IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 24 July 2023, accepted 5 August 2023, date of publication 15 August 2023, date of current version 21 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3305533

RESEARCH ARTICLE

Attention is Needed for RF Fingerprinting

HANQING GU[1], LISHENG SU[1,2], WEIFENG ZHANG[3], AND CHUAN RAN[4]

[1]Zhejiang JEC Electronic Company Ltd., Zhejiang 314036, China
[2]School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610056, China
[3]College of Information Science and Engineering, Jiaxing University, Jiaxing 314041, China
[4]IBM Corporation, Durham, NC 27703, USA

Corresponding author: Weifeng Zhang (zhangweifeng@zjxu.edu.cn)

ABSTRACT Radio Frequency (RF) fingerprinting is a novel solution for identifying a unique radio from a large pool of devices by analyzing the subtle characteristics that are inherent in the radio waveform. Deep convolutional neural networks have been widely used to handle the RF fingerprinting task because of their exceptional capacity for representation learning. However, there are still challenges in employing deep convolutional neural networks, such as how to enable the model learn more robust and discriminative RF fingerprints. This paper aims to explore new model architectures to learn robust RF fingerprints. Hence we proposes a novel Dual Attention Convolution module that simultaneously learns channel attention and spatial attention to tune the RF fingerprints, enhancing the convolutional layers' potential for representation learning. Our proposed module is lightweight and plug-and-play. A number of convolutional neural networks can be equipped with our module, which enables them to extract robust and discriminative RF fingerprints. Our approach has been extensively tested through experimental trials, and the results have demonstrated its effectiveness. It is shown that the performance of convolutional neural networks on RF fingerprinting can be improved 1.5% on average, and DAConv-ResNet50 which combined ResNet50 and our Dual Attention Convolution module can achieve 95.6% recognition accuracy on 10 USRP X310. Our source code is available at https://github.com/zhangweifeng1218/Adaptive_RF_Fingerprinting.

INDEX TERMS RF fingerprinting, channel attention, spatial attention, deep neural networks.

I. INTRODUCTION

Due to its low cost and flexible installation, smart wireless Internet of Things (IoT) devices are now widely employed in smart cities, internet of vehicles, and other industries [1]. However, it is important to note that since information is transferred through electromagnetic waves, wireless IoT networks are more vulnerable to passive attacks such as unauthorized listening, as well as active attacks such as information manipulation and denial of service attacks. These types of attacks can compromise the security and integrity of the network, potentially leading to serious consequences [2]. Therefore, it is crucial to implement robust security measures to protect wireless IoT networks from these types of threats. Access authentication and encrypted protocols are typically used as the network security mechanism in traditional wireless IoT networks. Recent studies, however, demonstrate that nefarious users can still access the network by stealing or fabricating identifying information [3].

Hence, RF fingerprinting as a new and efficient identification method for wireless IoT networks follows. The primary objective of RF fingerprinting is to identify transmitters by analyzing the subtle characteristics that are inherent in the radio waveform due to flaws in hardware circuits, which are commonly referred to as RF fingerprints. By leveraging these unique characteristics, RF fingerprinting can provide a highly accurate and reliable method for identifying and authenticating devices within a wireless IoT network. These fingerprints are challenging for malevolent users to imitate [4]. RF fingerprinting mainly has the following steps: the first step is to receive and preprocess the radio waveform, the second step is to extract and fuse subtle

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

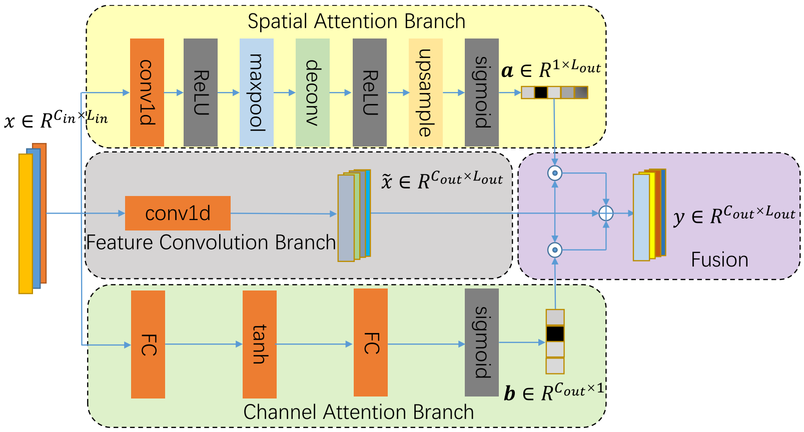


FIGURE 1. **Architecture of the proposed Dual Attention Convolution (DAConv). Our DAConv module is composed of four components: Spatial Attention Branch in yellow dashed box, Channel Attention Branch in green dashed box, Feature Convolution Branch in grey dashed box, and the Fusion block in purple dashed box. The dimensions of the key tensors are also marked in the diagram.**

characteristics from the received waveform, and the final step is to tell which device sends the radio waveform. Thus, RF fingerprinting is a typical classification task, in which how to extract the features that can reflect the subtle hardware differences or defects from the signal waveform is the key and difficult. Traditional RF fingerprinting approaches need manually design complex feature engineering according to specific communication system [5], [6]. For instance, Berik et al. [5] extracts frequency offset, instantaneous frequency, and higher-order moments as RF fingerprints to identify radio devices. However, such artificially designed fingerprints are susceptible to electromagnetic interference and is not universal. Since deep neural networks have the ability to learn features from the data automatically and has shown its outstanding representation learning ability in computer vision field, recent RF fingerprinting approaches have adopted deep convolutional neural networks to extract RF fingerprints. The research results demonstrate that, compared with traditional artificial feature engineering, deep convolutional neural network can automatically learn more effective and universal RF fingerprints [7], [8], [9], [10]. ORACLE [8], as a milestone research work on RF fingerprinting based on deep learning, has analyzed the physical mechanism of RF fingerprints and proposed a baseline convolutional neural network for RF fingerprinting. Since then, a large number of RF fingerprinting works based on deep convolutional neural networks have emerged [7], [9], [10], [11], [12], [13], [14], [15]. Deep learning has opened up new possibilities for enhancing the security and performance of wireless IoT networks. For example, literature [13] designed a modified VGG16 network to achieve RF fingerprinting under imapaired channels, while Tan et al. [14] proposed to combine RF and geomagnetic fingerprints to mine complementary information for realizing stable and reliable device identification. Yu et al. [9] proposed a multi-sampling network whose architecture is similar to AlexNet [16], to fingerprint devices using signals with different sampling rates. The above research shows that dynamic environment has great adverse effects on RF fingerprinting. They generally adopt signal preprocessing, such as denoising, channel estimation, data augmentation and other methods, to reduce the adverse effects. How to optimize the structure of neural networks to extract more discriminative and robust RF fingerprints is still a difficulty and blank in the research.

However, the convolutional networks adopted by them are relatively simple, and most of them are directly derived from the field of image recognition. Each convolution kernel in the network can be regarded as an adaptive feature learner, with different kernels learning different features that are important for recognition tasks. Unfortunately, the above simple networks ignore this point, and they only fuse multiple convolutional features by simple pooling operations. Researches in the field of computer vision have demonstrated that attention mechanism can effectively enhance the feature learning ability of deep neural networks [17], [18]. Thus, in this paper, we explore a novel RF fingerprinting method based on attention mechanism. To better understand the significance of RF fingerprints learned by convolution kernels, we have developed a flexible Dual Attention Convolution (DAConv) module. This module includes two branches: the Channel Attention Branch (CAB) and the Spatial Attention Branch (SAB). The CAB is designed to modulate the channels of RF fingerprints, while the SAB is designed to modulate the spatial dimensions of RF fingerprints. By using attention mechanisms in both branches, the DAConv module is able to enhance important features and weaken useless features, leading to more accurate and robust identification of transmitters within a wireless IoT network. Our DAConv module can be plugged into various convolutional neural networks, endowing them the ability to learn more discriminative and robust RF fingerprints.

The contributions of this paper can be summarized as follows:

- We propose an attention module named Dual Attention Convolution module, which helps the convolution operation to pay more attention to the significant parts of the input signal and suppress unnecessary information through channel and spatial attention, so as to enhance the representation learning ability of the convolutional layer.
- The proposed DAConv module can be directly used to replace the convolutional layer in traditional CNNs, such as AlexNet and ResNet, and effectively improve their performances in RF fingerprinting task under the premise of increasing a small amount of model parameters and computational consumption.
- We carry out extensive experiments on the actual collected RF fingerprinting datasets, which demonstrates that our DAConv has the potential to significantly improve the performance and effectiveness of deep convolutional neural networks for RF fingerprinting. In addition, we have published our experimental code and data to provide a basis for further research in the community.[1]

The rest of this paper is organized as follows: Firstly, the recent research progress in RF fingerprinting and attention mechanism is briefly reviewed in section II. Then our DAConv module is proposed in section III. We also introduce how to plug this block into traditional CNNs in this section. The signal collection and preprocessing method is also introduced. Then, the experimental setting is introduced in Section IV, whereby the experimental results and analysis are also presented. Finally, we summarizes this paper in section V.

II. RELATED WORK

A. RF FINGERPRINTING

RF fingerprinting is a form of signal intelligence that involves extracting features that are inherent to the hardware of a transmitter and inadvertently embedded in the transmitting waveform. These features are then used to help a passive receiver identify the transmitter. RF fingerprinting has received extensive attention from both industry and scholars in recent years, due to its potential to enhance the security and performance of wireless IoT networks. For an comprehensive survey paper on this task, please refer to [12].

Most of the traditional works have applied the feature extraction technique which is carefully customized at the physical layer to fingerprint wireless devices [5], [19], [20], [21]. For example, literature [19] extracted device-dependent radio features and designed a non-parametric Bayesian model to predict the number of radios, although their experiment was limited to only four ZigBee devices. Brik et al. [5] extracted several RF features, including frequency offset and constellation errors, as fingerprints to classify 130 Wi-Fi cards, achieving an impressive accuracy of 99%. However, their experiment was conducted in an ideal environment without dynamic signal-to-noise ratios (SNRs) and channel states, making their method impractical for real-world applications. Similarly, frequency offset and transients were also adopted in [21], but they only achieved about 47% accuracy in a non-controlled environment. More recently, Peng et al. [20] extracted modulation-specific features to fingerprint ZigBee devices and achieved about 95% accuracy on a 54-radio testbed. While these traditional approaches have shown promise, they are often limited by their reliance on carefully crafted feature extraction techniques and their inability to perform well in dynamic or non-controlled environments.

Existing feature-based fingerprinting techniques have a major flaw in that they are fundamentally designed for a particular wireless device, which restricts their usefulness in IoT settings where devices run under diverse standards. Recently, deep learning has been applied to several wireless communication problems, such as modulation classification [22], channel state estimation [23]. It has also been introduced into RF fingerprinting task. ORACLE [8], as a milestone in this direction, builded a baseline to solve RF fingerprinting using deep convolutional neural networks. In literature [10], extensive comparative experiments have been conducted to systematically analyze the influence of environmental factors such as signal-to-noise ratio (SNR), dynamic channel conditions, and the number of targets. By carefully controlling these factors and evaluating the performance of different algorithms under various conditions, researchers have been able to gain valuable insights into the strengths and limitations of different approaches to RF fingerprinting. Yu et al. [9] proposed a multi-sampling network to fingerprint devices using signals with different sampling rates. The above research shows that dynamic environment has great adverse effects on RF fingerprinting. They generally adopt signal preprocessing, such as denoising, channel estimation, data augmentation and other methods, to reduce the adverse effects. How to optimize the structure of neural networks to extract more discriminative and robust RF fingerprints is still a difficulty and blank in the research. In this paper, we present a novel model architecture for RF fingerprinting that leverages attention mechanisms, which have been widely used in computer vision. Specifically, we propose a Dual Attention Convolution module that can be integrated into any convolutional neural network to enhance its ability to learn discriminative and robust RF fingerprints.

B. ATTENTION MECHANISM IN NEURAL NETWORKS

The attention mechanism in neural networks is inspired by the human cognitive system and imitates the human cognitive process, mining and enhancing the key information in the input signals [24]. Attention mechanism can effectively

[1] Our code can be downloaded from https://github.com/zhangweifeng1218/Adaptive_RF_Fingerprinting

enhance the feature learning ability of neural networks, and has been widely and successfully applied in the fields of machine translation [25], visual understanding [17], [18], content generation [26], multi-agent system [27], *etc*. As the content of this paper is mainly inspired by the attention mechanism in the field of computer vision, so we briefly review the research status of attention mechanism in this field, which can be roughly divided into the following three categories: (1) Channel attention, which is designed to analyze the significance of each feature map channel. As a typical study, Squeeze and Excitation Attention (*SEAttention*) [28] proposed a general module for channel-wise feature recalibration. Their experimental results show that this module can be combined with ResNet [29] and other networks to improve the performance of image classification task. Subsequently, literature [30] improved the computational efficiency of the attention module to address the shortcomings of *SEAttention*. (2) Spatial attention, which focuses on evaluating attention scores from spatial patches of the feature maps rather than the channels. *CBAM* [31] used spatial attention to mine the correlation between spatial blocks of input image and compensate channel attention. *PiCANet* [32] evaluated the attention score for each pixel using its contextual information to enhance the visual representation of the image. (3) Self attention is an attention mechanism that encodes the relationships between all the input entities. Its main method is to first copy the input into three copies, called *key*, *query* and *value*, then calculate the similarity between *key* and *query*. And finally the *value* is adjusted based on the similarity. *Transformer* [33] is the representative model of the self attention and it is the cornerstone of current large scale AI models [34]. Most recently, attention has been introduced into deep networks for RF fingerprinting of real-world Bluetooth [35]. However, in their approach, the attention module and the convolution operation are completely separated. Inspired by the above studies, this paper introduces the attention mechanism into the task of RF fingerprinting, and designs an efficient dual attention block for this task. This block is able to selectively enhance important features and suppress irrelevant or noisy features, thus effectively improving the performance of deep convolutional neural networks in RF fingerprinting.

III. METHODOLOGY

In this section, we first revisit the channel attention and spatial attention in neural networks. Then we introduce our proposed dual attention convolution which combines channel attention, spatial attention and convolution operation. We also detail how to plug our module into existing CNNs. Finally, we introduce the dataset and its preprocessing and augmentation methods which will be used in our experiments.

A. PRELIMINARY

Let the input and output of one Conv1d block which has C_{out} convolution kernels be $x \in \mathbb{R}^{C_{in}\times L_{in}}$ and $x \in \mathbb{R}^{C_{out}\times L_{out}}$ respectively. In the field of deep learning, the C_{in} and C_{out} are usually called *channel*[2] dimension of the feature maps, and L_{in} and L_{out} are named *spatial* dimension. Extensive computer vision studies [36] have demonstrated the representation learning ability of convolutional neural networks fundamentally lies in their convolution kernels, each of which can be considered as a feature learner. These convolution kernels can be learned driven by the training data. However, traditional convolutional neural networks treat all convolution kernels in the same layer equally, assuming that features learned by different kernels are equally important for recognition. This approach is at odds with the way humans process information during recognition, where certain features may be more salient or relevant than others. Thus channel attention [28] emerged to evaluate the importance of different feature channels and modulate the features learned by different convolution kernels. Similarly, spatial attention [31] focuses on evaluating attention scores from spatial patches of the feature maps. The attention mechanism can be implemented in a variety of ways. At present, most of the attention modules in neural networks are implemented by neural networks. For instance, the channel attention score $\mathbf{A_c} \in \mathbb{R}^{C_{out}\times 1}$ and spatial attention score $\mathbf{A_s} \in \mathbb{R}^{1\times L_{out}}$ can be calculated based on the input:

$$\mathbf{A_c} = \mathcal{F}_{\phi}(x_{in}) \tag{1}$$

$$\mathbf{A_s} = \mathcal{F}_{\theta}(x_{in}) \tag{2}$$

where $\mathcal{F}_{\phi}$ is the neural network with parameter ϕ to estimate channel attention score, while $\mathcal{F}_{\theta}$ is the neural network with parameter θ to calculate spatial attention score.

B. DUAL ATTENTION CONVOLUTION

Results from previous studies have demonstrated that manually extracting and combining various RF fingerprints for particular radios can significantly boost performance [37]. Motivated by the fact that the attention mechanism can automatically perform feature selection, boosting the representation learning ability of convolutional networks [17], [18], we design a novel and flexible Dual Attention Convolution (DAConv) module to boost convolutional models for RF fingerprinting. Figure 1 shows the architecture of the proposed module. It contains the following components:

1) FEATURE CONVOLUTION BRANCH (FCB)

Since we extract features from raw IQ samples, the input of FCB is two-dimensional, denoted as $x \in \mathbb{R}^{C_{in}\times L_{in}}$. Thus we adopt one-dimensional convolution (*Conv1d*) in this branch, to learn a set of filters that capture important temporal patterns and variations in the signal. FCB is composed of C_{out} convolution kernels to extract RF features $\widetilde{x} \in \mathbb{R}^{C_{out}\times L_{out}}$

[2]The *"channel"* here has nothing to do with the concept of *"channel"* models in wireless communication.

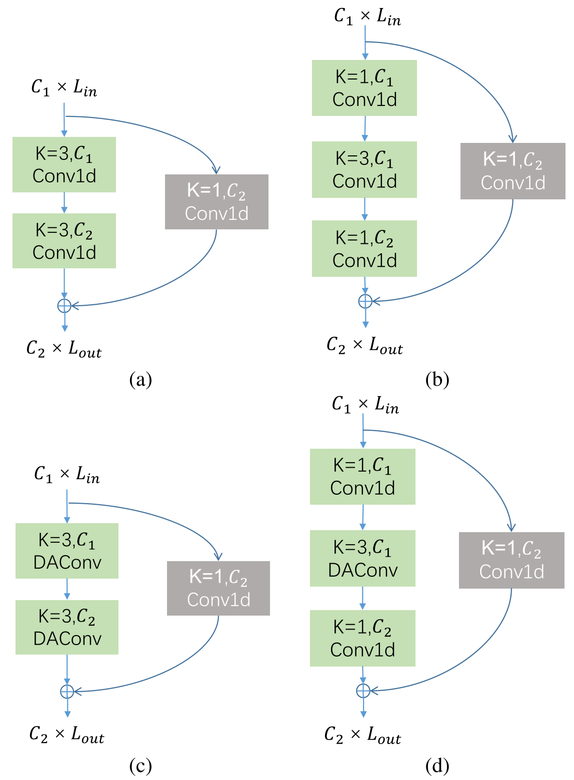


FIGURE 2. Building blocks: (a) BasicBlock. (b) Bottleneck. (c) DAConv-BasicBlock. (d) DAConv-Bottleneck. *K* denotes the kernel size, while C_1 and C_2 are the number of kernels. The *Conv1d* in grey box is optional.

from input, where L_{out} is the dimension of each output feature channel.

2) CHANNEL ATTENTION BRANCH (CAB)

It is designed to learn the attention map over channel dimension. Thus feature maps extracted by FCB will be assigned different weights. As shown in the green dashed box of Figure 1, CAB is composed of *fully connected layers* and several activation layers. The *sigmoid* layer transforms the output of the second fully connected layer into channel attention map $\mathbf{A_c} = \{\alpha_1, \ldots, \alpha_{C_{out}}\}^T \in \mathbb{R}^{C_{out} \times 1}$, in which:

$$\alpha_i = \sigma(\mathbf{W} \cdot Tanh(x \cdot \mathbf{U}^T))_i, i = 1, \ldots, C_{out} \tag{3}$$

where $\sigma(\cdot)$ denotes sigmoid function, $\mathbf{U} \in \mathbb{R}^{C_{out} \times L_{in}}$ and $\mathbf{W} \in \mathbb{R}^{1 \times C_{in}}$ are the weights of the fully connected layers respectively.

3) SPATIAL ATTENTION BRANCH (SAB)

Aiming to predict the spatial attention map, this branch incorporates a *downsampling - upsampling* structure. *Conv1d - ReLu -Maxpool* is the cascaded operation for *downsampling*, while the *upsampling* is composed of *deconv - ReLu - upsample*. This enables us to recover the original length of input signal while analyzing the spatial correlations. Then, a *sigmoid* activation layer is used to predict the spatial attention map $\mathbf{A_s} = \{\beta_1, \ldots, \beta_{L_{out}}\} \in \mathbb{R}^{1 \times L_{out}}$, shared by each channel of feature map.

4) FUSION

As shown in the purple dashed box of Figure 1, the feature maps learned by FCB are modulated by channel attention and spatial attention as follows:

$$x \leftarrow \tilde{x} + \tilde{x} \odot \mathbf{A_c}.repeat(1, L_{out}) + \tilde{x} \odot \mathbf{A_s}.repeat(C_{out}, 1) \tag{4}$$

where the first term is the feature maps learned by FCB, the second term denotes the features modulated by channel attention while the last term denotes the features modulated by spatial attention. $\odot$ denotes element-wise multiplication. $\mathbf{A_c}.repeat(1, L_{out})$ means copying tensor $\mathbf{A_c}$ along its second dimension L_{out} times which can be easily implemented using PyTorch [38] function *torch.Tensor.repeat*().

C. DEEP ATTENTION CONVOLUTIONAL NEURAL NETWORKS

From equation 4, we can see that the attention branches do not change the size of the extracted feature map. This characteristic allows our DAConv module to replace the convolutional layers in various types of convolutional neural networks and transform them into deep attention networks with little effort. For instance, our DAConv module can be plugged into ResNet [29] and transforms it to be DAConv-ResNet. Table 1 illustrates the architectures of ResNet34, ResNet50, DAConv-ResNet34 and DAConv-ResNet50. We can see that the overall architecture is not changed. Only the *Conv1d* module and build blocks are replaced or reformed with our DAConv modules. Figure 2 gives the detailed architectures of BasicBlock, Bottleneck, DAConv-BasicBlock and DAConv-Bottleneck which are the cornerstones of ResNet and DAConv-ResNet. Similarly, other deep convolutional neural networks, such as AlexNet [39], can also be transformed into deep attention networks by our DAConv. We will give their performance in section IV.

D. DATA COLLECTION

1) USRP OFDM

10 USRP X310 SDRs are used as transmitters and 1 USRP X310 SDR is used as receiver to sample IQ data. All transmitters, as depicted in Figure 3 (a), employ the *"gr-IEEE802-11"*[3] module based on GNURadio to produce baseband signals compliant with the IEEE 802.11a standard. The OFDM frames of the 10 transmitters have the same Short Training Sequence and Long Training Sequence. All the data in payload is random. Therefore, there is no field in the signal frame that can indicate the transmitter's identity. As shown in Figure 3 (b), the receiver is fixed, and each transmitter can be placed at location A or B. Thus we can collect IQ data from the same transmitter under different channel states. Finally, we collect 100 transmissions from each transmitter. Table 2 gives more details of the collected USRP OFDM dataset

[3]The source code of *"gr-IEEE802-11"* is available at: https://github.com/bastibl/gr-ieee802-11

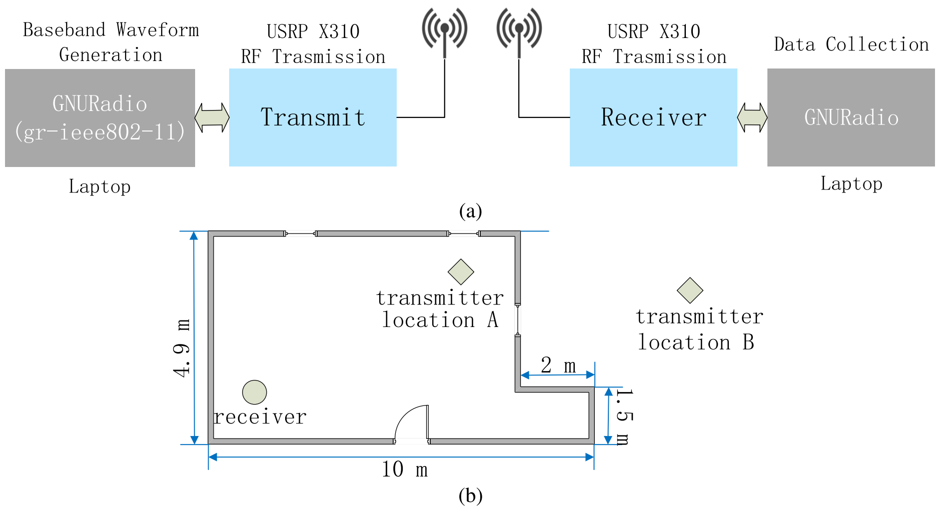


FIGURE 3. (a) Schematic diagram of signal collection. We generate baseband waveform using GNURadio software and transmit it using USRP X310 device. We also use another USRP X310 to receive the radio signals and store them in laptop. (b) Layout of signal acquisition experimental enviroment. The receiver is fixed at the left-bottom corner, while the transmitter can be located at location A or B to simulate different channel states.

TABLE 1. Architectures of our ResNet. Building blocks including BasicBlock, Bottleneck, DAConv-BasicBlock and DAConv-Bottleneck are shown in Figure 2. C is the number of kernels in conv1 layer while K denotes the kernel size. C_1 and C_2 are the number of kernels in the building blocks. *Block* × *N* denotes N same blocks are stacked. With the raw IQ data $x \in \mathbb{R}^{2\times512}$ as input, we give the output size of each convolution stage in the '*output size*' column.

layer name	output size	ResNet34	ResNet50	DAConv-ResNet34	DAConv-ResNet50
conv1	64×256	K=7, C=64, stride=2, Conv1d		K=7, C=64, stride=2, DAConv	
conv2_x	$C_2 \times 128$	1×3 max pool, stride=2			
		$C_1 = C_2 = 64$ BasicBlock ×3	$C_1 = 64, C_2 = 256$ Bottleneck ×3	$C_1 = C_2 = 64$ DAConv-BasicBlock×3	$C_1 = 64, C_2 = 256$ DAConv-Bottleneck ×3
conv3_x	$C_2 \times 64$	$C_1 = C_2 = 128$ BasicBlock ×4	$C_1 = 128, C_2 = 512$ Bottleneck ×4	$C_1 = C_2 = 128$ DAConv-BasicBlock×4	$C_1 = 128, C_2 = 512$ DAConv-Bottleneck ×4
conv4_x	$C_2 \times 32$	$C_1 = C_2 = 256$ BasicBlock ×6	$C_1 = 256, C_2 = 1024$ Bottleneck ×6	$C_1 = C_2 = 256$ DAConv-BasicBlock×6	$C_1 = 256, C_2 = 1024$ DAConv-Bottleneck ×6
conv5_x	$C_2 \times 16$	$C_1 = C_2 = 512$ BasicBlock ×3	$C_1 = 512, C_2 = 2048$ Bottleneck ×3	$C_1 = C_2 = 512$ DAConv-BasicBlock×3	$C_1 = 512, C_2 = 2048$ DAConv-Bottleneck ×3
	classes	average pool, 512-d fc, softmax			

TABLE 2. Summary of collected USRP OFDM signals.

Subset	Location	Date	Transmissions	Samples
1	A	Jan 6, 2021	10×100	87594
2	B	Jan 25, 2021	10×100	83860

and we also display several signal samples of two SDR transmitters at different location in Figure 4. The sampling rate is 5 M/s.

E. DATA PREPROCESSING AND AUGMENTATION

Numerous studies have shown that data augmentation is a successful method for enhancing the generalization capacity of deep learning models. The task of RF fingerprinting can also use this idea. We have developed the following data augmentation techniques for this study:

1) NORMALIZATION

As shown in Figure 4, signal samples from the same target that are gathered at different distances have varying amplitudes. Such amplitude characteristics should not be used as a basis for RF fingerprinting. Thus it is required to normalize the amplitude and the DC component is also need to be removed.

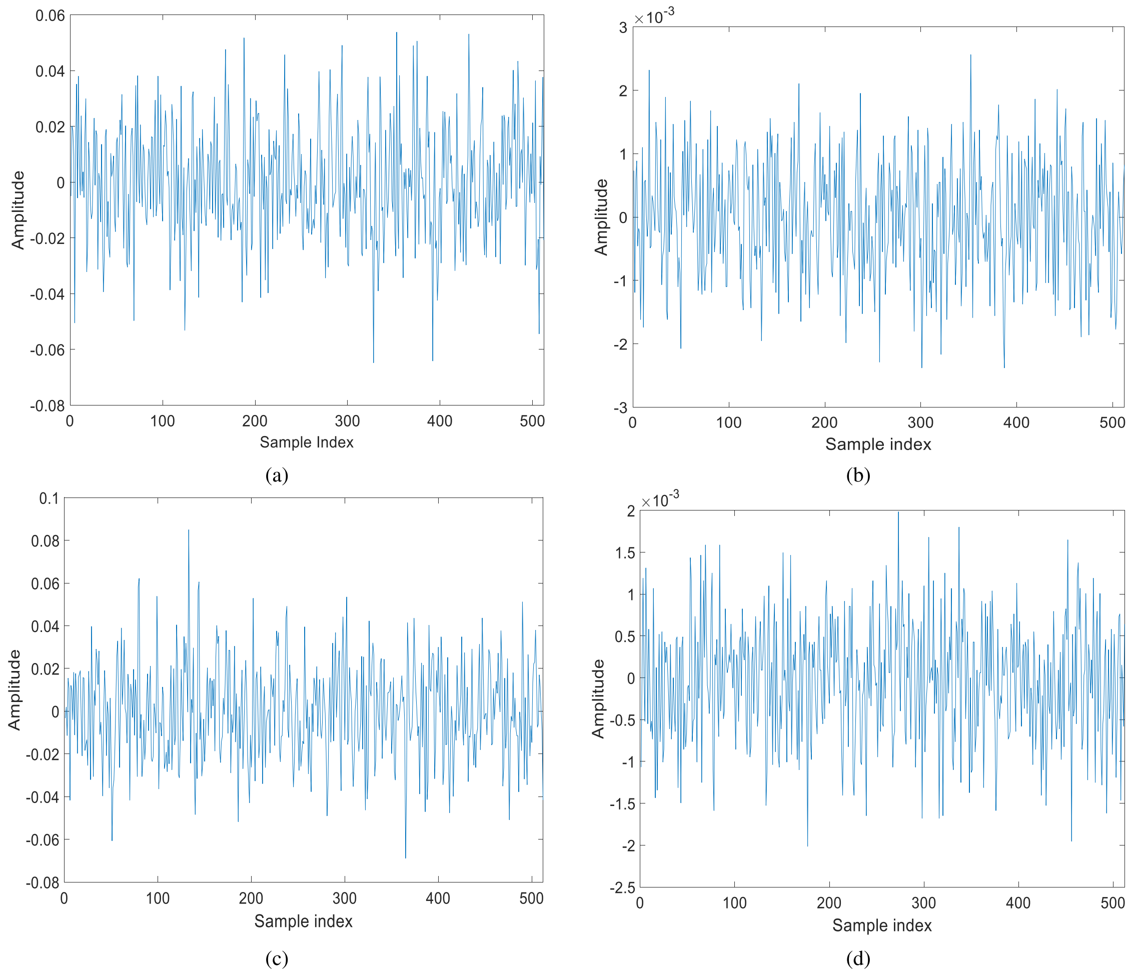


FIGURE 4. **Visualization of USPR OFDM signals: (a) Signal of USRP X310 #1 collected at location A. (b) Signal of USRP X310 #1 collected at location B. (c) Signal of USRP X310 #2 collected at location A. (d) Signal of USRP X310 #2 collected at location B. The sampling rate is 5 M/s.**

2) RANDOM SEGMENTATION

Since the collected transmission's length is not fixed. Thus, we need segment these transmissions into samples with fixed size to satisfy the specific requirements of the neural networks. To obtain samples $s \in \mathbb{R}^{2\times 512}$, a window of length 512 is used to slide on the transmission, and its sliding step is a random value drawn from uniform distribution $U(1, 256)$.

3) RANDOM NOISE

In the practical RF fingerprinting, it is impossible to collect signal samples with all possible SNRs to train the model. And the SNR of the signal to be identified is often changed. This variation of SNRs will degrade model's performance [10]. Therefore, in every training iteration, Gaussian white noises with random intensity are superimposed on the training samples. To be Specific, let $X = \{x_1, x_2, \ldots, x_N\}$, $x_i \in \mathbb{R}^{2\times 512}$ denotes the segmented samples. We add Gaussian white noises with random intensity to each x_i:

$$x_i \leftarrow \{x_i + n_1, x_i + n_2, \ldots, x_i + n_B\} \tag{5}$$

where n_i is the Gaussian white noise. In our following experiments, we set the $SNR_j = 20 \times \log(\frac{x_i}{n_j}) \in [0, 30]$ and $B = 5$.

IV. EXPERIMENTS

A. IMPLEMENTATION DETAILS

Since we have collected and preprocessed all the signal samples, now they can be used to train our models. We partition all signal transmissions in each dataset into a training set and a test set, with a ratio of 6:4. The transmissions in training set are normalized and random noise are added, thus increasing the diversity of training samples. The *MAX_EPOCH* in Algorithm 1 is set to 20 for all the models. We set the training batch size to 512, that is, 512 segmented samples are randomly selected and superimposed with Gaussian white noises with random intensities. We use Kaiming initialization [40] before training begins. This initialization method is designed to set the initial values of the model's parameters in a way that promotes learning efficiency during training. Then we use Adam optimizer [41] to optimize models' parameters. The learning rate is initialized to 2×10^{-4}, and warm-up trick is also adopted. To prevent overfitting during training, we add a Dropout layer (with a rate of 0.5) [42] after each fully connected layer in our model. Additionally, we apply $L2$ regularization to all of the network's learnable parameters. These techniques help to prevent the model from memorizing the training data and instead encourage it to learn more generalizable features that can be applied to new and unseen data. In training stage, the networks are optimized to minimize the following cross entropy loss function:

$$\mathcal{L}(\Theta) = -[\frac{1}{M}\sum_{i=1}^{B} y_i log(\hat{y}_i) + (1-y_i)log(1-\hat{y}_i)] \tag{6}$$

where M denotes the batch size, Θ is the set of learnable parameters. y_i is the ground-truth label of sample x_i, while $\hat{y}_i$ is the neural network's prediction. We have implemented our models using PyTorch.[4] We give the details about training and inferring of deep attention networks in Algorithm 1. It is worth to note that we do not add noise to test samples in inferring stage. All the following experiments are conducted on a computer with Intel Core i9 CPU and one single Nvidia RTX 3090 GPU.

Algorithm 1 Training and Inferring of Deep Attention Convolutional Neural Networks

Training of deep attention convolutional neural networks:
Randomly Initialize network weights Θ
for *epoch = 1 to MAX_EPOCHS* **do**
  for *iteration = 1 to STEPS* **do**
    Sample a batch of signal segments
    $X = \{x_1, x_2, \ldots, x_B\}$ and add random noise:
    $x_i \leftarrow \{x_i + n_1, x_i + n_2, \ldots, x_i + n_B\}$
    Input batch into the network and **Compute** loss:
    $\mathcal{L}(\Theta)$[standard forward pass]
    Compute gradients $\nabla\mathcal{L}(\Theta)$
    Update weights using Adam:
    $\Theta^* = \text{Adam}(\nabla\mathcal{L}(\Theta))$

Inferring using deep attention convolutional neural networks:
Freeze the networks with learned weights Θ^*
Get a testing signal
Normalize the signal and input it into the network
Predict the fingerprint type of the testing signal using trained networks

B. EVALUATION METRICS

To fairly compare our work with existing approaches, we use the identification accuracy which is widely adopted as the evaluation protocol. Recognition accuracy is the most commonly used criterion in classification tasks, which represents the proportion of correctly identified test samples to the total test samples. In addition, since the RF fingerprinting models discussed in this paper are all based on deep neural networks, the model size and computational cost of deep neural networks are important criteria to measure the practicality of the models. Therefore, we introduced several common protocols including floating-point operations per second (FLOPs), number of model parameters, training time for one iteration to compare the performance of different models.

C. COMPARISON WITH STATE-OF-THE-ARTS

Since our proposed attention module is specific for RF fingerprinting models based on CNNs, we adopt several State-Of-The-Art (SOTA) networks as baselines, including:

- **ORACLE** [8] is the milestone of RF fingerprinting using CNN. ORACLE is a simple convolutioanal neural network which has only two convolutional layers and two fully-connected layers.
- **AlexNet** in RF fingerprinting is modified from AlexNet [39] which is the first convolutional neural network to make a breakthrough in image recognition, thus exploding the craze of deep learning. This model is composed of 5 convolutional layers, three map pooling layers and several fully-connected layers. For more details about AlexNet in RF fingerprinting, please refer to [10] which takes this model as baseline and gives detailed architecture.
- **ResNet** in RF fingerprinting is a model family modified from popular residual network named ResNet [29]. By adding shortcut between the convolutional layers, the residual network optimizes the gradient flow in the network, thus effectively alleviating the gradient instability problem in deep network training, making it possible to train very deep networks. According to the number of layers, ResNet has several variants including ResNet34, ResNet50. Literature [16] adopts the ResNet50 while Literature [43] use ResNet34 to fingerprint Lora devices. We have demonstrated their detailed architectures in Table 1.

We can see that all the above models are inspired by popular computer vision models after proper modification. It is worth noting that since image data generally contains three channels and each channel is a $W \times H$ matrix, *Conv2d* is generally adopted by CNN models for image recognition. However, the signal IQ data only contains two channels, and each channel is a L-length vector. Therefore, *Conv1d* is usually required for CNN model used for RF fingerprinting, except that ORACLE uses *Conv2d* since it views the IQ data as a tensor with one channel and each channel is a $2 \times L$ matrix. In order to verify the effectiveness of our proposed DAConv module, we have implemented DAConv-ResNet34, DAConv-ResNet50 which have been described in section III-C, and we also have implemented

[4]Our code can be downloaded from https://github.com/zhangweifeng1218/Adaptive_RF_Fingerprinting

TABLE 3. Comparison of our models' performance with SOTA models. We show the mean testing accuracy, number of models' learnable parameters and training time cost per iteration.

Model	Accuracy (%)	FLOPs (G)	Parameters (M)	Training time (ms)
ORACLE [8]	75.3	0.0104	0.292	45.3
AlexNet [16]	79.6	0.0506	0.302	49.5
ResNet34 [43]	92.0	0.1793	7.229	89.7
ResNet50 [10]	94.2	0.3869	16.000	115.8
DAConv-AlexNet (ours)	81.5	0.0643	0.280	53.6
DAConv-ResNet34(ours)	93.4	0.2439	7.551	93.5
DAConv-ResNet50(ours)	95.6	0.4201	16.130	117.9

DAConv-AlexNet. We carry out a comprehensive comparison experiment to compare the performances of these State-Of-The-Art (SOTA) models with our models. The experimental results, including testing accuracy, model size, FLoating point OPerations (FLOPs) and training time per iteration, are summarized in Table 3. We also show the classification confusion matrix of each model in Figure 5.

From the results, we can see that: (1) DAConv-ResNet50, which is based on our proposed DAConv module and ResNet, achieves the best recognition accuracy with proper model size and training time cost. (2) Big models which have more learnable parameters have better performance than small models. The reason is that more learnable parameters means big models have larger capacity, thus then can better fit the RF fingerprinting space. For example, the smallest model ORACLE only achieves 75.3% accuracy with 0.292 M parameters, while ResNet34 achieves 92.0% with 7.229 M parameters. (3) All the existing RF fingerprinting models based on convolutional neural networks can be boosted by inserting our DAConv module into them. For instance, the best existing model ResNet50's performance can be significantly improved from 94.2% to 95.6% by using DAConv module. The comparison study mentioned above demonstrates how the DAConv method suggested in this research may aid standard deep convolutional neural networks in learning RF fingerprints more effectively. Figure 6 shows the RF fingerprints obtained by ResNet50 and DAConv-ResNet50 to illustratively demonstrate our claim. For ease of observation, we utilize the well-known t-Distributed Stochastic Neighbor Embedding (t-SNE) tool [44] to map the RF fingerprints into three dimensions. We can see that the RF fingerprints obtained by DAConv-ResNet50 clearly exhibit an intra-class compact and inter-class separation feature, which explains why DAConv-ResNet50 has better performance than ResNet50.

D. IMPACT OF ATTENTION

In this section, we try to explore the advantage of each component in our attention module. First, we analyze the impact of channel attention and spatial attention on the performance. We design two variant blocks: (1) ResNet50-w spatial attention: which has only the spatial attention branch. (2) ResNet50-w channel attention: in which only the channel attention branch is reserved. We compare their performances with DAConv-ResNet50 and ResNet50. The results are reported in Table 4. We can see that DAConv-ResNet50 degenerates into ResNet50 if the two attention branches are removed, and the accuracy on USRP OFDM drops from 95.6% to 94.2%. After the spatial attention branch is added to ResNet50, its accuracy slightly increases from 94.2% to 94.7%. Channel attention can provide more significant performance gains, which improve the accuracy from 94.2% to 94.9%. This experiment proves that channel attention and spatial attention are beneficial to improve the accuracy of ResNet in RF fingerprinting.

TABLE 4. Effectiveness of attention branches: Accuracies of the DAConv-ResNet50 and its variants.

Model	Accuracy (%)	Parameters (M)	Training time (ms)
ResNet50	94.2	16.000	115.8
ResNet50 -w spatial attention	94.7	16.018	116.3
ResNet50 -w channel attention	94.9	16.112	117.1
DAConv-ResNet50	95.6	16.130	117.9

E. IMPACT OF DYNAMIC CHANNEL STATE

Recent studies in RF fingerprinting have shown that the dynamic channel has a great negative impact on the models' performance, which is one of the key problems facing RF fingerprint urgently. And most existing approaches adopt kinds of signal preprocessing tricks to mitigate this negative impact [10], [11], [12], [16], [45]. How to improve the robustness of the algorithms to dynamic channel through model architecture optimization is still an open problem. Are the attention modules we proposed helpful in alleviating this problem? Thus we conduct the following experiment in this section. As shown in Figure 3, we collect signals at location A to obtain subset 1 in Table 2, while the samples in subset 2 is collected at location B. Thus, the channel states are different in these two subsets. We design two scenarios: (1) Simple training: we use the samples of subset 1 to train the models, and test them using the sample of subset 2. In this scenario, test samples and training samples are drawn from different subset and they have different channel states. It is challenging for models to fingerprinting. (2) Hybrid training: similar to the experiment in section IV-C, we concatenate subset 1 and subset 2, shuffle the samples before using 60% of them to train the models and the remaining 40% to evaluate them. The distribution of training and test data have different disstributions and we visualize their distributions by utilizing the t-SNE tool in Figure 7. We can see that training and test data have similar distribution in hybrid training scenario while Figure 7b indicates different distributions in simple training scenario. The RF fingerprinting results are shown in Figure 8. From the experimental results, two observations can be made: (1) Hybrid training is an effective training method to enhance models' robustness to dynamic channel. That is because hybrid training means the models can observe samples collected under different channel states during the

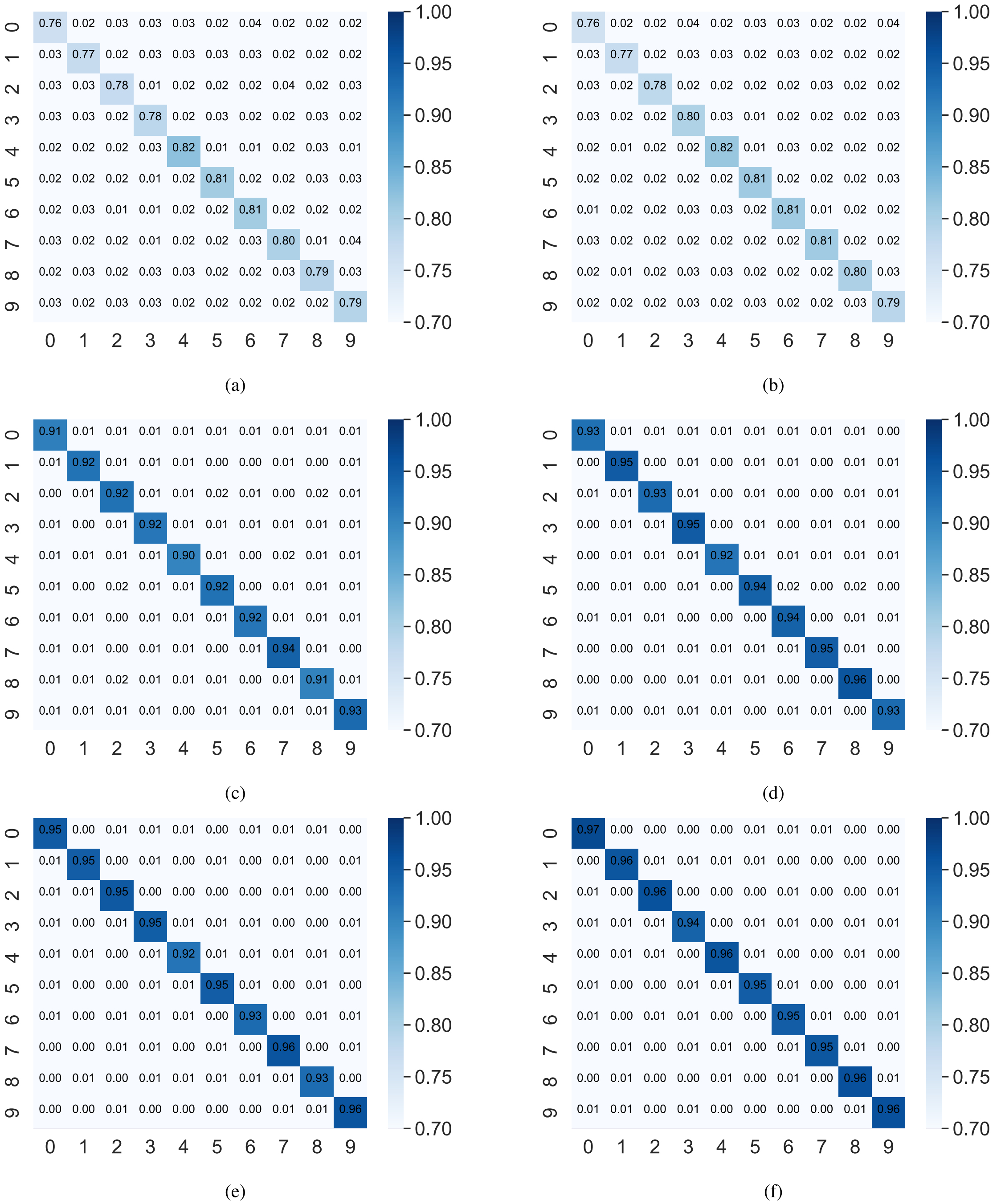
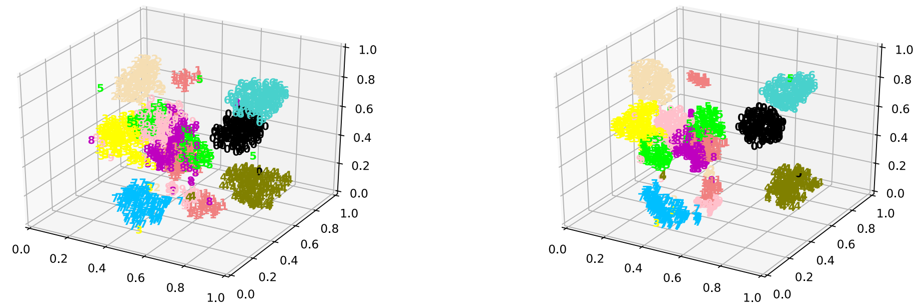


FIGURE 5. **Confusion matrix of models. (a) Result of AlexNet. (b) Result of DAConv-AlexNet. (c) Result of ResNet34. (d) Result of DAConv-ResNet34. (e) Result of ResNet50. (f) Result of DAConv-ResNet50.**

training period. However, the application scope of this training method is limited, because in practical engineering, it is difficult for us to collect the target signal under different channel states in advance. (2) The dynamic channel has brought challenges to all of the models, and their recognition accuracy has decreased significantly to varying degrees. For the traditional models such as AlexNet and VGG, when the training method is switched from the hybrid mode to the simple mode, their performances show an average decline of 12.25%. But for the models enhanced by DAConv, this decline is effectively controlled, and the performance decline is about 9.12%. This shows that the method proposed in this paper can alleviate the negative consequences of dynamic channel to a certain extent.

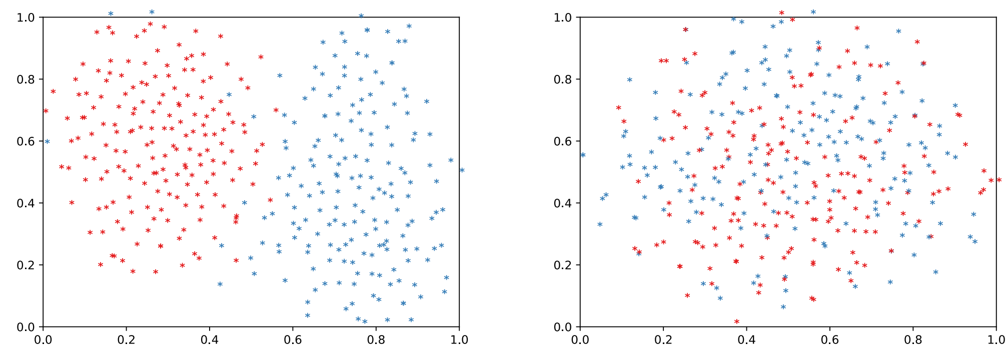
F. IMPACT OF SNR

Variation of signals' SNRs will degrade models' performance. Although we have randomly added Gaussian white



(a) (b)

FIGURE 6. (a) RF fingerprints extracted by ResNet50. (b) RF fingerprints extracted by DAConv-ResNet50. The output feature maps of the last convolutional layer are mapped into 3-dimensional space using t-SNE. Different colors represent different categories.



(a) (b)

FIGURE 7. t-SNE visualization of training and test data distribution of the two scenarios. (a) Distribution in simple training scenario. (b) Distribution in hybrid training scenario. Red points denote training set while blue points denote test set.

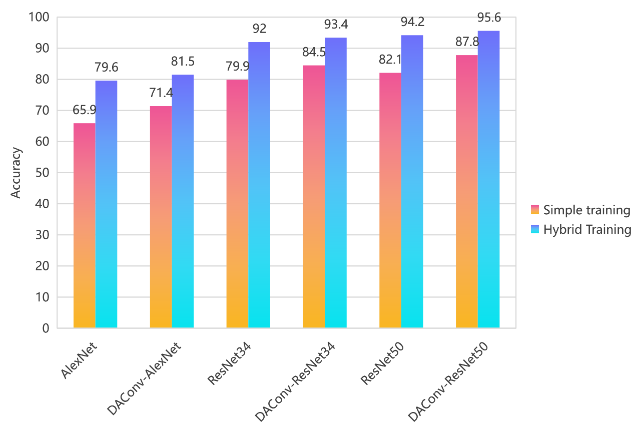


FIGURE 8. Impact of dynamic channel. The orange bars represent the results of simple training, while the blue bars represent the results of hybrid training.

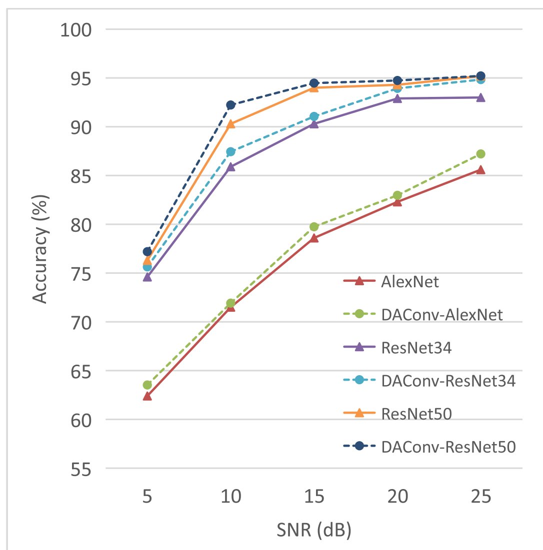


FIGURE 9. Impact of SNR. Testing accuracies of traditional models are given as dashed curves, while accuracies of DAConv enhanced models are given as solid curves.

noise to training samples in training stage to enhance the models' robustness to noise, the adverse effects of uncertain SNRs cannot be completely eliminated. Thus, we conduct a comparative experiment to explore the impact the SNR on our models and SOTA models in this section. In particular, we use the subset 1 in Table 2 to train the models in Table 3. Then in the testing stage, we collect new testing samples with

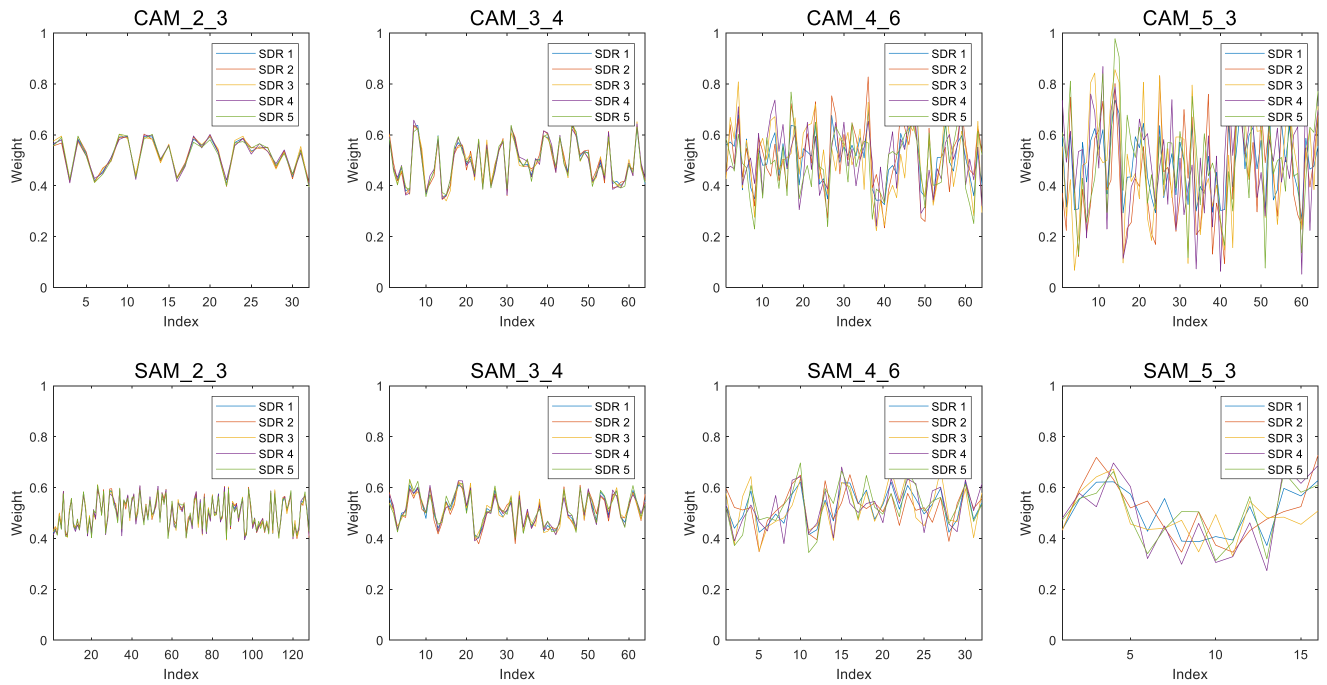


FIGURE 10. Visualization of attention weights in different convolution layers of DAConv-ResNet50. CAM_*i*_*j* and SAM_*i*_*j* respectively indicate the channel and spatial attention maps in *j*-th convolution block of *i*-th stage respectively. The top row is the weights of channel attentions and the bottom row represents the weights of spatial attentions.

various SNRs by adjusting the power amplifier in Figure 3. The testing results are shown in Figure 9. We can see that, for all models, the testing samples with higher SNR have higher recognition accuracies. Moreover, after the insertion of our DAConv module, all the models' ability of against noise has been effectively improved. Hence, our proposed convolution with attention can help models learn more discriminative and robust RF fingerprints.

G. VISUALIZATION

The previous sections demonstrate the effectiveness of our proposed attention module via comparative and ablation experiments. Although experiments have shown that deep networks such as AlexNet or ResNet can effectively improve the accuracy of RF fingerprinting by introducing our proposed attention module, due to the black box characteristics of neural networks, we do not know what knowledge the attention module has learned. The interpretability of deep neural networks is still an unsolved problem, but we can at least explore which convolutional features are strengthened or weakened by the attention module through visualizing the estimated attention weights of each attention module in the network. Hence, we select 500 signal segments from 5 devices (each device has 100 signal segments) and using the trained DAConv-ResNet50 to extract RF fingerprints. We compute the channel and spatial attention weights on average. Figure 10 visualizes the channel attention weights and spatial attention weights, denoted as CAM_i_j and SAM_i_j where i indicates i^{th} stage and j is the j^{th} convolution block in i^{th} stage. It is demonstrated that (1) The learned distributions of channel and spatial attention weights for different classes are very similar at the earlier layers. This may be caused by the fact that the earlier convolutional layers learn basic signal features which are almost similar for different classes. Such phenomenon is also observed in computer vision task [39]. (2) By focusing on CAM_5_3, we find that the learned weights are clearly different across various channels and classes, because convolution blocks in 5^{th} stage prefer to learn high-level class-specific semantic information. Thus our *DAConv* block can help CNN pay different attention to various channels in latter convolutional layers, resulting in obtaining more discriminative RF fingerprints.

V. CONCLUSION

We proposed a Dual Attention Convolution (DAConv) module, which is motivated by the popular attention mechanism used in computer vision. Our DAConv module combines spatial attention, channel attention and convolution operation. We show that DAConv module can help the convolutional layer to automatically evaluate and modulate feature maps extracted by different convolution kernels. By plugging our DAConv into deep convolutional neural networks such as AlexNet, ResNet, we construct DAConv-AlexNet, DAConv-ResNet34, and DAConv-ResNet50. The RF fingerprinting performance of all the CNNs can be improved by about 1.5 percentage points, while the increase in computation is negligible. And DAConv-ResNet50 which combined ResNet50 and our Dual Attention

Convolution module can achieve 95.6% recognition accuracy on 10 USRP X310.

Since our proposed attention module can replace the convolutional layer in CNNs, in future work, we will combine this module with other classical CNNs to verify the wide applicability of our DAConv module. At present, the CNNs designed in our experiment have a fixed limit on the length of the input signal segments. In our future work, we will also explore new CNN architecture which can take signal segments with any length as input.

REFERENCES

[1] M. Bhuiyan, M. Rahman, M. Billah, and D. Saha, "Internet of Things (IoT): A review of its enabling technologies in healthcare applications, standards protocols, security, and market opportunities," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 10474–10498, Jul. 2021.

[2] F. A. Alaba, M. Othman, I. A. T. Hashem, and F. Alotaibi, "Internet of Things security: A survey," *J. Netw. Comput. Appl.*, vol. 88, no. 15, pp. 10–28, 2021.

[3] R. Gerdes, T. E. Daniels, M. Mina, and S. Russell, "Device identification via analog signal fingerprinting: A matched filter approach," in *Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, 2006, pp. 78–88.

[4] X. Wang, P. Hao, and L. Hanzo, "Physical-layer authentication for wireless security enhancement: Current challenges and future developments," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 152–158, Jun. 2016.

[5] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proc. 14th ACM Int. Conf. Mobile Comput. Netw. (MobiCom)*, Sep. 2008, pp. 116–127.

[6] D. R. Reising, M. A. Temple, and M. J. Mendenhall, "Improved wireless security for GMSK-based devices using RF fingerprinting," *Int. J. Electron. Secur. Digit. Forensics*, vol. 3, no. 1, pp. 41–59, 2010.

[7] W. Lee, S. Y. Baek, and S. H. Kim, "Deep-learning-aided RF fingerprinting for NFC security," *IEEE Commun. Mag.*, vol. 59, no. 5, pp. 96–101, May 2021.

[8] K. Sankhe, M. Belgiovine, F. Zhou, S. Riyaz, S. Ioannidis, and K. Chowdhury, "ORACLE: Optimized radio classification through convolutional neural networks," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2019, pp. 370–378.

[9] J. Yu, A. Hu, G. Li, and L. Peng, "A robust RF fingerprinting approach using multisampling convolutional neural network," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6786–6799, Aug. 2019.

[10] T. Jian, B. C. Rendon, E. Ojuba, N. Soltani, Z. Wang, K. Sankhe, A. Gritsenko, J. Dy, K. Chowdhury, and S. Ioannidis, "Deep learning for RF fingerprinting: A massive experimental study," *IEEE Internet Things Mag.*, vol. 3, no. 1, pp. 50–57, Mar. 2020.

[11] W. Zhang, W. Zhao, X. Tan, L. Shao, and C. Ran, "Adaptive RF fingerprints fusion via dual attention convolutions," *IEEE Internet Things J.*, vol. 9, no. 24, pp. 25181–25195, Dec. 2022.

[12] A. Jagannath, J. Jagannath, and P. S. P. V. Kumar, "A comprehensive survey on radio frequency (RF) fingerprinting: Traditional approaches, deep learning, and open challenges," *Comput. Netw.*, vol. 219, pp. 109–125, Dec. 2023.

[13] O. M. Gul, M. Kulhandjian, B. Kantarci, A. Touazi, C. Ellement, and C. D'amours, "Secure industrial IoT systems via RF fingerprinting under impaired channels with interference and noise," *IEEE Access*, vol. 11, pp. 26289–26307, 2023.

[14] J. Tan, H. Wu, K.-H. Chow, and S.-H. G. Chan, "Implicit multimodal crowdsourcing for joint RF and geomagnetic fingerprinting," *IEEE Trans. Mobile Comput.*, vol. 22, no. 2, pp. 935–950, Feb. 2023.

[15] W. Wang, C. Luo, J. An, L. Gan, H. Liao, and C. Yuen, "Semi-supervised RF fingerprinting with consistency-based regularization," 2023, *arXiv:2304.14795*.

[16] K. Merchant, S. Revay, G. Stantchev, and B. Nousain, "Deep learning for RF device fingerprinting in cognitive communication networks," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 160–167, Feb. 2018.

[17] J. Yu, X. Jiang, Z. Qin, W. Zhang, Y. Hu, and Q. Wu, "Learning dual encoding model for adaptive visual understanding in visual dialogue," *IEEE Trans. Image Process.*, vol. 30, pp. 220–233, 2021.

[18] W. Zhang, J. Yu, W. Zhao, and C. Ran, "DMRFNet: Deep multimodal reasoning and fusion for visual question answering and explanation generation," *Inf. Fusion*, vol. 72, pp. 70–79, Aug. 2021.

[19] N. T. Nguyen, G. Zheng, Z. Han, and R. Zheng, "Device fingerprinting to enhance wireless security using nonparametric Bayesian method," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2011, pp. 1404–1412.

[20] L. Peng, A. Hu, J. Zhang, Y. Jiang, J. Yu, and Y. Yan, "Design of a hybrid RF fingerprint extraction and device classification scheme," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 349–360, Feb. 2019.

[21] T. D. Vo-Huu, T. D. Vo-Huu, and G. Noubir, "Fingerprinting Wi-Fi devices using software defined radios," in *Proc. 9th ACM Conf. Secur. Privacy Wireless Mobile Netw.*, Jul. 2016, pp. 3–14.

[22] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, 2016, pp. 3–14.

[23] Y. Zhang, Y. Mu, Y. Liu, T. Zhang, and Y. Qian, "Deep learning-based beamspace channel estimation in mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 12, pp. 2212–2215, Dec. 2020.

[24] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media*, vol. 8, no. 3, pp. 331–368, 2022.

[25] T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1412–1421.

[26] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.

[27] S. Iqbal and F. Sha, "Actor-attention-critic for multi-agent reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 2961–2970.

[28] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[30] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.

[31] S. Woo, J. Park, and J.-Y. Lee, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[32] N. Liu, J. Han, and M.-H. Yang, "PiCANet: Pixel-wise contextual attention learning for accurate saliency detection," *IEEE Trans. Image Process.*, vol. 29, pp. 6438–6451, 2020.

[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[34] R. Wang, D. Chen, Z. Wu, Y. Chen, X. Dai, M. Liu, Y.-G. Jiang, L. Zhou, and L. Yuan, "BEVT: BERT pretraining of video transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 14713–14723.

[35] A. Jagannath and J. Jagannath, "Embedding-assisted attentional deep learning for real-world RF fingerprinting of Bluetooth," *IEEE Trans. Cogn. Commun. Netw.*, vol. 9, no. 4, pp. 940–949, 2023.

[36] M. Hassanin, S. Anwar, I. Radwan, F. S. Khan, and A. Mian, "Visual attention methods in deep learning: An in-depth survey," 2022, *arXiv:2204.07756*.

[37] T. J. Bihl, K. W. Bauer, and M. A. Temple, "Feature selection for RF fingerprinting with multiple discriminant analysis and using ZigBee device emissions," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1862–1874, Aug. 2016.

[38] Facebook. *Pytorch*. Accessed: 2017. [Online]. Available: https://pytorch.org/

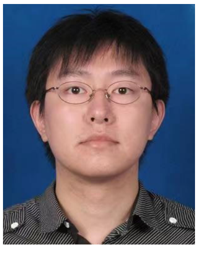
[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural network," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1106–1114.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[41] P. Diederik and B. Jimmy, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.

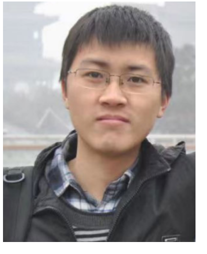
[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 51–60, 2015.
[43] G. Shen, J. Zhang, A. Marshall, and J. R. Cavallaro, "Towards scalable and channel-robust radio frequency fingerprint identification for LoRa," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 774–787, 2022.
[44] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 1, pp. 2579–2605, 2008.
[45] A. Weinand, M. Karrenbauer, R. Sattiraju, and H. D. Schotten, "Application of machine learning for channel based message authentication in mission critical machine type communication," 2017, *arXiv:1711.05088*.

HANQING GU received the B.S. degree in electronic engineering from the University of Electronic Science and Technology of China (UESTC). He is currently pursuing the Ph.D. degree with Zhejiang University. His current research interests include digital signal processing and intelligence algorithms and applications.



LISHENG SU received the B.S. and M.S. degrees in electronic science and technology from the University of Electronic Science and Technology of China (UESTC), in 2009, where he is currently pursuing the Ph.D. degree. His research interests include digital signal processing and intelligence algorithms and applications.

WEIFENG ZHANG received the B.S. degree in electronic information engineering from the Beijing University of Technology, in 2009, the M.S. degree in pattern recognition from Beihang University, in 2012, and the Ph.D. degree in computer science from Hangzhou Dianzi University, in 2019. He was a Senior Engineer with the Science and Technology on Communication Information Security Control Laboratory, from 2012 to 2018. Currently, he is an Associate Professor with Jiaxing University. His work has published in prestigious journals, such as IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON IMAGE PROCESSING, *Pattern Recognition*, and *Information Fusion*. His research interests include machine learning and its application on multimodal intelligence and communication.



CHUAN RAN received the B.S. degree in electronic information engineering from the Beijing University of Technology, in 2009, and the M.S. degree in computer science from the Beijing Institute of Technology, in 2012. He was a Senior Engineer with IBM, in 2012. Currently, he is a Senior Software Engineer with IBM Research Triangle Park. His research interests include cloud computing, artificial intelligence, and its engineering applications.