Innovation and Application Development of Signal Processing Technology Based on Artificial Intelligence (Deep Learning / Reinforcement Learning)

Zichen Wang

[1]School of Information and Communication Engineering, Dalian University of Technology, Dalian, China

1249948394@qq.com

Abstract. This paper systematically reviews the technological innovation and application development of artificial intelligence (deep learning/reinforcement learning) in signal processing. Firstly, it sorts out the evolutionary path of deep learning models (such as CNN and Transformer) in signal feature extraction. It also analyzes the technological breakthroughs and limitations from local perception to global modeling. Secondly, it discusses the collaborative application of few-shot and transfer learning in medical signal processing and AI-driven industrial radar and communication signal processing cases. Finally, it focuses on three core issues: the real-time bottleneck of AI models in edge devices, improving signal classification accuracy in few-shot scenarios, and designing signal processing models with strong interpretability. It also proposes solutions and future research directions. Studies have shown that different models must be selected according to signal characteristics (e.g., CNN is suitable for locally stationary signals, and Transformer is ideal for long-time series signals). Integrating physical perception and data-driven approaches can improve the system's robustness. This paper provides a theoretical reference for the interdisciplinary integration of signal processing and AI and helps the technical implementation in fields such as 5G/6G and innovative healthcare.

Keywords: Artificial intelligence; Signal processing; Deep learning; Few-shot learning; Interpretable models.

1. Introduction

Traditional signal processing methods face significant limitations in complex scenarios. For example, the Fourier transform (FFT) and empirical mode decomposition rely on manual feature extraction in vibration signal analysis, struggling with nonlinear and high-noise environments [1]. The plane wave assumption in traditional massive MIMO fails in extremely large-scale MIMO (XL-MIMO) near-field scenarios, unable to capture spherical wave characteristics and spatial non-stationarity [2]. In dysarthric speech recognition, traditional automatic speech recognition (ASR) systems, dependent on precise phoneme labeling, suffer performance drops with ambiguous pronunciation and scarce data [3].

AI technologies offer breakthroughs in these areas. For instance, convolutional neural networks (CNN) convert vibration signals into spatiotemporal graphs via short-time Fourier transform (STFT), enabling interpretable analysis with gradient-weighted class activation mapping (Grad-CAM) [1]. In XL-MIMO, the residual network-based P-MRDN channel estimation algorithm improves processing accuracy in complex channels [2]. The end-to-end model Speech Vision bypasses phoneme recognition by visualizing speech (spectrograms) and using a 2D CNN, enhancing recognition accuracy for 67% of dysarthric speakers [3].

This paper systematically reviews core AI technologies in signal processing and their cross-domain applications, analyzes current bottlenecks, and proposes solutions and future directions to promote interdisciplinary integration and accelerate technological advancements in 5G/6G, industrial IoT, innovative healthcare, and other fields.

2. Technological Evolution and Limitations of Deep Learning Models in Signal Feature Extraction

The evolution of deep learning models in signal feature extraction follows a path from local perception to global modeling. Convolutional neural networks (CNN) serve as the foundational architecture. Early CNNs proposed by LeCun achieved automatic feature extraction in handwritten digit recognition through local connectivity and weight-sharing mechanisms [4]. AlexNet expanded feature extraction capabilities to complex image scenarios using ReLU activation and GPU acceleration, with its multi-layer convolutional structure hierarchically extracting features from edges to semantics [4].

1DCNN is optimized for one-dimensional signals like vibrations. It applies convolutional kernels directly to raw time-domain signals to avoid information loss from Fourier transforms. Its convolutional layers function as adaptive filters, capturing typical frequency components. Matching the number of kernels to the signal’ s dominant frequencies improves feature extraction efficiency [5].

Transformers overcome the local receptive field limitations of CNNs with their self-attention mechanism, modeling long-range dependencies. Vision Transformer (ViT) embeds signal segments in vibration signal processing and captures cross-temporal feature correlations via multi-head attention but overlooks local trends [6]. TAPformer addresses this with dual patch- and signal-level attention modules, combining coarse global features with fine-grained local trends, achieving 93.54% accuracy in planetary gearbox fault diagnosis [6]. However, Transformers require large datasets for training and underperform CNNs in few-shot scenarios [7].

Current research gaps include CNN’ s limited dynamic feature capture for non-stationary signals, Transformer’ s high computational complexity hindering real-time applications, and both models’ lack of explicit physical modeling. Empirical results show 1DCNN achieves 100% accuracy in bearing fault diagnosis [5], while TAPformer demonstrates robustness across datasets [6]. This highlights the need to select models based on signal characteristics—CNNs for locally stationary signals and Transformers for long-sequence signals.

3. Applications and Case Studies

Few-shot learning (FSL) and transfer learning exhibit strong synergy in medical signal processing. Medical scenarios often suffer from scarce labeled data (e.g., rare disease imaging, niche modal signals), limiting model training. FSL enables rapid generalization to new tasks with few samples, while transfer learning bridges data gaps by transferring knowledge from general domains. For example, self-supervised learning (SSL) generates pseudo-labels from superpixels to train models without manual annotations, transferring feature extraction capabilities from natural images to medical image segmentation. In abdominal CT and cardiac MRI tasks, SSL-ALPNet achieves a Dice score of 76.81% in 1-shot scenarios, outperforming traditional annotation-dependent methods [8]. This approach reduces labeling costs and facilitates cross-modal knowledge transfer for early disease diagnosis [9].

In radar and communication signal processing, AI applications demonstrate the fusion of physics-aware and data-driven methods. For synthetic aperture radar (SAR) image processing, embedding physical layer knowledge (e.g., sensor characteristics, scattering mechanisms) into deep learning models enhances interpretability. Combining SAR signal time-frequency analysis with CNNs enables high-precision target recognition, such as bridges [10]. Multi-sensor fusion (radar, LiDAR, cameras) employs transfer learning for heterogeneous data alignment in connected vehicles. Millimeter-wave radar FMCW signal processing combined with visual feature transfer improves obstacle detection accuracy by over 20% in adverse weather [11]. These cases show that AI-driven signal processing enhances industrial system robustness and reduces reliance on large labeled datasets through physics-aware and data-driven synergy [10, 11].

In summary, FSL and transfer learning overcome data bottlenecks in healthcare, while radar and communication applications validate the role of physics-awareness in industrial reliability. Together, they advance AI adoption from general to specialized domains, offering paradigms for interdisciplinary integration [8, 9].

4. Challenges and Future Directions

4.1 Real-Time Bottlenecks of AI Models on Edge Devices

Edge devices are inherently constrained by limited computational resources, power budgets, and memory capacity, posing significant challenges to achieving real-time performance of complex AI models. Optimized architectures such as YOLO v5 integrated with Flip-Mosaic data augmentation and lightweight networks (e.g., MobileNet) have shown promise by balancing detection accuracy and computational efficiency, particularly in real-time vehicle detection scenarios [12]. Additionally, specialized hardware solutions like FPGA-based binary Deep Neural Network (DNN) accelerators demonstrate substantial improvements, reducing energy consumption by approximately seven times, highlighting the effectiveness of hardware-software co-design approaches [13].

However, these solutions remain somewhat limited by device heterogeneity, as no universally adaptable method currently addresses the diverse processing capabilities and energy constraints across different edge devices. Future research should prioritize developing dynamic task scheduling techniques capable of intelligently adjusting model complexity based on real-time device loads and contextual requirements. Federated learning approaches are also promising, as they minimize data transmission overhead and latency by performing decentralized model training directly on edge devices, thereby improving real-time responsiveness and data security simultaneously.

4.2 Improving Signal Classification Accuracy in Few-Shot Scenarios

Signal classification in scenarios with limited data availability, known as few-shot learning, is critical yet challenging, particularly in fields such as healthcare and industrial signal processing. Transfer learning has emerged as a valuable tool, leveraging pre-trained deep learning models to mitigate data scarcity. Recent advancements employing hybrid CNN-LSTM models in combination with continuous wavelet transform (CWT) have achieved significant accuracy (approximately 92%) in EEG-based Brain-Computer Interface (BCI) datasets by utilizing pre-trained networks like ResNet-50 and Inception-v3 [14]. Nevertheless, these methods heavily rely on extensive data augmentation strategies, limiting their efficiency in scenarios where synthetic data generation is less effective or insufficient.

Future studies should explore advanced meta-learning techniques combined with transfer learning, such as fine-tuning and domain-adaptive training methods, to enhance generalization capabilities further. Additionally, innovative approaches like cross-modal knowledge transfer, for instance between EEG and fMRI modalities, can potentially broaden the applicability and effectiveness of few-shot learning frameworks. Generative models, such as Generative Adversarial Networks (GANs), also represent a promising avenue for synthesizing high-quality, realistic data to augment limited datasets and facilitate robust few-shot training.

4.3 Designing Interpretable AI Models for Signal Processing

Interpretable AI models are increasingly important for validating and understanding predictions, particularly in safety-critical domains like healthcare, autonomous driving, and industrial monitoring. Conventional interpretable methods such as decision trees and linear models provide clarity in their decision-making processes but often lack the representational power needed for complex signal analysis tasks. Conversely, deep learning architectures like Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) offer high predictive accuracy but are considered black-box models due to their opaque internal decision mechanisms [15].

Hybrid modeling approaches integrating CNN-LSTM networks with attention mechanisms have shown potential by providing visualization of salient signal features, thereby bridging the gap between interpretability and performance [13]. Future research should concentrate on developing quantitative interpretability metrics, such as decision consistency and explanation stability, to objectively assess and benchmark interpretability. Additionally, exploring hybrid models that integrate neural networks with rule-based systems (e.g., CNN combined with interpretable rule lists) may offer a promising balance between interpretability and predictive accuracy, facilitating widespread adoption in critical real-world applications.

5. Conclusion

This paper systematically reviews AI’s core advancements, applications, and challenges in signal processing. Deep learning models (e.g., CNN, Transformer) overcome traditional limitations in nonlinear, high-noise scenarios, with CNNs excelling in locally stationary signals and Transformers in long-sequence analysis. Few-shot learning and transfer learning address data scarcity in healthcare, while AI-driven radar and communication applications validate physics-aware and data-driven fusion for industrial robustness.

Lightweight networks, cross-modal transfer, and hybrid models are solutions for edge device real-time performance, few-shot accuracy, and model interpretability. Future research should prioritize multimodal fusion, interpretability-performance balance, and adaptive few-shot strategies to deepen interdisciplinary integration and support 5G/6G, innovative healthcare, and industrial IoT advancements.

References

[1] Chen, H. Y., & Lee, C. H. (2020). Vibration signals analysis by explainable artificial intelligence (XAI) approach: Application on bearing faults diagnosis. IEEE Access, 8, 134246-134256.

[2] Wang, Z., Zhang, J., Du, H., Niyato, D., Cui, S., Ai, B., ... & Poor, H. V. (2024). A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications. IEEE Communications Surveys & Tutorials, 26(3), 1560-1605.

[3] Shahamiri, S. R. (2021). Speech vision: An end-to-end deep learning-based dysarthric automatic speech recognition system. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29, 852-861.

[4] Jogin, M., Madhulika, M. S., Divya, G. D., Meghana, R. K., & Apoorva, S. (2018, May). Feature extraction using convolutional neural networks (CNN) and deep learning. In 2018, the 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT) (pp. 2319-2323). IEEE.

[5] Huang, S., Tang, J., Dai, J., & Wang, Y. (2019). Signal status recognition based on 1DCNN and its feature extraction mechanism analysis. Sensors, 19(9), 2018.

[6] Zhang, R., Chu, F., Qiu, Y., He, D., & Wang, F. (2025). Trend-Aware Patch Feature Fusion Transformer for Fault Diagnosis of Vibration Signals. IEEE Sensors Journal.

[7] Zollanvari, A., Kunanbayev, K., Bitaghsir, S. A., & Bagheri, M. (2020). Transformer fault prognosis using deep recurrent neural network over vibration signals. IEEE Transactions on Instrumentation and Measurement, 70, 1-11.

[8] Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., & Rueckert, D. (2022). Self-supervised learning for few-shot medical image segmentation. IEEE Transactions on Medical Imaging, 41(7), 1837-1848.

[9] Song, Y., Wang, T., Cai, P., Mondal, S. K., & Sahoo, J. P. (2023). A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. ACM Computing Surveys, 55(13s), 1-40.

[10] Datcu, M., Huang, Z., Anghel, A., Zhao, J., & Cacoveanu, R. (2023). Explainable, physics-aware, trustworthy artificial intelligence: A paradigm shift for synthetic aperture radar. IEEE Geoscience and Remote Sensing Magazine, 11(1), 8-25.

[11] Butt, F. A., Chattha, J. N., Ahmad, J., Zia, M. U., Rizwan, M., & Naqvi, I. H. (2022). On the integration of enabling wireless technologies and sensor fusion for next-generation connected and autonomous vehicles. IEEE access, 10, 14643-14668.

[12] Zhang, Y., Guo, Z., Wu, J., Tian, Y., Tang, H., & Guo, X. (2022). Real-time vehicle detection based on improved yolo v5. Sustainability, 14(19), 12274.

[13] Hua, H., Li, Y., Wang, T., Dong, N., Li, W., & Cao, J. (2023). Edge computing with artificial intelligence: A machine learning perspective. ACM Computing Surveys, 55(9), 1-35.

[14] Khademi, Z., Ebrahimi, F., & Kordy, H. M. (2022). A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals. Computers in biology and medicine, 143, 105288.

[15] Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. Statistic Surveys, 16, 1-85