LPI Radar Signal Recognition through Vision Transformers and Integrated Pulses

Asfar Ahmad and Haejoon Jung

Kyung Hee Univ.

{asfarahmad, haejoonjung}@khu.ac.kr

Abstract

Low probability of intercept (LPI) radar signals employ low power and diverse modulation schemes to evade detection, posing major challenges for interception and classification. Conventional CNN-based approaches with time- frequency image (TFI) inputs struggle to capture long-range dependencies and often overlook the repetitive pulse structure inherent in radar transmissions. This paper presents PulseFormer, a transformer-based framework that integrates self-attention with multi-pulse aggregation for robust LPI signal recognition. Each intercepted pulse is converted into a TFI and encoded via a shared vision transformer backbone to exploit both intra- and inter-pulse dependencies. Experiments on twelve LPI modulation types under varying SNR conditions show that PulseFormer consistently outperforms CNN and standalone vit baselines, demonstrating superior robustness in low-SNR environments.

Ⅰ. Introduction

Low Probability of Intercept (LPI) radar systems are designed to minimize detectability by electronic surveillance sensors while maintaining reliable target detection. Unlike conventional radar waveforms, LPI signals employ low peak power, wide spectral spreading, and agile modulation schemes, making them difficult to intercept or classify using traditional receivers[1]. As modern electromagnetic environments grow increasingly congested, accurate recognition of LPI radar modulations has become essential for electronic warfare.

Deep learning has emerged as a powerful tool for automatic radar waveform recognition. Conventional approaches typically transform intercepted signals into two-dimensional time-frequency images (TFIs) and employ convolutional neural networks (CNNs) to classify modulation types based on learned spectral-temporal patterns[2].

To address these challenges, recent work has explored Pulse integration, a classical radar detection technique that coherently combines multiple pulses of the same modulation to enhance the effective SNR. In [3], Hwang *et al.* extended this concept to CNN-based LPI classification and demonstrated notable gains, yet confusion among structurally similar modulations such as P1 and P4 persisted, revealing the limitations of purely convolutional representations.

The Vision Transformer (ViT) [4] adapted from natural language processing, provides a promising alternative by modeling global dependencies through self-attention mechanisms. ViTs partition each TFI into patches and compute contextual relationships across the entire signal representation, offering improved robustness to noise and deformation. However, prior ViT-based methods have primarily focused on single-pulse analysis, neglecting the inherent repetitive nature of radar emissions.

In this work, we introduce PulseFormer, a transformer-based framework that integrates self-attention with multi-pulse aggregation for robust LPI radar signal classification. By encoding individual pulse TFIs through a shared ViT backbone and fusing their embeddings via a transformer-based integration module, PulseFormer captures both intra- and inter-pulse dependencies. Experimental evaluations on twelve representative LPI modulation types under varying SNR conditions show that PulseFormer consistently surpasses CNN and standalone ViT baselines, achieving superior performance particularly in low-SNR environments.

Ⅱ. System Model

In practical electronic-support environments, LPI radar signals are received as multiple low-power pulses transmitted in rapid succession. Each intercepted pulse is affected by random timing offsets and additive noise, making reliable classification challenging under non-coherent conditions. To address this, the proposed PulseFormer framework performs end-to-end processing from signal capture to modulation recognition, as illustrated in Fig.1.

In Phase 1, each intercepted pulse is first converted into a two-dimensional TFI using the Choi-Williams distribution (CWD). This representation effectively highlights spectral- temporal structures while suppressing cross- term interference, enabling robust visualization of modulation patterns such as LFM, Costas, and polyphase (P1- P4, T1- T4) schemes[3]. The resulting TFIs are fed into a shared Vision Transformer (ViT) backbone through a TimeDistributed layer. Unlike CNN, which primarily extracts local spatial features, the ViT employs patch

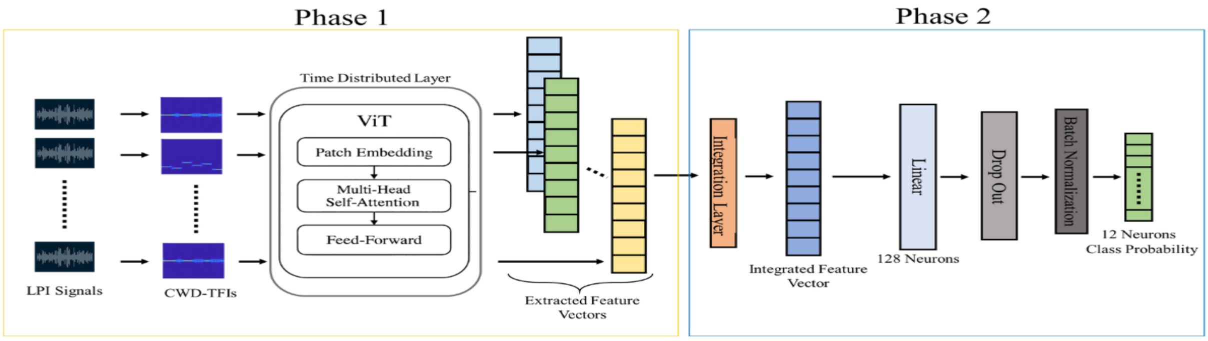

Figure 1. System Model

embedding and multi-head self-attention to capture long-range spectral dependencies and global contextual relationships across each pulse TFI.

In Phase 2, the extracted feature vectors from multiple pulses are aggregated through an average pooling-based pulse integration layer [3]. This module adaptively weights each pulse embedding according to its relevance, enhancing the effective SNR at the feature level. The integrated feature vector is then passed through a lightweight classification head that consists of a linear projection, dropout, batch normalization, and softmax layer to produce the final class probabilities corresponding to twelve LPI modulation types.

Overall, the proposed system leverages CWD-based TFI generation, ViT-driven feature encoding, and average pooling-based pulse integration to achieve robust and efficient classification of LPI radar waveforms in adverse electromagnetic conditions.

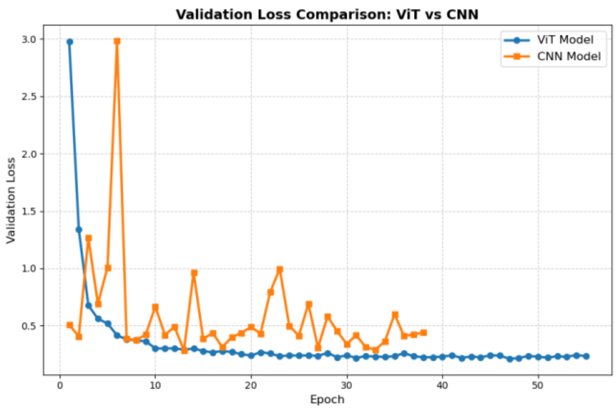

Figure 2. Validation Loss comparison

IV. Results

The performance of the proposed framework was evaluated and compared with a CNN-based model under identical experimental conditions. As illustrated in Fig. 2, the ViT backbone demonstrates a significantly faster convergence and lower validation loss across all training epochs compared to the CNN model. The reduced loss and smoother convergence trend of the ViT backbone confirm its superior ability to capture global dependencies and extract discriminative features from TFIs of LPI radar signals.

Overall, the results validate that the incorporation of the ViT into the classification framework enables more efficient training dynamics and improved generalization compared to conventional CNN-based models.

V. Conclusion

The proposed framework employing the ViT backbone demonstrated enhanced feature representation through global attention, making it more adaptable to diverse radar environments. The following framework can be extended by incorporating adaptive fusion or hybrid attention mechanisms to further improve performance under dynamic and low-SNR conditions.

ACKNOWLEDGEMENT

The work was supported by the MSIT, Korea, in part under the National Research Foundation of Korea grant (RS-2025-23323081), in part under the ITRC support programs (IITP-2025-RS-2021-II212046), and in part under the Convergence security core talent training business support program (IITP-2023-RS-2023-00266615) supervised by the IITP.

REFERENCES

[1] W. Tao, J. Kaili, L. Jingyi, J. Tingting, and T. Bin, "Research on lpi radar signal detection and parameter estimation technology," Journal of Systems Engineering and Electronics, vol. 32, no. 3, pp. 566– 572, 2021.

[2] C. Wang, J. Wang, and X. Zhang, "Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network," in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 2437– 2441.

[3] J. Hwang, Y. Kim, M. Yang, S. Cho, H. Jung, and D. Niyato, "Performance enhancement for cnn-based classification of low probability of intercept radar signals with pulse integration," IEEE Transactions on Vehicular Technology, pp. 1– 14, 2025.

[4] J. Kim, S. Cho, S. Hwang, and Y. Choi, "Automatic lpi radar waveform recognition using vision transformer," in 2023 IEEE International Radar Conference (RADAR), 2023, pp. 1– 6.