address high computational complexity and the challenge of explainability in deep learning-based AMR models [5]. This is particularly relevant in

AI-generated, for reference only.

optical networks, where machine learning techniques are increasingly employed to manage complexity and enhance performance [6]. The application of deep learning (DL) methods at the physical layer of wireless communication systems is explored to overcome the limitations of conventional machine learning (ML) algorithms [7].

The survey also tackles the challenges of wireless spectrum monitoring and modulation classification in distributed sensor networks, highlighting the inadequacies of existing methods that depend on expert features that are not robust under varying conditions [8]. The need for innovative approaches to bolster cybersecurity, particularly concerning remote computer virus radiation injection and its implications for communication systems, is underscored [9].

Additionally, this survey presents a comprehensive review of near-field communications (NFC), discussing the evolution of antenna technologies and their role in enhancing wireless network capacity through spatial degrees of freedom [10]. The integration of optical neural networks with fiber communication is examined to improve signal processing efficiency and reduce power consumption, addressing the limitations of existing methods that separate signal transmission and processing [11]. Furthermore, the enhancement of information transmission capacity in optical fibers through mode division multiplexing (MDM) systems is explored [12]. The concept of large intelligent surfaces (LIS) is discussed, envisioning a future where man-made structures are electronically active, surpassing contemporary massive MIMO technology [13]. This survey provides a holistic overview of these critical areas, emphasizing their collective role in enhancing the efficiency and reliability of communication systems.

1.2 Structure of the Survey

This survey is systematically organized to explore advanced methodologies and techniques in communication systems comprehensively. The paper begins with an **Introduction** that outlines the survey's scope and significance, emphasizing the critical roles of modulation recognition, model segmentation, inference acceleration, multi-node cooperation, signal processing, and distributed computing in modern communication networks. Following this, the **Background and Core Concepts** section delves into the fundamental principles underlying these methodologies, elucidating their interrelations and contributions to the efficiency of communication systems.

The survey progresses to a detailed examination of **Modulation Recognition Techniques**, exploring various strategies for recognizing modulation types, including innovative frameworks like noise-aware ensemble learning (NAEL), which dynamically adjusts neural network structures based on noise impact to improve recognition performance [14]. This section reviews traditional and hybrid approaches, as well as security aspects and robustness against adversarial attacks.

In **Model Segmentation for Optimized Computation**, the paper investigates strategies for partitioning complex models to enhance computational efficiency, highlighting resource allocation techniques and advanced approaches such as graph-structured and matrix factorization methods. The subsequent section, **Inference Acceleration Methods**, focuses on expediting inference processes, discussing optimizations in both hardware and software, and the role of end-to-end learning and FPGA-based approaches.

The survey then addresses **Multi-node Cooperation in Distributed Systems**, analyzing the significance of collaborative tasks across multiple nodes to bolster the reliability and performance of communication networks. Strategies discussed include optimizing collaborative tasks through innovative frameworks like HALP, which enhances inference acceleration in distributed edge computing by effectively partitioning tasks based on receptive fields and maximizing the parallelization of communication and computation. Additionally, the Galaxy system leverages hybrid model parallelism and heterogeneity-aware planning to facilitate efficient transformer inference across diverse edge devices, addressing challenges posed by limited computing resources and bandwidth constraints, ultimately leading to significant reductions in inference latency and improved service reliability [15, 16].

The section on **Signal Processing Advances** reviews recent progress in signal processing techniques that enhance communication systems, exploring interference management and innovative estimation methods. The section on **Distributed Computing Strategies** highlights the importance of leveraging distributed resources to improve data processing and transmission efficiency, discussing optimization strategies within distributed computing environments.

Finally, the **Conclusion** synthesizes the key findings of the survey, reflecting on the current state of the field and suggesting future research directions. This includes outlining potential challenges and exploring the applications of the discussed methodologies in modern communication systems. The survey's meticulously organized framework facilitates an in-depth exploration of advanced techniques and their significant implications for enhancing communication networks, particularly through the analysis of near-field communications, graph signal processing, and innovative methodologies presented in Multi Layer Analysis, collectively contributing to the evolution and capacity of wireless systems [17, 18, 10].The following sections are organized as shown in Figure 1.

2 Background and Core Concepts

2.1 Key Concepts and Interrelations

The integration of modulation recognition, model segmentation, inference acceleration, multi-node cooperation, signal processing, and distributed computing is crucial for enhancing the efficiency of modern communication systems. Modulation recognition facilitates adaptive communication systems and is essential in contexts like software-defined radios and military applications [19]. Challenges arise in recognizing mixed signals within shared channels, which traditional machine learning frameworks struggle to address due to their limited adaptability in dynamic environments [20, 21]. Moreover, existing methods often fail to decode compressible signals without prior knowledge of encoding-decoding schemes, necessitating advanced approaches [22].

Signal processing is indispensable for extracting features amidst noise and channel variations. The co-existence of radar and communication systems requires sophisticated techniques to enhance spectrum efficiency and hardware utilization [23]. The classification and reconstruction of multicomponent signals with varying energy and spectral characteristics are complex tasks [24]. In distributed settings, innovative methods like Chebyshev polynomial-based approaches address the challenges of limited communication range and bandwidth [25]. The integration of sensing technologies with communication systems underscores the challenge of achieving high data transmission rates alongside accurate sensing in intelligent applications [26].

Model segmentation and inference acceleration are pivotal for optimizing computational efficiency. Partitioning models facilitates resource allocation, while hardware and software optimizations enhance real-time processing [23]. These advancements are essential in managing the high bandwidth and power demands of conventional systems, particularly with the increasing complexity of convolutional neural networks [12]. The deployment of deep learning models for signal modulation recognition is challenged by high computational complexity and large model sizes, necessitating strategies like layer pruning [27].

Multi-node cooperation enhances network reliability through collaborative tasks across nodes, crucial for in-network computation and federated learning, especially in satellite networks with heterogeneous objectives and limited resources [23]. The survey categorizes research into resource-level, node-level, and infrastructure-level cooperation, highlighting their interrelations and challenges.

Distributed computing strategies leverage computational resources across nodes for efficient data processing and transmission. Optimizing downlink user rates in large-scale distributed antenna systems underscores the importance of virtual cells in enhancing performance. This approach addresses the computational demands and resource limitations of neural network-assisted channel estimation in complex environments. Employing a scalable framework based on virtual cells allows users to select nearby base-station antennas, improving user rates while managing computational complexity in advanced beamforming techniques [28, 7, 29, 30]. Innovative methodologies, including automated Bayesian signal processing algorithm derivation with ForneyLab, demonstrate advancements in this domain.

The interplay among these domains collectively advances modern communication systems, ensuring robustness and adaptability. These concepts address existing methodological limitations, paving the way for innovative solutions that enhance network efficiency and reliability. Extending digital signal processing (DSP) to graph signal processing (GSP) involves adapting principles to finite graphs while managing unique boundary conditions. This evolution highlights the sophistication of signal processing techniques designed to tackle challenges in analyzing signals on non-Euclidean domains represented by weighted graphs. As GSP develops, it integrates DSP concepts with novel frameworks

AI-generated, for reference only.

to process irregular data structures across applications like data science, neuroscience, and multimedia processing [17, 31].

2.2 Integration of Sensing and Communication

The integration of sensing technologies with communication systems is pivotal for developing next-generation networks, particularly with the progression toward 5G and 6G. This synthesis supports applications requiring simultaneous data transmission and environmental sensing, such as intelligent transportation systems, where Integrated Sensing and Communication (ISAC) systems are increasingly crucial [32]. ISAC systems optimize hardware resources, enhancing radar and communication functionalities within a single platform, thus improving performance and spectrum efficiency [33].

Effectively merging radar sensing with communication functionalities necessitates advanced signal processing techniques to mitigate interference and enable seamless spectrum sharing. Interference alignment addresses theoretical and practical challenges in real-world applications [34]. Waveform design and interference mitigation are integral for coexistence without compromising performance. Machine learning, particularly deep neural networks, plays a significant role in cognitive radio tasks like modulation recognition, adapting to dynamic environments and improving signal classification accuracy [35].

In resource-constrained environments like Low Earth Orbit (LEO) satellite networks, integrating federated learning with communication systems is advantageous. This approach enables distributed learning across nodes without centralized data aggregation, preserving privacy and reducing communication overhead [36]. Learnable distortion correction modules in cognitive radio setups address signal distortion challenges due to wireless channels, enhancing modulation recognition [37].

Sophisticated automation strategies, including resource allocation, dynamic traffic routing, and security management, are vital for managing complex system interactions [38]. Advanced statistical techniques, such as approximating complex posterior distributions with simpler parametric ones, enhance the integration of sensing and communication [39]. Graph Signal Processing (GSP) techniques, involving graph signal transformation and reconstruction, manage data across distributed systems, addressing limited communication range and bandwidth constraints [25].

The integration of sensing and communication improves data transmission efficiency and reliability while opening new avenues for innovation in applications ranging from intelligent transportation to space exploration, where processing and analyzing data from multiple cosmic messengers are crucial [40]. This comprehensive integration ensures robust, adaptable communication systems capable of addressing modern wireless environment challenges.

3 Modulation Recognition Techniques

Category	Feature	Method
Innovative Modulation Recognition Frameworks	Model Optimization	LPM[27]

Table 1: Summary of an innovative modulation recognition framework highlighting the use of the Layer Pruning Method (LPM) for model optimization. This table categorizes the framework into distinct features and methods, emphasizing advancements in modulation recognition techniques.

The advancement of modulation recognition techniques is driven by the need to adapt to complex, dynamic signal environments. Table 1 provides a concise overview of the innovative modulation recognition frameworks, focusing on model optimization strategies such as the Layer Pruning Method (LPM). This section explores emerging frameworks that enhance classification accuracy and robustness through advanced methodologies and deep learning integration.

3.1 Innovative Modulation Recognition Frameworks

Cutting-edge modulation recognition frameworks are essential for improving communication systems' adaptability and performance in complex signal environments. Innovations such as the TLDNN architecture leverage global feature extraction and temporal modeling to advance modulation classification.

AI-generated, for reference only.

Incremental Learning (IL) addresses real-time application challenges by mitigating catastrophic forgetting and ensuring robust performance, even against adversarial perturbations [41, 42, 43]. Deep learning techniques, including CNNs, have significantly improved recognition accuracy while reducing parameter counts, with architectures like ResNet, DenseNet, and CLDNNs exemplifying deep networks' capacity to learn from raw data, enhancing both recognition and decoding tasks.

A hybrid framework, 'model-based deep learning,' merges traditional model-based methods with data-driven approaches, enhancing interpretability and reliability by leveraging domain-specific insights [39, 27, 44, 23, 45]. This framework includes strategies like Learned Optimization, Deep Unfolding, and DNN-Aided Inference, offering varying parameterization levels for diverse signal environments. Additionally, robust inverse transformation methods and energy-weighted distances enhance graph reconstruction, improving modulation recognition.

Data augmentation techniques, such as segment substitution and signal mixing, bolster the robustness and generalization of deep learning models for Automatic Modulation Recognition (AMR), addressing limited training data challenges and capturing long temporal dependencies and global correlations [41, 46]. Techniques like the linear Gaussian orthonormal measurement model enable efficient estimation with mixed-resolution data, enhancing modulation recognition.

The integration of Nonuniform Compressive Samples (NCS) with Nth Power Nonlinear Transformation (NPT) represents a novel approach to spectrum utilization, facilitating accurate modulation recognition in wideband environments by leveraging digital signals' phase sparsity and enabling sub-Nyquist sampling. This strategy reduces ADC burdens and enhances spectrum reconstruction, improving AMR performance [47, 48, 49]. Multi-view CNNs and wide linear CNNs aim to learn features directly from multi-antenna signals, enhancing modulation classification by exploiting spatial diversity.

Deep learning benchmarks demonstrate the ability to classify complex signal data without relying on manually crafted features, highlighting these technologies' transformative potential in modulation recognition. The proposed layer pruning method streamlines deep learning models by selectively preserving performance-contributing layers [27]. Collectively, these frameworks advance modulation recognition by addressing traditional methods' limitations, paving the way for more robust and efficient communication systems. The integration of deep learning, hybrid models, and advanced signal processing techniques enhances modulation recognition's adaptability and performance in dynamic wireless environments.

As shown in Figure 2, this figure illustrates the key innovative frameworks in modulation recognition, categorized into deep learning techniques, hybrid frameworks, and data augmentation methods. Each category highlights specific advancements and methods that enhance the adaptability and performance of communication systems in dynamic environments. The examples highlight innovative frameworks, with the first image detailing parameters for simulating radio frequency signals. The second image compares time plots for various modulation techniques, aiding in distinguishing unique characteristics. The third image illustrates a neural network architecture with ReLU and weight layers, showcasing a sophisticated approach to signal processing and modulation recognition [49, 50, 51].

3.2 Traditional and Hybrid Techniques

Traditional modulation recognition techniques rely on feature-based methodologies, extracting signal characteristics for classification. These methods often require high sampling rates and are limited by handcrafted features, which may inadequately capture modern communication complexities. Likelihood-based approaches highlight traditional methods' challenges in dynamic environments [20]. Integrating feature extraction methods with support vector machines (SVM) has improved recognition accuracy, showcasing the potential of combining classical techniques with machine learning.

Hybrid approaches address limitations by combining traditional methods with advanced neural network architectures. Integrating architectures like CLDNNs, LSTMs, and ResNet, alongside strategies for optimizing training efficiency, such as dimensionality reduction and subsampling, has improved classification accuracy, especially in high-dimensional data scenarios with limited computational resources. Each architecture's strengths—CLDNNs for sequential data, LSTMs for temporal dependencies, and ResNet for deeper network training—help mitigate challenges posed by increased model complexity and training time [52, 39]. This hybrid strategy effectively leverages convolutional and recurrent architectures to tackle modulation recognition challenges.

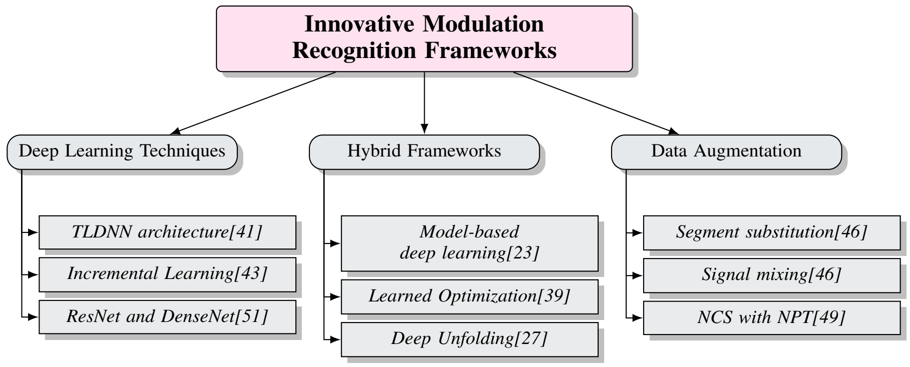


Figure 2: This figure illustrates the key innovative frameworks in modulation recognition, categorized into deep learning techniques, hybrid frameworks, and data augmentation methods. Each category highlights specific advancements and methods that enhance the adaptability and performance of communication systems in dynamic environments.

Deep residual networks have been applied to radio signal classification, significantly outperforming traditional methods and enabling end-to-end feature learning. This approach allows automatic feature extraction from raw data, reducing reliance on handcrafted features and improving classification performance. Hybrid techniques in signal processing utilize specialized neural networks for specific SNR ranges, enhancing performance and reducing computational complexity compared to a single network trained across a broad SNR spectrum. By employing distinct networks tailored to different SNR levels, researchers have achieved substantial efficiency and effectiveness improvements, underscoring the importance of selecting appropriate architectures for specific applications [28, 27, 52, 45, 18]. This strategy exemplifies the potential of combining deep learning with traditional methods to enhance recognition performance in complex scenarios.

Data augmentation techniques enhance deep learning models' robustness for AMR. Signal mixing methods generate new training data, improving model generalization across varying conditions. Such techniques address real-world signal environment variability and complexity. Learnable distortion correction modules, utilizing neural networks to estimate and rectify carrier frequency offsets and phase noise, significantly enhance CNN performance for modulation recognition by addressing signal distortions introduced by wireless channels [27, 48, 49, 1]. This hybrid approach improves recognition accuracy, allowing end-to-end training using modulation scheme labels without requiring true frequency or phase offsets.

The integration of traditional and hybrid techniques in modulation recognition enhances communication systems' adaptability and robustness. By merging traditional techniques with advanced neural network architectures, such as the TLDNN framework combining transformer and LSTM models, these approaches overcome classical methods' shortcomings. They enhance recognition efficiency and accuracy by capturing global correlations and temporal dependencies in modulated signals. Incorporating robust data augmentation strategies, like segment substitution, improves model generalization, especially in challenging scenarios such as few-shot learning, setting a new performance standard in AMR tasks [41, 51, 53].

3.3 Adversarial Robustness and Security

Ensuring the security and robustness of modulation recognition systems against adversarial attacks is crucial for communication networks' integrity. Adversarial perturbations, small, crafted modifications to input signals, pose a significant threat to AMC systems by inducing misclassification and undermining reliability. These perturbations exploit DNN vulnerabilities, raising concerns about model predictions' security and trustworthiness. Adversarial training has been proposed to enhance AMC model robustness, leading to improved performance and interpretability. Robust models demonstrate greater resilience against attacks and provide insights into relevant signal features, ensuring effective communication in dynamic environments like Cognitive Radio Networks [54, 55]. Despite their capabilities, DNNs are vulnerable to perturbations, compromising classification accuracy and reliability.

Developing robust models capable of withstanding adversarial attacks is essential. Current benchmarks often lack robustness in low SNR conditions and require preset formulations, introducing recognition biases [20]. The DNN-MR method enhances robustness against adversarial examples by effectively managing complex modulation schemes, outperforming traditional methods in resilience and accuracy. Spectrum-focused frequency adversarial attacks (SFFAA) improve attack effectiveness by concentrating adversarial energy within the frequency spectrum, enhancing performance while maintaining low detectability.

Frameworks designed to test and enhance state-of-the-art modulation recognition models' robustness are crucial for fortifying systems against adversarial threats. Such frameworks ensure AMC systems maintain high accuracy and reliability under adversarial conditions, safeguarding communication networks. Incorporating multiple antennas in modulation recognition systems enhances noise tolerance and classification accuracy, leveraging spatial diversity to provide robust defense against adversarial conditions. Techniques like adversarial training and innovative deep learning architectures enhance model robustness and interpretability, ensuring decisions are based on relevant signal statistics rather than spurious correlations [54, 56, 42, 41].

CNNs in space-time architectures have proven effective in achieving high classification accuracy for automatic modulation recognition, even in low SNR environments, without prior knowledge of channel coefficients or noise power. This capability is advantageous in noncooperative contexts, where traditional methods may struggle, as demonstrated by algorithms identifying modulation types and channel coding from received signals under challenging conditions [57, 19]. This approach enhances modulation recognition systems' security and robustness by enabling adaptation to and mitigation of adversarial perturbation effects, ensuring reliable performance across diverse and dynamic signal environments.

Enhancing adversarial robustness and security in modulation recognition systems is imperative for maintaining communication networks' effectiveness and integrity. By addressing traditional methods' limitations and harnessing advanced deep learning architectures' capabilities, these systems can improve resilience against adversarial attacks, ensuring robust and reliable performance in increasingly complex and dynamic wireless communication environments, where challenges such as unknown channel models and high-speed processing demands are prevalent [7, 58].

In recent years, the optimization of computation strategies in communication systems has garnered significant attention from researchers. This optimization is crucial for improving both the efficiency and effectiveness of various methodologies employed within the field. As illustrated in Figure 3, the hierarchical structure of these optimized computation strategies is clearly delineated. This figure categorizes advanced methodologies into two primary areas: resource allocation and model segmentation. Within this framework, it further distinguishes between graph-structured and matrix factorization approaches, thereby highlighting their respective contributions to enhancing computational performance. Such a structured representation not only aids in understanding the complex interrelations among these strategies but also underscores the importance of selecting appropriate methodologies tailored to specific communication challenges.

4 Model Segmentation for Optimized Computation

4.1 Resource Allocation and Model Segmentation

Optimizing computational efficiency in communication systems, particularly in complex signal environments, necessitates effective resource allocation and model segmentation. The CNN-STBC method exemplifies leveraging deep learning for enhanced computational efficiency through feature extraction from Space-Time Block Coded (STBC) signals, highlighting deep learning's role in resource management [19].

The RF1024 dataset, encompassing real RF data across eight modulation classes, is crucial for developing robust modulation recognition models, thus optimizing resource allocation strategies [20]. Advanced sampling methods like CaSCADE optimize resource allocation by recovering carrier frequencies and directions of arrival (DOAs) from compressed samples, reducing high-rate data streaming needs [37]. Additionally, integrating compressive sensing with eigenvalue decomposition enhances signal component separation, further improving resource allocation efficiency [24].

Graph-based techniques, such as multidimensional scaling, transform graphs into signals, facilitating signal property insights and reconstruction from processed signals [25]. This underscores graph signal processing's potential to enhance model segmentation and resource utilization.

In distributed learning environments, the SLIM-KL method optimizes Gaussian processes' hyperparameters using a quantized ADMM scheme with a distributed successive convex approximation algorithm, enhancing computational efficiency through effective resource allocation [36]. The fiber neural network, utilizing light transmission for direct computations on transmitted signals, further exemplifies the potential of optical systems in optimizing communication network resources [11].

The modified UNet architecture for 1D convolution predicts segmentation labels and estimates FFT representations, optimizing computational efficiency [59]. Dimensionality reduction techniques applied to high-dimensional Wi-Fi signal data enhance performance by reducing noise and redundancy before classification with SVM [60]. Layer pruning further optimizes efficiency by removing redundant layers, minimizing model size and computational demands [27].

These methodologies illustrate the potential of integrating advanced computational techniques, such as model segmentation and approximate computing, with innovative architectural designs. This integration enhances resource allocation and model segmentation in communication systems, leading to optimized performance and energy efficiency. Leveraging edge computing and DNN partitioning facilitates real-time inference by distributing computational tasks between mobile devices and cloud resources, minimizing latency and enhancing user experience. Graph signal processing offers new avenues for analyzing and processing data in irregular domains, enriching modern communication systems' capabilities [61, 62, 17, 63]. Future research should explore novel strategies to advance these processes further.

4.2 Graph-Structured and Matrix Factorization Approaches

Graph-structured and matrix factorization approaches for model segmentation hold significant potential for optimizing computational efficiency in communication systems. Graph-structured methods leverage graphs' mathematical properties to capture irregular interactions among diverse data points, facilitating complex model segmentation through graph signal processing (GSP) and quiver signal processing (QSP). These methods enable hidden structure extraction in applications like data science, neuroscience, and image processing [17, 64, 25].

Integrating GSP with model segmentation offers advantages in distributed computing environments, providing a robust framework for analyzing signals on non-Euclidean domains represented by weighted graphs. This integration enhances the management of heterogeneous multidimensional information in networks, paving the way for innovative graph-based deep learning approaches and advanced graph neural networks [17, 64].

Matrix factorization methods decompose large matrices into smaller components, improving computational efficiency by significantly reducing data dimensionality. This enables effective feature extraction, essential for tasks like noise removal and analyzing non-stationary signals, as demonstrated by data-driven signal decomposition methods such as empirical mode decomposition and variational mode decomposition [35, 17, 65, 66]. Their application in communication systems optimizes resource allocation, especially with high-dimensional data.

The Mobility Cost-Aware Segmentation Algorithm (MCSA) illustrates matrix factorization's integration with model segmentation to optimize resource allocation in edge intelligence systems. By considering user mobility, energy consumption, and resource costs, MCSA enhances segmentation efficiency, ensuring effective computational resource allocation [67].

The Xampling framework compresses high-bandwidth signals in the analog domain before sampling, allowing for input subspace detection prior to digital processing, optimizing segmentation by reducing data processing volume [68].

In FPGA-based machine learning research, categorizing current research into distinct categories highlights diverse model segmentation approaches and their potential to optimize computational efficiency [69].

The Interleaved Operator (IOP) strategy offers a novel model segmentation approach by minimizing required communication connections for processing, enhancing computational efficiency by reducing communication overhead [70].

Augmenting radio signals through wavelet decomposition and reconstruction exemplifies matrix factorization techniques' application in enhancing training datasets. By decomposing IQ sequences and reconstructing new samples, this method improves robustness and accuracy in model segmentation processes [71].

These graph-structured and matrix factorization approaches highlight innovative model segmentation strategies that enhance communication systems' efficiency and effectiveness. Integrating GSP and matrix factorization techniques effectively addresses irregular data environments' complexities. GSP facilitates analyzing and processing signals on non-Euclidean domains, while matrix factorization enhances communication networks' robustness through data-driven statistical model learning, improving inference processes' efficiency and reliability. These methodologies tackle complex data challenges and pave the way for developing resilient and efficient communication systems [72, 17, 64, 25]. Future research should continue exploring novel strategies that optimize these processes, further advancing communication systems.

5 Inference Acceleration Methods

5.1 Hardware and Software Optimizations

Enhancing inference processes in communication systems requires optimizing both hardware and software components to boost network efficiency and performance. FPGA-based accelerators have significantly improved energy efficiency and real-time processing capabilities. For example, fiber neural networks in the optical domain facilitate real-time signal processing, enhancing efficiency and reducing power consumption compared to traditional electronic methods [11]. This underscores the potential of optical systems in optimizing hardware resources for inference acceleration.

The HALP method demonstrates distributed inference by partitioning tasks among edge devices, enabling collaborative CNN inference and minimizing communication overhead [73]. This highlights distributed computing's role in enhancing scalability and efficiency in communication networks.

Software optimizations complement hardware advancements, enhancing adaptability and efficiency in inference processes. The SynFlow method accelerates inference by simplifying the pruning process through training-free performance estimation, reducing model complexity [27]. Additionally, the Dynamic Feature Extraction Method (DFEM) adaptively selects features, improving model performance and reducing computational costs, crucial for maintaining high classification accuracy in noisy environments [60].

In radar and communication integration, the Joint Radar and Communication (JRC) method leverages millimeter-wave frequencies for high data rates and precise target detection [3]. This illustrates the potential of advanced signal processing techniques to enhance radar and communication functionalities within a unified framework.

The AID method evaluates algorithmic complexity across dimensional configurations, providing a framework for reconstructing messages from non-random signals [22]. This analysis enhances the robustness and accuracy of inference processes by optimizing complex signal decoding.

These hardware and software optimizations address the challenges of complex signal environments, paving the way for more robust and efficient communication networks. By employing advanced computational techniques and methodologies, these optimizations significantly improve inference efficiency in communication systems, as evidenced by advancements such as distributed CNNs in edge computing. This approach maximizes parallelization between communication and computing, ensuring high service reliability and reduced inference times, achieving performance improvements of 1.7-2.0x for single tasks and 1.7-1.8x for multiple tasks per batch. Moreover, integrating artificial intelligence and deep learning has revolutionized processing diverse datasets in fields like Multi-Messenger Astrophysics, highlighting advanced algorithms and high-performance computing's role in driving innovation and timely insights across various domains [16, 74].

5.2 End-to-End Learning and FPGA-Based Approaches

End-to-end learning and FPGA-based methods are crucial for enhancing inference efficiency in communication systems. Recent research highlights their capacity to accelerate machine learning algorithms and optimize deep neural network performance. End-to-end learning streamlines wireless signal classification by eliminating complex feature extraction, while FPGA implementations boost inference speed and reduce latency, making them valuable in edge computing environments where real-time performance is critical [75, 69, 27, 76, 63]. These approaches leverage deep learning architectures and specialized hardware to optimize performance, reduce latency, and improve real-time processing capabilities.

End-to-end learning frameworks, particularly those using recurrent neural networks (RNNs), have shown potential in decoding convolutional codes, demonstrating adaptability and efficiency in complex signal environments compared to traditional maximum likelihood (ML) Viterbi decoders [52]. The exploration of denoising neural networks (NNs) for channel estimation underscores strategic NN configuration, including quantization and pruning, to achieve complexity reductions while maintaining performance [28].

FPGA-based approaches provide essential hardware acceleration that complements end-to-end learning, enabling real-time processing through parallel execution capabilities. While FPGAs face architectural limitations in accommodating complex machine learning models, they offer trade-offs between inference and training acceleration [69]. Utilizing FPGA hardware acceleration, as demonstrated in methods employing shallow CNNs trained via knowledge distillation, enhances real-time processing by capitalizing on FPGA's parallel processing capabilities and reduced latency [77].

The HALP method also exemplifies distributed inference by allowing simultaneous processing of CNN layers across multiple edge servers, effectively managing communication to reduce total inference time [16]. This underscores distributed computing strategies' importance in enhancing scalability and efficiency in communication networks [73].

The RFS method segments the input tensor along the largest dimension to facilitate parallel processing while maintaining accuracy, showcasing an innovative optimization approach for inference processes [78]. This segmentation strategy is particularly beneficial in distributed systems, where efficient resource allocation and task distribution are critical for performance optimization.

The Galaxy framework orchestrates collaborative inference across heterogeneous edge devices using a hybrid model parallelism architecture, optimizing resource allocation and enhancing inference efficiency [15]. This collaborative approach exemplifies distributed systems' potential in leveraging FPGA-based acceleration to improve performance across diverse network environments.

Additionally, experiments simulating six digital signals (2ASK, 4ASK, BPSK, QPSK, 2FSK, and 4FSK) in mixed signal scenarios with varying signal-to-noise ratios (SNR) underscore feature extraction methods' robustness in modulation recognition [79]. These findings highlight adaptive learning strategies' importance in managing complex signal environments.

Integrating end-to-end learning with FPGA-based methods presents a robust framework for accelerating inference processes in communication systems. By combining deep learning architectures' strengths with specialized hardware, these approaches address complex signal environments' challenges, paving the way for more efficient and effective communication networks. Future research should focus on developing innovative strategies that enhance communication process optimization, particularly by leveraging advancements in machine learning, semantic signal processing, and graph signal processing. These strategies can lead to more efficient information understanding and transmission, addressing critical challenges such as bandwidth limitations and power consumption in communication systems. Integrating artificial intelligence and domain-specific algorithms will enable better analysis and processing of diverse data sources, ultimately driving significant progress in communication systems [17, 80, 81, 82, 74].

6 Multi-node Cooperation in Distributed Systems

6.1 Collaborative and Distributed Systems

Collaborative and distributed systems are pivotal for enhancing communication network performance and reliability through efficient multi-node cooperation. These systems enable the development of lightweight, adaptive AI models that can be retrained in real-time, crucial for tasks such as signal processing and environmental reconstruction in dynamic communication settings [83]. Distributed signal processing techniques, particularly those based on Chebyshev polynomials, effectively manage large-scale networks by minimizing communication demands and increasing robustness against noise and signal degradation, especially in sparse environments [84].

The Asynchronous Distributed Alternating Direction Method of Multipliers (AD-ADMM) exemplifies how asynchronous updates in distributed systems enhance time efficiency by reducing waiting times and optimizing computational resource utilization [85]. The HALP method further optimizes distributed deep learning inference by minimizing communication overhead through overlapping computations handled by the host edge server, thereby improving performance in time-sensitive applications [16, 73].

In multi-node scenarios, maintaining accuracy and robustness in challenging conditions is achieved through incremental refinement of segmentation predictions, as demonstrated by the multi-stage learning approach for radar pulse segmentation [86]. Deep neural network architectures that capture both spatial and temporal features enhance modulation recognition accuracy in distributed systems, surpassing traditional methods [2]. Graph-based signal processing techniques further enhance the ability to manage noise and signal degradation effectively [25].

Collaborative and distributed systems advance communication networks by enabling efficient resource allocation, optimizing task distribution, and enhancing performance in multi-node environments. Techniques such as hybrid model parallelism and seamless collaboration among edge devices significantly improve inference acceleration for deep learning tasks. Studies indicate that collaborative edge AI frameworks can reduce latency by up to 2.5 times through optimized inference task partitioning and load balancing across heterogeneous devices. These systems address network variability and resource constraints, ensuring high service reliability and improved user experiences in real-time applications [15, 16, 63, 73, 61]. Such advancements pave the way for more robust communication networks, encouraging future research into novel strategies that enhance collaborative and distributed system capabilities.

6.2 Collaborative Task Optimization

Collaborative task optimization in distributed systems is essential for improving communication network efficiency and effectiveness. Advanced methodologies, such as those by O'Shea and Hoydis, demonstrate the potential of convolutional neural networks in enhancing modulation classification within cognitive radio systems [50]. This approach highlights the need for scalable and flexible frameworks that optimize collaborative tasks across nodes, enhancing performance in dynamic signal environments.

The Galaxy framework exemplifies a resource-efficient method for collaborative task optimization, utilizing hybrid model parallelism to orchestrate inference across heterogeneous edge devices, significantly reducing latency for Transformer inference [15]. The coexistence of radar and communication systems introduces unique challenges that require collaborative design strategies, emphasizing the need for further research into hardware implementations to optimize collaborative tasks across nodes [87].

Alvarado et al.'s distributed DC programming approach provides algorithms with provable convergence, suitable for large-scale applications across diverse fields [88]. The integration of dithering and Bayesian methods in resource allocation, as explored by Berman et al., showcases efficient utilization of both analog and quantized measurements, improving estimation accuracy and reducing computational complexity [89]. Shuman et al.'s work on distributed signal processing using Chebyshev polynomial-based methods suggests potential for future research to enhance adaptability and robustness in various contexts [84].

The strategies outlined in these studies reveal innovative methodologies for optimizing collaborative tasks across network nodes, significantly enhancing the efficiency and effectiveness of distributed systems in communication networks. For example, the HALP scheme accelerates CNN inference by 1.7-2.0 times through effective task collaboration in edge computing. Adaptive dynamic pruning in federated learning reduces communication costs by 57

7 Signal Processing Advances

7.1 Interference Management and Signal Processing

Effective interference management is crucial for optimizing signal processing in communication systems, especially amidst increasing demands for high data rates and reliable connectivity. Techniques such as interference alignment and robust modulation recognition significantly mitigate interference in wireless environments, enhancing radio resource utilization and data transmission rates. The LinksIQ approach exemplifies robust modulation recognition, accurately identifying transmission patterns under imperfect conditions, thereby ensuring reliable signal detection in complex settings [34, 90].

The Quiver Signal Processing (QSP) framework offers an innovative approach to interference management, adeptly handling heterogeneous data types and providing a generalized methodology for graph neural networks. This is particularly beneficial in environments with diverse signal sources and interference patterns, where conventional techniques may falter [64]. By leveraging QSP, communication systems can enhance interference mitigation and signal clarity, improving overall network performance.

Algebraic Signal Processing Theory (ASPT) provides a robust framework for interference management, offering mathematical derivations that compare favorably with traditional techniques. This approach enables precise characterization and mitigation of interference effects in complex environments, facilitating targeted strategies that enhance signal quality and reliability [91].

Integrating advanced interference management techniques with signal processing methodologies equips communication systems to adapt to dynamic signal environments. Frameworks such as Quantized Spectrum Processing (QSP) and Adaptive Spectrum and Power Transmission (ASPT) significantly improve interference mitigation capabilities. For instance, the LinksIQ system achieves a 43

7.2 Innovative Estimation Techniques

Innovative estimation techniques are pivotal for enhancing signal processing capabilities in communication systems, especially in complex and dynamic environments. Recent advancements, such as large dimensional random matrix theory and Approximate Message Passing with Input noise (AMPI), are transforming traditional methodologies to effectively address issues like signal detection in colored noise and hardware impairments, thereby improving performance in applications such as wireless communications, radar, and bioinformatics [82, 92, 93, 94].

The Xampling framework represents a significant advancement in signal acquisition and processing, enabling reduced sampling rates and computational loads while maintaining compatibility with existing digital signal processing (DSP) algorithms [68]. This versatility makes Xampling a valuable tool in various applications, facilitating efficient signal estimation without compromising data recovery quality.

The Adaptive Message Passing Iteration (AMPI) technique demonstrates substantial improvements in data detection and signal recovery by effectively incorporating input noise into the estimation process [93]. This approach achieves near-optimal performance with low computational complexity, underscoring the potential of innovative estimation methods to enhance signal processing tasks in noisy environments.

Utilizing complex radio frequency (RF) samples in time-series data is essential for estimating the time of arrival of radio packets, benefiting applications such as wildlife tracking systems [95]. This dataset provides insights into the temporal dynamics of RF signals, facilitating precise time estimation and improving signal processing task accuracy.

Recent studies highlight the critical role of advanced methodologies, such as data-driven signal decomposition approaches and modern random matrix theory, in enhancing signal processing capabilities within communication systems. These techniques improve signal analysis accuracy by effectively decomposing non-stationary signals and addressing noise and artifact removal challenges, while also adapting to large-scale data complexities, facilitating faster and more efficient real-time processing [65, 94]. Leveraging frameworks like Xampling and AMPI, along with complex RF datasets, these techniques address dynamic and noisy environment challenges, paving the way for more robust and efficient communication networks. Future research should continue to explore novel strategies to optimize these processes, advancing the field of signal estimation and processing.

8 Distributed Computing Strategies

8.1 Leveraging Distributed Resources for Enhanced Data Processing

Integrating distributed resources significantly enhances data processing efficiency in communication systems by addressing the limitations of traditional models that focus on exact message replication, which often leads to high bandwidth and power consumption. By employing semantic information processing, these systems improve data comprehension while reducing resource demands. Techniques such as model segmentation between cloud and mobile devices and virtual cell frameworks in large-scale distributed antenna systems adapt to user mobility and varying resource constraints, thereby boosting efficiency and performance [80, 30, 81, 67, 61]. Distributing computational tasks across multiple nodes enhances processing speed and resource utilization, leading to improved overall network performance.

The decomposition framework by Alvarado et al. illustrates the application of distributed strategies across domains like signal processing, communications, and networking [88]. This framework partitions complex tasks into manageable sub-tasks processed concurrently across nodes, alleviating bottlenecks and optimizing resource allocation.

Efficient resource management in distributed computing environments is crucial for maintaining performance and reliability. Distributed algorithms facilitate dynamic resource allocation based on real-time network conditions, enhancing data processing across nodes by prioritizing data meaning over exact replication, thus optimizing bandwidth and power usage [96, 29, 81].

Distributed resources also enable advanced signal processing techniques, including graph signal processing and matrix factorization, which manage large-scale datasets by segmenting computational tasks between mobile devices and cloud services. This minimizes data transfer, optimizes workloads, and accelerates processing speed and accuracy. Graph signal processing (GSP) enhances data analysis on non-Euclidean domains, providing innovative tools for data handling in diverse applications [61, 17, 35, 18].

Strategically deploying distributed resources is essential for optimizing data processing in communication systems, addressing challenges like data explosion, privacy concerns, and mobile device limitations. By leveraging distributed frameworks and algorithms, these systems achieve improvements in performance, scalability, and adaptability, paving the way for robust communication networks. Future research should explore innovative strategies to further optimize distributed resource use, advancing data processing in communication systems [15, 75, 81, 67, 61].

8.2 Optimization Strategies in Distributed Computing

Optimizing distributed computing environments is vital for enhancing the efficiency of communication systems, especially given the increasing data processing demands. Recent advancements, such as the HALP task collaboration scheme, have significantly improved inference acceleration for distributed CNNs by maximizing parallelization between communication and computing. Integrating cloud and mobile computing for machine learning has shown that fine-grained task distribution effectively balances computational loads, reduces user wait times, and optimizes workloads [61, 16]. Effective optimization strategies focus on resource allocation, task distribution, and minimizing communication overhead to enhance scalability and reliability.

Advanced distributed algorithms, such as those by Alvarado et al., play a crucial role in optimizing environments by providing frameworks for task decomposition and resource allocation [88]. These algorithms partition tasks into components processed concurrently across nodes, enhancing system efficiency and reducing processing time.

Incorporating machine learning into distributed strategies optimizes resource use by enabling adaptive task scheduling and workload balancing. Integrating deep learning models, such as CNNs, allows for automated feature extraction, improving data processing accuracy and speed [50]. These models can be distributed across nodes for parallel processing, leveraging the computational power of networks to accelerate data analysis.

Graph signal processing techniques offer additional optimization by transforming data into graph structures for efficient processing in distributed environments. This approach simplifies data representation and enhances data segmentation and reconstruction [25]. Leveraging graph-based data structure enables significant improvements in processing speed and accuracy.

Collaborative task optimization strategies, as demonstrated in the Galaxy framework, highlight hybrid model parallelism's potential to orchestrate inference across heterogeneous devices [15]. This approach optimizes resource allocation by dynamically distributing tasks based on each device's capabilities, minimizing latency and enhancing performance.

Optimizing distributed environments requires a multifaceted approach incorporating advanced algorithms, machine learning, and graph signal processing. By prioritizing semantic fidelity over message replication, systems enhance performance, scalability, and adaptability, addressing bandwidth and power consumption limitations. Techniques like model segmentation in cloud computing, collaborative edge computing, and neural network compression contribute to optimizing resource allocation and improving accuracy, paving the way for advanced communication infrastructures. Future research should explore innovative strategies to further enhance distributed system capabilities [61, 16, 29, 81].

9 Conclusion

9.1 Future Directions and Challenges

The evolution of communication systems hinges on advancements in key domains, with modulation recognition remaining central to improving network accuracy and reducing errors. Future work should focus on refining neural architectures and exploring automated feature selection and dynamic extraction methods to validate their efficacy in real-world scenarios. Expanding transformation techniques to encompass a broader range of modulation types is expected to enhance classification outcomes. In model segmentation, optimizing algorithms to handle diverse noise profiles across signal processing applications is crucial. Innovations in convolutional neural networks, such as integrating residual and Long Short-Term Memory networks, promise to elevate communication capabilities. Additionally, refining hybrid models to balance accuracy with computational demands is imperative.

Inference acceleration can gain from optimizing decentralized frameworks and exploring non-convex objectives to improve performance and convergence. Enhancing optical mesh architectures to minimize crosstalk in high-dimensional optical systems presents a promising research avenue. Investigating advanced pruning techniques for real-time applications also offers significant potential. Multi-node cooperation in distributed systems should prioritize robust algorithms that operate with imperfect channel information and integrate interference alignment with new technologies, addressing scalability in extensive networks. Enhancing the resilience of distributed systems like HALP in dynamic environments and optimizing task partitioning remains vital.

Signal processing advancements should emphasize noise resilience and feedback optimization in photonic computing, with potential benefits for communication systems. Integrating interference management with emerging technologies to boost performance in dynamic conditions, refining waveform design, and exploring cognitive spectrum management capabilities are promising directions. The future of communication systems will be shaped by optimizing existing methods, exploring innovative frameworks, and addressing challenges in dynamic signal environments, leveraging advanced computational techniques to forge robust and efficient networks.

9.2 Applications in Modern Communication Systems

The methodologies discussed in this survey have profound implications for modern communication systems, notably enhancing efficiency, adaptability, and performance. FPGA-based acceleration is crucial for optimizing machine learning applications, particularly in inference tasks, with CNNs highlighting FPGA's role in achieving real-time processing and energy efficiency, essential for next-generation systems. Integrating model-based approaches with deep learning frameworks advances signal processing tasks, fostering robust and efficient systems that ensure network reliability. In integrated sensing and communication systems, index modulation techniques optimize spectrum usage and improve radar and communication coexistence, warranting experimental validation for real-world effectiveness.

Fully dense neural networks excel in automatic modulation recognition, achieving superior recognition rates and enhancing classification accuracy in complex environments, thus boosting system efficiency. Developing multi-standard transceivers is vital for efficient data exchange in wireless communication, with protocol selection based on application requirements optimizing performance across diverse settings. Real-time event recognition systems for long-distance fiber monitoring demonstrate efficacy in smart IoT applications, enhancing processing speed and accuracy, crucial for managing extensive communication infrastructures.

Distributed signal processing techniques, particularly Chebyshev polynomial-based methods, significantly enhance communication efficiency and accuracy, proving effective in denoising and semi-supervised learning applications. Goal-oriented semantic signal processing frameworks are vital for IoT, autonomous vehicles, and smart cities, ensuring efficient data processing and system performance. Advanced techniques also enhance automatic modulation classification and specific emitter identification, improving accuracy in complex scenarios. Lastly, a unifying framework for message-passing algorithms supports various wireless communication scenarios, enhancing robustness and adaptability. Collectively, these methodologies underscore the transformative potential of advanced computational techniques in modern communication systems, paving the way for more efficient, reliable, and adaptable networks.

References

[1] Shangao Lin, Yuan Zeng, and Yi Gong. Learning of time-frequency attention mechanism for automatic modulation recognition, 2022.

[2] Xiaoyu Liu, Diyu Yang, and Aly El Gamal. Deep neural network architectures for modulation classification, 2018.

[3] Kumar Vijay Mishra, Bhavani Shankar M. R., Visa Koivunen, Björn Ottersten, and Sergiy A. Vorobyov. Toward millimeter wave joint radar-communications: A signal processing perspective, 2019.

[4] Deep cooperation in isac system:.

[5] Fuxin Zhang, Chunbo Luo, Jialang Xu, Yang Luo, and FuChun Zheng. Deep learning based automatic modulation recognition: Models, datasets, and challenges, 2022.

[6] I. P. Vieira, T. C. Pita, and D. A. A. Mello. Modulation and signal processing for leo-leo optical inter-satellite links, 2023.

[7] Tianqi Wang, Chao-Kai Wen, Hanqing Wang, Feifei Gao, Tao Jiang, and Shi Jin. Deep learning for wireless physical layer: Opportunities and challenges. *China Communications*, 14(11):92–111, 2017.

[8] Timothy James O'Shea, Tamoghna Roy, and T Charles Clancy. Over-the-air deep learning based radio signal classification. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):168–179, 2018.

[9] Ruochen Wu. Research on virus cyberattack-defense based on electromagnetic radiation, 2023.

[10] Yuanwei Liu, Chongjun Ouyang, Zhaolin Wang, Jiaqi Xu, Xidong Mu, and A. Lee Swindlehurst. Near-field communications: A comprehensive survey, 2024.

[11] Yubin Zang, Zuxing Zhang, Simin Li, Fangzheng Zhang, and Hongwei Chen. Fiber neural networks for the intelligent optical fiber communications, 2024.

[12] Kaihang Lu, Zengqi Chen, Hao Chen, Wu Zhou, Zunyue Zhang, Hon Ki Tsang, and Yeyu Tong. Empowering high-dimensional optical fiber communications with integrated photonic processors, 2023.

[13] Sha Hu, Fredrik Rusek, and Ove Edfors. Beyond massive mimo: The potential of data transmission with large intelligent surfaces. *IEEE Transactions on Signal Processing*, 66(10):2746–2758, 2018.

[14] Do-Hyun Park, Min-Wook Jeon, Jinwoo Jeong, Isaac Sim, Sangbom Yun, Junghyun Seo, and Hyoung-Nam Kim. Efficient radar modulation recognition via a noise-aware ensemble neural network, 2024.

[15] Shengyuan Ye, Jiangsu Du, Liekang Zeng, Wenzhong Ou, Xiaowen Chu, Yutong Lu, and Xu Chen. Galaxy: A resource-efficient collaborative edge ai system for in-situ transformer inference, 2024.

[16] Nan Li, Alexandros Iosifidis, and Qi Zhang. Distributed deep learning inference acceleration using seamless collaboration in edge computing, 2022.

[17] Geert Leus, Antonio G. Marques, José M. F. Moura, Antonio Ortega, and David I Shuman. Graph signal processing: History, development, impact, and outlook, 2023.

[18] Luca Pinello. Multi layer analysis, 2011.

[19] Wenjun Yan, Qing Ling, and Limin Zhang. Convolutional neural networks for space-time block coding recognition, 2020.

[20] Umar Khalid, Nazmul Karim, and Nazanin Rahnavard. Rf signal transformation and classification using deep neural networks, 2022.

[21] Aymeric Dieuleveut, Gersende Fort, Eric Moulines, and Hoi-To Wai. Stochastic approximation beyond gradient for signal processing and machine learning, 2023.

[22] Hector Zenil and Felipe S. Abrahão. Decoding geometric properties in non-random data from first information-theoretic principles, 2024.

[23] Nir Shlezinger and Yonina C. Eldar. Model-based deep learning, 2023.

[24] Andjela Draganic. Analysis of non-stationary multicomponent signals with a focus on the compressive sensing approach, 2019.

[25] Ronan Hamon, Pierre Borgnat, Patrick Flandrin, and Céline Robardet. From graphs to signals and back: Identification of network structures using spectral analysis, 2016.

[26] Dao Thanh Hai. Optical networking in future-land: From optical-bypass-enabled to optical-processing-enabled paradigm, 2022.

[27] Yao Lu, Yutao Zhu, Yuqi Li, Dongwei Xu, Yun Lin, Qi Xuan, and Xiaoniu Yang. A generic layer pruning method for signal modulation recognition deep learning models, 2024.

[28] Michel van Lier, Alexios Balatsoukas-Stimming, Henk Corporaaal, and Zoran Zivkovic. Optcomnet: Optimized neural networks for low-complexity channel estimation, 2020.

[29] Jiajia Guo, Jinghe Wang, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. Compression and acceleration of neural networks for communications, 2019.

[30] Junyuan Wang and Lin Dai. Downlink rate analysis for virtual-cell based large-scale distributed antenna systems, 2015.

[31] John Shi and Jose M. F. Moura. Gsp = dsp + boundary conditions – the graph signal processing companion model, 2024.

[32] Bhavani Shankar M. R., Kumar Vijay Mishra, and Mohammad Alaee-Kerahroodi. Emerging prototyping activities in joint radar-communications, 2022.

[33] Zhiqing Wei, Hanyang Qu, Yuan Wang, Xin Yuan, Huici Wu, Ying Du, Kaifeng Han, Ning Zhang, and Zhiyong Feng. Integrated sensing and communication signals toward 5g-a and 6g: A survey, 2023.

[34] Nima Najari Moghadam, Hamed Farhadi, Per Zetterberg, Majid Nasiri Khormuji, and Mikael Skoglund. Interference alignment: Practical challenges and test-bed implementation, 2014.

[35] Sulaiman Aburakhia, Abdallah Shami, and George K. Karagiannidis. On the intersection of signal processing and machine learning: A use case-driven analysis approach, 2024.

[36] Richard Cornelius Suwandi, Zhidi Lin, Feng Yin, Zhiguo Wang, and Sergios Theodoridis. Sparsity-aware distributed learning for gaussian processes with linear multiple kernel, 2025.

[37] Shahar Stein, Or Yair, Deborah Cohen, and Yonina C. Eldar. Cascade: Compressed carrier and doa estimation, 2016.

[38] Dan Zhang, Xiaohang Song, Wenjin Wang, Gerhard Fettweis, and Xiqi Gao. Unifying message passing algorithms under the framework of constrained bethe free energy minimization, 2021.

[39] Qianru Zhang, Meng Zhang, Tinghuan Chen, Zhifei Sun, Yuzhe Ma, and Bei Yu. Recent advances in convolutional neural network acceleration. *Neurocomputing*, 323:37–51, 2019.

[40] Jing Ma and Ying Jun Zhang. On capacity of wireless ad hoc networks with mimo mmse receivers, 2008.

[41] Yunpeng Qu, Zhilin Lu, Rui Zeng, Jintao Wang, and Jian Wang. Enhancing automatic modulation recognition through robust global feature extraction, 2024.

[42] Javier Maroto, Gérôme Bovet, and Pascal Frossard. On the benefits of robust models in modulation recognition, 2021.

[43] Ali Owfi, Ali Abbasi, Fatemeh Afghah, Jonathan Ashdown, and Kurt Turck. Dynamic online modulation recognition using incremental learning, 2023.

[44] Sixing Yu, Phuong Nguyen, Ali Anwar, and Ali Jannesari. Heterogeneous federated learning using dynamic model pruning and adaptive gradient, 2023.

[45] Josiah W. Smith. Complex-valued neural networks for data-driven signal processing and signal understanding, 2023.

[46] Xinjie Xu, Zhuangzhi Chen, Dongwei Xu, Huaji Zhou, Shanqing Yu, Shilian Zheng, Qi Xuan, and Xiaoniu Yang. Mixing signals: Data augmentation approach for deep learning based modulation recognition, 2024.

[47] Zhengli Xing, Jie Zhou, Jiangfeng Ye, Jun Yan, Lin Zou, and Qun Wan. Automatic modulation recognition of psk signals using nonuniform compressive samples based on high order statistics, 2014.

[48] Yuan Zeng, Meng Zhang, Fei Han, Yi Gong, and Jin Zhang. Spectrum analysis and convolutional neural network for automatic modulation recognition. *IEEE Wireless Communications Letters*, 8(3):929–932, 2019.

[49] Kumar Yashashwi, Amit Sethi, and Prasanna Chaporkar. A learnable distortion correction module for modulation recognition, 2018.

[50] Timothy J O'Shea, Johnathan Corgan, and T. Charles Clancy. Convolutional radio modulation recognition networks, 2016.

[51] Nathan E West and Tim O'shea. Deep architectures for modulation recognition. In *2017 IEEE international symposium on dynamic spectrum access networks (DySPAN)*, pages 1–6. IEEE, 2017.

[52] Daniel Tandler, Sebastian Dörner, Sebastian Cammerer, and Stephan ten Brink. On recurrent neural networks for sequence-based processing in communications, 2019.

[53] Nathan E West and Timothy J. O'Shea. Deep architectures for modulation recognition, 2017.

[54] Javier Maroto, Gérôme Bovet, and Pascal Frossard. Safeamc: Adversarial training for robust modulation recognition models, 2021.

[55] Ya Tu, Yun Lin, Jin Wang, and Jeong-Uk Kim. Semi-supervised learning with generative adversarial networks on digital signal modulation classification. *Computers, Materials & Continua*, 55(2), 2018.

[56] Lei Li, Qihang Peng, and Jun Wang. Deep modulation recognition with multiple receive antennas: An end-to-end feature learning approach, 2020.

[57] Sharan Ramjee, Shengtai Ju, Diyu Yang, Xiaoyu Liu, Aly El Gamal, and Yonina C Eldar. Fast deep learning for automatic modulation classification. *arXiv preprint arXiv:1901.05850*, 2019.

[58] Tianqi Wang, Chao-Kai Wen, Hanqing Wang, Feifei Gao, Tao Jiang, and Shi Jin. Deep learning for wireless physical layer: Opportunities and challenges, 2017.

[59] Akila Pemasiri, Zi Huang, Fraser Williams, Ethan Goan, Simon Denman, Terrence Martin, and Clinton Fookes. Automatic radar signal detection and fft estimation using deep learning, 2024.

[60] Shujie Hou, Robert C. Qiu, Zhe Chen, and Zhen Hu. Svm and dimensionality reduction in cognitive radio with experimental validation, 2011.

[61] Ruiqi Xu and Tianchi Zhang. Combining cloud and mobile computing for machine learning, 2024.

[62] Vasileios Leon, Muhammad Abdullah Hanif, Giorgos Armeniakos, Xun Jiao, Muhammad Shafique, Kiamal Pekmestzi, and Dimitrios Soudris. Approximate computing survey, part ii: Application-specific architectural approximation techniques and applications, 2023.

[63] En Li, Liekang Zeng, Zhi Zhou, and Xu Chen. Edge ai: On-demand accelerating deep neural network inference via edge computing. *IEEE transactions on wireless communications*, 19(1):447–457, 2019.

[64] Alejandro Parada-Mayorga, Hans Riess, Alejandro Ribeiro, and Robert Ghrist. Quiver signal processing (qsp), 2020.

[65] Thomas Eriksen and Naveed ur Rehman. Data-driven signal decomposition approaches: A comparative analysis, 2022.

[66] Alireza Roodaki, Julien Bect, and Gilles Fleury. Summarizing posterior distributions in signal decomposition problems when the number of components is unknown, 2011.

[67] Xin Yuan, Ning Li, kang Wei, Wenchao Xu, Quan Chen, Hao Chen, and Song Guo. Mobility and cost aware inference accelerating algorithm for edge intelligence, 2023.

[68] Moshe Mishali, Yonina C. Eldar, and Asaf Elron. Xampling: Signal acquisition and processing in union of subspaces, 2011.

[69] Feng Yan, Andreas Koch, and Oliver Sinnen. A survey on fpga-based accelerator for ml models, 2024.

[70] Zhibang Liu, Chaonong Xu, Zhizhuo Liu, Lekai Huang, Jiachen Wei, and Chao Li. Cooperative inference with interleaved operator partitioning for cnns, 2024.

[71] Tao Chen, Shilian Zheng, Kunfeng Qiu, Luxin Zhang, Qi Xuan, and Xiaoniu Yang. Augmenting radio signals with wavelet transform for deep learning-based modulation recognition, 2023.

[72] Nir Shlezinger, Nariman Farsad, Yonina C. Eldar, and Andrea J. Goldsmith. Data-driven factor graphs for deep symbol detection, 2020.

[73] Zhongtian Dong, Nan Li, Alexandros Iosifidis, and Qi Zhang. Design and prototyping distributed cnn inference acceleration in edge computing, 2022.

[74] E. A. Huerta and Zhizhen Zhao. Advances in machine and deep learning for modeling and real-time detection of multi-messenger sources, 2021.

[75] Tomer Gafni, Nir Shlezinger, Kobi Cohen, Yonina C. Eldar, and H. Vincent Poor. Federated learning: A signal processing perspective, 2021.

[76] Merima Kulin, Tarik Kazaz, Ingrid Moerman, and Eli de Poorter. End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications, 2017.

[77] Zhongyao Luo, Hao Wu, Zhao Ge, and Ming Tang. Real-time event recognition of long-distance distributed vibration sensing with knowledge distillation and hardware acceleration, 2024.

[78] Nan Li, Alexandros Iosifidis, and Qi Zhang. Receptive field-based segmentation for distributed cnn inference acceleration in collaborative edge computing, 2022.

[79] Rong Han and Zihuai Lin. Feature extraction, modulation and recognition of mixed signal based on svm, 2022.

[80] Mert Kalfa, Mehmetcan Gok, Arda Atalik, Busra Tegin, Tolga M. Duman, and Orhan Arikan. Towards goal-oriented semantic signal processing: Applications and future challenges, 2021.

[81] Guangming Shi, Dahua Gao, Xiaodan Song, Jingxuan Chai, Minxi Yang, Xuemei Xie, Leida Li, and Xuyang Li. A new communication paradigm: from bit accuracy to semantic fidelity, 2021.

[82] Emil Björnson, Yonina C. Eldar, Erik G. Larsson, Angel Lozano, and H. Vincent Poor. 25 years of signal processing advances for multiantenna communications, 2023.

[83] Bohao Lu, Zhiqing Wei, Huici Wu, Xinrui Zeng, Lin Wang, Xi Lu, Dongyang Mei, and Zhiyong Feng. Deep learning based multi-node isac 4d environmental reconstruction with uplink-downlink cooperation. *IEEE Internet of Things Journal*, 2024.

[84] David I Shuman, Pierre Vandergheynst, Daniel Kressner, and Pascal Frossard. Distributed signal processing via chebyshev polynomial approximation, 2017.

[85] Tsung-Hui Chang, Wei-Cheng Liao, Mingyi Hong, and Xiangfeng Wang. Asynchronous distributed admm for large-scale optimization- part ii: Linear convergence analysis and numerical performance, 2015.

[86] Zi Huang, Akila Pemasiri, Simon Denman, Clinton Fookes, and Terrence Martin. Multi-stage learning for radar pulse activity segmentation, 2023.

[87] Le Zheng, Marco Lops, Yonina C. Eldar, and Xiaodong Wang. Radar and communication co-existence: an overview, 2019.

[88] Alberth Alvarado, Gesualdo Scutari, and Jong-Shi Pang. A new distributed dc-programming method and its applications, 2013.

[89] Itai E. Berman and Tirza Routtenberg. Resource allocation and dithering of bayesian parameter estimation using mixed-resolution data, 2020.

[90] Wei Xiong, Karyn Doke, Petko Bogdanov, and Mariya Zheleva. Linksiq: Robust and efficient modulation recognition with imperfect spectrum scans, 2020.

[91] Markus Püschel and José M. F. Moura. Algebraic signal processing theory, 2019.

[92] N. Raj Rao and Jack W. Silverstein. Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples, 2009.

[93] Ramina Ghods, Charles Jeon, Arian Maleki, and Christoph Studer. Optimal data detection and signal estimation in systems with input noise, 2020.

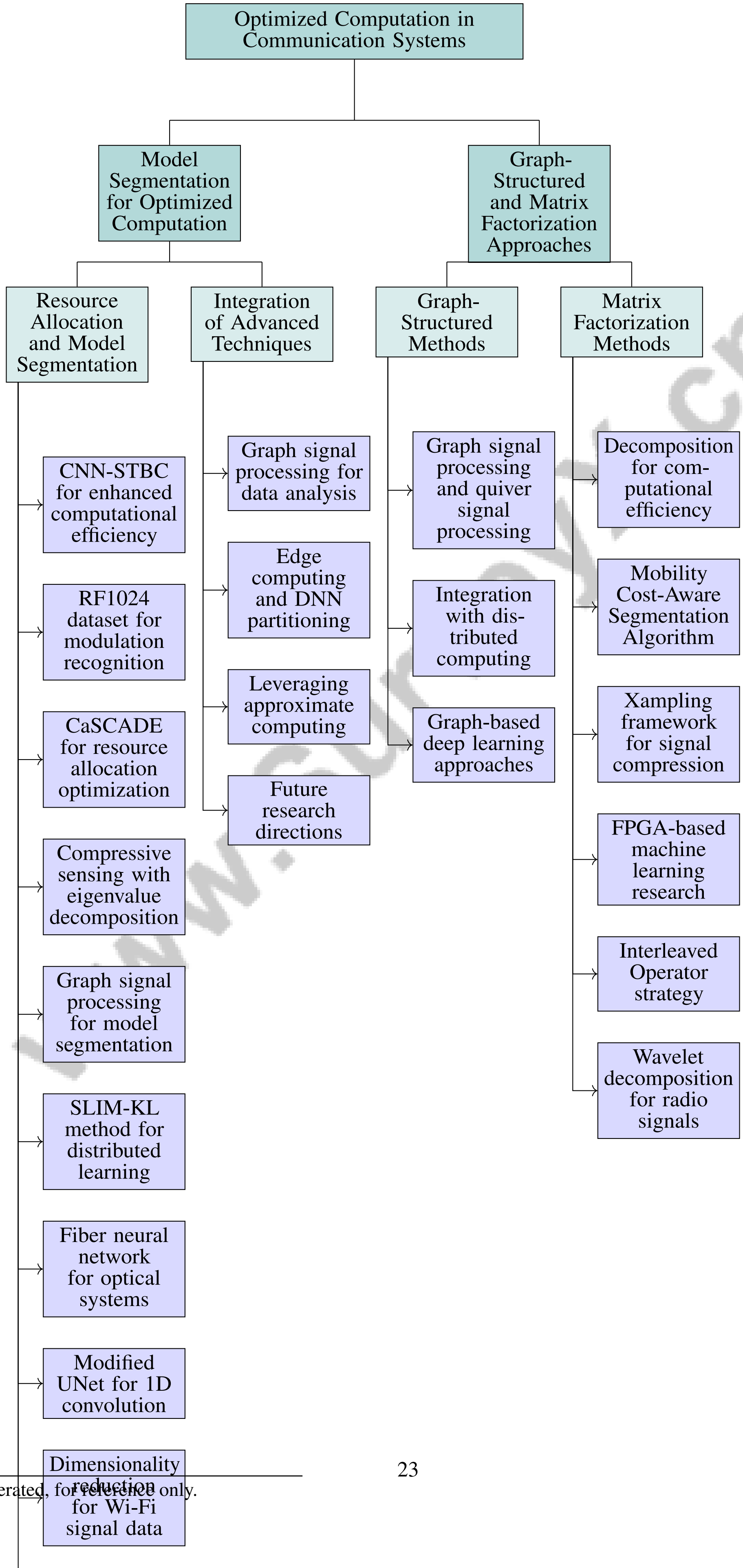
[94] Romain Couillet and Merouane Debbah. Signal processing in large systems: a new paradigm, 2012.

[95] Yaniv Rubinpur and Sivan Toledo. Signal processing for a reverse-gps wildlife tracking system: Cpu and gpu implementation experiences, 2021.

[96] Shashank Jere, Karim Said, Lizhong Zheng, and Lingjia Liu. Towards explainable machine learning: The effectiveness of reservoir computing in wireless receive processing, 2023.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.


Optimized Computation in Communication Systems
Model Segmentation for Optimized Computation
Graph-Structured and Matrix Factorization Approaches
Resource Allocation and Model Segmentation
Integration of Advanced Techniques
Graph-Structured Methods
Matrix Factorization Methods
CNN-STBC for enhanced computational efficiency
RF1024 dataset for modulation recognition
CaSCADE for resource allocation optimization
Compressive sensing with eigenvalue decomposition
Graph signal processing for model segmentation
SLIM-KL method for distributed learning
Fiber neural network for optical systems
Modified UNet for 1D convolution
Dimensionality reduction for Wi-Fi signal data
Graph signal processing for data analysis
Edge computing and DNN partitioning
Leveraging approximate computing
Future research directions
Graph signal processing and quiver signal processing
Integration with distributed computing
Graph-based deep learning approaches
Decomposition for computational efficiency
Mobility Cost-Aware Segmentation Algorithm
Xampling framework for signal compression
FPGA-based machine learning research
Interleaved Operator strategy
Wavelet decomposition for radio signals