An Efficient Deep Learning Model for Automatic Modulation Classification

Xuemin LIU [1], Yaoliang SONG [1], Jiewei ZHU [2], Feng SHU [3], Yuwen QIAN [1]

[1] School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Xiaolingwei Street 200, 210094 Nanjing City, China
[2] China United Network Communication Group Co., Ltd. Suzhou Branch, Binhe Street 1300, 215100 Suzhou City, China
[3] School of Information and Communication Engineering, Hainan University, Renmin Avenue 58, 570228 Haikou City, China

liuxm@njust.edu.cn, ylsong@njust.edu.cn

Submitted August 29, 2024 / Accepted November 2, 2024 / Online first November 12, 2024

Abstract. *Automatic Modulation Classification (AMC) has emerged as a critical research domain with wide-ranging applications in both civilian and military contexts. With the advent of artificial intelligence, deep learning techniques have gained prominence in AMC due to their unparalleled ability to automatically extract relevant features. However, most contemporary AMC models rely heavily on downsampling strategies to increase the receptive field while reducing computational complexity. Empirical evidence indicates that progressive downsampling substantially reduces the spatial resolution of feature maps, leading to poor generalization, particularly for closely related modulation schemes. To address these challenges, this paper proposes a novel Multiscale Dilated Pyramid Module (MDPM). In contrast to traditional downsampling techniques, MDPM mitigates resolution loss and retains a broader range of features, facilitating more comprehensive recognition. Furthermore, the multiscale features captured by MDPM enhance the robustness of the model to noise, thereby improving classification performance in noisy environments. The model's efficiency is further optimized through the integration of group convolutions and channel shuffle techniques. Extensive experimental results and evaluations confirm that the MDPM-based approach surpasses state-of-the-art methods, underscoring its significant potential for practical deployment. The signal database and model can be freely accessed at https://pan.baidu.com/s/1g_HQXcRXshrT8nwKUNDYrQ?pwd=9ug6.*

Keywords

Automatic modulation classification, deep learning, spatial resolution, multiscale dilated pyramid module, group convolution

1. Introduction

The competition among various electromagnetic systems, including broadcasting, detection, and navigation equipment, has intensified due to the scarcity of spectrum resources. Recently, Automatic Modulation Classification (AMC), a pivotal component of blind signal processing [1], has garnered considerable attention for its capacity to automatically and rapidly identify emitter waveforms.

Much like in image classification, the identification of discriminative features is crucial for successful signal classification. Early research [2–4] on signal classification primarily focused on selecting handcrafted features from the time, frequency, time-frequency, and ambiguity domains. With the advent of artificial intelligence, deep convolutional neural networks (CNNs) [5] have gained prominence in signal classification. Studies [6–8] have leveraged CNNs to identify radio signals from raw in-phase and quadrature (I/Q) samples, demonstrating superior recognition performance over methods that rely on handcrafted features, owing to the enhanced representational learning capabilities of CNNs. However, the instability of I/Q samples in noisy environments results in suboptimal classification performance. To mitigate this, an attention-based residual shrinkage network [9] has been proposed, which emphasizes informative features from the more stable high-order spectra (HOS) in noisy environments. Despite its potential, the computationally intensive nature of HOS limits its practicality for real-world applications.

Other approaches have centered on time-frequency images (TFIs) [10], which offer richer signal characteristics. For example, LeNet5 [11] has been used to recognize radar signal types, while additional denoising preprocessing [12] has been applied to simple CNNs to improve recognition performance. These methods typically adopt classical forward CNN architectures and repeated down-

DOI: 10.13164/re.2024.0713

sampling strategies to learn increasingly abstract data representations. However, the reduction in feature resolution hampers performance in classifying similar modulation schemes.

To address the challenge of distinguishing similar modulation schemes, a cascaded CNN network [13] has been adopted to handle easy-to-recognize and difficult-to-recognize modulation schemes in two stages. Another approach [14] suggests using a pre-trained autoencoder to enhance system robustness, while a dual-branch CNN network [15] aggregates features from complementary feature domains to improve signal recognition. [21], [22] introduce deep attention mechanisms to concentrate computing resources on the most informative parts, further improving the accuracy of recognition systems. Although these approaches have achieved more generalized recognition performance, the increased complexity and computational demands of these models pose challenges for deployment on resource-constrained devices, such as automotive radar systems.

Considering the aforementioned issues, this paper develops a novel multiscale dilated pyramid module (MDPM) to establish an efficient and generalized AMC system. The main contributions of this study are as follows:

- The MDPM is developed as a powerful alternative to downsampling operations, avoiding resolution loss and maintaining more distinctive features among various signals, thereby facilitating a more generalized AMC.
- The multiscale feature aggregation mechanism of the MDPM enhances robustness to noise and further improves recognition performance in noisy environments.
- The MDPM employs computationally efficient group convolutions, reducing computational demands and complexity, making it suitable for deployment on resource-constrained embedded devices for real-time signal recognition.

The structure of this paper is as follows. Section 2 presents an illustrative description of the signal model and time-frequency analysis tools. Section 3 provides a detailed description of the proposed MDPM. Section 4 demonstrates the generalized and efficient recognition performance. Finally, Section 5 summarizes the entire work.

2. Problem Formulation

2.1 Signal Model

The modern electromagnetic environment is populated with a variety of radio signals originating from both radar equipment and communication devices. This paper comprehensively considers typical radar and communication signals.

Radar signals consist of no modulated signal (NS), phase modulation signals: P1-P4 codes, Binary Phase-Shift Keying (BPSK), Frank code, frequency modulated (FM) signals: Linear Frequency Modulation (LFM), Sinusoidal Frequency Modulation (SFM), Even Quadratic Frequency Modulation (EQFM), combined modulation signals: LFM-BPSK, SFM-BPSK.

Communication signals include discrete frequency code signals: Binary Frequency-Shift Keying (BFSK), Quadrature Frequency-Shift Keying (QFSK), discrete amplitude code signals: Binary Amplitude-Shift Keying (BASK), Quadrature Amplitude-Shift Keying (QASK), Quadrature Amplitude Modulation (QAM) signals: 16QAM, 64QAM. The mathematical expressions of the above electromagnetic signals are available in the literature [6–13].

After down-converting and sampling, the intercepted discrete-time signal $i[l]$ can be written as below:

$$i[l] = a[l]\mathrm{e}^{\mathrm{j}\theta[l]} + w[l], \quad l = 1,2,...,L \tag{1}$$

where $a[l]\mathrm{e}^{\mathrm{j}\theta[l]}$ is the pure radar or communication signal, and $w[l]$ denotes noise, which is generally assumed as the additive white Gaussian noise. $a[l]$ represents the amplitude. $\theta[l]$ denotes the phase. L represents the length of the received discrete-time signal.

Strong noise will severely pollute the extracted features, causing confusion among different signals [23]. To validate the effectiveness of the methods in intense noise environments, noisy signals with various noise levels are constructed. The signal-to-noise ratio (SNR) is adopted to represent the noise levels, which is defined as follows:

$$\mathrm{SNR}=10\times\log_{10}\frac{\sigma_{\mathrm{S}}}{\sigma_{\mathrm{N}}} \tag{2}$$

where σ_{S} and σ_{N} represent the power of the pure signal and the noise, respectively.

2.2 Time-Frequency Representation

To simplify the computational cost, the short-time Fourier transform (STFT) is adopted to extract the time-frequency representation, which can be written as below [11]:

$$\aleph[u,k] = \sum_{l=0}^{L-1} i[l]\mho[l-u]\mathrm{e}^{-\mathrm{j}2\pi nk/L} \tag{3}$$

where $\mho[l]$ denote the Hamming window function.

3. Methodology

Existing signal recognition networks overly depend on the downsampling strategy. While the repeated use of downsampling can effectively expand receptive fields and

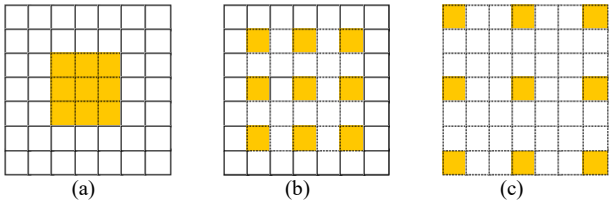

Fig. 1. Standard convolution and dilated convolution. (a) Standard convolution (dilated convolution has an expansion rate of 1). (b) Dilated convolution whose expansion rate is 2. (c) Dilated convolution whose expansion rate is 3.

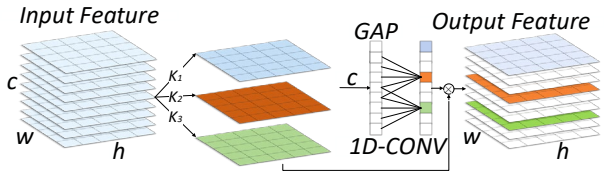

Fig. 2. The flowchart of MDPM.

decrease the size of feature maps, it drastically reduces their resolution. This reduction leads to confusion between closely modulated schemes. This research rethinks the downsampling strategy and proposes a novel MDPM to address these issues.

3.1 Dilated Convolution

For a signal with a length $L = 1024$, the size of its TFI is $L \times L$, while CNN can hardly take such a big size TFI as the input. Considering the informative shape and envelope features, the interpolation algorithm is adopted to decrease the size of TFIs, which can maintain the shape and envelope features in TFIs. Then, CNNs can extract high-level features from the resized TFIs. This pre-interpolation step reduces the number of network parameters and increases inference speed, offering significant benefits over downsampling within the network. Almost all studies have adopted this strategy [9–15].

In AMC tasks, the receptive field is crucial for capturing global feature representation. Generally, increasing the depth of the network and applying progressive downsampling can expand the filters view field. Although this approach has led to outstanding performance in TFI classification, the further loss of spatial information hampers recognizing TFIs of signals. Consequently, methods that adopt progressive downsampling uncritically have poor generalization, especially in closely modulated schemes.

Dilated convolution [16], [17] is a generalized form of traditional convolution operation that exponentially expands the receptive fields without extra computational burden and avoids resolution loss. The mathematical expression of dilated convolution is [16], [17]:

$$A(i,j) = \sum_{m=1}^{S}\sum_{n=1}^{S} F_{\text{in}}\left(i + r\cdot m, j + r\cdot n\right) kernel[m,n]\,. \quad (4)$$

Here, $A(i,j)$ is the operation results of the input $F_{\text{in}}(i,j)$ by dilated convolution with the kernel size $S \times S$. r denotes expansion rate. If $r = 1$, equation (4) denotes a traditional convolution. Figure 1 illustrates dilated convolution with different rates. Figure 1(a) shows the standard convolution, which is equivalent to dilated convolution with an expansion rate of 1. Figure 1(b) shows the atrous convolution whose expansion rate is 2. Figure 1(c) shows the atrous convolution whose expansion rate is 3. Dilated convolution with an expansion rate greater than 1 has larger receptive fields while saving computation costs.

Inspired by dilated convolutions, this work adopts dilated convolutions to preserve spatial resolution effectively and implement the exponential expansion of receptive fields.

3.2 Multiscale Dilated Pyramid Block

As illustrated in Fig. 2, adaptive multiscale dilated pyramid (AMDP) extracts multiscale features simultaneously.

Each dilated convolution branch independently captures features at different scales. These features are then concatenated to fuse multiscale spatial features, that is

$$F = Cat([DilatedConv(K_i, K_i)(F_{\text{in}})])\,. \quad (5)$$

Here, $i = 1,2,\ldots,N$ is the branch index. F represents the multiscale feature map and F_{in} represents the input feature map. *Cat* denotes matrix concatenate operation. $DilatedConv(K_i, K_i)$ represents the dilated convolution with the kernel size (K_i, K_i).

Furthermore, this work employs the channel attention mechanism to adaptively recalibrate the channel-wise relationship [18], enhancing the allocation of the most informative feature maps. First, a global average pooling layer integrates spatial information to form the channel descriptor $A_{\text{channel}}^{1,1,c}$:

$$A_{\text{channel}}^{1,1,c} = \frac{1}{h \times w}\sum_{i=1,j=1}^{h,w} F \quad (6)$$

where h, w, and c denote the length, width, and the channel numbers of the features, respectively.

Then, one-dimensional (1D) convolution $conv_{1\text{d}}$ is used to capture local attention E_c from $A_{\text{channel}}^{1,1,c}$:

$$E_c = conv_{1\text{d}}\left(A_{\text{channel}}^{1,1,c}\right). \quad (7)$$

Using 1D convolution instead of a fully connected layer [20] greatly reduces parameters and calculations. Finally, an adaptive channel-wise mapping U is realized by weighting the input feature with channel-wise attention:

$$U = E_c \times F\,. \quad (8)$$

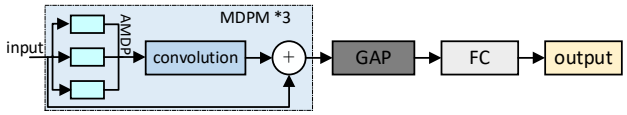

Fig. 3. Architecture of the proposed MDPNet.

The channel attention mechanism enables the model to focus on the important features, which enhances the representational capability of the model efficiently.

As shown in Fig. 3, AMDP and a 3 × 3 convolution further constitute MDPM as

$$F_{\text{out}} = Conv(AMDP(F_{\text{in}})) + F_{\text{in}}. \quad (9)$$

The designed MDPM provides a richer receptive field and better multiscale learning ability. Finally, three MDPMs, one global average pooling layer (GAP), and one fully connected layer (FC) build up the proposed multiscale dilated pyramid network (MDPNet).

3.3 Group Convolution

Since the size of feature maps remains constant during capturing features, the computation cost inevitably increases. The group convolution (GC) technique is adopted in MDPM to reduce computing costs. The group convolution *Convg* separates its input into several groups and then computes feature maps by utilizing only input channels within the same group as

$$Convg = Cat(conv(F_{\text{in}}^{g})) \quad (10)$$

where F_{in}^{g} denotes the gth group feature map of input, $g = 1,2,\ldots,G$ represents the group index.

Assuming that the size of the input feature map is $C_{\text{in}}{\cdot}W_{\text{in}}{\cdot}H_{\text{in}}$ and that of the output feature map is $C_{\text{out}}{\cdot}W_{\text{out}}{\cdot}H_{\text{out}}$, the computational cost $MACC_{\text{s}}$ and the number of parameters $Parm_{\text{s}}$ of a standard convolution is calculated as [20]

$$MACC_{\text{s}} = K \times K \times C_{\text{in}} \times C_{\text{out}} \times W_{\text{out}} \times H_{\text{out}}, \quad (11)$$

$$Parm_{\text{s}} = K \times K \times C_{\text{in}} \times C_{\text{out}} \quad (12)$$

where K symbolizes the spatial dimension of the kernel.

Under the same conditions, group convolution, whose group is g, has a computational complexity $MACC_{\text{g}}$ and storage size $Parm_{\text{g}}$ as below

$$\begin{aligned} MACC_{\text{g}} &= K \times K \times (C_{\text{in}} / g) \times \left(C_{\text{out}} / g\right) \\ &\quad \times W_{\text{out}} \times H_{\text{out}} \times g \\ &= K \times K \times C_{\text{in}} \times C_{\text{out}} \times W_{\text{out}} \times H_{\text{out}} / g, \end{aligned} \quad (13)$$

$$Parm_{\text{g}} = K \times K \times C_{\text{in}} \times C_{\text{out}} / g. \quad (14)$$

Compared with standard convolution, group convolution gets a reduction of $1/g$ in the computational cost and memory space, respectively. Group convolution guarantees the real-time performance of MDPM and eases the storage burden. Inspired by ShuffleNet [20], channel shuffle is deployed to strengthen the connection between groups.

4. Experiments

This section begins by detailing the dataset, which contains 18 types of signals. This section then presents the classification results of MDPNet and compares them with other state-of-the-art algorithms. Finally, this section includes ablation studies to verify the effectiveness of the designed MDPM.

4.1 Dataset and Experiment Parameters

Experiments were conducted on 18 types of radar and communication signals to validate the effectiveness of MDPNet, whose parameters are randomly generated in the range of Tab. 1. In the training set, 200 samples of each signal per 2 dB were simulated from –4 dB to 10 dB, with parameters randomly disturbed within the above-mentioned ranges. The training dataset contains 28,800 samples. The testing set (TestingData1) includes 100 samples of each signal from –8 dB to 10 dB, with parameters randomly disturbed within the same ranges. To assess robustness, another testing set (TestingData2) considers carrier frequency offsets. In TestingData2, the carrier frequency of FM signals is disturbed within $(0.075\sim0.1 \cup 0.4\sim0.425){\cdot}f_{\text{s}}$, with bandwidth ranges within $(0.1\sim0.15){\cdot}f_{\text{s}}$. Other signals have a carrier frequency range of $(0.05\sim0.1 \cup 0.4\sim0.45){\cdot}f_{\text{s}}$. TestingData2 also contains 100 samples of each signal from –8 dB to 10 dB.

The optimizer used in the experiments is Adam, with a batch size of 128. The initial learning rate is 0.01, dropping by 90% after 20 epochs. A total of 60 epochs were run

Signal	Parameter	Ranges
NS	Carry frequency (f_{c})	$(0.1–0.4)\cdot f_{\text{s}}$
LFM, SFM, EQFM	f_{c} Bandwidth Δf	$(0.1–0.4)\cdot f_{\text{s}}$ $(0.1–0.2)\cdot f_{\text{s}}$
BFSK	$f_{\text{c1}}, f_{\text{c2}}$ T_{s}	$(0.1–0.4)\cdot f_{\text{s}}$ $(1/32–1/8)\cdot N$
QFSK	$f_{\text{c1}}, f_{\text{c2}}, f_{\text{c3}}, f_{\text{c4}}$, T_{s}	$(0.1–0.4)\cdot f_{\text{s}}$ $(1/32–1/8)\cdot N$
BASK, QASK	f_{c} T_{s}	$(0.1–0.4)\cdot f_{\text{s}}$ $(1/32–1/8)\cdot N$
BPSK	f_{c} Barker codes T_{s}	$(0.1–0.4)\cdot f_{\text{s}}$ [5,7,11,13] $(1/64–32)\cdot N$
P1-4, Frank code	f_{c} T_{s} Phase number	$(0.1–0.4)\cdot f_{\text{s}}$ $(1/100–1/64)\cdot N$ [4,5,6,7]
32QAM, 64QAM	f_{c} T_{s}	$(0.1–0.4)\cdot f_{\text{s}}$ $(1/32–1/8)\cdot N$
LFM-BPSK, SFM-BPSK	f_{c} Δf Barker codes	$(0.1–0.4)\cdot f_{\text{s}}$ $(0.1–0.2)\cdot f_{\text{s}}$ [5,7,11,13]

Here, f_{s} denotes the sampling frequency, N is the length of the discrete signal, and T_{s} represents the modulation period.

Tab. 1. Simulate signal parameters.

(all 60 epochs were executed). Network training is conducted in PyTorch 1.10 supported by RTX 3070 GPU. Evaluation indicators include recognition accuracy, run time, and network parameters.

4.2 Recognition Performance

As demonstrated in Fig. 4, the recognition performance on TrainingData is slightly better than that on TestData1, indicating the excellent fitting of the model. The system also identifies TestData2 with nearly the same accuracy as TestData1, demonstrating the robustness of the proposed architecture to carrier frequency mismatches.

Figure 5 illustrates the identification performance of the proposed MDPNet for 18 different radio signals. The recognition results of all signals increases with SNR. All signals maintain nearly 100% recognition precision when the SNR is over 0 dB. This high accuracy results primarily from the intuitive signal characteristics captured in time-frequency images and the powerful learning capabilities of deep neural networks, which significantly enhance recognition accuracy compared to traditional methods. However, the accuracy for SFM, P1, QASK, and 32QAM declines more rapidly than other signals when the SNR is below 0 dB, as intense noise obscures some critical phase information, leading to signal confusion.

The confusion matrix of the proposed method at the SNR of −2 dB is exhibited in Fig. 6. It can be manifested that the recognition system can accurately identify various signals. P3 code signals have a precision of 92%, which is the worst performance of all signals. The recognition mistakes mainly occur between the signal pairs with similar TFIs, such as SFM and SFM-BPSK signals, QASK and BASK signals, FRANK and P3 code signals, due to lost fine frequency details in high noise levels. The process of TFI resizing also makes small frequency jump blurred. The losses and blurs of this small frequency information finally lead to confusion between signals.

To further highlight the performance of the proposed methods, comparative experiments with other state-of-the-art algorithms were conducted. In [4], HOG-based descriptors were extracted from STFT spectrums to classify radar signals. [6] utilized 1D-ResNet to recognize signals from I/Q samples. In [10], [11], CNNs were employed to distinguish TFIs of signals. As shown in Fig. 7, the recognition precision of the traditional machine learning method [4] is low due to the limitations of hand-crafted features. The accuracy of 1D-ResNet [6] is significantly higher than that of the machine learning-based method. However, the classification effect is considerably reduced when the SNR is lower than 0 dB due to limited robustness with I/Q samples. While deep learning methods [10], [11] based on 2D CNNs can recognize various radar signals with relatively high accuracy, the progressive downsampling strategies used in [10], [11] dramatically decrease the spatial resolution of feature maps, resulting in poor overall recognition precision. In comparison, the proposed MDPNet can accurately recognize all types of signals, outperforming all existing methods. This performance boost stems from the

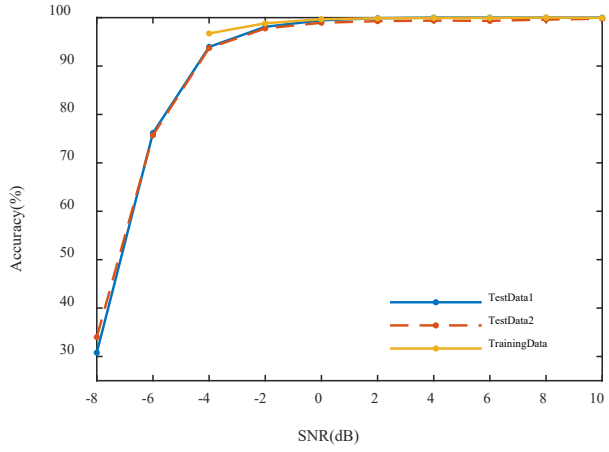

Fig. 4. Overall recognition results of different datasets.

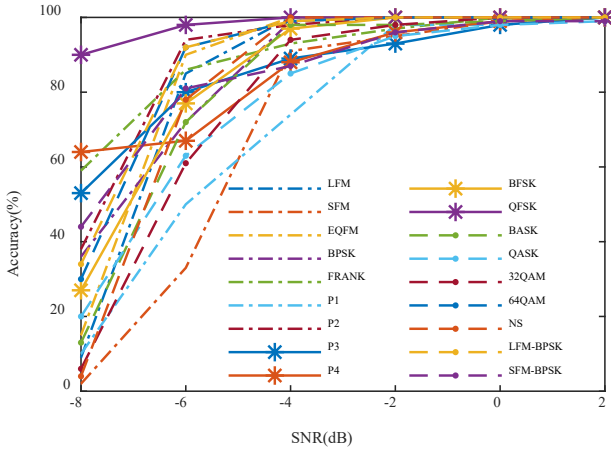

Fig. 5. Recognition accuracy of each signal.

	A	B	C	D	E	F	G	H	I	J	K	L	F	N	O	P	Q	R
A	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
B	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
C	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	99	0	1	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	1	99	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	1	2	92	2	1	2	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	98	2	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	98	2	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	1	99	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	98	2	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	5	95	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
Q	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0
R	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95

Fig. 6. Confusion matrix for the proposed method at the SNR of –2 dB. A, B, C, D, E, F, G, H, I, J, K, L, F, N, O, P, Q, and R represent LFM, SFM, EQFM, FRANK, P1, P2, P3, P4, BFSK, QFSK, BASK, QASK, 32QAM, 64QAM, NS, LFM-BPSK, and SFM-BPSK, respectively.

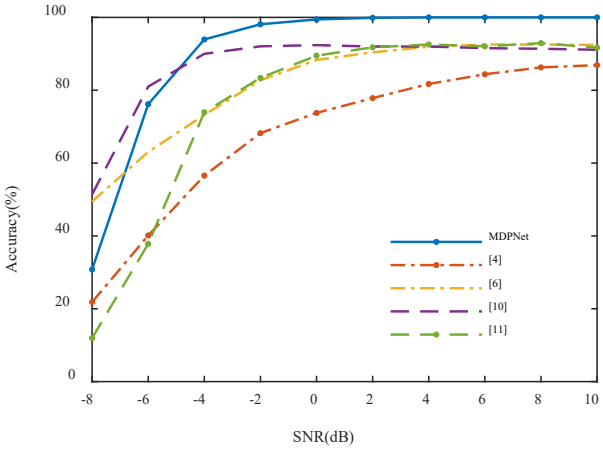

Fig. 7. Comparison with other methods.

Model	Time [ms]	Parameters
HOG [4]	94.8	/
1D-ResNet [6]	29	212,490
LeNet-5 [11]	659.9	117,986
LPINet [10]	251.9	230,866
MDPNet	61.9	48,952

(the '/' in the parameter column for HOG indicates that this parameter is not applicable in our implementation for the respective model.)

Tab. 2. Comparison of different methods.

proposed MDPM module, which preserves feature map resolution and prevents loss of critical features during downsampling, while also broadening the receptive fields to capture global context.

Table 2 reveals the run time and model size of each method. The run time of the proposed system is 61.9 ms, which is reduced by 10.66× and 4.06× compared to [10], [11], respectively. This indicates that the proposed system can efficiently complete recognition tasks. Moreover, MDPNet has 48,952 parameters, 4.71 times fewer parameters than [10]. The compact size of MDPNet ensures it can be deployed on most devices. Comparative experiments demonstrate that MDPNet achieves superior accuracy and efficiency, significantly outperforming previous methods.

4.3 Ablation Experiments

This section conducts ablation studies to validate the effectiveness of each component in MDPNet. To examine the impact of dilated convolution on recognition results, we designed a baseline network (baselineNet) based on the ResNet-18 block with a progressive downsampling strategy. As shown in Fig. 8, baselineNet significantly degrades classification performance due to feature resolution loss. In contrast, MDPNet avoids this issue, enabling accurate signal recognition. The ablation studies confirm the effectiveness of each module in MDPNet, demonstrating the contributions of individual components to the overall performance and efficiency of the network.

To investigate the effect of multiscale convolution within the MDPM, we conducted experiments with varying numbers of branches in MDPNet. Results indicate that as

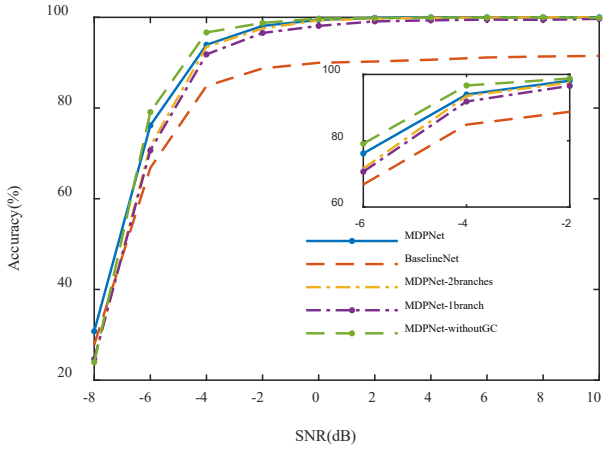

Fig. 8. Effect of dilated convolution and multiscale convolution.

Model	MACC	Parameters
MDPNet	190,743,680	48,952
MDPNet without GC	450,266,240	112,304

Tab. 3. Complexity and size of different models.

the number of branches increases, MDPNet's recognition accuracy improves accordingly. These additional scale features help compensate for information loss due to noise, contributing to robust recognition performance under low SNRs.

This section also designed a version of MDPNet without GC and channel shuffle to further examine their impact. As shown in Fig. 8, the recognition performance of MDPNet without GC is slightly higher than that with GC and channel shuffle, likely because GC restricts feature associativity to some extent. However, as shown in Tab. 3, incorporating GC reduces MDPNet's computational requirements and parameters by more than half. Therefore, the MDPNet configuration with GC strikes an optimal balance between recognition performance and computational efficiency.

5. Conclusion

Signal classification in modern, complex electromagnetic environments remains a critical challenge. In this work, we present an efficient and generalized automatic modulation classification (AMC) method based on the combination of short-time Fourier transform and deep learning techniques. A novel Multi-scale Dilated Pyramid Module (MDPM) was introduced to replace the traditional progressive downsampling strategy, preserving essential features for classification. This approach allows MDPM to retain feature resolution while capturing multi-scale information, significantly improving the model's robustness, particularly in low signal-to-noise ratio (SNR) conditions. Moreover, the compact structural design of MDPM ensures low computational costs, making it ideal for real-time applications with constrained resources.

While the proposed MDPNet method has demonstrated superior performance at reduced computational costs, several potential avenues for future research remain:

(1) Further investigation into the generalizability of this model across a broader range of environments is necessary. Future work could explore how the model performs in more dynamic, rapidly changing electromagnetic environments or under various channel conditions such as fading and interference.

(2) Integrating this method with other deep learning architectures, such as transformer models, could enhance its feature extraction capabilities and allow for improved classification accuracy, especially for complex modulation schemes.

(3) The development of more energy-efficient models for deployment on edge devices and low-power platforms would expand the real-world applicability of the proposed method.

References

[1] WANG, Q., DU, P., YANG, J., et al. Transferred deep learning based waveform recognition for cognitive passive radar. *Signal Processing*, 2019, vol. 155, p. 259–267. DOI: 10.1016/j.sigpro.2018.09.038

[2] LUNDEN, J., KOIVUNEN, V. Automatic radar waveform recognition. *IEEE Journal of Selected Topics in Signal Processing*, 2007, vol. 1, no. 1, p. 124–136. DOI: 10.1109/JSTSP.2007.897055

[3] XU, C., ZHANG, J., ZHOU, Q., et al. Recognition of radar signals based on AF grids and geometric shape constraint. *Signal Processing*, 2019, vol. 157, p. 30–44. DOI: 10.1016/j.sigpro.2018.11.004

[4] CHEN, K., CHEN, S., ZHANG, S., et al. Automatic modulation recognition of radar signals based on histogram of oriented gradient via improved principal component analysis. *Signal, Image and Video Processing*, 2023, vol. 17, no. 6, p. 3053–3061. DOI: 10.1007/s11760-023-02526-x

[5] LEE, Y. S., BANG, C. C. Framework for the classification of imbalanced structured data using under-sampling and convolutional neural network. *Information Systems Frontiers*, 2022, vol. 24, p. 1795–1809. DOI: 10.1007/s10796-021-10195-9

[6] O'SHEA, T. J., ROY, T., CLANCY, T. C. Over-the-air deep learning based radio signal classification. *IEEE Journal of Selected Topics in Signal Processing*, 2018, vol. 12, no. 1, p. 168–179. DOI: 10.1109/JSTSP.2018.2797022

[7] WANG, Y., LIU, M., YANG, J., et al. Data-driven deep learning for automatic modulation recognition in cognitive radios. *IEEE Transactions on Vehicular Technology*, 2019, vol. 68, no. 4, p. 4074–4077. DOI: 10.1109/TVT.2019.2900460

[8] ZHANG, F., LUO, C., XU, J., et al. An efficient deep learning model for automatic modulation recognition based on parameter estimation and transformation. *IEEE Communications Letters*, 2021, vol. 25, no. 10, p. 3287–3290. DOI: 10.1109/LCOMM.2021.3102656

[9] CHEN, K., ZHU, L., CHEN, S., et al. Deep residual learning in modulation recognition of radar signals using higher-order spectral distribution. *Measurement*, 2021, vol. 185, p. 1–8. DOI: 10.1016/j.measurement.2021.109945

[10] HUYNH-THE, T., DOAN, V. S., HUA, C. H., et al. Accurate LPI radar waveform recognition with CWD-TFA for deep convolutional network. *IEEE Wireless Communications Letters*, 2021, vol. 10, no. 8, p. 1638–1642. DOI: 10.1109/LWC.2021.3075880

[11] ZHANG, M., DIAO, M., GUO, L. Convolutional neural networks for automatic cognitive radio waveform recognition. *IEEE Access*, 2017, vol. 5, p. 11074–11082. DOI: 10.1109/ACCESS.2017.2716191

[12] CHEN, K., ZHANG, J., CHEN, S., et al. Automatic modulation classification of radar signals utilizing X-net. *Digital Signal Processing*, 2022, vol. 123, p. 1–10. DOI: 10.1016/j.dsp.2022.103396

[13] HOANG, L. M., KIM, M., KONG, S. H. Automatic recognition of general LPI radar waveform using SSD and supplementary classifier. *IEEE Transactions on Signal Processing*, 2019, vol. 67, no. 13, p. 3516–3530. DOI: 10.1109/TSP.2019.2918983

[14] CHEN, K., WANG, L., ZHANG, J., et al. Semantic learning for analysis of overlapping LPI radar signals. *IEEE Transactions on Instrumentation and Measurement*, 2023, vol. 72, p. 1–15. DOI: 10.1109/TIM.2023.3242013

[15] ZHANG, Z., WANG, C., GAN, C., et al. Automatic modulation classification using convolutional neural network with features fusion of SPWVD and BJD. *IEEE Transactions on Signal and Information Processing over Networks*, 2019, vol. 5, no. 3, p. 469 to 478. DOI: 10.1109/TSIPN.2019.2900201

[16] YU, F., KOLTUN, V., FUNKHOUSER, T. Dilated residual networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu (HI, USA), 2017, p. 636–644. DOI: 10.1109/CVPR.2017.75

[17] CHEN, L., PAPANDREOU, G., SCHROFF, F., et al. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint,* 2017, p. 1–14. DOI: 10.48550/arXiv.1706.05587

[18] WANG, Q., WU, B., ZHU, P., et al. ECA-Net: efficient channel attention for deep convolutional neural networks. In *2020 IEEE Conference on Computer Vision and Pattern Recognition*. Seattle (WA, USA), 2020, p. 11531–11539. DOI: 10.1109/CVPR42600.2020.01155

[19] HU, J., SHEN. L., ALBANIE, S., et al. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, vol. 42, no. 8, p. 2011–2023. DOI: 10.1109/TPAMI.2019.2913372

[20] ZHANG, X., ZHOU, X., LIN, M., et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In *2018 IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City (UT, USA), 2018, p. 6848–6856. DOI: 10.1109/CVPR.2018.00716

[21] CHEN, K., ZHANG, J., CHEN, S., et al. Deep metric learning for robust radar signal recognition. *Digital Signal Processing*, 2023, vol. 137, p. 1–8. DOI: 10.1016/j.dsp.2023.104017

[22] WEI, S., QU, Q., ZENG, X., et al. Self-attention Bi-LSTM networks for radar signal modulation recognition. *IEEE Transactions on Microwave Theory and Techniques*, 2021, vol. 69, no. 11, p. 5160–5172. DOI: 10.1109/TMTT.2021.3112199

[23] KUBANKOVA, A., KUBANEK, D. Extended method of digital modulation recognition and its testing. *Radioengineering*, 2011, vol. 20, no. 1, p. 25–30. ISSN: 1210-2512

About the Authors …

Xuemin LIU is a lecturer at Nanjing University of Science and Technology, focusing on communication signal processing, circuits and systems.

Yaoliang SONG is a Professor at Nanjing University of Science and Technology, focusing on modern signal processing technology, radar signal design and detection technology, etc.

Jiewei ZHU graduated from Changshu Institute of Technology with a bachelor's degree in 2017 and from Nanjing Univ. of Science and Technology with a master's degree in 2019. He is currently a researcher at China Unicom Suzhou Branch, focusing on electronic countermeasures.

Feng SHU is a Professor at Hainan University, focusing on intelligent wireless communication, intelligent wireless information security transmission, etc.

Yuwen QIAN is an Associate Professor at Nanjing University of Science and Technology, focusing on wireless communication, quantum covert communication, intelligent communication, etc.