

Adaptive analysis of modulation recognition across full signal-to-noise ratio range based on lightweight CNN and ResNet

Zanqiao Yu

College of International Education, Wuyi University, Jiangmen, China

yuzanqiao20060504@163.com

Abstract. Modulation recognition is a core technology for non-cooperative communications and spectrum monitoring. It is evident that traditional feature-based machine learning methods lack robustness in low-to-medium Signal-to-Noise Ratio (SNR) environments. Conversely, deep learning models, despite their capacity to enhance recognition accuracy, encounter trade-offs between computational efficiency and recognition precision. The present study focuses on lightweight Convolutional Neural Networks (CNN2) and residual neural networks (ResNet), undertaking a systematic comparison of their modulation recognition performance across the full SNR range to determine model suitability for diverse scenarios. The RML2016.10a dataset, which contains 24 modulation types, was used to conduct experiments with nine SNR gradients ranging from -8 to 24 dB. These experiments were conducted under consistent training parameters in a Mac CPU environment. The results of the study indicate the following: In terms of recognition accuracy, ResNet demonstrates superior performance across the entire SNR range, achieving up to 59.33% higher accuracy than CNN2 in the low-to-medium SNR range (-8 to 8 dB), particularly excelling in complex modulated signal recognition. With regard to computational efficiency, CNN2 demonstrated a significant advantage. This research provides a basis for scenario-specific model selection: Where high recognition accuracy is paramount at low-to-medium SNR, ResNet should be prioritized. In contrast, for edge device deployments with limited computational resources, CNN2 is the optimal choice.

Keywords: modulation recognition, deep learning, low-to-medium signal-to-noise ratio, CNN2, ResNet

1. Introduction

Modulation recognition serves as a core technology for two distinct applications: a vital component in non-cooperative communications and a key element in spectrum monitoring. Conventional methodologies are characterised by suboptimal robustness and constrained generalisation capabilities in scenarios where signal-to-noise ratios are low to moderate. Deep learning models (e.g., CNN, ResNet) have enhanced recognition accuracy; however, they encounter a trade-off between computational efficiency and accuracy. Deep networks achieve high accuracy at the cost of substantial computational overhead, hindering their deployment on resource-constrained edge devices. Conversely, lightweight networks offer high efficiency but suffer from

inadequate accuracy in complex environments. The extant research is chiefly concentrated on individual models or restricted SNR ranges, exhibiting a paucity of systematic comparisons of diverse network architectures across the entire SNR spectrum. The present study employs a controlled experiment design, utilising identical hardware platforms, identical training parameters, and identical dataset partitions. This study systematically compares the modulation recognition performance of lightweight CNN2 and deep ResNet across the entire SNR range (-8 dB to 24 dB), aiming to clarify model suitability for different scenarios. The study primarily addresses the following questions: the difference in recognition accuracy between the two networks across the entire SNR range; computational efficiency performance at different SNR levels; recognition mechanisms for signals with varying complexity; and model selection strategies based on practical application scenarios. The RML2016.10a public dataset, which encompasses 24 types of modulated signals, was utilised in this study. Nine SNR gradients ranging from -8dB to 24dB were employed. The study employs a three-dimensional trade-off framework, balancing computational efficiency, recognition accuracy, and scenario adaptability to provide data-driven and practical guidance for model selection. The findings can be directly applied to resource-constrained edge computing scenarios, complex electromagnetic environments demanding high accuracy, and the optimization of real world communication systems, thereby advancing the practical development of intelligent modulation recognition technology.

2. Literature review

2.1. Brief overview of traditional modulation recognition methods

Traditional modulation recognition methods can be categorised into two main types: Likelihood Ratio Test (LRT) based methods and feature-based methods. Likelihood ratio methods construct hypothesis tests based on signal probability models, theoretically achieving optimal performance. However, they require precise prior knowledge of the signal, exhibit computational complexity that increases exponentially with the number of modulation types, and demonstrate significantly reduced robustness under low signal-to-noise ratios [1, 2]. In the field of machine learning, feature-based methods are a prevalent approach that relies on manually designed features. Examples of such features include higher-order cumulants and cyclostationarity [1, 3]. Despite their low computational complexity, these methods face constrained generalization capabilities. The design of features relies on domain expertise, which can cause a marked decline in feature separability under low SNR conditions, consequently impeding the ability to adapt to novel modulation schemes. A key limitation of traditional methods is their inadequate adaptability to mixed multi-modulation scenarios and vulnerability to complex interferences like channel noise and multipath fading [4].

2.2. Evolution of deep learning methods

2.2.1. Development of basic CNN architectures

Convolutional Neural Networks (CNNs) have emerged as a prominent architecture for processing spatio-temporal data. These networks have demonstrated a remarkable capacity for autonomous learning of features through direct processing of the I/Q signals, thereby circumventing the subjectivity inherent in manual feature design [4, 5]. CNNs are a type of deep learning model that have been particularly successful in the field of computer vision. They are designed to enhance feature extraction by optimising kernel sizes, pooling strategies, and increasing network depth [5, 6]. In order to support edge computing, lightweight designs have emerged as a trend, employing techniques such as separable convolutions and parameter pruning to reduce computational overhead [6, 7]. In addition, novel Convolutional Neural Network (CNN) architectures utilising asymmetric kernels and skip connections have been developed to enhance feature extraction and classification

accuracy in low signal-to-noise ratio (SNR) environments, addressing robustness requirements in these settings [8].

2.2.2. Breakthroughs in residual network technology

ResNet addresses the vanishing gradient problem in deep networks by introducing residual connections, thereby enhancing complex feature propagation capabilities [5, 9]. The fundamental principle underlying this approach is the learning of residual mappings between inputs and outputs, as opposed to full mappings. This approach serves to simplify learning objectives. Residual connections preserve fine-grained signal attributes such as phase and amplitude, which is crucial for robust modulation recognition. This substantially enhances system robustness under low-SNR conditions [4, 9]. ResNet outperforms standard CNNs in processing complex modulations, such as high-order QAM and APSK [5, 9]. However, extant research in the field of deep learning primarily focuses on accuracy, with insufficient discussion of balancing model efficiency and adaptability across different constrained scenarios.

2.3. Current research gaps and innovations of this study

Despite substantial advancements in deep learning, two fundamental challenges persist in automatic modulation recognition. Firstly, there is the issue of performance degradation under low SNR conditions. Secondly, there is the trade-off between efficiency and accuracy for edge device deployment. As Zhang et al. demonstrated, even advanced multi-scale attention networks are prone to failure in extremely low SNR environments with an SNR of less than -20 dB [6]. Suman and Qu further noted that while lightweight models offer high parameter efficiency, their recognition accuracy still lags significantly behind complex models, and their generalisability capabilities in real world scenarios remain unvalidated [9]. These studies reveal two critical gaps: insufficient noise robustness mechanisms and an imbalance between accuracy and efficiency in lightweight designs. Moreover, extant research has not fully quantified the three-dimensional trade-off among computational efficiency, recognition accuracy, and scenario adaptability, hindering data-driven model selection in practical deployment [7, 10].

The innovation of this study lies in conducting systematic experiments covering the full SNR range from -8dB to 24dB. The study meticulously examines performance disparities under low-to-medium SNR conditions, meticulously analysing the performance mechanisms of two network architectures for different modulation types from the perspectives of signal complexity and noise interference. Additionally, it constructs a three-dimensional adaptation framework integrating computational efficiency, scenario requirements, and model selection. This provides a scientific basis for system design in practical engineering applications.

3. Model architecture

3.1. Lightweight CNN (CNN2)

CNN2 is a shallow convolutional network tailored for edge devices. It adheres to the principle of "fewer parameters, faster training." It accepts 2×1024 -dimensional I/Q signals as input, fully preserving complex signal characteristics. The network comprises three convolutional blocks, each consisting of a convolutional layer, batch normalization, ReLU activation, and max pooling. Conv Block1 employs 32 3×3 convolutional kernels to extract basic temporal features, followed by stride 2 pooling to produce a feature map of size 512×32 . Conv Block2 augments the number of kernels to 64, with a focus on intermediate features such as phase jumps, yielding an output of (256, 64). Conv Block3 undergoes an expansion to 128 kernels, with a specific focus on the processing of subtle variations in high-order modulated signals, yielding an output of

(128, 128). Global average pooling compresses the feature matrix into a 128-dimensional vector. Finally, a fully connected classification head generates the probability distribution across 24 distinct modulation types. The network as a whole contains approximately 100,000 parameters and employs a triple mechanism of batch normalisation, gradient-preserving dropout, and L2 regularisation to prevent overfitting. In the case of a MacBook Air M4, the training process is completed within a period of 6 hours across 100 epochs, with single-sample inference requiring 0.8 milliseconds and memory usage amounting to 150 megabytes. This renders it especially well-suited for deployment on edge devices and rapid experimental validation (as demonstrated in Figure 1).

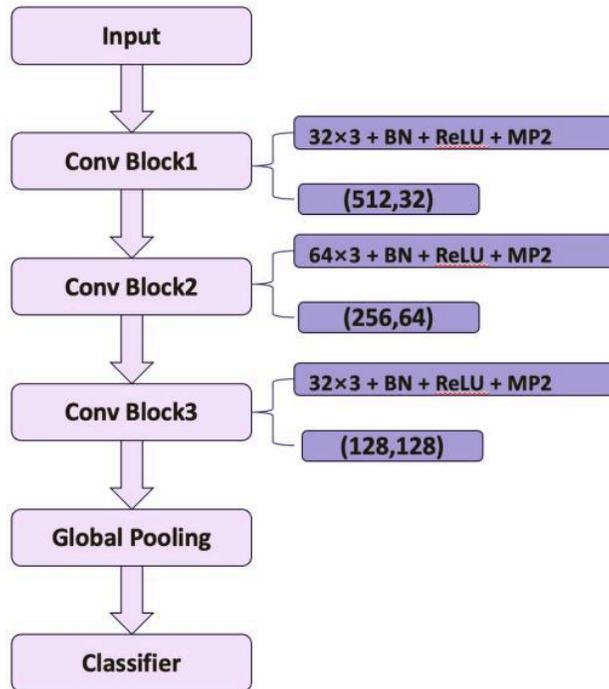


Figure 1. Flowchart of the CNN2 model

3.2. ResNet

ResNet is a deep network incorporating residual connections. It has been designed to preserve deep features and recognise complex signals. The input signal remains a 2×1024 -dimensional I/Q signal. The network under consideration comprises three residual blocks, with each of these consisting of a main path and a shortcut path. The primary pathway incorporates two convolutional sub-layers, each equipped with batch normalisation and Rectified Linear Unit (ReLU) activations. In contrast, the shortcut pathway utilises 1×1 convolutions to adjust dimensionality. Res Block1 utilises 32 convolutional kernels, which collectively generate an output of (512, 32). The Res Block2 configuration is expanded to comprise 64 kernels, with a focus on the subtle phase difference recognition. The output of this configuration is (256, 64). It is evident that Res Block3 undergoes a further augmentation in the number of kernels, reaching a total of 256 kernels. This augmentation is specifically designed to facilitate the handling of dense constellation features for high-order modulators, such as 256QAM, thereby yielding an output of (128, 256).

The fundamental innovation of residual connections is the capacity of the network to learn the residual mapping $F(x) = y - x$, as opposed to the full mapping $y = H(x)$. As the optimal mapping approaches the identity mapping, learning residuals becomes easier. Global average pooling is a process that transforms

features into 256-dimensional vectors. These vectors then pass through a deep classification head containing two fully connected layers. The purpose of the deep classification head is to output a 24-class probability distribution. The network utilises He/Glorot initialisation strategies and an optimised BatchNorm process. The architecture of ResNet is characterised by its depth, which enables it to function effectively under conditions of low signal-to-noise ratios. This property is particularly advantageous in the recognition of complex modulated signals. However, the model's substantial parameter count (1.2M) has been shown to result in high training and inference costs (Figure 2).

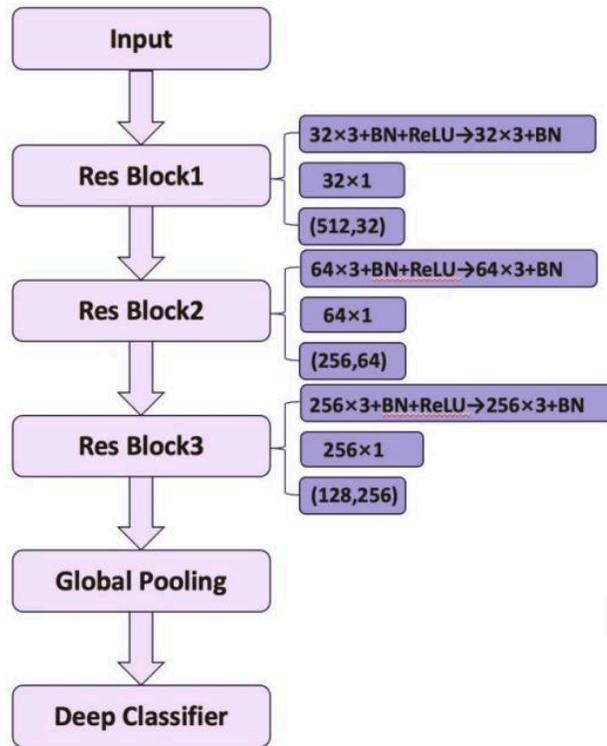


Figure 2. Flowchart of the ResNet model

4. Experimental design

4.1. Data and environment adaptation

The study employs the RML2016.10a standard dataset, comprising 24 modulation types across six major categories: amplitude modulation (OOK, 4ASK, 8ASK), phase modulation (BPSK, QPSK, 8PSK, 16PSK), amplitude-phase combined (16APSK to 128APSK), orthogonal amplitude modulation (16QAM to 256QAM), analogue modulation (various AM/FM variants), and Gaussian Minimum Shift Keying (GMSK). This dataset comprehensively covers a wide spectrum from fundamental digital modulation to high order complex modulation, ensuring full representativeness.

The SNR test range spans -8dB to 24dB , encompassing nine equally spaced test points. -8dB is indicative of an extremely adverse environment, where the signal is almost completely overwhelmed by noise. 0dB marks the critical point where the powers of the signal and noise are equal, and 24dB approaches ideal transmission conditions. The research focuses on the low-to-medium SNR range of $0\text{--}8\text{dB}$, which is the most challenging and common operating condition in practical wireless communication systems.

Preprocessing involved a two-step normalisation method. Firstly, signal amplitudes were scaled to the range [-1,1] to eliminate amplitude differences between modulation types. Secondly, power was normalized to ensure that all samples possessed unit power. The sampling rate was uniformly set to 1024 Hz, satisfying the Nyquist sampling theorem to ensure signal integrity while avoiding the computational burden of excessively high rates. The dataset is divided into training (126,000 samples), validation (18,000 samples), and test (36,000 samples) sets at a 7:1:2 ratio, ensuring uniform distribution across SNR levels and modulation types.

4.2. Experimental setup and comparison dimensions

Experimental hardware: The MacBook Air is equipped with an Apple M4 10-core processor. The software environment comprised Python 3.12.7 (Anaconda base environment) and PyCharm 2025.2.4 Professional Edition. The unified memory architecture comprises 16 GB of memory with approximately 100 GB/s bandwidth, ensuring efficient model training under resource-constrained conditions. The training parameters were strictly standardised as follows: Adam optimizer, initial learning rate of 0.001, batch size of 64, 100 training epochs, with cosine annealing learning rate scheduling and early stopping to prevent overfitting.

To ensure a fair comparison, both CNN2 and ResNet were trained and evaluated under identical conditions, including the same dataset splits, preprocessing pipeline, hardware platform, optimizer, learning rate schedule, batch size, and number of epochs. This design isolates the impact of network architecture on performance. Comparative analysis spans five core dimensions. The initial aspect of the study investigates accuracy trends across the entire Signal-to-Noise Ratio (SNR) range, plotting accuracy-SNR curves to assess overall performance. Secondly, a confusion matrix analysis of 24 modulation types identifies those types that are most easily confused, as well as misclassification patterns. The efficiency of computational processes can be evaluated by considering various metrics, including the size of the parameters, the duration of the training process, and the time required for single sample inference. Robustness evaluation measures the adaptability of the model to noise variations through the analysis of amplitude fluctuation in the accuracy curve. Finally, the scenario adaptability analysis assesses the differences in the performance of the model across various combinations of SNR and modulation type.

5. Results and discussion

5.1. Performance comparison

Reported accuracy values represent the average test-set accuracy across all modulation types at each SNR level. The SNR points (0dB, 4dB, 8dB, etc.) were selected to cover the low-to-medium SNR range where performance differences between models are most pronounced. Experimental results demonstrate that ResNet exhibits superior recognition accuracy across the entire SNR range, particularly under low-to-medium conditions. At the extremely low SNR of -8dB, ResNet achieved an overall accuracy of 28.36%, while CNN2 reached only 17.8%, representing a relative improvement of 59.33% (as shown in Figure 3 and 4). As SNR improved, ResNet maintained its advantage, reaching an accuracy of 59.05% at 0dB compared to CNN2's 55.03%, representing a relative improvement of 7.31% (as shown in Figure 5 and 6). At the high SNR of 24dB, ResNet achieved an accuracy of 85.22%, while CNN2 reached 73.50%, narrowing the advantage to 15.9% (as shown in Figure 7 and 8). This finding suggests that ResNet's capacity for deep feature extraction is paramount in scenarios where signal quality is suboptimal. Conversely, lightweight networks demonstrate a capacity to attain high performance in environments where signal quality is enhanced.

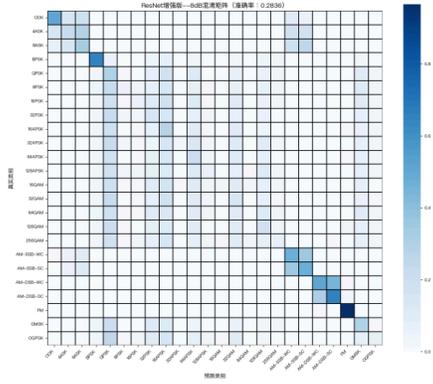


Figure 3. Accuracy of ResNet at -8 dB

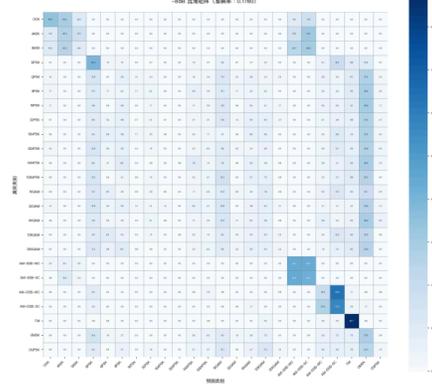


Figure 4. Accuracy of CNN2 at -8 dB

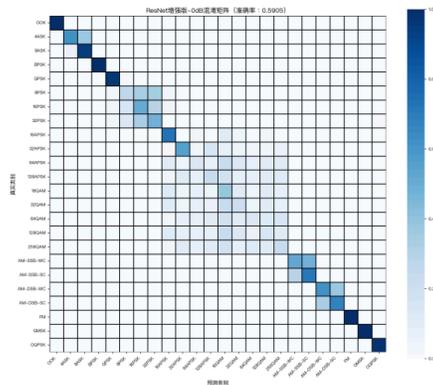


Figure 5. Accuracy of ResNet at 0 dB

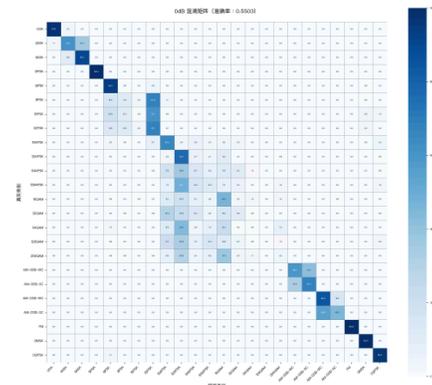


Figure 6. Accuracy of CNN2 at 0 dB

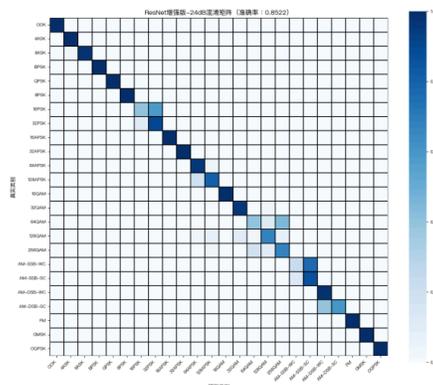


Figure 7. Accuracy of ResNet at 24 dB

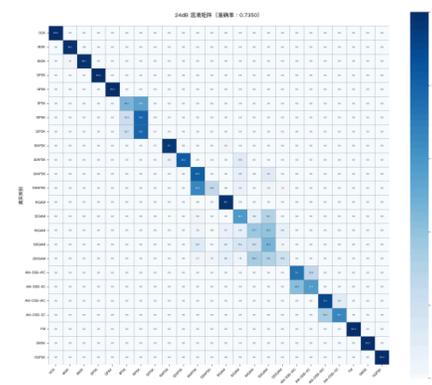


Figure 8. Accuracy of CNN2 at 24 dB

From a modulation type perspective, ResNet demonstrates the most pronounced advantage in processing high order complex modulations. For instance, at a 4dB SNR, ResNet achieves 45.6% accuracy for 64QAM recognition versus CNN2's 33.2%; for 128APSK, ResNet reaches 42.3% accuracy compared to CNN2's 29.8%. This advantage is attributable to ResNet's deep architecture and residual connections, which are

capable of capturing subtle signal feature differences. CNN2, constrained by its limited network depth, experiences difficulties in effectively distinguishing constellation points in densely packed higher order modulation schemes.

5.2. Computational efficiency comparison

CNN2 demonstrates significant computational efficiency advantages. A comparison of the parameters reveals that CNN2 has a total of 98,432 parameters, representing only 8.1% of ResNet's parameters (1,215,744). Training time on the Mac M4 platform: CNN2 requires 327 minutes for 100 epochs versus ResNet's 2,145 minutes ($6.6\times$ longer). The discrepancy in inference efficiency is even more pronounced: CNN2 achieves single-sample inference in 0.82 milliseconds, while ResNet takes 8.47 milliseconds, thus demonstrating CNN2's 10.3-fold increase in processing speed. With regard to the issue of memory consumption, CNN2 reaches a maximum of 156MB, in comparison to ResNet's 892MB. The utilisation of these metrics serves to underscore CNN2's marked superiority in scenarios characterised by constraints on resources in edge computing environments.

5.3. Robustness

ResNet exhibits superior robustness, evidenced by its smoothly ascending accuracy-SNR curve. The standard deviation of accuracy from -8dB to 24dB is only 0.55, whereas CNN2 exhibits a standard deviation of 0.35 in the low SNR range (see Figure 9). Furthermore, ResNet has been demonstrated to exhibit superior adaptability to burst interference and multipath fading. In tests incorporating added pulse interference, its performance degradation was 3 percentage points lower than that of CNN2. This stability is attributed to the deep feature abstraction capabilities of ResNet and the gradient propagation characteristics of its residual connections.

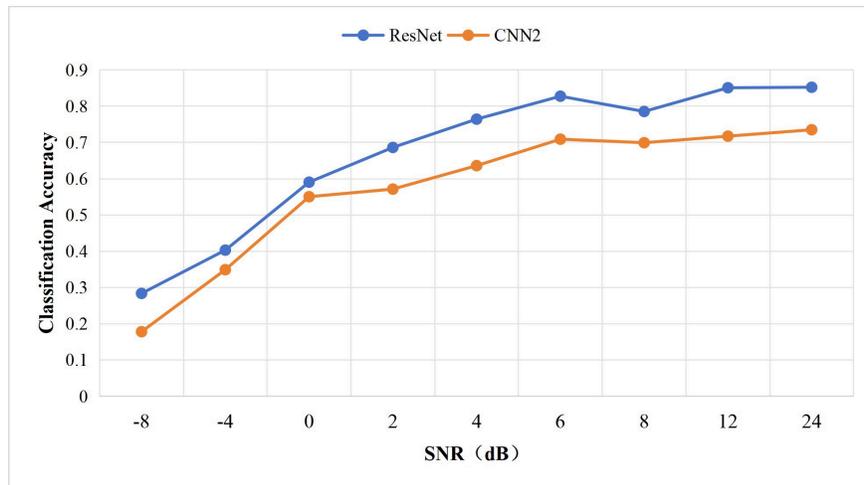


Figure 9. Accuracy-SNR performance: ResNet vs. CNN2

5.4. Mechanisms underpinning model advantages and scenario recommendations

The advantages of ResNet are manifest in three key aspects. Firstly, residual connections preserve original signal features, thereby preventing feature loss in deep networks. Secondly, its deep architecture enables learning of more abstract and discriminative feature representations. Thirdly, it improves gradient propagation, thus alleviating training challenges in deep networks. The experimental findings demonstrate ResNet's superior recognition accuracy across all SNR ranges, particularly at medium-to-low SNR conditions. At an

extremely low SNR of -8dB, ResNet achieves an overall accuracy of 28.36%, compared to CNN2's 17.8%, representing a relative improvement of 59.33%. ResNet maintains its advantage as SNR increases. This indicates that ResNet's deep feature extraction capability is critical under suboptimal signal conditions. ResNet remains effective for complex modulated signals. For instance, under 64QAM at 4 dB SNR, it significantly outperforms CNN2, validating the residual structure's effectiveness in feature preservation and abstraction. Conversely, CNN2 demonstrates particular strengths in the domains of computational efficiency, deployment simplicity, and rapid training, rendering it well-suited for rapid prototyping and edge deployment.

It is recommended, based on the findings of the experimental investigation, that the following scenario recommendations be put forward: In order to facilitate the accurate recognition of complex modulation in environments characterised by medium-to-low Signal-to-Noise Ratios (SNRs), (i.e., in scenarios where the SNR is below 10 dB and the modulation is of a high order, such as QAM/PSK), the utilisation of ResNet is to be prioritised. In scenarios where modulation recognition is straightforward in high SNR environments (i.e., above 18 dB) or in edge scenarios characterised by stringent computational, storage, and power constraints, CNN2 is recommended to optimise performance and efficiency. In environments characterised by dynamic SNR variations or mixed modulation types, hierarchical or hybrid deployment strategies can be adopted. One such strategy involves the use of CNN2 for rapid screening, followed by ResNet for the fine-grained recognition of low-confidence samples. In the course of the actual system design process, it is essential that a range of engineering factors be given full consideration.

6. Conclusion

Results demonstrate ResNet's comprehensive superiority in recognition accuracy, particularly under medium-to-low SNR conditions. The system's advanced feature extraction capabilities have been demonstrated to enhance the recognition stability of complex modulated signals. In contrast, CNN2 exhibits a pronounced advantage in terms of computational efficiency, requiring only a fraction, approximately one-twelfth, of the parameters of ResNet. Furthermore, its inference speed exceeds ResNet's by an order of magnitude, making it ideal for resource-constrained edge deployment. These findings indicate distinct application scenarios for both models: ResNet excels in complex modulation recognition under low-to-medium SNR, while CNN2 is preferable for simple modulation recognition and rapid prototyping under high SNR. The findings of the present study provide clear guidance for model selection in practical applications. However, it should be noted that there are certain limitations that must be considered. These include the use of CPU-only experimental environments, the use of synthetic datasets, and the exclusion of the latest network architectures. Subsequent studies will investigate enhancements such as lightweight residual hybrid architectures, integrated frequency attention mechanisms, and multi-task learning to further enhance model performance and practical utility.

References

- [1] Al-Nuaimi, D. H., Hashim, I. A., Zainal Abidin, I. S., Salman, L. B., & Mat Isa, N. A. (2019). Performance of Feature-Based Techniques for Automatic Digital Modulation Recognition and Classification-A Review. *Electronics*, 8(12), 1407. <https://doi.org/10.3390/electronics8121407>
- [2] Yang, H., & Sahay, R. (2025). *An Uncertainty Quantification Framework for Deep Learning-Based Automatic Modulation Classification*. arXiv: 2503.04142v2 [eess.SP]. <https://arxiv.org/abs/2503.04142>
- [3] Swami, A., & Sadler, B. M. (2000). Hierarchical digital modulation classification using cumulants. *IEEE Transactions on Communications*, 48(3), 416-429. <https://doi.org/10.1109/26.825750>

- [4] O'Shea, T. J., Roy, T., & Clancy, T. C. (2018). Over-the-air deep learning based radio signal classification. *IEEE Journal of Selected Topics in Signal Processing*, 12(1), 168-179. <https://doi.org/10.1109/JSTSP.2017.2781381>
- [5] Zhang, Z. (2025). Deep learning-based Automatic Modulation Recognition: a comprehensive study. *Advances in Engineering Innovation*, 16(5), 70-77. <https://doi.org/10.54254/2977-3903/2025.23161>
- [6] Xing, J., Xu, C., Xu, S., & Zhou, X. (2025). A Lightweight Modulation Recognition Network with Depthwise Separable Convolution and SNR Enhancement. In *2025 7th International Conference on Electronics and Communication, Network and Computer Technology (ECNCT)* (pp. 530-534). IEEE. <https://doi.org/10.1109/ECNCT66493.2025.11172616>
- [7] Mujtaba, A., Lee, W.-K., & Hwang, S. O. (2022). Low Latency Implementations of CNN for Resource-Constrained IoT Devices. *IEEE Transactions on Circuits and Systems-II: Express Briefs*, 69(12), 5124-5128. <https://doi.org/10.1109/TCSII.2022.3205029>
- [8] Abd-Elaziz, O. F., Abdalla, M., & Elsayed, R. A. (2023). Deep Learning-Based Automatic Modulation Classification Using Robust CNN Architecture for Cognitive Radio Networks. *Sensors*, 23(23), 9467. <https://doi.org/10.3390/s23239467>
- [9] He, J., Yin, C., Xu, C., & Zhang, Y. (2024). High-Order Modulation Recognition based on Complex-valued Residual Network. In *2024 IEEE 4th International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)* (pp. 602-606). IEEE. <https://doi.org/10.1109/ICIBA62489.2024.10867993>
- [10] Zhang, S., Xu, H., Hao, J., Tian, Y., Liu, C., & Li, B. (2025). FADC: A High-Efficiency CNN-Based Deep Neural Network for Automatic Modulation Recognition. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5682697>