IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 October 2023, accepted 4 November 2023, date of publication 14 November 2023,
date of current version 22 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3332643

RESEARCH ARTICLE

Dilated CNN Design Approach for Extracting Multi-Scale Features in Radar Emitter Classification

ENZE GUO[1,3], HAO WU[2], MING GUO[1], YINAN WU[2], AND JIAN DONG[3]
[1]The 63893 Forces of PLA, Luoyang 471000, China
[2]The 63896 Forces of PLA, Luoyang 471000, China
[3]School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing 100081, China

Corresponding author: Jian Dong (radarvincent@bit.edu.cn)

This work was supported in part by 111 Project of China under Grant B14010, in part by the Key Laboratory of Satellite Information Intelligent Processing and Application Technology under Grant 2022-ZZKY-JJ-11-01.

ABSTRACT Radar emitter classification plays an increasingly significant role in the electronic reconnaissance system. Due to many convolutional neural network (CNN)-based approaches suffer from insufficient spatial receptive fields and inadequate feature representation, the classification accuracy is poor in low signal-to-noise ratio (SNR) conditions. Therefore, in this paper, we stress the importance of multi-scale dilated convolutions for target feature extraction, and propose two novel CNN architecture design approaches called multi-scale dilated residual network (MDRN). By combining multi-scale dilated convolutions with residual architecture, MDRN not only has a larger receptive field, but also can learn more diverse features, thereby improving the ability to process time-frequency images (TFI) under high-noise energy conditions. Moreover, compared with the original residual model, MDRN does not increase any parameter complexity or floating-point operations per second (FLOPS). Experiments on the TFI classification task show that the proposed MDRN has superior performance over state-of-the-art CNN models.

INDEX TERMS Intra-pulse modulation classification, dilated convolution, feature maps, multi-scale feature fusion, time-frequency image.

I. INTRODUCTION

The intra-pulse modulation of radar transmitter signals is an essential component of electronic support measurement systems, electronic intelligence systems, and radar warning receivers. By precisely identifying the intra-pulse modulation of these signals, electronic support measure (ESM) systems can assess the level of threat posed by incoming radar emitter and develop appropriate countermeasures to protect against potential threats [1]. Therefore, radar intra-pulse signal modulation classification is critical for ensuring the security and safety of military. However, as the contemporary electromagnetic environment becomes increasingly complex, the intra-pulse modulation of radar transmitters has correspondingly become more intricate, posing significant challenges for classification tasks. In this article, we are committed to proposing an effective approach to accomplish pulse modulation of radar signals in complex electromagnetic environments [2].

The Pulse Description Word [3] is a commonly used, early method that extracted pulse interval characteristics, which has found extensive application in low electromagnetic density environments and in traditional radar systems. However, in modern radar systems, with their complex and varied signal waveforms, effective classification of signals solely based on the interval features is not feasible. To identify radar radiation emitter signals in these environments, it is necessary to explore the detailed structural feature information inside each pulse [4], as modulation signals within each pulse possess their unique electrical signal structure, including intentional and unintentional modulation. Time-frequency analysis can better handle non-stationary signals

The associate editor coordinating the review of this manuscript and approving it for publication was Chengpeng Hao.

by simultaneously providing time and frequency information and facilitating more comprehensive extraction of signal feature information. Therefore, time-frequency analysis is one of the most effective ways to extract intra-pulse features of radar signals. Naturally, a range of time-frequency analysis techniques, including short-time Fourier transform (STFT) [4], wavelet transform [5], Wigner-Ville distribution, and Choi-Williams distribution [6], have been proposed.

Traditional machine learning methods combined with time-frequency analysis have been an effective approach for intra-pulse modulation classification, such as Support Vector Machine (SVM) [7], Decision Tree [8], and Random Forest [9]. However, these approaches generally lack the capacity of automation, robust and generalization [10], [11], [12]. This is because these methods require manual design of feature extractors, and it is difficult to extract the high-level features of the target, therefore, the performance drops rapidly in low signal-to-noise ratio (SNR) environment [13].

In recent years, significant breakthroughs have been achieved in various fields such as image classification and recognition [14], natural language processing [15], object detection [16], using deep learning (DL) algorithms. The typical approach involves transforming raw one-dimensional data into two-dimensional time-frequency images (TFI) through time-frequency analysis, which is then used to drive a convolutional neural network (CNN) for automatic training.

Compared with the hand-crafted methods, the DL-based methods improved the information completeness, efficiency, and noise-robust [17], [18], therefore, it has state-of-the-art performance, intelligence and generalization. This is of great significance for improving the intelligence and stability of ESM. Top researchers worldwide increasingly employ DL-based approaches to tackle bottleneck challenges in radar signal processing [19]. Correspondingly, These methods have been shown to achieve high accuracy (up to 90%) in classifying radar emitter signals even in environments with high electromagnetic density and noise. For examples, the classic open-source CNN models including GoogleNet [20], ResNet [21] and MobileNetV2 [22] have achieved remarkable performance in the classification of time-frequency images due to its strong generalization abilities.

Despite the significant progress achieved by the CNN-based methods, there are still some deficiencies that need to be addressed. One of the main concerns is the performance limitation of these CNN models, which is determined by their architecture and size [23]. To achieve state-of-the-art classification accuracy, some complex and large-scale CNN models have been utilized. The large-scale CNN models require significant storage and computational resources (floating-point operations per second (FLOPS)), which makes it difficult to apply CNN models on portable devices and reduces their real-time performance. This is undoubtedly unfavorable for the application of deep learning algorithms in the rapidly changing modern electromagnetic field environment. Therefore, how to improve the model performance while maintaining the model complexity has become an important research direction in the current field.

The analysis of the differences between time-frequency images and conventional natural images (such as CIFAR-100, ImageNet, and FLOWERS102 dataset) should be the starting point of our work, as the effective extraction and transmission of data features determine the accuracy of CNN models [24]. Therefore, in practical applications, it is necessary to propose the adjustments and optimization methods of model based on the characteristics of specific dataset, building upon classic CNN architectures, to achieve state-of-the-art performance.

Compared with natural images, the scale of features in time-frequency images is relatively larger, and the subtle features such as shapes and textures are less. Additionally, noise energy can disrupt the structural features of time-frequency images to some extent, which is also a key factor affecting the performance of time-frequency image classification. To address this issue, we aim to enhance and optimize the CNN model structure from two directions: one is to enable the model to extract more comprehensive large-scale feature information; the other is to enable the model to mitigate the adverse impact of noise.

Aiming at the aforementioned objectives, we propose a novel design idea of the Dilated Convolution Neural Networks (DCNN) inspired by the idea of dilated residual network [25], multi-scale fusion [26], and denoise [27], [28]. On the one hand, in this research, we employed dilated convolution to increase the receptive field and capture more comprehensive structural features. On the other hand, we adopted dilated convolution with different dilation rates in multiple convolutional layers to combine features of different abstraction levels and effectively suppress the adverse impact of noise.

To be detailed, the main contributions of this paper are summarized as follows:

(1) we have conducted a comprehensive analysis of the distinguishing features between time-frequency images and conventional natural images. Based on this analysis, we have explored targeted improvement directions from a fundamental mechanism perspective, paving the way for proposing CNN improvement methods suitable for time-frequency image recognition.
(2) A design approach of dilated DCNN suitable for time-frequency image recognition has been proposed. Compared to standard CNN models, the DCNN model possesses a larger receptive field, which enables it to capture structural features of time-frequency images more effectively. Moreover, the model suppresses the adverse impact of noise by combining abstract features extracted using different dilation scales of convolution. The DCNN model achieves superior recognition performance without adding any additional model complexity and computational resources.
(3) The transferability of our proposed design method is effectively validated in this study. We evaluate the method on four different types of networks,

including residual networks constructed with basic block, residual networks constructed with linear bottleneck structures, GoogleNet constructed with the Inception structure, and MobileNetV2 constructed with Inverted linear bottleneck structures. Our experimental results demonstrate that our proposed design method exhibits favorable transferability, as the improved dilated models outperform their counterpart models with simple modifications.

(4) The proposed method has been evaluated on a dataset of 14 categories of intra-pulse modulation of radar signals commonly used in modern radar systems. Additionally, various levels of environmental noise were incorporated, posing further challenges to classification accuracy. The experimental results demonstrate that the improved models using our proposed method achieve state-of-the-art accuracy and exhibit excellent transferability.

The reminder of this paper is organized as follows. The design method of DCNN is proposed in Section II, including the design ideas, specific optimization operation and the analysis of the method's transferability. Furthermore, Section III conducts a series of comparative experiments and profoundly analyze the experiment results. Finally, the conclusion of this article is drawn in Section IV.

II. METHODOLOGY

To begin with, we compare and analyze the differences between time-frequency images and natural images, which provides a theoretical foundation for model selection and design. Subsequently, we review the principles of dilated convolution and explore its suitability for recognition of time-frequency images. Accordingly, an improved dilated residual network that incorporates multi-scale fusion naturally follows. Finally, we will substantiate the transferability of our approach by validating it using several open-source models.

A. THE COMPARISON AND ANALYSIS OF TFI AND NATURAL IMAGES

Since 2012, a number of CNN models have emerged, such as GoogleNet, ResNet, and MobileNet, which were originally designed for computer vision tasks and focused on natural images. Although these open-source models exhibit strong generalization and have achieved considerable success in TFI recognition, they were not specifically designed for this task. Therefore, it is essential to analyze and compare the similarities and differences between TFI and natural images in order to develop an improved CNN model that is tailored for TFI classification.

Natural images refer to various images that we encounter in our daily lives, including landscapes, people, animals, buildings, and so on. Color, texture, and shape are considered the most fundamental and effective features of natural images [25], as they represent a significant portion of the information contained in the image. In contrast, TFI are constructed using time-frequency analysis methods applied to signals. In such images, time is represented along the horizontal axis while the vertical axis represents frequency. The intensity of the signal is represented using different colors or brightness levels based on its frequency. Thus, TFI reflect the frequency characteristics of the signal over time. The features of time-frequency images are usually composed of two dimensions, time and frequency, and signal components at different frequencies are often considered as features.

Therefore, there are significant differences in the expression of features between the two types of images, mainly in the following aspects:

(1) Spatial correlation of features. Time-frequency images have clear boundaries, with distinct signal and non-signal regions, strong continuity between adjacent pixel values, and representation information that is reflected in a strong spatial structure connection method between pixel points. In contrast, natural images exhibit spatial correlation between adjacent pixels, where this correlation is localized to only surrounding pixels and not the entire image [26].

(2) Characteristics of features. Natural images have more spatial features such as edges, textures, shapes, and complex structures and diversities. They also contain various detailed information, such as edges and corners, which are critical for distinguishing different objects or scenes [29]. In contrast, TFI have fewer and simpler detail and texture features than natural images.

(3) Natural images generally have higher clarity, better image quality, and a higher fidelity of feature information [25]. However, actual radar signals often contain large amounts of noise, resulting in fuzzy TFI where the feature information of the images is destroyed to varying degrees.

In light of the above analysis, we should focus on enhancing the ability of the original model in two areas: firstly, reinforcing the model's capacity to extract comprehensive structural feature information; secondly, decreasing the model's susceptibility to noise.

B. MULTI-SCALE DILATED RESIDUAL NETWORK

Convolutional neural networks (CNNs) used for natural image classification gradually reduce the resolution of feature maps through multiple pooling or adjusting convolution stride, abstracting and compressing the target's feature information until a tiny feature map containing representative information is output (7×7 is typical) [25]. Given natural images contain a significant amount of subtle information crucial for classification, smaller convolution kernels can locally perceive images and extract their detailed features. Using smaller convolution kernels also results in lower computational complexity, enabling faster operation [26], [29], therefore, they are better suited for natural image recognition. However, it is evident that smaller receptive fields of convolution kernels are inadequate to extract comprehensive spatial structural features [30], which is necessary for TFI. Conversely, using larger convolution kernels would

exponentially increase model and computational complexity, adversely affecting CNN's operating efficiency.

The use of dilated convolution has emerged as a key approach to address this challenge. The reason why dilated convolution can lead to improved model performance is that it allows for the expansion of the receptive field [31]. Formally, we denote I as input feature maps and w as a convolution kernel, respectively. Standard convolution $*$ is defined as follows:

$$(I^{*}w)_{(t)} = \sum_{p+q=t} I_{(p)}w_{(q)} \tag{1}$$

where subscript (.) is the position of the input feature maps. Dilated convolution is a generalization of standard convolution which is formulated as follow:

$$(I \quad *lw)_{(t)} = \sum_{p+lq=t} I_{(p)}w_{(q)} \tag{2}$$

where l is the dilation factor. We denote $*l$ as dilated convolution with a factor l. This means one pixel in w corresponds to l pixels (skip of l-1 pixels) in I, as in [32]. The skipped positions are filled with holes with a value of 0. The receptive field of $*l$ with $k\times k$ kernel is $(k\text{-}1)\times l+1$, while the receptive field of $*$ is k.

Therefore, the range of the receptive field can be expanded, enabling larger feature information coverage. Furthermore, dilated convolution expands the receptive field by increasing 'holes' without increasing model size and computational complexity [33]. The dilated convolutions with different rates are presented in Fig. 1, where the red markers represent the actual values and the spaces between them denote 'holes'. Specifically, Fig. 1 (a)-(c) correspond to the traditional convolution, dilated convolution with rate 2, and dilated convolution with rate 3, respectively. It is noticeable that compared with the traditional convolution that only has a 3-by-3 receptive field, the dilated convolutions have expanded their receptive fields to 5-by-5 and 7-by-7, respectively, providing advantages for extracting image features.

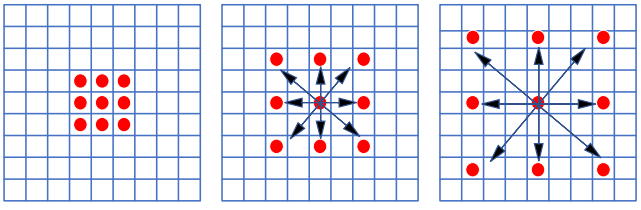

FIGURE 1. Dilated convolutions of different dilation factors. (a) The traditional convolution of dilation rate is 1. (b) The dilated convolution of dilation rate is 2. (c) The dilated convolution of dilation rate is 3.

Compared to traditional convolutional approaches, dilated convolution is widely used for tasks requiring large spatial feature extraction, such as image and video processing [26], [30], [31], [34], where larger scale feature information must be extracted, and semantic segmentation [27], [28], [29], [35], which involves processing greater ranges of feature information.

SNR is an another critical factor that significantly affects the accuracy of model classification, whereby radar signals are more readily identifiable in high SNR environments compared to low SNR environments. This disparity can be attributed to the well-established fact that noise can corrupt the genuine characteristics of radar signals, causing a wide range of difficulties during signal identification. Driven by the significance of SNR, a question arises as to how to extract features in low SNR environments. To address this problem, an effective enhancement method called multi-scale dilated convolution fusion technology has been proposed, and has been demonstrated significant advantages in various applications such as SAR image recognition [28], image semantic segmentation task [29] and audio source separation task [36]. Compared to traditional CNN, the improved CNN model with this technology is better equipped to identify complex features from noisy data. The primary reason behind this improvement is that traditional CNN extract features using a fixed convolution kernel of single scale size which cannot fully exploit feature information at different abstraction levels [37], and are susceptible to noise interference. However, by implementing multiple scales of convolution kernels, the multi-scale dilated convolution fusion technology can effectively capture features at different abstraction levels [38] and integrate the feature maps obtained from these multiple scales into more comprehensive representations of input images. This approach reduces the model's sensitivity to noise and suppresses noise interference in feature extraction to some extent. And the experimental results will demonstrate the effectiveness of this method, as it significantly improves the performance of CNNs and enhances feature extraction in low SNR environments.

Based on the above comprehensive analysis, we propose a new design method for a multi-scale dilated residual network to extract more complete spatial structural features of time-frequency images while mitigating the adverse effects of noise, without increasing model complexity and computational cost. The starting point of our construction is a set of residual network architectures proposed by He et al [39], which have been widely applied to time-frequency image recognition due to their well-designed model architecture. The residual network architecture consists of five groups of convolutional layers, denoted as conv1∼conv5, with the output feature map size of each group of convolutional layers being 112×112, 56×56, 28×28, 14×14, and 7×7, respectively. The architecture of ResNet-18 is presented in Fig. 2 (a). The yellow marks denote shortcut connections, the relationship between the input and output of the residual block can be written as follow:

$$y = x + F(x) \tag{3}$$

where x represent input, y is the output, and the function $F(x)$ denotes residual mapping. Through a multilayer network

structure, the output can be represented as follow:

$$x_O = x_l + \prod_{i=l}^{o-1} F(x_i) \tag{4}$$

Thus, through residual block, each layer parameters can play a role. We denote loss as loss function of networks, the gradient of the reverse process can be represented as follow:

$$\frac{\partial loss}{\partial x_l} = \frac{\partial loss}{\partial x_L} \bullet \frac{\partial x_L}{\partial x_l} = \frac{\partial loss}{\partial x_L}(1 + \frac{\partial}{\partial x_l} \prod_{i=l}^{o-1} F(x_i)) \tag{5}$$

Shortcut connections ensures that gradients do not vanish during the multiplication process. And another residual gradient $\frac{\partial}{\partial x_l} \prod_{i=l}^{o-1} F(x_i)$ needs to pass through weight layers.

In the design of convolutional neural network models, two critical aspects are where to use dilated convolutions and how to choose the dilation rate. Our design follows the principle of using standard convolutions in the lower layers and dilated convolutions in the upper layers. This is because at the lower layers, the model mainly extracts ''low-level'' features such as edges and corners, where small convolution kernels are more effective. However, as the depth of the network increases, the abstract ability of the network gradually improves. At the top layers, the convolution kernels need to extract ''high-level'' features such as textures, shapes, and global information, thus requiring larger receptive fields. Moreover, the selection of dilation rates should be proportional to the output feature map size [40]. If the output feature map is small, smaller dilation rates should be used, and vice versa. This is because improper use of dilated convolutions can lead to the problem of grid artifacts. Due to the insertion of zero values in the input signal by dilated convolutions, there may exist large gaps between input signals, resulting in discontinuities in information, thereby forming a grid-like artifact in the output feature map. As the resolution of the feature map decreases, the discontinuity caused by too large dilation rates will become more apparent [36]. Finally, in situations where model performance is similar, lower dilation rates should be preferred wherever possible. We should avoid unnecessary modifications as much as possible, clearly understand the purpose and underlying mechanisms behind each modification, and improve the interpretability of the model after the enhancement.

In accordance with the design principles outlined above, we adopted a highly flexible approach and propose two specific design methods that mainly modify the conv3 and conv4 convolutional layers. The first approach is relatively simple, involving the replacement of standard convolutions with dilated convolutions, as shown in Fig. 2 (b). Since the output feature map resolution of conv4_x halves compared to that of conv3_x, the corresponding dilation rate also halves accordingly. It should be noted that compared to the original residual network, the multi-scale dilated residual network (MDRN) models proposed in this paper maintain the same computational complexity and volume. The only difference between the two architectures is the replacement of standard convolution with dilated convolution in conv3_x and conv4_x.

To explore the impact of the dilation factor on classification performance, we provide six different dilation-scaled MDRN models named MDRN-1 to MDRN-6, as shown in Table 1.

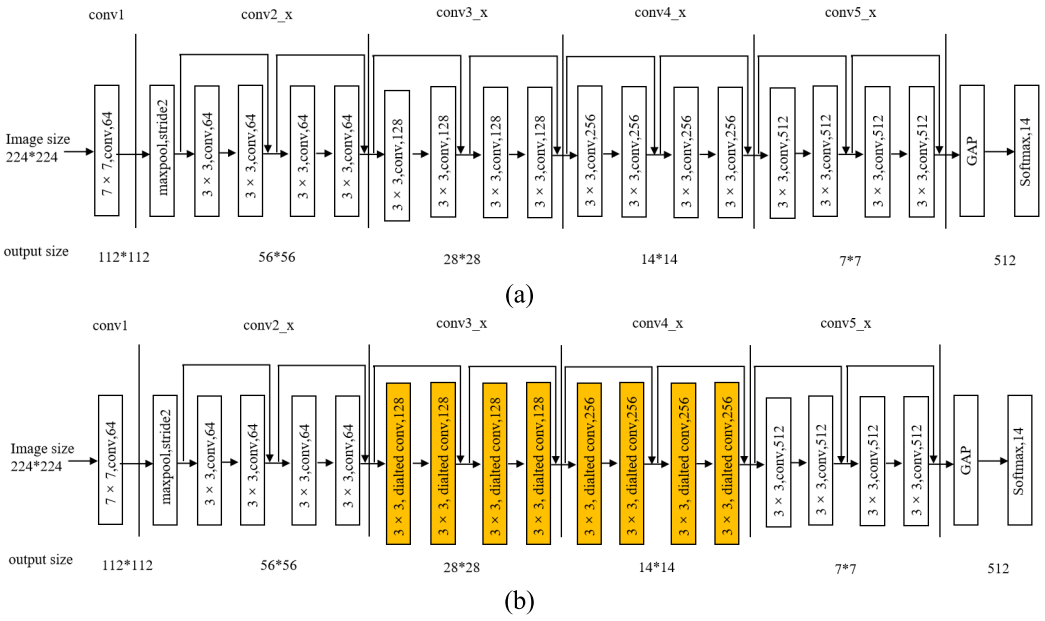

FIGURE 2. The architecture of ResNet and its corresponding MDRN, the yellow mark indicates the part that has been modified. (a) The architecture of ResNet-18. (b) The architecture of MDRN.

TABLE 1. Comparison of structural parameter configuration between mdrn and resnet models.

	input size	ResNet-18	MDRN-1	MDRN-2	MDRN-3	MDRN-4	MDRN-5	MDRN-6
conv3_x	28×28	d=1	d=2	d=4	d=6	d=8	d=10	d=12
conv4_x	14×14	d=1	d=1	d=2	d=3	d=4	d=5	d=6

TABLE 2. Comparison of structural parameter configuration between mdrn and resnet models.

	input	MDRN-A				MDRN-B				MDRN-C				MDRN-D			
	size	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
conv3_x	28×28	d=4	d=5	d=6	d=7	d=5	d=6	d=7	d=8	d=6	d=7	d=8	d=9	d=7	d=8	d=9	d=10
conv4_x	14×14	d=1	d=2	d=3	d=4	d=2	d=3	d=4	d=5	d=3	d=4	d=5	d=6	d=4	d=5	d=6	d=7

The second design method is more complex, which inspired by the design idea of the inception structure proposed in GoogleNet model [41]. Our design idea is to construct an inception structure instead of the standard convolutional structure, which involves multiple dilation-scaled convolutions, to extract multi-scale features. To be specific, we evenly divided the original number of convolutions into four parts and used different dilation-scaled convolutions for feature extraction in each part before fusing the extracted features. The convolutional structure before and after modification is shown in Fig. 3. It is worth noting that the modifications to the architecture did not change the complexity or volume of the model. The inception structure can be formulated as follow:

$$y_p = \text{concat}(\sigma(W^*_{d1}x_p), \sigma(W^*_{d2}x_p), \sigma(W^*_{d3}x_p), \sigma(W^*_{d4}x_p)) \tag{6}$$

where W_{d1}, W_{d2}, W_{d3}, W_{d4} are the parameter of dilated convolutions with different dilation rates, respectively; σ is a composite of batch normalization (BN) and rectified linear unit (ReLu). Compared to the first simple approach, it is clear that the MDRN models improved using second design method have more dilation scales and a more complex structure.

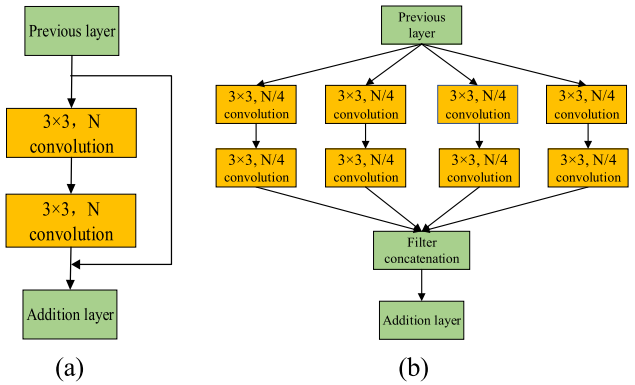

FIGURE 3. The structure of basic block and its improved structure using method 2. (a) The structure of basic block. (b) Improved structure using method 2.

We also added four new MDRN models, named MDRN-A to MDRN-D, as shown in Table 2.

C. THE TRANSFERABILITY OF OUR METHOD

To evaluate the transferability and effectiveness of our proposed method in optimizing not only the MDRN model but also other related models, we conducted comparative experiments on three widely-used classic CNN models, namely ResNet50 [39], GoogleNet [41], and MobileNetV2 [42]. These models are commonly employed in TFI classification tasks and thus serve as control groups for our study. To demonstrate the reliability and effectiveness of our proposed method, we modified these three models using the first type of improvement strategy presented in this paper and designated them as ResNet50-A, GoogleNet-A, and MobileNetV2-A, respectively, which served as the experimental group.

The basic building blocks of three typical models are Inception, Bottleneck, and Inverted residuals and linear bottlenecks, as shown in Fig. 4. In this study, we improved the original CNN model by utilizing dilated convolution with an expansion factor of 6 at the output feature map resolution of 28×28, and an expansion factor of 3 at the output feature map resolution of 14×14, while no operation was conducted on other positions. It is noteworthy that our improved model maintains the same architecture, number of convolutional kernels, complexity, and output feature map resolution as the original CNN model, except for using dilated convolution.

As shown in Fig. 4(a), the model adopting the inception structure was adjusted by altering the expansion factors of 3×3 and 5×5 convolutional kernels. The Bottleneck structure depicted in Fig. 4(b) is the fundamental unit of ResNet50, ResNet101, and ResNet152 models, which differs from the basic block structure in ResNet-18. Here, only the expansion factor of the 3×3 convolutional kernel was adjusted. Additionally, the Inverted residuals and linear bottlenecks structure, which is the fundamental unit of the lightweight CNN MobileNetV2 model, was studied in our research, as shown in Fig. 4(c). We only adjusted the parameters of the 3×3 depthwise convolutional kernel.

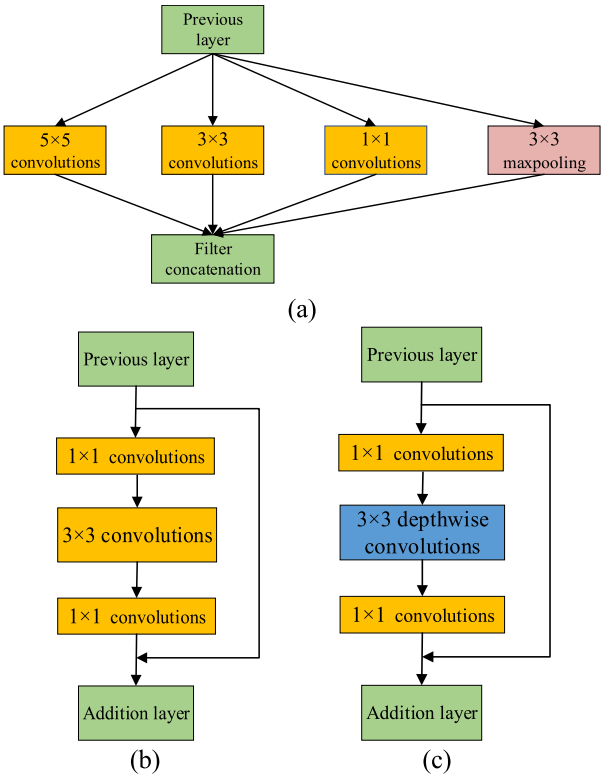

FIGURE 4. The basic building blocks of three typical models. (a) Inception. (b) Bottleneck. (c) Inverted residuals and linear bottlenecks.

III. EXPERIMENT AND RESULTS DISCUSSION

This section presents experiments conducted to demonstrate the efficiency and transferability of the proposed methods. Firstly, the experimental dataset is introduced and constructed. Subsequently, compared to the original ResNet models, we evaluate the performance of the improved models (MDRN). Additionally, we compare the accuracy of different classifiers under varying SNR. Lastly, we test the transferability of our proposed method based on three typical CNN models. Notably, to ensure fairness and rationality in the comparative experiment, we adopted the same experimental setup for both the control group and the experimental group. Specifically, we used identical dataset and training methods to train and test these models.

Before proceeding with the discussion, we introduce the definition of ''accuracy'' as a measurement metric. Accuracy is the ratio of ''correct'' to ''incorrect'' predictions made by a model,which is the quality of being right or accurate, or a way of measuring it [43]. This is because CNN model architectures are usually complex, multilayered and nonlinear, making it difficult to see what happens inside them and how they make decisions. To address this dilemma, scholars in this field usually adopt ''accuracy'' to evaluate the performance of CNN models. For example, in the popular ImageNet (ImageNet Large-Scale Visual Recognition Challenge) competition, ''Top-5 accuracy'' and ''Top-1 accuracy'' are common metrics used to evaluate the performance of the model.

A. DATASET INTRODUCTION AND CONSTRUCTION

The whole dataset contains radar signals with 14 modulation models, including continuous wave signal (CW), linear frequency modulation (LFM), non-linear frequency modulation (NLFM), frequency modulated continuous wave signal (FMCW), quaternary frequency shift keying signal (QFSK), binary frequency shift keying signal (BFSK), frequency hopping coded signal (COSTAS), binary phase shift keying signal (BPSK), quaternary phase shift keying signal (QPSK) and polyphase codes signals (FRANK, P1, P2, P3, P4). To facilitate computer processing, the intercepted radar signal is usually describe into

$$x(l) = s(l) + n(l) \tag{7}$$

where x is the actual received signal, s is the original signal without noise, n is the noise, $l \in N^*$ denotes the sample point. Gaussian white noise is adopted in current work. And the signal can be formulated as follow:

$$\begin{aligned} s(l) &= Ae^{j\phi} \\ \phi &= \frac{2\pi l \delta(l)}{f_s} + \theta(l) + \theta_0 \end{aligned} \tag{8}$$

where A is the signal amplitude, δ and θ are the frequency modulation function and phase modulation function used for frequency coding and phase coding respectively, θ_0 is initial phase.

In order to better approximate real-world scenarios, Gaussian white noise with different SNRs ranging from -16 to 0 dB in 2 dB increments is added to each original signal. Each modulation scheme generates 1200 samples at each SNR level. Based on the ratio of 4:1:1, the samples are divided into training sets, validation sets, and testing sets, resulting in 100800, 25200, and 25200 samples, respectively. The pulse width and the sampling frequency fs of the simulated signal are fixed at 2 μs and 400MHz, respectively. The configuration of more parameters are detailed in Table 3

Radar signals are non-stationary signals, and time-frequency analysis is an important method for describing the frequency variation of non-stationary signals over time. Time-frequency analysis can be used to reveal the modulation characteristics of radar radiation source signals. Among them, Choi-Williams distribution (CWD) is a widely used and effective method for time-frequency analysis. The CWD uses an exponential weighted kernel to suppress cross-terms and has high resolution capability and identification accuracy for signals of different times or frequencies. As radar signals contain a large amount of noise energy, cross-terms are generated during the time-frequency transformation process, which interferes with the features of the real radar signals. In this context, the CWD can effectively suppress cross-terms and minimize damage to the features of radar signal images. Therefore, in this paper, the CWD is adopted for the time-frequency transformation of radar signals. Mathematically, the CWD can be expressed

TABLE 3. Signal parameter settings.

Modulation Type	Parameter	Ranges
CW	Carrier frequency f_0	[1/10, 1/4]f_s
LFM, NLFM	Carrier frequency f_0	[1/10, 1/4]f_s
	Bandwidth B	[1/10, 1/4]f_s
BPSK	Carrier frequency f_0	[1/10, 1/4]f_s
	Barker Code Length	[7, 11, 13]
QPSK	Carrier frequency f_0	[1/10, 1/4]f_s
	Phase coding sequence length	[5, 7]
FMCW	Carrier frequency f_0	[1/20, 1/5]f_s
	Bandwidth B	[1/20, 1/5]f_s
	Cycle T	1 μs
FRANK, P1, P2, P3, P4	Carrier frequency f_0	[1/10, 1/4]f_s
	Step frequency N	8
BFSK	Carrier frequency f_0	[1/20, 1/4]f_s
	Barker Code Length	11
QFSK	Carrier frequency f_0	[1/10, 1/4]f_s
	Frequency coding sequence length	6
COSTAS	Carrier frequency f_0	[1/10, 1/4]f_s
	Frequency coding sequence length	6

as follows:

$$\mathrm{CWD}(t,w)=\iint \frac{1}{\sqrt{4\pi\tau^2/\sigma}}\exp[-\frac{(t-u)^2}{4\tau^2/\sigma}]s(t+\frac{\tau}{2}) \times s^*(t+\frac{\tau}{2})\mathrm{e}^{-jw\tau}dud\tau \quad (9)$$

where t and w denote time and angular frequency, σ is the attenuation coefficient. To acquire the high time-frequency resolution and the suppression of cross terms, the configuration of σ is 1.

The CWD transformation can create 14 distinct types of radar signals, as revealed in Fig. 5, when the SNR is 0dB. Analysis of the resulting time-frequency images shows that the structural characteristics of the signals are partially disrupted, with some subtle features obscured by noise energy. This increases the difficulty of correctly classifying the radar signals. For instance, CW, LFM, NLFM, FMCW, QFSK, BFSK, and COSTAS waveforms exhibit clear and distinctive features that enable distinguishing them from other signals. However, the characteristics of BPSK and QPSK waveforms are less apparent and have some similarities. Furthermore, it is challenging to distinguish between P1 and P2, or FRANK and P3 at low SNR conditions, when multiple phase-encoded signals with the same step frequency (FRANK, P1, P2, P3 and P4) are set.

To facilitate the training and testing of the model, it is necessary to preprocess TFI. This paper adopts a simple preprocessing method to ensure that the original features of TFI are not destroyed. Only grayscale processing and bicubic interpolation operations are performed on TFI, and the data samples with a size of 224 × 224 are obtained in the end.

B. EXPERIMENT RESULTS OF MDRN

Given the significant differences between TFI and conventional images, we proposed the MDRN design idea of using dilated convolutions in high-resolution space to enhance the model's feature extraction capability under low signal-to-noise ratio conditions. To verify the superiority of our method, we tested the accuracy of the models listed in Table 1 and Table 2. As shown in Fig. 6, we plotted the average classification rate of ten models. The experimental results of our study are consistent with our previous predictions, suggesting that the reasonable use of dilated convolutions can improve model classification accuracy. The ResNet-18 model was selected as the benchmark model, and its classification accuracy was the lowest, at 92.15%. Any MDRN model improved from the benchmark outperformed the ResNet-18 model, with MDRN-3 performing the best, achieving a classification accuracy of 93.12%. It should be noted that MDRN models improved by strategy 1 (MDRN-1 to MDRN-6) showed significantly better classification accuracy than MDRN models improved by strategy 2 (MDRN-A to MDRN-D). This suggests that the inception structure combining multiple dilation convolution scales is not suitable for TFI recognition. One possible reason is that excessive dilation convolutions with different feature extractions under high noise energy would introduce errors during feature fusion.

It can be seen in Fig. 6 that the classification accuracy changes with the variation of the dilation rate in design method one and two. This is because as the dilation factor increases, the model can extract more comprehensive representations of images, but this can also exacerbate the grid artifacts [36], which can blur the extracted features. Therefore, Therefore, it is the model with an appropriately chosen dilation factor, rather than the model with the maximum dilation factor, that will achieve the best performance.

In order to further compare the stability and robustness of the proposed methods, we will analyze the classification accuracy of models for every signals under different SNR conditions, especially in low SNR environment. As MDRN-1 to MDRN-6 adopt the same improvement method, their performance under different signal-to-noise ratio conditions for different signals should be roughly similar, with differences only in terms of performance levels. Similar behaviors can be seen in MDRN-A to MDRN-D. Consequently, in order to visually and efficiently demonstrate the superiority of the proposed method, we have selected the best models based on the two proposed modifications, namely MDRN-3 and MDRN-B, to compare with the benchmark model ResNet-18.

Fig. 7 shows the recognition accuracy of the three models for various signals under different SNR conditions. The results indicate that under each SNR condition, the accuracy of the two MDRN models is higher than or equal to that of the

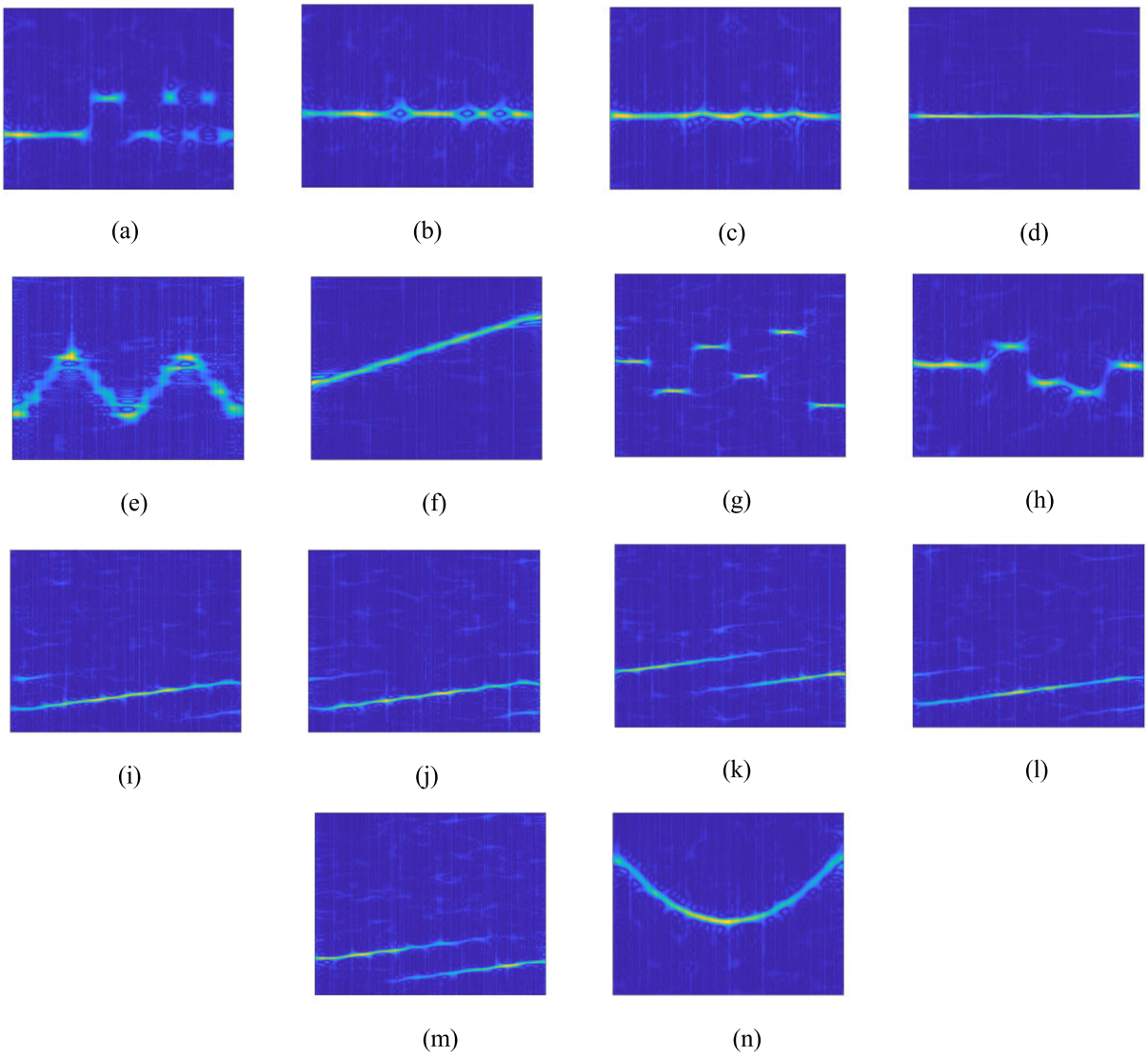

FIGURE 5. The TFI of 14 types intra-pulse modulation signals by CWD transformation at the condition of SNR is 0dB. (a) BFSK. (b) BPSK. (c) QPSK. (d) CW. (e) FMCW. (f) LFM. (g) COSTAS. (h) QFSK. (i) P1. (j) P2. (k) P3. (l) P4. (m) FRANK. (c) NLFM.

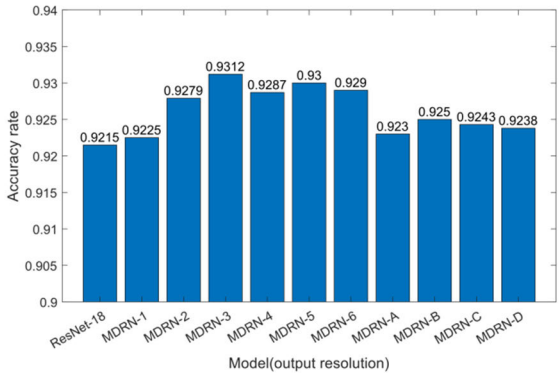

FIGURE 6. The accuracy of models of different dilation degree.

corresponding ResNet models with no increase in complexity. Notably, when the SNR is below −10dB, the accuracy of ResNet-18 declines faster than that of the two MDRN models, indicating that the classification performance advantage of MDRN models becomes more apparent as SNR decreases. For instance, when the SNR is −10dB, the accuracy of MDRN-3 and MDRN-B is approximately 0.5% and 0.4% higher than that of ResNet-18, respectively. Furthermore, when the SNR is −16dB, the accuracy difference between them increases to 4.1% and 2.5%, respectively.

Three models perform well in the classification of relatively easy signals. For example, when the SNR is greater than −10dB, the classification accuracy of BFSK, BPSK, CW, COSTAS, FRANK, LFM, NLFM, P3, QFSK and QPSK signals reaches nearly 100% for all three models. However, for signals that are more difficult to distinguish, the performance of the three models decreases. Specifically, the overall performance of the three models in classifying FMCW, P1, P2, and P4 signals is inferior to that of easily identifiable signals. When the SNR decreases, MDRN models demonstrate superior classification accuracy to ResNet-18 for all signals,

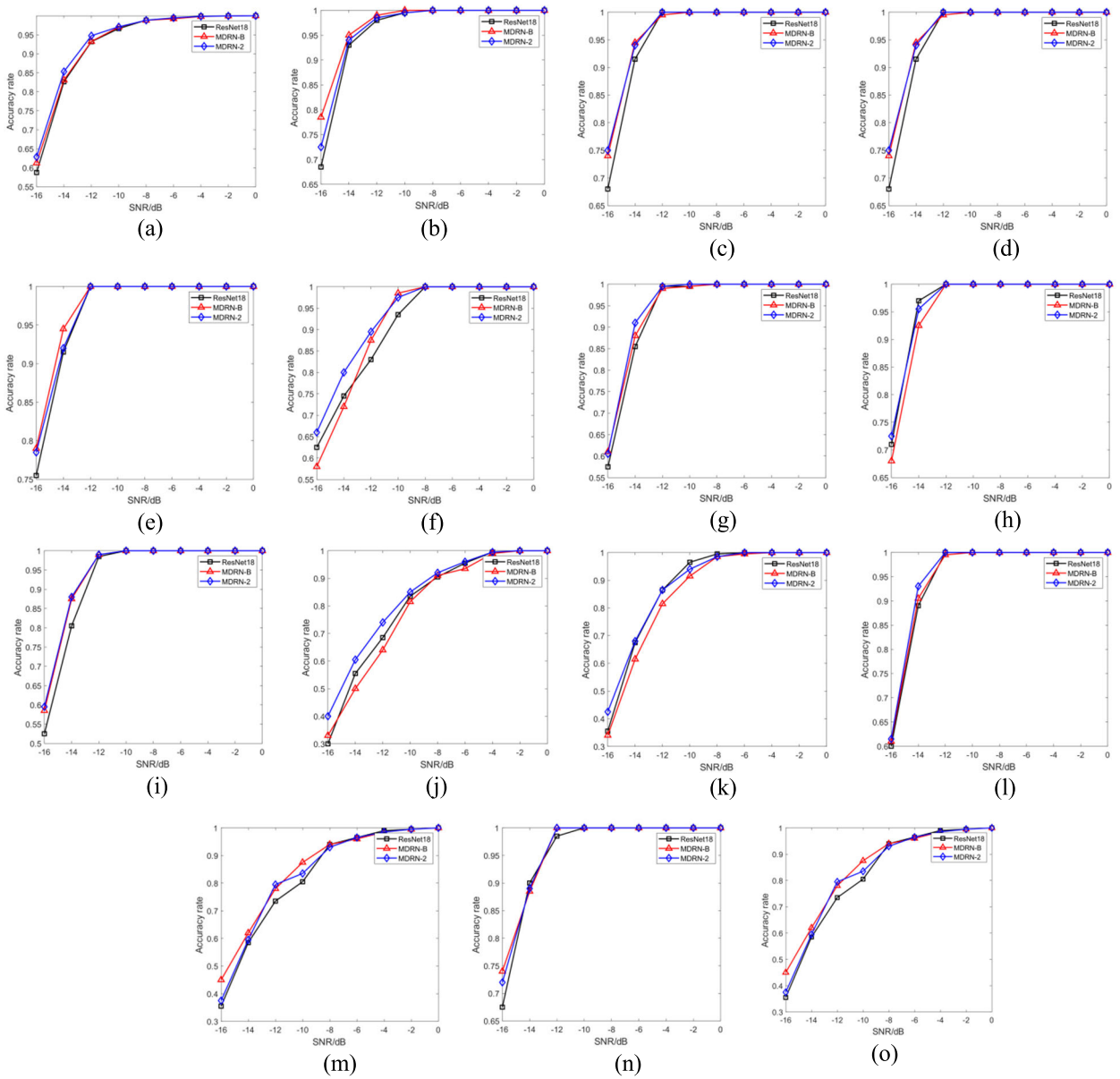

FIGURE 7. The performance comparison of three models for fourteen signals classification. (a) The overall average accuracy of three models. (b) BFSK. (c) BPSK. (d) COSTAS. (e) CW. (f) FMCW. (g) FRANK. (h) LFM. (i) NLFM. (j) P1. (k) P2. (l) P3. (m) P4. (n) QFSK. (o) QPSK.

particularly MDRN-3. This advantage is even more evident in recognizing confusing signals. For example, when the SNR is as low as −16dB, the recognition rate of MDRN-3 for FMCW, P1, P2, and P4 signals is over 8% higher than that of ResNet-18.

C. TRANSFERABILITY OF MDRN

In Section III-B, we successfully tested the superiority of the MDRN model. The purpose of this section is to evaluate the transferability of our proposed method to other baseline models. To this end, we selected three classic open-source models, namely ResNet50, GoogleNet, and MobileNetV2, and improved them using the approach described in this paper. Specifically, we respectively named the corresponding improved models ResNet50-A, GoogleNet-A, and MobileNetV2-A and evaluated their performance through testing.

As shown in Fig. 8, dilated models have better performance than their corresponding original models. For instance, ResNet50-A achieved a higher accuracy rate of 0.87% compared with ResNet50, GoogleNet-A outperformed GoogleNet by 0.95% in terms of accuracy, and MobileNetV2-A improved the accuracy of MobileNetV2 by 1.21%. It should be emphasized that traditional models are hardly able to achieve high accuracy rates comparable to those of dilated models, particularly when they have similar levels of depth and complexity. This further confirms our analysis in Section II-C, which suggests that our approach can

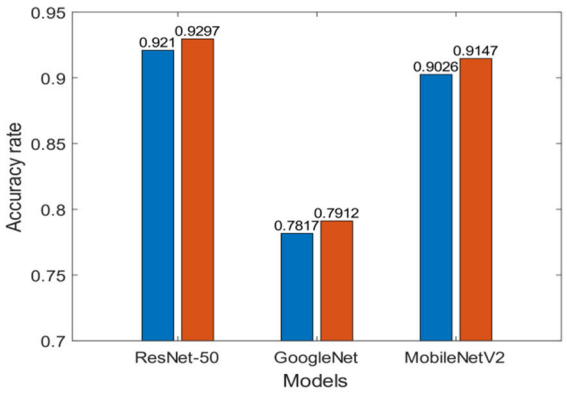

FIGURE 8. Comparison of accuracy between three classic models before and after improvement.

not only be applied to residual networks but also transferred well to other CNN models.

We observed that the classification accuracies of three typical models, GoogleNet, ResNet-50, and MobileNetV2, were lower than or comparable to that of ResNet-18, despite having deeper and more complex architectures. Therefore, based on our experience, it can be concluded that CNN models built on basic blocks (the fundamental building blocks of ResNet-18) are more suitable for TFI classification than those constructed respectively with Inception (the basic structure of GoogleNet), Bottleneck (the basic structure of ResNet-50), and Inverted residuals and linear bottlenecks (the basic structure of MobileNetV2). This suggests that, in the context of TFI classification, higher depth and complexity do not necessarily lead to better performance, and that simpler basic block structures may be better suited for this task.

IV. CONCLUSION

In this paper, we showed the importance of multi-scale dilated convolutions in learning representation features under high noise conditions and proposed a novel CNN design approach for TFI classification. By increasing the model's receptive field and incorporating richer representation features, we can extract more complete global features of the target and suppress noise. Experimental results demonstrate that our simple and flexible transformation yields higher model performance compared to original CNN models. For example, MDRN-3 has an overall average accuracy that is 0.97% higher than ResNet-18's. Additionally, we find that our design method has good transferability and can be easily applied to other types of models. For instance, the transformed models have overall average accuracy that are 0.87%, 0.95%, and 1.21% higher than the original models GoogleNet, ResNet-50, and MobileNetV2, respectively. This study provides a novel CNN design approach that will further facilitate the convergence of deep learning algorithms and TFI classification.

REFERENCES

[1] X. Wang, "Electronic radar signal recognition based on wavelet transform and convolution neural network," *Alexandria Eng. J.*, vol. 61, no. 5, pp. 3559–3569, May 2022.

[2] S. Yuan, P. Li, and B. Wu, "Towards single-component and dual-component radar emitter signal intra-pulse modulation classification based on convolutional neural network and transformer," *Remote Sens.*, vol. 14, no. 15, p. 3690, Aug. 2022.

[3] S. Zhao, W. Wang, D. Zeng, X. Chen, Z. Zhang, F. Xu, X. Mao, and X. Liu, "A novel aggregated multipath extreme gradient boosting approach for radar emitter classification," *IEEE Trans. Ind. Electron.*, vol. 69, no. 1, pp. 703–712, Jan. 2022.

[4] M. Du, X. He, X. Cai, and D. Bi, "Balanced neural architecture search and its application in specific emitter identification," *IEEE Trans. Signal Process.*, vol. 69, pp. 5051–5065, 2021.

[5] M. R. Azimi-Sadjadi, D. Yao, Q. Huang, and G. J. Dobeck, "Underwater target classification using wavelet packets and neural networks," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 784–794, May 2000.

[6] J. Lunden and V. Koivunen, "Automatic radar waveform recognition," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 124–136, Jun. 2007.

[7] S. Liu, X. Yan, P. Li, X. Hao, and K. Wang, "Radar emitter recognition based on SIFT position and scale features," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 65, no. 12, pp. 2062–2066, Dec. 2018.

[8] V. Iglesias, J. Grajal, P. Royer, M. A. Sanchez, M. Lopez-Vallejo, and O. A. Yeste-ojeda, "Real-time low-complexity automatic modulation classifier for pulsed radar signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 108–126, Jan. 2015.

[9] I. Jordanov, N. Petrov, and A. Petrozziello, "Supervised radar signal classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Vancouver, BC, Canada, Jul. 2016, pp. 1464–1471.

[10] Z. Jing, P. Li, B. Wu, S. Yuan, and Y. Chen, "An adaptive focal loss function based on transfer learning for few-shot radar signal intra-pulse modulation classification," *Remote Sens.*, vol. 14, no. 8, p. 1950, Apr. 2022.

[11] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, Feb. 2019.

[12] P. Lang, X. Fu, M. Martorella, J. Dong, R. Qin, X. Meng, and M. Xie, "A comprehensive survey of machine learning applied to radar signal processing," 2020, *arXiv:2009.13702*.

[13] X. Zhang, J. Zhang, T. Luo, T. Huang, Z. Tang, Y. Chen, J. Li, and D. Luo, "Radar signal intrapulse modulation recognition based on a denoising-guided disentangled network," *Remote Sens.*, vol. 14, no. 5, p. 1252, Mar. 2022.

[14] P. Chikontwe, S. Kim, and S. H. Park, "CAD: Co-adapting discriminative features for improved few-shot classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 14534–14543.

[15] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 9992–10002.

[16] A. Gupta, S. Narayan, K. J. Joseph, S. Khan, F. S. Khan, and M. Shah, "OW-DETR: Open-world detection transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 9225–9234.

[17] Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe, and J. Morimoto, "Deep learning, reinforcement learning, and world models," *Neural Netw.*, vol. 152, pp. 267–275, Aug. 2022.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[19] Z. Qu, X. Mao, and Z. Deng, "Radar signal intra-pulse modulation recognition based on convolutional neural network," *IEEE Access*, vol. 6, pp. 43874–43884, 2018.

[20] T. Wan, K.-L. Jiang, H. Ji, and B. Tang, "Deep learning-based LPI radar signals analysis and identification using a Nyquist folding receiver architecture," *Defence Technol.*, vol. 19, pp. 196–209, Jan. 2023.

[21] K. Chen, S. Zhang, L. Zhu, S. Chen, and H. Zhao, "Modulation recognition of radar signals based on adaptive singular value reconstruction and deep residual learning," *Sensors*, vol. 21, no. 2, p. 449, Jan. 2021.

[22] B. Lay and A. Charlish, "Classifying LPI signals with transfer learning on CNN architectures," in *Proc. Sensor Signal Process. Defence Conf. (SSPD)*, Sep. 2020, pp. 1–5.

[23] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of image classification algorithms based on convolutional neural networks," *Remote Sens.*, vol. 13, no. 22, p. 4712, Nov. 2021.

[24] Y. Ma, D. Tsao, and H.-Y. Shum, "On the principles of parsimony and self-consistency for the emergence of intelligence," *Frontiers Inf. Technol. Electron. Eng.*, vol. 23, no. 9, pp. 1298–1323, Sep. 2022.

[25] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 636–644.
[26] G.-R. You, Y.-R. Shiue, C.-T. Su, and Q.-L. Huang, "Enhancing ensemble diversity based on multiscale dilated convolution in image classification," *Inf. Sci.*, vol. 606, pp. 292–312, Aug. 2022.
[27] M. Lan, Y. Zhang, L. Zhang, and B. Du, "Global context based automatic road segmentation via dilated convolutional neural network," *Inf. Sci.*, vol. 535, pp. 156–171, Oct. 2020.
[28] Q. Liu, M. Kampffmeyer, R. Jenssen, and A.-B. Salberg, "Dense dilated convolutions' merging network for land cover classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6309–6320, Sep. 2020.
[29] E. Yvinec, A. Dapogny, M. Cord, and K. Bailly, "RED: Looking for redundancies for data-free structured compression of deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 20863–20873.
[30] Q. Chen, W. Zhang, N. Zhou, P. Lei, Y. Xu, Y. Zheng, and J. Fan, "Adaptive fractional dilated convolution network for image aesthetics assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14102–14111.
[31] R. Xue, X. Bai, and F. Zhou, "Spatial–temporal ensemble convolution for sequence SAR target classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1250–1262, Feb. 2021.
[32] Z. Zhang, X. Wang, and C. Jung, "DCSR: Dilated convolutions for single image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1625–1635, Apr. 2019.
[33] L. Wu, X. Zhang, K. Wang, X. Chen, and X. Chen, "Improved high-density myoelectric pattern recognition control against electrode shift using data augmentation and dilated convolutional neural network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2637–2646, Dec. 2020.
[34] C. Schmidt, A. Athar, S. Mahadevan, and B. Leibe, "D2Conv3D: Dynamic dilated convolutions for object segmentation in videos," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Snowmass Village, CO, USA, Jan. 2022, pp. 1929–1938.
[35] J. Wang, H. Xiong, H. Wang, and X. Nian, "ADSCNet: Asymmetric depthwise separable convolution for semantic segmentation in real-time," *Appl. Intell.*, vol. 50, no. 4, pp. 1045–1056, Nov. 2019.
[36] Y. Chen, Q. Guo, X. Liang, J. Wang, and Y. Qian, "Environmental sound classification with dilated convolutions," *Appl. Acoust.*, vol. 148, pp. 123–132, May 2019.
[37] C. Zhuang, Z. Lu, Y. Wang, J. Xiao, and Y. Wang, "ACDNet: Adaptively combined dilated convolution for monocular panorama depth estimation," in *Proc. 36th AAAI Conf. Artif. Intell.*, Vancouver, BC, Canada, Feb. 2022, pp. 3653–3661.
[38] V. Chalavadi, P. Jeripothula, R. Datla, S. B. Ch, and M. C. Krishna, "mSODANet: A network for multi-scale object detection in aerial images using hierarchical dilated convolutions," *Pattern Recognit.*, vol. 126, Jun. 2022, Art. no. 108548.
[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[40] N. Takahashi and Y. Mitsufuji, "Densely connected multidilated convolutional networks for dense prediction tasks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 993–1002.
[41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.
[42] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 4510–4520.
[43] A. Somani, A. Horsch, and D. K. Prasad, "Deep learning glossary," in *Interpretability in Deep Learning*, 1st ed. Norway, Europe: Springer, 2023, pp. 25–28.

ENZE GUO received the B.S. and M.S. degrees from Army Engineering University, in 2015 and 2022, respectively. He is currently an Assistant Engineer with The 63893 Forces of PLA. His research interests include automatic pattern recognition, image processing, and radar signal processing.

HAO WU received the B.S. degree from the Changchun University of Science and Technology and the M.S. degree from Army Engineering University, in 2018. He is currently an Assistant Engineer with The 63896 Forces of PLA. His research interests include deep learning, radar counter-measurement and counter counter-measurement, and radar signal processing.

MING GUO received the B.S. degree from Air Force Engineering University, in 2005. He is currently an Engineer with The 63893 Forces of PLA. His research interests include waveform design, radar counter-measurement and counter counter-measurement, and radar signal processing.

YINAN WU received the B.S. degree from the Military Non Commissioned Officer College, in 2015. His research interests include waveform design, radar counter-measurement and counter counter-measurement, and radar signal processing.

JIAN DONG received the B.Eng. degree in signal and information processing from Army Engineering University, in 2006, and the Ph.D. degree in navigation, guidance and control from the Beijing Institute of Technology (BIT), Beijing, China, in 2009. He is currently an Associate Professor of BIT. His research interests include radar signal processing, radar counter-measurement and counter counter-measurement, inverse synthetic aperture radar imaging, waveform design, automatic target recognition, and micro-Doppler feature extraction.