Multi-Channel Attentive Feature Fusion for Radio Frequency Fingerprinting

Yuan Zeng, *Member, IEEE*, Yi Gong, *Senior Member, IEEE*, Jiawei Liu, *Student Member, IEEE*, Shangao Lin, Zidong Han, Ruoxiao Cao, Kaibin Huang, *Fellow, IEEE*, and Khaled B. Letaief, *Fellow, IEEE*,

Abstract—Radio frequency (RF) fingerprinting is a promising device authentication technique for securing the Internet of Things. It exploits the intrinsic and unique hardware impairments of the transmitters for device identification. Recently, due to the superior performance of deep learning (DL)-based classification models on real-world datasets, DL networks have been explored for RF fingerprinting. Most existing DL-based RF fingerprinting models use a single representation of radio signals as the input, while the multi-channel input model can leverage information from different representations of radio signals and improve the identification accuracy of RF fingerprints. In this work, we propose a multi-channel attentive feature fusion (McAFF) method for RF fingerprinting. It utilizes multi-channel neural features extracted from multiple representations of radio signals, including in-phase and quadrature (IQ) samples, carrier frequency offsets, fast Fourier transform coefficients and short-time Fourier transform coefficients. The features extracted from different channels are fused adaptively using a shared attention module, where the weights of neural features are learned during the model training. In addition, we design a signal identification module using a convolution-based ResNeXt block to map the fused features to device identities. To evaluate the identification performance of the proposed method, we construct a WiFi dataset, named WiFi Device Identification (WFDI), using commercial WiFi end-devices as the transmitters and a Universal Software Radio Peripheral (USRP) platform as the receiver. Experimental results show that the proposed McAFF method significantly outperforms the single-channel-based as well as the existing DL-based RF fingerprinting methods in terms of identification accuracy and robustness.

Index Terms—Radio frequency fingerprinting, feature fusion, convolutional neural network, attention mechanism

I. INTRODUCTION

With the recent advancement in electronic devices and beyond fifth-generation (B5G) mobile systems, Internet of Things (IoT) technology has been widely developed that plays an increasingly important role in our daily lives, such as smart healthcare, smart cities and intelligent transportation [1]–[3]. The evolution of IoT tends to enable ubiquitous connectivity, where billions of tiny embedded wireless devices, enabling sensing, computing, and communications, are deployed everywhere. However, wireless security is a critical challenge for IoT applications, since mobile devices in wireless communication networks are vulnerable to malicious attacks when operating in an open environment. Impersonation attack is one of the most common and threatening malicious attacks in wireless communication networks. Device authentication, which validates whether the devices or users are legitimate, can protect the wireless devices from impersonation attacks and improve the security of IoT [4]. Traditional authentication methods employed software addresses, such as Internet Protocol (IP) or Media Access Control (MAC), as identity and used classical cryptography-based authentication techniques [5]. Cryptography algorithms usually rely on complicated mathematical operations or protocols [6]. However, due to limitations in power consumption and computation resources, such complex cryptography algorithms may be undesirable in many IoT applications, such as smart cities [7] and intelligent industries [8]. An alternative scheme for wireless security is radio frequency (RF) fingerprinting, which uses waveform-level imperfections imposed by the RF circuit to obtain a fingerprint of the wireless device [9]. The imperfections normally include in-phase and quadrature (IQ) imbalance, phase offset, frequency offset, sampling offset, and phase noise [10], which are usually unique and can hardly be imitated by adversarial devices.

RF fingerprinting is a standard pattern recognition problem and usually consists of two stages: learning and inference. During the learning phase, signals are gathered from end-devices and processed to learn a classification model at a server. After that, the server uses the trained model to authenticate end-devices during the inference phase. Compared with conventional cryptography-based security schemes, RF fingerprinting, by operating at the physical layer, does not impose any additional computational burden and power consumption on end-devices. On the other hand, the receiver/server that is usually equipped with sufficient computational resources is capable of gathering signals from multiple end-devices and performing signal learning and inference. Such arrangement of computational resources and capabilities among the receiver/server and end-devices is particularly desirable for many IoT applications, since most end-devices are low-cost with limited computational and energy resources.

Y. Zeng is with the Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology (SUSTech), Shenzhen, 518055, P. R. China (e-mail: zengy3@sustech.edu.cn).

Y. Gong, S. Lin and Z. Han are with the Department of Electrical and Electronic Engineering, SUSTech, Shenzhen, P. R. China (e-mail: gongy@sustech.edu.cn, linsa2019@mail.sustech.edu.cn, hanzd@sustech.edu.cn).

J. Liu and K. Huang are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: liujw@eee.hku.hk, huangkb@eee.hku.hk).

R. Cao and K. B. Letaief are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong (e-mail: rcaoah@connect.ust.hk, eekhaled@ust.hk).

Traditional RF fingerprinting methods are based on carefully hand-crafting specialized feature extractors and classifiers. However, manually extracting handcrafted features requires prior comprehensive knowledge of the communication system, such as channel state and communication protocol. In addition, it is difficult to extract features of hardware imperfection accurately, since the hardware imperfections are interrelated. By performing feature extraction and signal classification at two separate stages, traditional RF fingerprinting methods may work well when the extracted features are distinguishable for classifiers. The identification performance of classifiers is highly dependent on the quality of extracted features, and an efficient feature extraction method for one scenario might easily fail in another.

Recently, deep learning (DL)-based RF fingerprinting methods have been proposed to fingerprint radios through learning of the hardware impairments [9], [11]. Different neural network architectures, such as convolutional neural networks (CNNs) [12], recurrent neural networks (RNNs) [13] and long short-term memory (LSTM) [14], have been explored for automatic RF fingerprinting and have shown better RF device identification performance than traditional RF fingerprinting methods. Early DL-based methods mainly used the received IQ samples as the input of the DL network [15]–[17]. More recently, a few DL-based RF fingerprinting methods explored the specific characteristics of radio signals, such as bispectrum [16], Hilbert spectrum [17] and differential constellation trace figure [18], for RF fingerprinting. The frequency features of radio signals have also been explored for assisting radio signal identification in recent studies [15], [19], [20], where it is shown that the fast Fourier transform (FFT) and short-time Fourier transform (STFT) coefficients are effective representations of the radio signals, since FFT extracts frequency features of the radio signals and STFT reflects how the instantaneous frequency varies over time. Utilizing multiple signal representations as the input data may further improve the performance of RF fingerprinting models, since different signal representations can provide different views of the radio signals. In [21], a hybrid classification framework was proposed to weight multiple features of radio signals with pre-calculated weights for RF fingerprinting. Obtaining an efficient fusion of different representations is a complex task, and pre-calculated weights may produce sub-optimal identification results in practical applications. Therefore, we face the problem of designing an effective learning method to learn multi-channel fusion of different signal representations in an end-to-end manner for better RF device identification.

In general, training and test data are collected at different time and locations. However, hardware impairments may change across days due to temperature or voltage oscillations, and wireless channels may become unstable due to time-varying fading and noise. Noise-robustness is a general and crucial problem in wireless communication systems. Although a lot of progress has been made in RF fingerprinting on radio signals under high signal-to-noise ratio (SNR) conditions, the identification performance still drops sharply in low SNR conditions, such as electromagnetic inference and environment noises. A few recent DL-based methods [11],

[22] proposed model-based augmentation strategies to improve the robustness of the DL-based RF fingerprinting models and validated the proposed strategies using simulated RF devices. Though these methods provided possible solutions to improving identification accuracy when real RF devices with similar nonlinear characteristics are used, identifying the fingerprints of real RF devices from noisy radio signals is still a challenge task.

To address the problem of improving RF fingerprint identification performance in noisy environments with varying levels of additive noise, this paper presents a multi-channel attentive feature fusion (McAFF) model to adaptively fuse multi-channel neural features for RF fingerprinting. We first represent the baseband signal by raw IQ samples, carrier frequency offset (CFO), FFT coefficients, and STFT coefficients. Then, neural features extracted from multi-channel inputs are fused in an end-to-end learning manner. In addition, a shared attention module is designed to learn weights regarding feature importance for RF fingerprinting. The entire model is trained in a supervised manner.

Although a few datasets have been constructed for evaluating DL-based RF fingerprint identification performance in recent studies [9], [15], [20], they are generally not open-source. In addition, WiFi signals in those datasets were either simulated using software or collected from non-commercial end-devices. To evaluate the performance of the proposed McAFF method for commercial WiFi end-devices, we design a RF signal acquisition system that uses a Universal Software Radio Peripheral (USRP) software defined radio (SDR) platform as the receiver and 72 commercial WiFi end-devices as the target devices. We collect WiFi signals transmitted using the standard IEEE 802.11n protocol over several experiment days and use the legacy-short training field (L-STF) sequences collected in the uplink frames to construct a dataset for RF fingerprinting, named WiFi Device Identification (WFDI). Extensive experiments are conducted to evaluate the identification performance of the proposed McAFF model. Experimental results demonstrate the effectiveness of the proposed McAFF model on WiFi device identification. The main contributions of our work are summarized as follows.

- We propose a novel McAFF model that exploits the complementary information from different representations of radio signals. It optimizes the combination of multi-channel features and utilizes a convolution layer to attend to important features, thus leading to superior RF fingerprinting performance. In contrast to fusing the multiple signal representations using fixed or pre-calculated weights, the proposed McAFF model learns to fuse multiple signal representations adaptively.
- We design a signal identification module that uses a convolution-based ResNeXt block to map the fused neural features to device identities for automatic RF fingerprinting. Moreover, we construct a new dataset with data collected from 72 commercial WiFi end-devices to evaluate the identification performance of the proposed method.
- We provide an extensive ablation study of different signal representations and analyze the rationality of the

proposed attention module through comparative experiments. Experimental results demonstrate that the proposed McAFF model outperforms the baseline models in terms of identification accuracy. In addition, we analyze the impact of different data scales and collection dates on the identification performance. Experimental results show that the proposed McAFF model is more robust against noisy radio signals and changes in data scale and collection time.

The rest of the paper is organized as follows. Related work is given in Section II. The problem statement is given in Section III. In Section IV, we briefly introduce the signal pre-processing and representation algorithms. The McAFF model is proposed in Section V. In Section VI, experiments are conducted to generate the WFDI dataset and evaluate the WiFi end-device identification performance of the proposed method. Finally, we draw our contribution in Section VII.

II. RELATED WORK

A. RF Fingerprinting

A variety of approaches have been proposed for RF fingerprinting. Early RF fingerprinting methods usually perform feature extraction and classification separately. Existing hand-tailored feature extraction techniques can be roughly divided into three categories: transient-based (TB) techniques, spectrum-based (SB) techniques, and modulation-based (MB) techniques. TB methods extract the signal variations in the trun-on/off transient, such as the transient signal envelope [23] and phase offset [24]. These methods use time domain signal processing algorithms to extract features from received signals. SB methods extract frequency domain features, such as power spectrum density (PSD) [25], [26], Hilbert spectrum [17], and time-frequency statistics [27]. MB features including IQ offset [28], clock skew [29], [30], CFO [31], [32], sampling frequency offset [33], etc., can be extracted from the received baseband signal. During the classification process, the authenticator feeds the extracted features to train the classifier and then infers the device identity. Typical classification algorithms include support vector machine (SVM) [28] and K-nearest neighbor (KNN) [34]. The identification performance of these classifiers relies heavily on feature extraction, and hand-tailored feature extraction techniques require comprehensive knowledge of communication technologies and protocols, which limits their practical applications.

In the past few years, DL, which utilizes a cascade of multiple layers of nonlinear processing units for feature extraction and transformation, has achieved great success in image classification and speech recognition. Recently, it has also been used for physical-layer communications, including channel estimation [35], [36], modulation recognition [19], [37], [38] and RF fingerprinting [20], [39], [40]. Compared to traditional RF fingerprinting methods, DL-based methods use deep neural networks (DNNs) to automatically extract more distinguishable and high-level features by learning a mapping strategy from training data. A CNN-based RF fingerprinting model, which operates on the error signal obtained after subtracting an estimated ideal signal from frequency-corrected data, was proposed in [12]. Riyaz et al. [11] evaluated the effectiveness of an optimized CNN architecture on RF fingerprinting and introduced artificial impairments to improve the identification accuracy. Cekic et al. [22] investigated the identification performance of a complex-valued DNN on WiFi and ADS-B signals. Sankhe et al. [9] used a 16-node USRP SDR testbed and an external database of more than 100 WiFi devices to generate a WiFi dataset and introduced a robust CNN architecture for device identification using raw IQ samples. In [41], an impairment hopping spread spectrum method using a random pseudo-noise binary sequence was proposed. Shen et al. [20] evaluated the impact of different signal representations, including spectrogram, IQ samples, and FFT coefficients on a CNN-based RF fingerprinting model, and designed a hybrid classifier to improve the RF fingerprint identification performance on LoRa devices. Shawabka et al. [15] investigated the identification performance of CNN-based RF fingerprinting models on a large-scale dataset with WiFi and ADS-B devices.

B. Multi-Channel Signal Classification

Our work is closely related to multi-channel signal classification, which utilizes multiple signal representations to provide classifiers with richer information from different perspectives. Peng et al. [21] proposed combining four modulation features, including differential constellation trace figure, CFO, modulation offset, and IQ offset, with the pre-calculated weights according to the estimated SNR, and they evaluated this feature fusion method on RF fingerprinting using a KNN classifier. In [42], a multi-channel learning framework was proposed to exploit the complementary information for automatic modulation recognition. Lin et al. [43] introduced a dual-channel spectrum fusion module that includes an original signal channel and a filtered signal channel for modulation recognition. Inspired by existing multi-channel-based classifiers, in this paper, we use four signal representations, namely IQ samples, CFO, FFT coefficients, and STFT coefficients, as the input of our RF fingerprinting model. Moreover, features extracted from these four signal representations are adaptively weighted for RF fingerprinting.

C. Attention-Based Fusion

Recently, attention mechanisms [44] that allow the network to automatically find target-related features and enhance them while diminishing other features have achieved great performance improvement on various natural language processing and computer vision tasks [45]–[47]. Fan et al. [48] proposed a deep attention-based fusion algorithm to dynamically control the weights of the spatial and spectral features. Xu et al. [49] presented a co-attention mechanism to guide the fusion of RGB and infrared multi-spectral information for semantic segmentation. In wireless communications, attention mechanisms have been investigated for DL-based modulation classification models to improve classification performance. O'Shea et al. [50] designed an attention mechanism to learn a localization network for blindly synchronizing and normalizing radio signals. Lin et al. [51] proposed a time-frequency attention mechanism to learn which channel, time, and frequency information
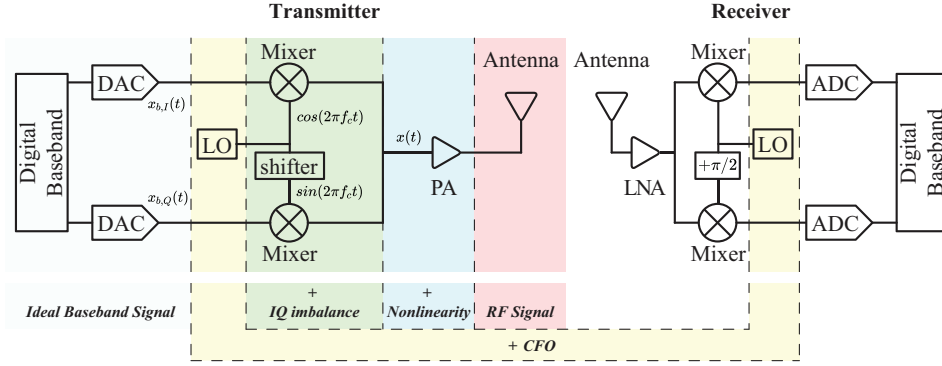
Fig. 1. The origins of signal distortions in a communication system.

is more important for modulation recognition. In this paper, we design a shared attention module to adaptively fuse neural features extracted from multi-channel inputs.

III. PROBLEM STATEMENT

In wireless communication systems, the received signal is typically down-converted to baseband, filtered, and sampled. The complex baseband samples are processed to extract the information bits. The baseband samples are not only degraded by non-ideal channel and noise but also affected by non-idealities of RF and analog parts, including sampling clock offset, power amplifier (PA), phase noise, and CFO. In this section, we introduce several classic RF hardware impairments, including IQ imbalance, CFO, and PA nonlinearity, in the form of mathematical models, as well as the system model of RF fingerprinting.

A. Mathematical Model of RF Impairments

Let $x_b(t)$ denote the transmitted baseband complex signal at time t, the in-phase (I) and quadrature (Q) signals of which are denoted by $x_{b,I}(t)$ and $x_{b,Q}(t)$, respectively. $x_b(t)$ can be expressed as

$$x_b(t) = x_{b,I}(t) + jx_{b,Q}(t) = r(t)e^{j\phi(t)}, \qquad (1)$$

where $j = \sqrt{-1}$ denotes the imaginary unit. $r(t)$ and $\phi(t)$ denote the amplitude and phase of $x_b(t)$, respectively. The ideal transmitted RF signal $x(t)$ without hardware impairment is modeled as

$$x(t) = x_{b,I}(t)\cos(2\pi f_c t) + x_{b,Q}(t)\sin(2\pi f_c t), \qquad (2)$$

where f_c is the carrier frequency. The RF signal is often distorted in a practical communication system due to the inevitable hardware imperfections, as shown in Fig. 1, where the IQ imbalance, CFO, and nonlinear distortions of the power amplifier are three typical RF impairments.

IQ Imbalance: In the RF chain of a transmitter, the I and Q baseband signals are up-converted to the cosine and sine carrier waves using two independent quadrature mixers. The cosine and sine carrier waves are generated by the local oscillator (LO) of the transmitter. The ideal sine carrier wave is a copy of the cosine carrier wave delayed by $\pi/2$. However, quadrature mixers are often impaired by gain and phase

mismatches between the parallel sections of the RF chain dealing with the I and Q signal paths. The gain mismatch causes amplitude imbalance, and phase deviation from $\pi/2$ results in phase imbalance. Assume that the gains of the I and Q branches are g_I and g_Q, respectively, and the phase error is ε_p. Then the transmitted IQ signals with IQ imbalance can be modeled as

$$g_I \cos(2\pi f_{c,t} t - \varepsilon_p/2) \text{ and } g_Q \sin(2\pi f_{c,t} t + \varepsilon_p/2), \quad (3)$$

where $f_{c,t}$ is the center frequency that depends on the LO of the transmitter.

Non-Linear Distortions of Power Amplifier: Power amplifiers are used in the RF chain to amplify the pass-band signal for transmission. In practice, the analog amplification process in PA incurs nonlinear distortions in both amplitude and phase. These effects are directly applied to the pass-band signal and can be equivalently modeled on baseband signals. An empirical baseband model of PA was presented in [52]. Let $g(t)$ and $\epsilon_\phi(t)$ denote the amplitude and phase offset of the PA output, respectively. They can be modeled as

$$g(t) = \frac{\alpha_a r(t)}{1 + \beta_a [r(t)]^2}, \qquad (4)$$

and

$$\epsilon_\phi(t) = \frac{\alpha_\phi r(t)^2}{1 + \beta_\phi [r(t)]^2}, \qquad (5)$$

where $\alpha_a, \alpha_\phi, \beta_a$ and β_ϕ are fitting parameters extracted from measurement results. The equivalent baseband signal with PA nonlinear distortion, denoted by $\tilde{x}_b(t)$, is then given by

$$\tilde{x}_b(t) = g(t)e^{j[\phi(t)+\epsilon_\phi(t)]}. \qquad (6)$$

Carrier Frequency Offset: The carrier frequencies used for up-conversion and down-conversion are generated in the transmitter and receiver using their own LOs, respectively, which can hardly oscillate at an identical frequency. Let $y(t)$ denote the ideal signal received at the receiving antenna. After down-conversion, the received baseband signal is given by

$$y_b(t) = y(t)e^{-j2\pi f_{c,r} t}, \qquad (7)$$

where $f_{c,r}$ is the local carrier frequency of the receiver. Let ϵ_f denote the carrier frequency offset between the transmitter and receiver, i.e., $f_{c,r} = f_{c,t} + \epsilon_f$. Eq. (7) can be rewritten as

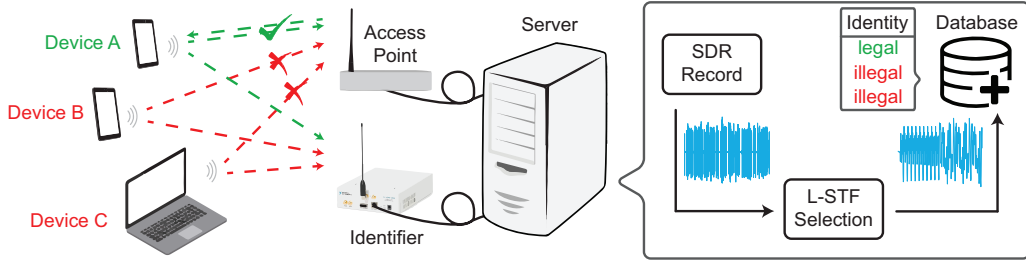$$y_b(t) = y(t)e^{-j2\pi(f_{c,t}+\epsilon_f)t}, \qquad (8)$$

Fig. 2. Illustration of a RF fingerprinting system.

and the carrier waves generated by the receiver's LO can be modeled as

$$\cos(2\pi(f_{c,t} + \varepsilon_f)t) \quad \text{and} \quad \sin(2\pi(f_{c,t} + \varepsilon_f)t). \quad (9)$$

B. RF Fingerprinting System Model

RF fingerprinting refers to the task of identifying radio devices by analyzing their transmitted RF signals and extracting RF fingerprints associated with their hardware impairments. An RF fingerprinting system is illustrated in Fig. 2. Specifically, the user device transmits data to the access point (AP) using the standard IEEE 802.11n protocol. The AP is configured to work at the legacy or mixed transmission mode such that the L-STF sequences in the uplink frames are identical. At the same time, the uplink signal is captured by a receiver equipped at the edge server. Throughout this paper, the L-STF sequence is used for signal representation and dataset generation. This work aims to identify RF devices using the received baseband signals (corresponding to the L-STF sequences) at the receiver. Given the received baseband signal $y_b(t)$, the server performs a series of signal processing and classification algorithms on $y_b(t)$ to identify the RF devices.

IV. SIGNAL PREPROCESSING AND REPRESENTATION

A discrete-time version of the baseband signal $y_b(t)$ in Eq. (8), denoted by $y(n)$, is obtained by sampling the continuous-time signal with a sampling rate f_s, associated with a sampling period $T_s = 1/f_s$. That is,

$$y(n) = y_b(t)|_{t=nT_s}. \quad (10)$$

In preprocessing, we first perform energy normalization on $y(n)$ to remove energy differences between RF devices. Then, we adopt a group of signal representation methods, including CFO estimation, FFT and STFT, to reveal the non-idealities of RF signals.

A. Energy Normalization

The energy of the received signal, despite being used as a feature in conventional classification algorithms, is susceptible to transmission distance and channel conditions. Since RF fingerprinting aims at identifying devices by learning hardware imperfections imposed by the RF circuit, rather than environment differences, we perform energy normalization using the root-mean-square of the amplitude of the signal $y(n)$. The normalized version of $y(n)$ is given by

$$\tilde{y}(n) = \frac{y(n)}{\sqrt{\frac{1}{N_s} \sum_{n=0}^{N_s-1} y^2(n)}}, \quad (11)$$

where N_s is the number of signal samples.

B. Signal Representation

IQ Samples: Let $\tilde{y}_I(n)$ and $\tilde{y}_Q(n)$ denote the in-phase and quadrature samples of $\tilde{y}(n)$, respectively. Define $\tilde{Y}_I = [\tilde{y}_I(0), \cdots, \tilde{y}_I(N_s - 1)]^T$ and $\tilde{Y}_Q = [\tilde{y}_Q(0), \cdots, \tilde{y}_Q(N_s - 1)]^T$. The IQ samples are concatenated to represent $\tilde{y}(n)$. That is, $\mathbf{R}_{IQ} = [\tilde{Y}_I, \tilde{Y}_Q]$. The size of $\mathbf{R}_{IQ}$ is $N_s \times 2$.

CFO Estimation: CFO is caused by the difference between the LOs of the transmitter and receiver, and results in the phase offset of the received baseband IQ samples, see Eq. (8). This work adopts a coarse estimation algorithm introduced in [53] to estimate CFO, that is,

$$\hat{\epsilon}_f = \frac{1}{16} \angle \left(\sum_{n=0}^{N_s-17} \tilde{y}^*(n)\tilde{y}(n+16) \right), \quad (12)$$

where $\angle(\cdot) \in [-\pi, \pi]$ denotes the angle of a complex variable, and $(\cdot)^*$ denotes the complex conjugation. The phase difference between $\tilde{y}(n)$ and $\tilde{y}(n+16)$ indicates the accumulated CFO over 16 samples. Then, we construct a vector using the phase information of the CFO estimation to represent the signal $\tilde{y}(n)$, that is,

$$\mathbf{R}_{CFO} = [\angle(\tilde{y}^*(0)\tilde{y}(16)), \cdots, \angle(\tilde{y}^*(N_s - 17)\tilde{y}(N_s - 1))]^T. \quad (13)$$

The size of $\mathbf{R}_{CFO}$ is $(N_s - 16) \times 1$. After that, we zero-pad $\mathbf{R}_{CFO}$ to expand its size to $N_s \times 1$.

FFT Coefficients: Given the signal $\tilde{y}(n)$, we use FFT to compute the discrete Fourier transform (DFT) of it, i.e.,

$$Y_{FFT}(f) = \sum_{n=0}^{N_s-1} \tilde{y}(n)e^{-j2\pi fn/N_s}, f = 0, \cdots, N_s - 1. \quad (14)$$

We use the FFT coefficients, defined by $\mathbf{R}_{FFT} = [Y_{FFT}(0), \cdots, Y_{FFT}(N_s - 1)]^T$, as a frequency-domain representation of the signal $\tilde{y}(n)$.

STFT Coefficients: STFT is a sequence of DFTs of windowed signals. Unlike FFT where DFT provides the averaged frequency information over the entire signal, STFT computes the time-localized frequency information for situations where frequency components of a signal vary over time [54]. Let
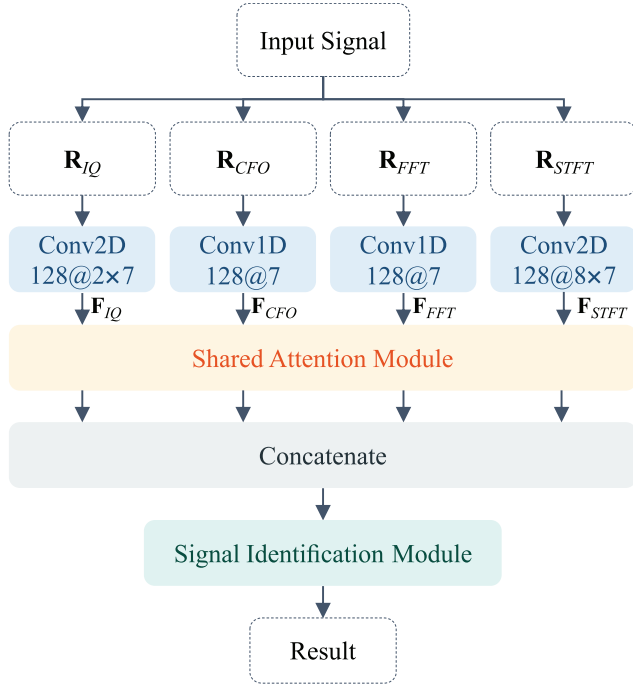
Fig. 3. Overview of the proposed McAFF model.

$w(\cdot)$ denote a window function of length J and let K be the window shift. The signal $\tilde{y}(n)$ is windowed and transformed into the frequency domain with a DFT, i.e.,

$$Y_{STFT}(m, f) = \sum_{n=mK}^{mK+J-1} \tilde{y}(n)w(n - mK)e^{-j\omega_f(n-mK)},$$
(15)

where $Y_{STFT}(m, f)$ denotes the STFT coefficients at discrete-time index m and frequency-bin index f, and $\omega_f = 2\pi f/J$ is the discrete frequency variable at frequency-bin index f. We use the STFT coefficients $Y_{STFT}(m, f), \forall m, \forall f$, as another representation of the signal $\tilde{y}(n)$, denoted as $\mathbf{R}_{STFT}$.

V. FEATURE FUSION AND SIGNAL IDENTIFICATION

This section introduces the proposed McAFF model for RF fingerprinting. The model consists of two functions: multi-channel feature fusion and signal identification. Instead of simply stacking multi-channel inputs including IQ samples, CFO, FFT coefficients and STFT coefficients, a deep attention fusion algorithm is designed to combine them deeply. It uses a shared attention module to adaptively fuse multi-channel neural features, followed by a ResNeXt block [55] based signal identification module. Given the multi-channel inputs, denoted as $\mathbf{R} = [\mathbf{R}_{IQ}; \mathbf{R}_{CFO}; \mathbf{R}_{FFT}; \mathbf{R}_{STFT}]$, the McAFF model $\mathcal{F}$ is trained to minimize the cross-entropy loss function, that is,

$$\min \mathbb{E}_{\mathbf{R},\bar{y}} \left(\mathcal{L}(\mathcal{F}(\mathbf{R}), \bar{y}) \right),$$
(16)

where $\mathbb{E}(\cdot)$ denotes the expectation operation and $\mathcal{L}(\mathcal{F}(\mathbf{R}), \bar{y})$ is the cross-entropy loss between the estimation result $\mathcal{F}(\mathbf{R})$ and the true label $\bar{y}$. Fig. 3 illustrates an overview of the proposed McAFF model.

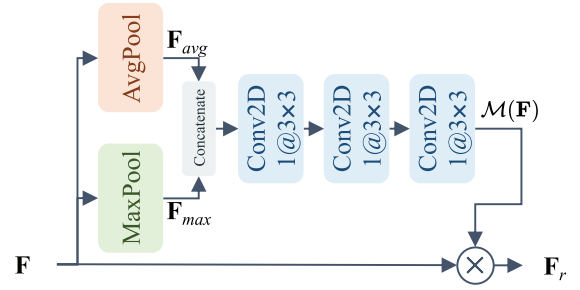Fig. 4. Structure of the proposed shared attention module.

A. Multi-Channel Feature Fusion

As shown in Fig. 3, the multi-channel inputs ($\mathbf{R}_{IQ}$, $\mathbf{R}_{CFO}$, $\mathbf{R}_{FFT}$, $\mathbf{R}_{STFT}$) are fed to four separate convolutional layers to acquire four neural feature maps, denoted as $\mathbf{F}_{IQ}$, $\mathbf{F}_{CFO}$, $\mathbf{F}_{FFT}$ and $\mathbf{F}_{STFT}$, respectively. Specifically, the settings of the convolutional layers used to process $\mathbf{R}_{CFO}$ and $\mathbf{R}_{FFT}$ are the same, i.e., one dimensional (1D) convolutional layer with 128 channels and a filter size of 1×7. A 2D convolutional layer with 128 channels and a filter of size 2×7 is used to process $\mathbf{R}_{IQ}$, and a 2D convolutional layer with 128 channels and a filter size of 8×7 is used to process $\mathbf{R}_{STFT}$.

The shared attention module is used to refine the intermediate neural feature maps, as illustrated in Fig. 4. Given a feature map $\mathbf{F} \in \mathbb{R}^{C \times H \times W}$ with C being the number of feature map channels and $H \times W$ the size of feature maps, the shared attention module first aggregates spatial information by performing average-pooling and max-pooling in parallel, generating two different spatial context features, denoted by $\mathbf{F}_{avg} \in \mathbb{R}^{H \times W}$ and $\mathbf{F}_{max} \in \mathbb{R}^{H \times W}$, respectively. The two context features are then concatenated in the channel dimension and convolved by three 2D convolutional layers to infer the attention map $\mathcal{M}(\mathbf{F}) \in \mathbb{R}^{1 \times H \times W}$, expressed as

$$\mathcal{M}(\mathbf{F}) = \sigma \left(v_3^{3 \times 3} ([\mathbf{F}_{avg}; \mathbf{F}_{max}]) \right),$$
(17)

where σ is the Sigmoid function and $v_3^{3 \times 3}$ denotes three cascaded convolutional layers. The attention module $\mathcal{M}$ aims to adaptively control the weight of each point on the input feature map $\mathbf{F}$. Finally, the refined feature map $\mathbf{F}_r \in \mathbb{R}^{C \times H \times W}$ is given by

$$\mathbf{F}_r = \mathcal{M}(\mathbf{F}) \otimes \mathbf{F},$$
(18)

where $\otimes$ is the element-wise multiplication. During multiplication, the important parts of the input feature map are enhanced, and the rest ones are faded out according to the learned attention weights $\mathcal{M}(\mathbf{F})$.

With $\mathbf{F} \in \{\mathbf{F}_{IQ}, \mathbf{F}_{CFO}, \mathbf{F}_{FFT}, \mathbf{F}_{STFT}\}$, the refined versions of the intermediate neural feature maps are denoted as $\mathbf{F}_{r,IQ}$, $\mathbf{F}_{r,CFO}$, $\mathbf{F}_{r,FFT}$, and $\mathbf{F}_{r,STFT}$, respectively. Note that the module $\mathcal{M}$ is shared during training and inference. After that, the refined neural feature maps are concatenated to generate the fused feature map $\mathbf{F}_{r,c} \in \mathbb{R}^{4C \times H \times W}$, given by

$$\mathbf{F}_{r,c} = [\mathbf{F}_{r,IQ}; \mathbf{F}_{r,CFO}; \mathbf{F}_{r,FFT}; \mathbf{F}_{r,STFT}].$$
(19)

The fused feature map $\mathbf{F}_{r,c}$ is then fed to the signal identification module for RF fingerprinting, as shown in Fig. 3.
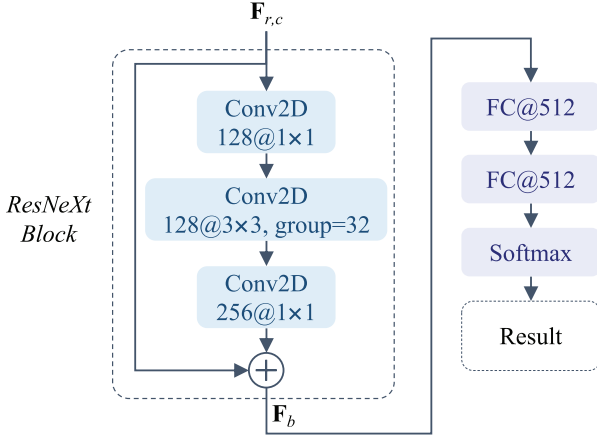
Fig. 5. Structure of the proposed signal identification module.

B. Signal Identification

The signal identification module maps refined features to device identities. As illustrated in Fig. 5, the signal identification module consists of a convolution-based ResNeXt block [55], two fully connected (FC) layers and a softmax layer. ResNeXt block is designed to build a simple but highly modularized network architecture for classification. It aggregates a set of transformations with the same topology and exploits the split-transform-merge strategy in an easy and extensible way. Compared with ResNet [56], ResNeXt block-based neural network is able to improve classification accuracy by increasing the size of the set of transformations. Let $\mathbf{x} = [x_1, \cdots, x_D]$ denote a D-channel input vector to a neuron and w_i denote the weight of a filter for the ith channel. A simple neuron in deep neural networks performs the inner product of the input vector and its corresponding weights, which can be thought of as a form of aggregating transformation

$$\sum_{i=1}^{D} w_i x_i. \tag{20}$$

Such inner product operation is the elementary transformation in fully-connected and convolutional layers. Inspired by the inner production operation, ResNeXt presents a more generic operation by replacing the elementary transformation $w_i x_i$ with an arbitrary function $\Gamma_i(\mathbf{x})$, that is,

$$\sum_{i=1}^{G} \Gamma_i(\mathbf{x}), \tag{21}$$

where G is the size of the set of transformations to be aggregated. As shown in Fig. 5, the ResNeXt block processes the input feature map $\mathbf{F}_{r,c}$ and generates the output $\mathbf{F}_b$ as

$$\mathbf{F}_b = \mathbf{F}_{r,c} + \sum_{i=1}^{G} \Gamma_i(\mathbf{F}_{r,c}). \tag{22}$$

The ResNeXt block consists of three convolution-based operations: a single 2D convolutional layer with 128 channels and filter size of 1×1, a grouped convolutional layer performing $G = 32$ groups of convolutions with 128 channels and filter size of 3×3, and a single 2D convolutional layer with 256
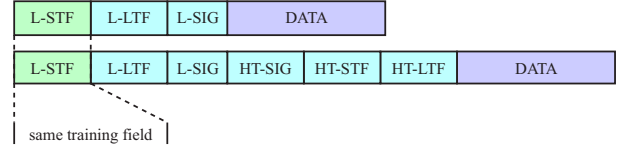

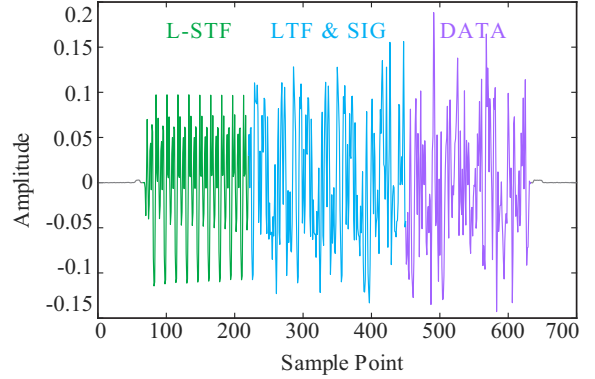Fig. 6. WiFi frame structures at the legacy (upper one) and mixed transmission modes [59].

Fig. 7. An uplink frame in the legacy transmission mode.

channels and filter size of 1×1. The ResNeXt block is followed by two fully-connected layers with 512 neurons each and finally an output layer with the softmax function. Each convolutional and fully-connected layer is followed by batch normalization [57] and Rectified Linear Unit (ReLU) [58].

VI. EXPERIMENTS

A. Experimental Setup and Dataset

To evaluate the RF fingerprinting performance, we design a signal acquisition system (see Fig. 2). The system consists of a wireless AP, WiFi end-devices that transmit WiFi signals to the AP, a National Instrument (NI) USRP-2922 SDR platform that collect WiFi signals from the end-devices, and a server processing the received signals collected at the SDR. The RF signal acquisition system (except the server) is placed in an anechoic chamber, which helps to reduce the impact of the wireless environment. The transmission inside the anechoic chamber is not affected by the external RF activities, and the cones deployed in the anechoic chamber absorb signals generated internally, thus preventing multi-path effects.

The WFDI dataset is constructed, according to Section IV, by preprocessing the uplink signal of each individual device from the population of 72 commercial WiFi end-devices. The communication between the WiFi device and the AP follows the standard IEEE 802.11n protocol [59]. The AP is configured to work at the legacy or mixed transmission mode such that the L-STF sequences in the uplink frames sent by different devices are identical. The frame structures at the legacy and mixed transmission modes are illustrated in Fig. 6.

At the SDR receiver, the uplink RF signal sent by the WiFi device is down-converted to baseband and then sampled at a rate of 20 MSa/s for one consecutive second. The one-second signal contains more than 256 uplink frames from the device and approximately the same number of downlink frames from

the AP, since the network protocol requires acknowledgement for each frame. In our experiments, the AP is configured to transmit at a lower power than the device, therefore the downlink frames can be easily filtered out by power-based frame detection. An uplink frame collected at the SDR receiver in the legacy transmission mode is illustrated in Fig. 7.

In the one-second signal, the standard frame detection algorithm is used to collect $N_f = 256$ complete uplink frames. The basic idea is find the peak location of the autocorrelation of L-STF sequences. As prescribed in the IEEE 802.11n protocol, the L-STF sequence is 8 μs long in each frame, which corresponds to 160 samples at the 20 MSa/s sampling rate. In each frame, therefore, the first $N_s = 160$ samples are selected to construct the dataset for this device. The rest samples in the frame are discarded since our RF fingerprinting algorithm generates device fingerprint based on the L-STF sequence only. Exploring the utilization of the rest samples in the frame can be considered in our future RF fingerprinting algorithm design.

The above signal collection experiment is repeated for each device on $N_d = 8$ different days at a room temperature of approximately $24°C$. The WFDI dataset is stored in a matrix form with the dimension of $N_n \times N_d \times N_f \times N_s \times N_c$, where $N_n = 72$ is the number of devices and $N_c = 2$ represents the I and Q channels.

B. Implementation Details

For each device, WDFI consists of $N_t = N_d \times N_f$ frames of data samples and each frame here contains N_s consecutive complex data samples. To evaluate the identification performance on a training dataset of different sizes, we randomly select a certain percentage of data from the N_t frames. All the experiments are implemented using Keras with Tensorflow backbone and trained on a PC with an NVIDIA TITAN X GPU. The proposed RF fingerprinting model is trained with a batch size of 256 and optimized with the Adam optimizer [60]. The learning rate is 0.001 with a decay factor of 0.1 and decays when the validation loss does not drop within 10 epochs. The training is terminated when the validation loss does not drop within 15 epochs. The RF fingerprinting model with the smallest validation loss is saved and used for test. In addition, the window length $J = 120$ and window shift $K = 5$ are selected for the STFT coefficients calculation.

C. Experimental Results

We conduct experiments to evaluate the identification performance of the proposed McAFF model. The other four RF fingerprinting models that use single-channel signal representation for IQ samples, CFO, FFT and STFT coefficients individually are compared with the proposed McAFF model.

Ablation study of signal representations: We do ablation studies to illustrate the impact of different signal representation on RF device identification accuracy. The validation losses of different RF fingerprinting models are shown in Fig. 8, where the RF fingerprinting models using single-channel signal representation for IQ samples, CFO, FFT, and STFT coefficients are referred to as MIQ, MCFO, MFFT,

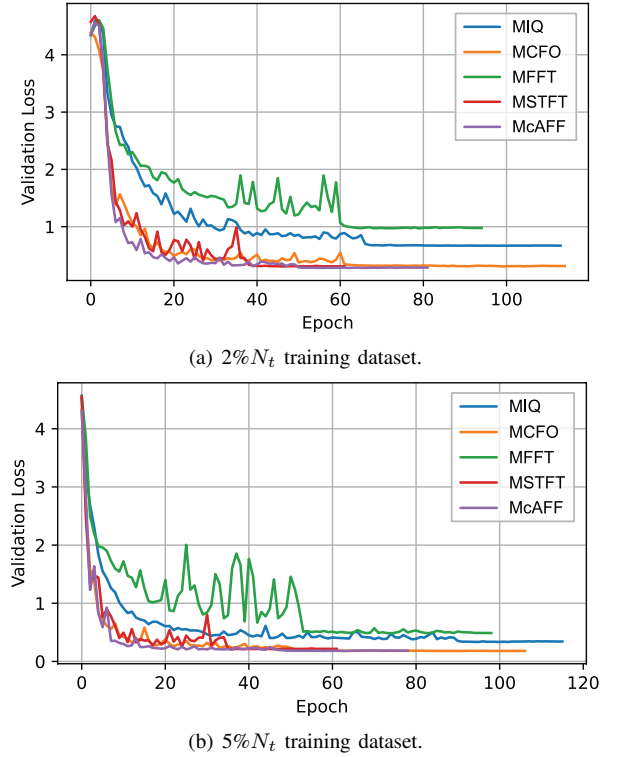(a) $2\%N_t$ training dataset.

(b) $5\%N_t$ training dataset.

Fig. 8. Validation loss versus training epochs.

TABLE I
RF FINGERPRINT IDENTIFICATION PERFORMANCE COMPARISON.

	$1\%N_t$	$5\%N_t$	$15\%N_t$	$25\%N_t$	$50\%N_t$
MIQ	64.52%	91.17%	95.95%	97.21%	98.29%
MCFO	86.69%	96.06%	97.92%	98.38%	98.94%
MFFT	47.37%	86.43%	93.27%	94.3%	96.17%
MSTFT	89.94%	95.18%	96.81%	97.41%	98.01%
McAFF	**90.00%**	**96.11%**	**97.94%**	**98.41%**	**99.01%**

and MSTFT, respectively. It shows how the validation cross-entropy loss varies when increasing the number of training epochs. Compared with the single-channel RF fingerprinting models, i.e., MIQ, MCFO, MFFT and MSTFT, the proposed McAFF model not only converges faster, but also converges to a lower validation loss. We also observe that all the models converge to a lower validation loss when the training dataset size increases from $2\%N_t$ to $5\%N_t$. Table I further shows the test performance of RF fingerprinting for different training dataset sizes. When the training dataset size is $50\%N_t$, all the models can provide excellent accuracies (over 96%) for device identification. The proposed McAFF model outperforms all the other models under the same training dataset sizes. It is also more robust to the change of training dataset sizes.

"Train and Test Same Days" scenario: We split our WFDI dataset into 8 sub-datasets according to the collection date and name the sub-dataset collected on day i as $D_i, i = 1, 2, \cdots, 8$. We train the McAFF model on each sub-dataset and get a model for each day. Table II shows the identification results of the proposed method on different sub-datasets. It is observed

TABLE II
RF FINGERPRINT IDENTIFICATION PERFORMANCE OF THE PROPOSED
McAFF MODEL IN THE "TRAIN AND TEST SAME DAYS" SCENARIO.

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8
99.17%	98.25%	99.26%	99.62%	99.01%	99.6%	98.21%	99.87%

TABLE III
RF FINGERPRINT IDENTIFICATION PERFORMANCE COMPARISON IN THE
"TRAIN AND TEST DIFFERENT DAYS" SCENARIO.

	1_train/7_test	2_train/6_test	3_train/5_test	4_train/4_test
MIQ	58.78%	75.81%	79.76%	82.81%
MCFO	62.44%	75.26%	78.9%	81.18%
MFFT	70.22%	81.39%	84.58%	86.45%
MSTFT	**76.10%**	85.43%	86.52%	88.06%
McAFF	75.35%	**86.04%**	**87.72%**	**88.67%**

Fig. 9. RF fingerprint identification performance versus SNR. The training dataset size is $50\% N_t$.

that the proposed McAFF model works pretty well when training data and test data are collected on the same day. Specifically, the proposed McAFF model achieves the highest identification accuracy of 99.85% on D_8 and the lowest one of 98.21% on D_7. The average identification accuracy of the proposed McAFF model over eight days is 99.12%.

"Train and Test Different Days" scenario: In this experiment, the training data and test data are collected on different days. Table III shows the experimental results. When data collected from 1 day is used for training and data collected from the remaining 7 days is used for test (referred to as 1_train/7_test in Table III), the proposed method has slightly worse identification accuracy than MSTFT. It is observed that by using more training data collected from more days (the other data from the remaining days is used for test), all the models achieve better performance and the identification accuracies are improved significantly. In general, the proposed McAFF method still achieves the best identification performance. This is not unexpected since the proposed method uses multi-channel inputs and adaptively fuses the multi-channel neural features. Therefore, it is more robust to different data collection dates.

Impact of noise: The impact of noise on identification performance is shown in Fig. 9. The additive noise is generated in a simulated way such that 14 different SNR levels from 4 dB to 30 dB are tested. It is shown that the identification accuracy of all the RF fingerprinting methods increases with SNR. At high SNRs, the performance approaches that in Table I, which is based on almost clean WiFi signals collected in an anechoic chamber. The performance of a recent state-of-the-art: ResNet50-IQ [15] is also plotted in Fig. 9. ResNet50-IQ directly used ResNet50 [61] as the signal identification network with IQ information as the input. It is clear that the proposed McAFF method significantly outperforms ResNet50-IQ, as well as the other single-channel RF fingerprinting methods. Closer observations show that MSTFT and MIQ have similar identification performance and outperform MFFT at all SNR levels. When the SNR is lower than 20 dB, ResNet50-IQ has better performance than both MCFO and MFFT, and the performance improvement of MFFT over MCFO is significant. When the SNR level is increased above 20 dB,
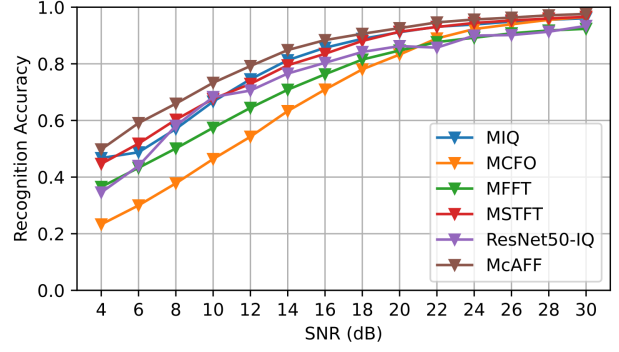
MCFO starts to perform better than ResNet50-IQ and MFFT, and ResNet50-IQ shows similar performance as MFFT. Fig. 10 further shows the RF fingerprint identification performance at different SNR levels when the training and test data are collected on different days. Compared with the single-channel RF fingerprinting models, the proposed method has the best identification performance at all SNR levels. The experimental results demonstrate that the proposed method that uses multi-channel attention fusion is robust against the noisy signals.

Effect of feature attention: In Fig. 11, we compare the proposed McAFF model with a simplified version that concatenates multi-channel features without using the feature attention module in Fig. 4. It is clear to see that the proposed McAFF model shows constantly improved identification performance at different SNRs. Fig. 11 demonstrates the effectiveness of the proposed shared attention module since it is able to dynamically control the weights of multi-channel neural features under various conditions.

VII. CONCLUSION

In this paper, we proposed an effective McAFF model for RF fingerprinting. The proposed model consists of two sequential parts: multi-channel feature fusion for feature learning and fusion, and signal identification for RF device identification. The multi-channel feature fusion part first computes the representations of the input WiFi signals through CFO estimation, FFT and STFT operations. Then, a shared attention module is used to adaptively control the weight of the neural features learned from the four channel inputs, including IQ samples, CFO, FFT and STFT coefficients. The model is trained to minimize the cross-entropy loss function. To evaluate the effectiveness of the proposed McAFF model for real-world RF signals, we designed a dataset (WFDI) captured from 72 WiFi end-devices over the course of 8 days in an anechoic chamber. We conducted an extensive evaluation of the impact of signal representation on RF fingerprint identification performance, and analyzed the effectiveness of the proposed shared attention module in terms of fingerprinting accuracy in scenarios where training and test data were collected on different days. In addition, we evaluated the identification performance of the proposed McAFF model on a simulated noisy dataset. Experimental results demonstrated that the proposed
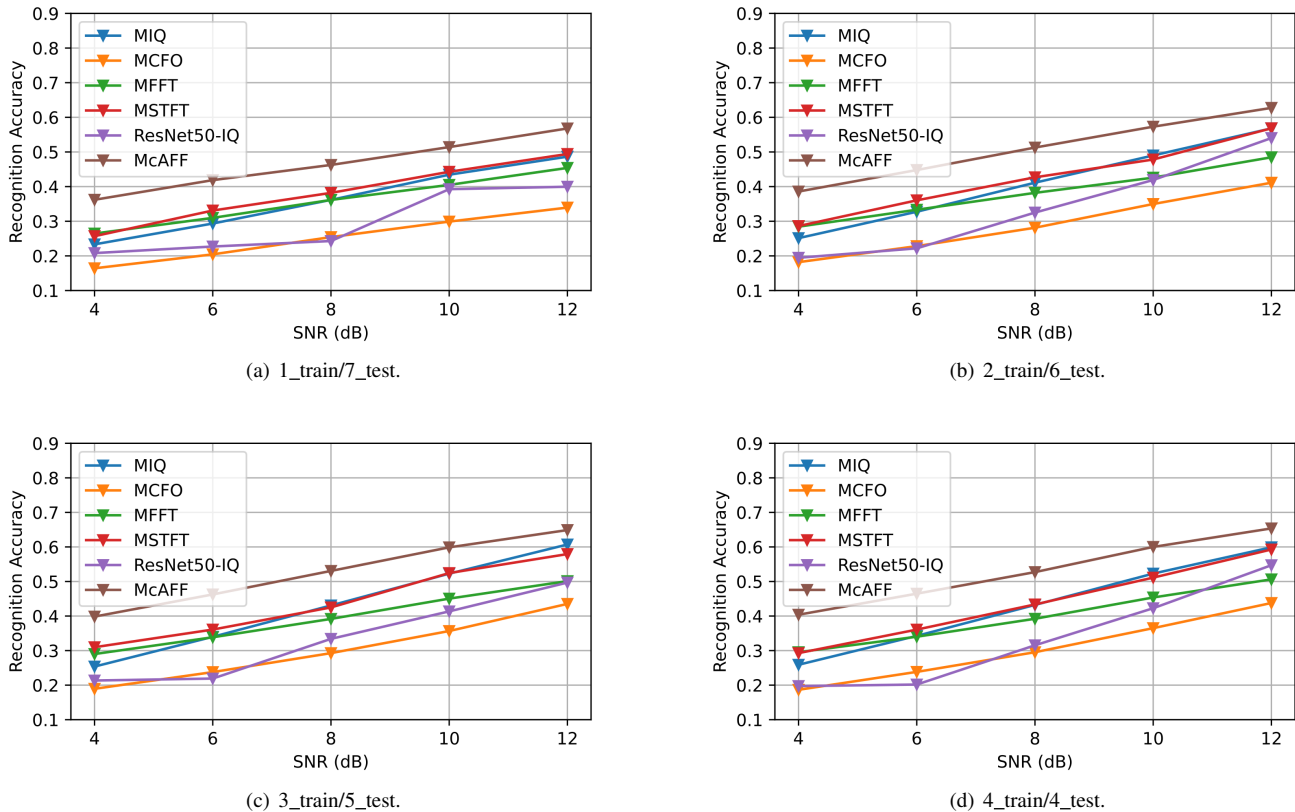
Fig. 10. RF fingerprint identificaiton performance versus SNR in the "Train and Test Different Days" scenario.
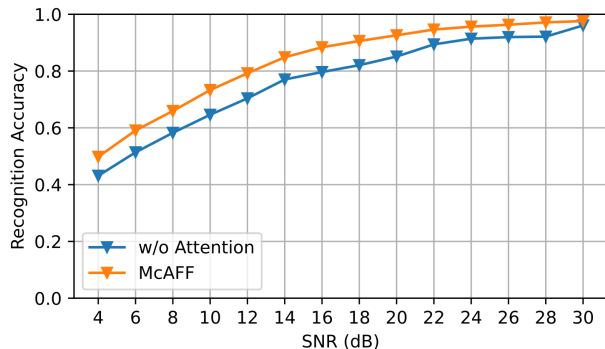

Fig. 11. Effect of feature attention. The training dataset size is $50\% N_t$.

RF fingerprinting model with multi-channel feature fusion achieves much better identification performance on datasets of different sizes and different collection dates, and is more robust in noisy environments. Learning to fuse multi-channel features is a straightforward approach to improving the design of a feature-based RF fingerprinting system. In this work, the RF signal acquisition system was placed in an anechoic chamber with simulated noise. Exploring and improving the proposed method for RF fingerprinting in a more dynamic communication environment is an interesting direction for future research.

REFERENCES

[1] T. Qiu, N. Chen, K. Li, M. Atiquzzaman, and W. Zhao, "How can heterogeneous internet of things build our future: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2011–2027, 3rd Quart. 2018.

[2] S. B. Baker, W. Xiang, and I. Atkinson, "Internet of things for smart healthcare: Technologies, challenges, and opportunities," *IEEE Access*, vol. 5, pp. 26 521–26 544, Nov. 2017.

[3] N. Sharma and R. D. Garg, "Real-time iot-based connected vehicle infrastructure for intelligent transportation safety," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–9, Apr. 2023.

[4] F. Restuccia, S. D'Oro, and T. Melodia, "Securing the internet of things in the age of machine learning and software-defined networking," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4829–4842, Jun. 2018.

[5] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A survey on wireless security: Technical challenges, recent advances, and future trends," *Proc. IEEE*, vol. 104, no. 9, pp. 1727–1765, Sep. 2016.

[6] D. He and S. Zeadally, "An analysis of RFID authentication schemes for internet of things in healthcare environment using elliptic curve cryptography," *IEEE Internet Things J.*, vol. 2, no. 1, pp. 72–83, Feb. 2014.

[7] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[8] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 289–299, Aug. 2014.

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

[10] E. O. Johnson, "Physical limitations on frequency and power parameters of transistors," in *IRE Int. Conv. Rec.*, vol. 13, New York, NY, USA, Mar. 1965, pp. 27–34.

[11] S. Riyaz, K. Sankhe, S. Ioannidis, and K. Chowdhury, "Deep learning convolutional neural networks for radio identification," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 146–152, Sep. 2018.

[12] 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.

[13] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.

[14] R. Das, A. Gadre, S. Zhang, S. Kumar, and J. M. Moura, "A deep learning approach to IoT authentication," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[15] A. Al-Shawabka, F. Restuccia, S. D'Oro, T. Jian, B. C. Rendon, N. Soltani, J. Dy, S. Ioannidis, K. Chowdhury, and T. Melodia, "Exposing the fingerprint: Dissecting the impact of the wireless channel on radio fingerprinting," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Toronto, ON, Canada, Jul. 2020, pp. 646–655.

[16] J. Gong, X. Xu, and Y. Lei, "Unsupervised specific emitter identification method using radio-frequency fingerprint embedded InfoGAN," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2898–2913, Mar. 2020.

[17] Y. Pan, S. Yang, H. Peng, T. Li, and W. Wang, "Specific emitter identification based on deep residual networks," *IEEE Access*, vol. 7, pp. 54 425–54 434, Apr. 2019.

[18] L. Peng, J. Zhang, M. Liu, and A. Hu, "Deep learning based RF fingerprint identification using differential constellation trace figure," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1091–1095, Jan. 2020.

[19] Y. Zeng, M. Zhang, F. Han, Y. Gong, and J. Zhang, "Spectrum analysis and convolutional neural network for automatic modulation recognition," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 929–932, Jun. 2019.

[20] G. Shen, J. Zhang, A. Marshall, L. Peng, and X. Wang, "Radio frequency fingerprint identification for LoRa using spectrogram and CNN," in *Proc. Conf. Comput. Commun. (INFOCOM)*, Vancouver, BC, Canada, May 2021, pp. 1–10.

[21] 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. 2018.

[22] M. Cekic, S. Gopalakrishnan, and U. Madhow, "Robust wireless fingerprinting: Generalizing across space and time," *arXiv preprint arXiv:2002.10791*, 2020.

[23] H. J. Patel, M. A. Temple, and R. O. Baldwin, "Improving ZigBee device network authentication using ensemble decision tree classifiers with radio frequency distinct native attribute fingerprinting," *IEEE Trans. Rel.*, vol. 64, no. 1, pp. 221–233, Mar. 2014.

[24] D. A. Knox and T. Kunz, "Practical RF fingerprints for wireless sensor network authentication," in *n Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Limassol, Cyprus, Aug. 2012, pp. 531–536.

[25] W. C. Suski II, M. A. Temple, M. J. Mendenhall, and R. F. Mills, "Using spectral fingerprints to improve wireless network security," in *Proc. Global Telecommun. Conf. (GLOBECOM)*, New Orleans, LO, USA, Nov.-Dec. 2008, pp. 1–5.

[26] A. C. Polak and D. L. Goeckel, "Identification of wireless devices of users who actively fake their RF fingerprints with artificial data distortion," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 5889–5899, Nov. 2015.

[27] 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 Sec.*, vol. 11, no. 8, pp. 1862–1874, Aug. 2016.

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

[29] T. Kohno, A. Broido, and K. C. Claffy, "Remote physical device fingerprinting," *IEEE Trans. Depend. Secure Comput.*, vol. 2, no. 2, pp. 93–108, Jun. 2005.

[30] S. Jana and S. K. Kasera, "On fast and accurate detection of unauthorized wireless access points using clock skews," *IEEE Trans. Mobile Comput.*, vol. 9, no. 3, pp. 449–462, Mar. 2009.

[31] M. Leonardi, L. Di Gregorio, and D. Di Fausto, "Air traffic security: Aircraft classification using ADS-B message's phase-pattern," *Aerospace*, vol. 4, no. 4, p. 51, 2017.

[32] J. Hua, H. Sun, Z. Shen, Z. Qian, and S. Zhong, "Accurate and efficient wireless device fingerprinting using channel state information," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Honolulu, HI, USA, Apr. 2018, pp. 1700–1708.

[33] 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. (WISEC)*, Darmstadt, Germany, Jul. 2016, pp. 3–14.

[34] I. O. Kennedy, P. Scanlon, F. J. Mullany, M. M. Buddhikot, K. E. Nolan, and T. W. Rondeau, "Radio transmitter fingerprinting: A steady state frequency domain approach," in *Proc. Veh. Technol. Conf. (VTC)*, Calgary, BC, Canada, May 2008, pp. 1–5.

[35] R. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Mag.*, vol. 4, no. 2, pp. 4–22, Apr. 1987.

[36] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for mimo detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702–1715, Feb. 2020.

[37] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, Athens, Greece, Aug. 2016, pp. 213–226.

[38] Y. Wang, G. Gui, T. Ohtsuki, and F. Adachi, "Multi-task learning for generalized automatic modulation classification under non-gaussian noise with varying SNR conditions," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3587–3596, Jun. 2021.

[39] N. Soltanieh, Y. Norouzi, Y. Yang, and N. C. Karmakar, "A review of radio frequency fingerprinting techniques," *IEEE J. Radio Freq. Identificat.*, vol. 4, no. 3, pp. 222–233, Sep. 2020.

[40] Y. Li, Y. Ding, J. Zhang, G. Goussetis, and S. K. Podilchak, "Radio frequency fingerprinting exploiting non-linear memory effect," *IEEE Trans. Cognit. Commun. Netw.*, pp. 1–1, Oct. 2022.

[41] K. Sankhe, M. Belgiovine, F. Zhou, L. Angioloni, F. Restuccia, S. D'Oro, T. Melodia, S. Ioannidis, and K. Chowdhury, "No radio left behind: Radio fingerprinting through deep learning of physical-layer hardware impairments," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 1, pp. 165–178, Mar. 2019.

[42] J. Xu, C. Luo, G. Parr, and Y. Luo, "A spatiotemporal multi-channel learning framework for automatic modulation recognition," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1629–1632, Jun. 2020.

[43] S. Lin, Y. Zeng, and Y. Gong, "Modulation recognition using signal enhancement and multistage attention mechanism," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9921–9935, Nov 2022.

[44] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 30, Dec. 2017.

[45] N. Moritz, T. Hori, and J. Le Roux, "Semi-supervised speech recognition via graph-based temporal classification," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Toronto, ON, Canada, Jun. 2021, pp. 6548–6552.

[46] Y. Zhao and D. Wang, "Noisy-reverberant speech enhancement using denseunet with time-frequency attention," in *Proc. Annu. Conf. Int. Speech Commun. Assoc (INTERSPEECH),*, Shanghai, China, Oct. 2020, pp. 3261–3265.

[47] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, Sep. 2018, pp. 3–19.

[48] C. Fan, B. Liu, J. Tao, J. Yi, and Z. Wen, "Spatial and spectral deep attention fusion for multi-channel speech separation using deep embedding features," *arXiv preprint arXiv:2002.01626*, 2020.

[49] J. Xu, K. Lu, and H. Wang, "Attention fusion network for multi-spectral semantic segmentation," *Pattern Recognit. Lett.*, vol. 146, pp. 179–184, Jun. 2021.

[50] T. J. O'Shea, L. Pemula, D. Batra, and T. C. Clancy, "Radio transformer networks: Attention models for learning to synchronize in wireless systems," in *Proc. 50th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 662–666.

[51] S. Lin, Y. Zeng, and Y. Gong, "Learning of time-frequency attention mechanism for automatic modulation recognition," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 707–711, Apr. 2022.

[52] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 4, pp. 1150–1163, Apr. 2005.

[53] E. Sourour, H. El-Ghoroury, and D. McNeill, "Frequency offset estimation and correction in the IEEE 802.11 a WLAN," in *Proc. IEEE 60th Veh. Technol. Conf. (VTC)*, vol. 7, Los Angeles, CA, USA, Sep. 2004, pp. 4923–4927.

[54] N. Kehtarnavaz, "CHAPTER 7 - frequency domain processing," in *Digital Signal Processing System Design (Second Edition)*. Burlington: Academic Press, 2008, pp. 175–196.

[55] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1492–1500.

[56] 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.

[57] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn. (ICML)*, F. R. Bach and D. M. Blei, Eds., Lille, France, Jul. 2015, pp. 448–456.

[58] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," *CoRR*, vol. abs/1803.08375, 2018.

[59] "IEEE Standard for Information Technology - Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE Std 802.11-2020 (Revision of IEEE Std 802.11-2016)*, pp. 1–4379, Jun. 2021.

[60] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–15.

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

Zidong Han (Graduate Student Member, IEEE) received the B. S. degree in Information Engineering from the Shanghai Jiao Tong University (SJTU) in 2012 and the M. S. degree in Electrical Engineering from the National University of Singapore (NUS) in 2013. He is currently with the Southern University of Science and Technology as a Research Scholar. He is also a Senior Engineer (Information Technology) certified by the Human Resources and Social Security Department of Guangdong Province. His research interests include wireless communication and edge intelligence.

Yuan Zeng (Member, IEEE) received her Ph.D. degree in Signal and Information Processing from Delft University of Technology, the Netherlands, in 2015. She was a research fellow with the Delft University of Technology from 2015 to 2018. She is currently a researcher with Southern University of Science and Technology, Shenzhen, China. Her main research interests are intelligent signal processing, including wireless sensor networks, view synthesis and image restoration.

Ruoxiao Cao (Graduate Student Member, IEEE) received his B.Eng. degree from Southern University of Science and Technology (SUSTech), Shenzhen, China, in 2021. He is currently pursuing the Ph.D. degree in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST), Hong Kong SAR, China. His research interests include signal processing for millimeter-wave/terahertz communications and ultra-massive MIMO.

Yi Gong (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees from the Southeast University and the Ph.D. degree from the Hong Kong University of Science and Technology, all in electrical engineering. He was with the Hong Kong Applied Science and Technology Research Institute, Hong Kong, and Nanyang Technological University, Singapore. He is currently a Professor with the Southern University of Science and Technology, Shenzhen, China. His research interests include cellular networks, mobile computing, and signal processing for wireless communications and related applications. He was on the Editorial Board of the IEEE Transactions on Wireless Communications and the IEEE Transactions on Vehicular Technology.

Jiawei Liu received his B.Eng. degree from Southern University of Science and Technology, Shenzhen, China, in 2020. He is currently working toward his Ph.D. in the Department of Electrical and Electronics Engineering at the University of Hong Kong (HKU), Hong Kong. His research interests include wireless communication system design and in-memory computing hardware.

Kaibin Huang (Fellow, IEEE) received the B.Eng. and M.Eng. degrees from the National University of Singapore and the Ph.D. degree from The University of Texas at Austin, all in electrical engineering. He is a Professor and an Associate Head at the Dept. of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong. He received the IEEE Communication Society's 2021 Best Survey Paper, 2019 Best Tutorial Paper, 2019 Asia–Pacific Outstanding Paper, 2015 Asia–Pacific Best Paper Award, and the best paper awards at IEEE GLOBECOM 2006 and IEEE/CIC ICCC 2018. He received the Outstanding Teaching Award from Yonsei University, South Korea, in 2011. He has been named as a Highly Cited Researcher by the Clarivate Analytics in 2019-2022. He is a member of the Engineering Panel of Hong Kong Research Grants Council (RGC) and a RGC Research Fellow. He served as the Lead Chair for the Wireless Communications Symposium of IEEE Globecom 2017 and the Communication Theory Symposium of IEEE GLOBECOM 2023 and 2014, and the TPC Co-chair for IEEE PIMRC 2017 and IEEE CTW 2023 and 2013. He is also an Executive Editor of IEEE Transactions on Wireless Communications, and an Area Editor for both IEEE Transactions on Machine Learning in Communications and Networking and IEEE Transactions on Green Communications and Networking. Previously, he served on the Editorial Boards for IEEE Journal on Selected Areas in Communications and IEEE Wireless Communication Letters. He has guest edited special issues of IEEE Journal on Selected Areas in Communications, IEEE Journal of Selected Areas in Signal Processing, and IEEE Communications Magazine. He is also a Distinguished Lecturer of the IEEE Communications Society and the IEEE Vehicular Technology Society.

Shangao Lin received the B.Eng. degree in communication engineering from the Dalian Maritime University in 2018. He obtained his master's degree in electronics science and technology at the Southern University of Science and Technology in 2022. His research interests include deep learning, signal processing, and wireless communications.

Khaled B. Letaief (Fellow, IEEE) is an internationally recognized leader in wireless communications and networks. He is a Member of the United States National Academy of Engineering, Fellow of IEEE, Fellow of the Hong Kong Institution of Engineers, Member of India National Academy of Sciences, and Member of the Hong Kong Academy of Engineering Sciences. He is also recognized by Thomson Reuters as an ISI Highly Cited Researcher and was listed among the 2020 top 30 of AI 2000 Internet of Things Most Influential Scholars.

Dr. Letaief is the recipient of many distinguished awards and honors including the 2022 IEEE Communications Society Edwin Howard Armstrong Achievement Award; 2021 IEEE Communications Society Best Survey Paper Award; 2019 IEEE Communications Society and Information Theory Society Joint Paper Award; 2016 IEEE Marconi Prize Paper Award in Wireless Communications; 2011 IEEE Communications Society Harold Sobol Award; 2010 Purdue University Outstanding Electrical and Computer Engineer Award; 2007 IEEE Communications Society Joseph LoCicero Publications Exemplary Award; and over 19 IEEE Best Paper Awards.

He has also been a dedicated teacher committed to excellence in teaching and scholarship. He received the Mangoon Teaching Award from Purdue University; Teaching Excellence Appreciation Award by the School of Engineering at HKUST (4 times); and the Michael G. Gale Medal for Distinguished Teaching (Highest university-wide teaching award and only one recipient/year is honored for his/her contributions). Since 1993, he has been with the Hong Kong University of Science & Technology (HKUST) where he has held many administrative positions, including Acting Provost, Head of the Electronic and Computer Engineering department, Director of the Wireless IC Design Center, founding Director of Huawei Innovation Laboratory, and Director of the Hong Kong Telecom Institute of Information Technology. While at HKUST he has also served as Chair Professor and Dean of Engineering. Under his leadership, the School of Engineering has dazzled in international rankings (rising from # 26 in 2009 to # 14 in the world in 2015 according to QS World University Rankings.) From September 2015 to March 2018, he joined HBKU as Provost to help establish a research-intensive university in Qatar in partnership with strategic partners that include Northwestern University, Carnegie Mellon University, Cornell, and Texas A&M. Dr. Letaief is well recognized for his dedicated service to professional societies and IEEE where he has served in many leadership positions. These include founding Editor-in-Chief of the prestigious IEEE Transactions on Wireless Communications. He also served as President of the IEEE Communications Society (2018-19), the world's leading organization for communications professionals with headquarter in New York City and members in 162 countries. He is currently serving as member of the IEEE Board of Directors.

Dr. Letaief received the BS degree with distinction in Electrical Engineering from Purdue University at West Lafayette, Indiana, USA, in December 1984. He received the MS and Ph.D. Degrees in Electrical Engineering from Purdue University, in Aug. 1986, and May 1990, respectively. He has also received a Ph.D. Honoris Causa from the University of Johannesburg, South Africa, in 2022.