IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 17, 2019, accepted August 18, 2019, date of publication August 30, 2019, date of current version September 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2938533

Deep Neural Networks for CSI-Based Authentication

QIAN WANG, HANG LI, DOU ZHAO, ZHI CHEN, (Senior Member, IEEE), SHUANG YE, AND JIANSHENG CAI

National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Zhi Chen (chenzhi@uestc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61571089, and in part by the National Key Research and Development Program of China under Grant 2018YFC0807101.

ABSTRACT From the viewpoint of *physical-layer authentication*, spoofing attacks can be foiled by checking channel state information (CSI). Existing CSI-based authentication algorithms mostly require a deep knowledge of the channel variation to deliver decent performance. In this paper, we investigate CSI-based authenticators that can spare the effort to predetermine channel properties by utilizing deep neural networks (DNNs). First, we propose a convolutional neural network (CNN)-enabled authenticator that is able to extract the local features in CSI. Next, the recurrent neural network (RNN) is employed to capture the dependencies between different frequencies in CSI. In addition, we propose to use the convolutional recurrent neural network (CRNN)—a combination of the CNN and the RNN—to learn local and contextual information in CSI for user authentication. Finally, experiments based on Universal Software Radio Peripherals (USRPs) are conducted to demonstrate the performance of the proposed methods on real-world channel estimates. According to the experimental results, the proposed DNNs-enabled schemes can significantly outperform the dynamic time warping (DTW) technique and a heuristic Neyman-Pearson (NP) test in the aspects of false alarm and miss detection. Besides, the hybrid of the CNN and the RNN can further promote the authentication accuracy.

INDEX TERMS Physical layer authentication, CNN, RNN, CRNN, machine learning.

I. INTRODUCTION

A. PHYSICAL LAYER AUTHENTICATION

With the rapid development of wireless communication, an enormous amount of private and confidential information, e.g., financial data, medical records, and customer files, will be transmitted via the wireless medium [1]. The sharp increase in demand for wireless security continuously requests more advanced authentication schemes. Traditionally, authentication mechanisms are performed above the physical layer by using secret keys to identify wireless transmitters. Despite their effectiveness, they are faced with two main challenges: On the one hand, the high key management overhead results in concerns such as excessive latencies. On the other hand, the time required to crack a key has been remarkably shortened with the growing processing power; see a recent overview [2] and the references therein. The idea of *physical-layer authentication* is to validate a wireless transmitter by verifying the physical-layer attributes of the wireless transmission. In comparison to conventional secret key-based authentication schemes, *physical-layer authentication* needs no key distribution and management. Besides, it is extremely difficult to impersonate a wireless transmission's physical-layer features. Thanks to these facts, *physical-layer authentication* is deemed as a promising technique to make the unrivalled security service a reality. Some of the existing *physical-layer authentication* approaches rely on the analog front-end imperfections, which are *device-specific* characteristics caused by manufacturing variability [3]. *Device-specific* characteristics, such as in-phase/quadrature imbalance [4], the power amplifier characteristics [5], and the carrier frequency offset [6], have relatively stable nature. However, the difference of the targeted hardware features between devices is usually too small in practice, which will be further influenced by noise and interference [2]. Another class of physical-layer features used for authentication purposes are channel-based characteristics, like channel state information (CSI) [7]–[9] and received

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

signal strength (RSS) [10]. CSI is hard to predict due to the presence of rich scatters and reflectors in a general wireless communication environment. Besides, it is safe to say that users located at different places have uncorrelated channels. These facts make CSI a *location-specific* characteristic that has aroused great interest for user authentication.

Paper [7] studied CSI-based authentication in a time-variant wireless environment, wherein the channel variation was modeled with the assumption of a first-order autoregressive model. The authors in [8] studied frequency-selective channels, in which the terminal mobility-caused channel variation was modeled as a first-order autoregressive model, and the environment changes and the estimation errors were modeled as independent complex Gaussian processes. Moreover, a two-dimensional (the dimensions of channel amplitude and path delay) quantization method was proposed in [9] to preprocess the channel variations, wherein the temporal processes were still modeled as autoregressive models. To sum up, existing CSI-based approaches [7]–[9] formulated the authentication process as binary hypothesis testing by exploiting the correlation between CSI at adjacent times. All these works designed algorithms that would look for predetermined features in CSI. To do this, the system operator needs to possess sufficient channel information such as the channel model and the channel variation pattern. This kind of authentication system will be vulnerable to small ambiguities in the a priori messages.

B. RELATED WORK

Given the significant efficiency of machine learning techniques in objects recognition, it is straightforward to consider the exploitation of machine learning for facilitating *physical-layer authentication*. Recent years, huge strides have been made in making the recognition of objects more accurate. Current approaches to object identification make essential use of neural networks. *Deep neural networks* (DNNs) have been extensively studied and found to be very effective in learning high-level features from raw data for objects identification [11], [12]. The CNN was first proposed for digit recognition [13] and later became one of the most widely applied DNNs. It usually utilizes multiple convolutional layers that can successively generate deeper- level abstractions of the input data. [14] improved the CNN-based attention models by incorporating multi-level saliency predictions within a single network. Through using continuous deep Q-learning, a hyperparameter optimization algorithm is proposed in [15] for object tracking. Reference [16] proposed a triplet loss that can achieve more powerful feature than original logistic loss in tracking object. Authors in [17] proposed the quadruplet network, which is armed with multi-tuples for training so as to accurately mine the potential connections among instances and derive more robust representations for one-shot learning. As a type of neural networks different from CNNs, *recurrent neural networks* (RNNs) were mainly designed for sequence modeling [18]. RNNs employ feedback loops which allow connections from previous states to the subsequent ones and thus are able to represent advanced patterns of dependencies in the sequence. Convolutional recurrent neural networks (CRNNs) [19], [20] are emerging deep models, which are made up of multiple convolutional layers (together with pooling layers) and a few recurrent layers so that they can exploit not only the representation power of CNNs but also the contextual information modeling ability of RNNs.

Recently, machine learning techniques have found their applications in the realm of *physical-layer authentication*. Particularly, authors in [21] investigated the RSS-based authentication game in a dynamic environment, in which *reinforcement learning* was utilized to achieve the optimal test threshold in the hypothesis test. Paper [3] used time-domain complex baseband error signals to train a CNN so that user identities can be derived according to *device-specific* imperfections. The logistic regression model was utilized in [22] to exploit the received signal strength indicators measured at multiple landmarks for user identification. Although the spatial resolution of the transmitter can be enhanced through using multiple landmarks, their deployment will raise the system overhead and more pressingly, the communication between the landmarks and the security agent will be confronted with severe security threats. Thankfully, CSI contains much more *location-specific* information than RSS and can thus be reliable enough without assistances such as landmarks.

C. OUR WORK

In this paper, we establish deep neural networks (DNNs)-enabled authenticators that connect a transmitter's CSI to its estimated identity. The key of implementing CSI-based authentication lies in the correlation between the channel observations for the same user at different times. However, this correlation can be weakened by factors such as environment changes and practical imperfections. Fortunately, the CNN can be invariant to the transformations of the channel observations resulting from these factors; hence we propose to exploit the CNN to extract the deep features in CSI for user authentication. Also, we try to analyze CSI from a sequential point of view. In this way, CSI is seen as a data sequence and we utilize the RNN to model the dependencies between different frequencies in CSI. As a matter of fact, the CNN and the RNN possess different modeling abilities. More concretely, the CNN is good at representing locally invariant information while the RNN is better at contextual information modeling [23]. Owing to this observation, we propose to use the CRNN [19], [20] for CSI-based authentication. Accordingly, it is expected that the CRNN can have advantages over the CNN and the RNN in modeling channel features for authentication, which is confirmed by the simulation results (cf. Section IV).

To be more specific, this work considers a three-node wireless system consisting of a service agent, a legitimate user and a potential attacker. The agent aims to validate the access right of the user via examining its CSI. Different from existing works [7]–[9], we make no assumptions on the channel model or the channel variation pattern since our interest lies on data driven self-adaptive algorithms. As a merit of that, the proposed approaches apply where the channel properties remain unknown. Our main contribution includes the following aspects:

1) To begin with, we build a CNN-enabled classifier. The main components of this classifier are convolutional layers, which are able to generate deep-level *feature maps* through locally convolving small subregions of its input. Also, the network employs pooling layers to subsample the output of the convolutional layers so as to reduce the computational complexity and avoid over-fitting. At the end of the network, we use fully connected layers, with one logistic layer on the top, to collect the early extracted features and learn the user identity.
2) Next, we establish a RNN-enabled authenticator that analyzes CSI from a sequential perspective. This authenticator is composed of several recurrent layers and a few fully connected layers. The recurrent layers use feedback loops to capture the spectral dependencies in CSI, which are then input to the following fully connected layers such that object recognition can be implemented.
3) Further, we propose a CRNN-enabled approach that works in the following way: The first part of the proposed authenticator is a CNN, which is used for extracting middle-level features. Then, the output features of the CNN are fed into recurrent layers so that the contextual information of CSI can be well captured. Finally, fully connected layers are employed to perform classification.
4) Last, the experimental results based on Universal Software Radio Peripherals (USRPs) [24] show that the proposed DNNs-enabled algorithms can achieve significant performance gains over the dynamic time warping (DTW) technique [25] and a heuristic Neyman-Pearson (NP) test. In addition, the RNN-based approach can achieve better accuracy when compared with the CNN-enabled authenticator, while the CRNN-enabled one owns the highest accuracy among the proposed DNNs-enabled algorithms.

II. SYSTEM MODEL AND PROBLEM STATEMENT

Consider a typical "Alice-Bob-Eve" network shown in Fig. 1, in which Bob is entasked with the job of providing services for both Alice and Eve, while Eve is unauthorized as far as the secure service intended for Alice is concerned. We assume that Bob is equipped with M_B antennas, both Alice and Eve have M_A antennas, and CSI is measured at N tones. In our setup, Alice, Bob, and Eve are geographically placed at different locations. Also, suppose that Alice, Bob, and Eve stay stationary in a time-variant communication environment. This is common in practical scenarios where one may put one's cellphone/laptop on the phone stand/desk while using it, and the service agent is stationary by nature. As an untrusted user, Eve may impersonate Alice by forging the digital credential of Alice, like the password, the IP address, and the MAC address, in attempts to illegitimately acquire confidential information intended for Alice or send false messages to Bob. Once Eve successfully obtains the illegal advantages, the following attacks will be devastating so that the authentication process is of paramount importance for the network security.

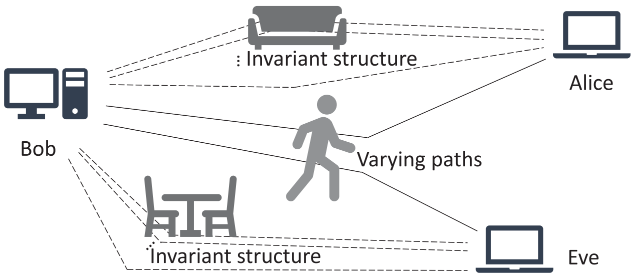


FIGURE 1. Illustration of our considered three-node communication system.

In our consideration, Bob aims to foil the spoofing attacks launched by Eve through establishing a physical-layer authentication mechanism. Specifically, Bob intends to verify whether the transmitter who uses Alice's digital credential is Alice or not by carefully checking its CSI according to the historical CSI of Alice and Eve. To do this, it is supposed that Bob can record historical CSI and the corresponding user identities. This is possible when both Alice and Eve are network users. Generally, device mobility and environment changes are the factors that give rise to the channel variation. Since all the communication nodes are assumed to be static in this work, it is safe to say that environment changes are the only reasons that bring about the channel variation. Environment changes such as the movement of objects can affect part of the existing paths while other paths stay invariant. Take an indoor environment for example, there may be moving people and objects, but the ceiling, the floor, walls, and furniture will always stay still. This fact indicates that the channel between two static terminals has an invariant structure, which forms the foundation of our application of deep learning into the CSI-based authentication problem.

A CSI-based authenticator, which maps CSI to the transmitter's authenticated identity, can be written as

$$\hat{I}_t = f_a(H(t)), \tag{1}$$

in which $H(t) \in \mathcal{C}^{M_B \times M_A \times N}$ denotes the communication channel observed at Bob in slot t, and $f_a(\cdot)$ is the authenticator whose output, i.e., $\hat{I}_t$, is the estimated identity of the transmitter at time t. In our settings, $\hat{I}_t = 1$ indicates that the estimated transmitter at time t is Eve, and $\hat{I}_t = 0$ means that Alice is the estimated transmitter at time t. Additionally, $I_t = 1$ denotes that the transmitter at time t is Eve, and $I_t = 0$ represents that Alice is the transmitter at time t. With no information about the underlying channel properties, one can not predetermine a decent $f_a(\cdot)$. Instead, we propose to model $f_a(\cdot)$ with neural networks, which are efficient models for statistical pattern recognition.

III. DEEP NEURAL NETWORKS-BASED AUTHENTICATOR

The kernel of CSI-based authentication lies in the correlation between the channel observations for the same transmitter-receiver pair at different times. Nevertheless, this correlation can be terribly hurt by adverse factors like environment changes and practical imperfections. Consequently, the proposed neural networks should be insensitive to the signal transformations brought by the adverse elements. Toward this end, we introduce DNNs to construct CSI-based authenticators that can capture the invariant channel structure from varying historical CSI. Specifically, the introduced DNNs are the CNN in Section III-A.1, the RNN in Section III-A.2, and the CRNN in Section III-A.3. These DNNs count upon different mechanisms to model the underlying channel features. The details will be elaborated in the following.

A. INTRODUCED NETWORK STRUCTURES

1) CONVOLUTIONAL NEURAL NETWORKS

Due to the fact that the channel observations for the same user at different times can be seen as transformed versions of each other, the ability to recognize the channel after complex transformation is vital to an authenticator. Thankfully, the great success of the CNN in digit recognition [13] suggests that the CNN is able to be invariant to transformations like scaling and shifting. Accordingly, the CNN is introduced to extract the invariant channel features from varying channel observations. As shown in Fig. 2, the architecture of our employed CNN is consisted of convolutional layers, pooling layers, and fully connected layers. Being the core components of the CNN, convolutional layers utilize the following mechanisms:

- Local Receptive Fields - The input of the convolutional layer is divided into local receptive fields (small subregions), each of which is connected to a single neuron of the next layer. As a benefit of this, the number of connections, as well as the number of parameters, is drastically cut down in the convolutional layer.
- Weight Sharing - Each locally applied connection is essentially a filter. A convolutional layer employs multiple filters, which are reused over all the local receptive fields. The reuse of filters leads to the sharing of an identical set of weights among different connections.

Invoking the above mechanisms, a convolutional layer organizes its input units into *feature maps*, i.e.,

$$\boldsymbol{F} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_N) = f_f(\boldsymbol{s} * \{\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \ldots, \boldsymbol{\phi_d}\}), \qquad (2)$$

where $\boldsymbol{f}_n \in \mathcal{R}^d, n = 1, 2, \ldots, Q$, represents a *feature map*, $f_f(\cdot)$ is an activation function, $\boldsymbol{s}$ is the input vector, $*$ denotes a convolution, and $\{\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \ldots, \boldsymbol{\phi_d}\}$ is a set of filters. Units in a *feature map* take input from a local receptive field of $\boldsymbol{s}$, and different receptive fields share the same filters.

Notice that before feeding the network input into the first convolutional layer, we organize it into $2M_AM_B$ N-dimensional "channels",[1] each of which corresponds to the real or imaginary part of the channel between a pair of transmitting and receiving antennas. To be specific, the set of all the filters is repeatedly applied to all these "channels" and the results for different "channels" are added before input to the activation function $f_f(\cdot)$. There are numerous nonlinear activation functions applicable to the neural network framework, such as the hyperbolic tangent function, the softmax, and the ReLU, in which the ReLU is the most widely applied in the CNN and thus is chosen to be the activation function $f_f(\cdot)$ in (2).

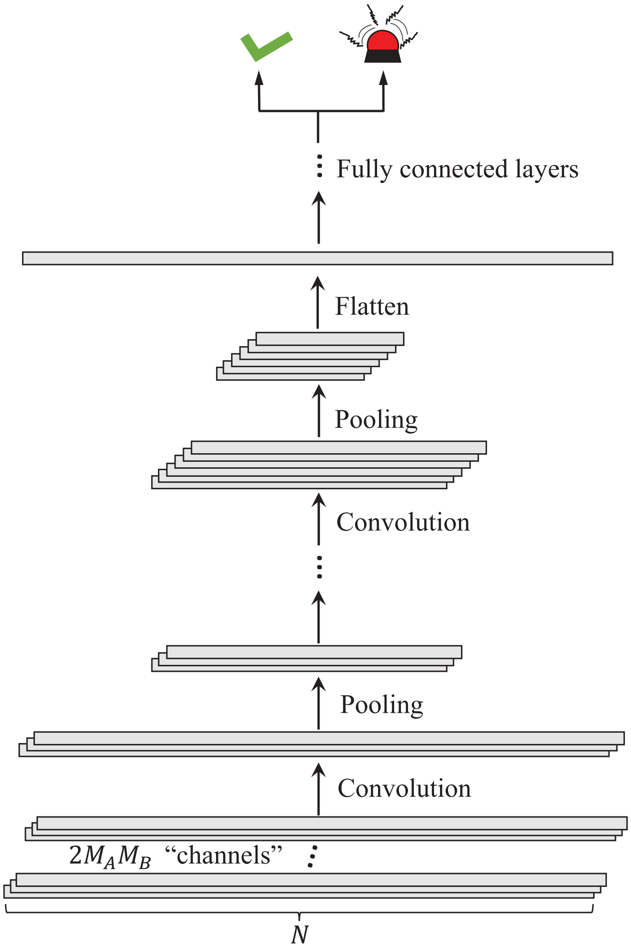


FIGURE 2. Illustration of our employed architecture of the CNN.

We perform pooling to summarize the *feature maps* created in the convolutional layer. Specifically, each pooling unit takes input from a region in the corresponding *feature map*, which is called a *pooling window*. The commonly used pooling operations are *max pooling*, *average pooling*, and *stochastic pooling*. By utilizing pooling, the neural network can achieve more compact representations that are more robust to noise and interference. In our architecture, the size

[1] When a neural network is utilized to analyze an image, there will be three input "channels" corresponding to the red, green, and blue elements of the input image, respectively.

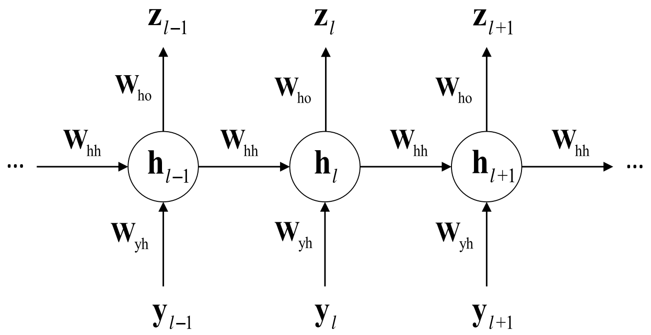


FIGURE 3. The structure of a recurrent layer.

of the *pooling window* is set as 1 × 3, on which a max operation is implemented. One can see that the features extracted by the convolutional layers and the pooling layers contain multiple "channels", which are flattened before fed into the fully connected layers, i.e., multiple "channels" are used in series. The final layers of the CNN are fully connected layers whose activation functions are chosen as rectified linear units (ReLUs), with one logistic layer on the top producing the authentication result.

2) RECURRENT NEURAL NETWORKS

Because there exist spectral dependencies in the channel, our goal in this subsection is to analyze the channel from a sequential point of view. Towards this end, we use the RNN to capture the contextual information in the channel for the purpose of authentication. The RNN utilizes feedback loops in its recurrent layers to connect the previous states with the current ones [18]. A graphical illustration of a recurrent layer is shown in Fig. 3. When a sequence is processed with length L, its hidden feature $\boldsymbol{h}_l$ and the predicted output z_l at stage $l \in [1, \ldots, L]$ are derived as

$$\boldsymbol{h}_l = f_h(\boldsymbol{W}_{hh}\boldsymbol{h}_{l-1} + \boldsymbol{W}_{yh}\boldsymbol{y}_l), \tag{3a}$$

$$z_l = f_z(\boldsymbol{W}_{hz}\boldsymbol{h}_l), \tag{3b}$$

respectively, in which $\boldsymbol{y}_l$ denotes the lth input, $\boldsymbol{W}_{hh}$, $\boldsymbol{W}_{yh}$, and $\boldsymbol{W}_{hz}$ are transformation matrices, and $f_h(\cdot)$ and $f_z(\cdot)$ are activation functions. With the existence of feedback loops, a recurrent layer is able to memorize the historical information so that they can discover meaningful connections between a single data and its context. In our architecture, the recursive function $f_h(\cdot)$ and the activation function $f_z(\cdot)$ are chosen to be the hyperbolic tangent function and the logistic sigmoid, respectively. It should be pointed out that as a $M_B \times M_A \times N$ complex channel, the network input is flattened before it is fed into the first layer of the RNN.

The RNN we use in this work first employs several recurrent layers to capture spectral dependencies in its input. Then, fully connected layers are utilized to implement classification based on the early extracted features that contain the contextual information.

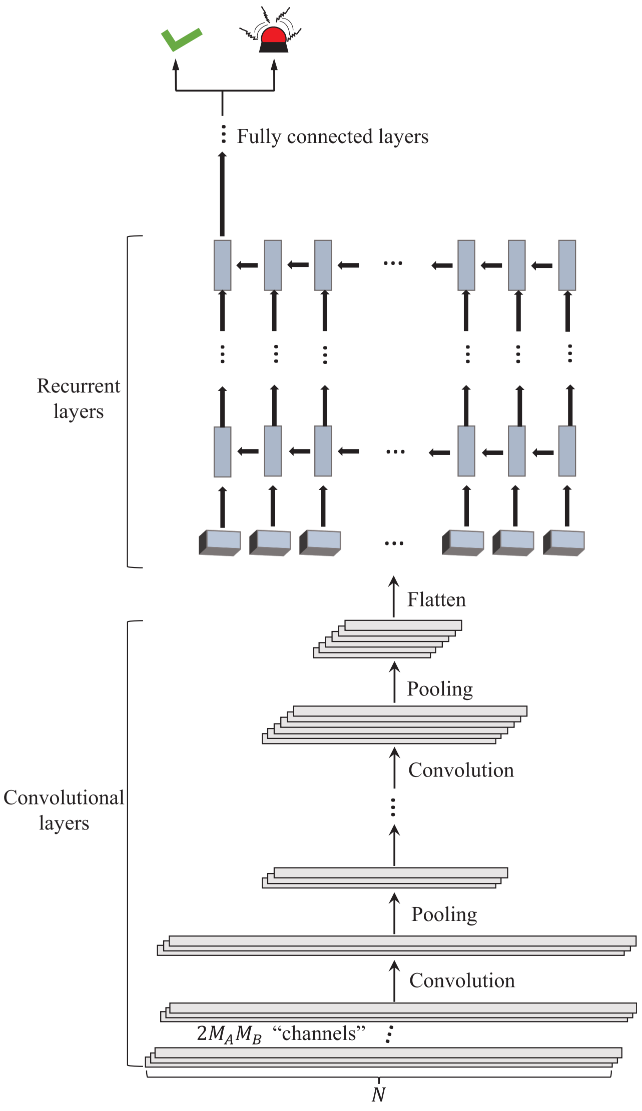


FIGURE 4. Illustration of our employed architecture of the CRNN.

3) CONVOLUTIONAL RECURRENT NEURAL NETWORKS

So far, we have introduced the CNN and the RNN for CSI-based authentication. It is easy to notice that these two DNNs have distinct modeling abilities since they rely on different mechanisms. To be specific, the CNN is good at capturing locally invariant features while the RNN is adept at contextual information extraction [23]. In this subsection, we propose to utilize a hybrid of the CNN and the RNN, i.e., the CRNN, that combines the abilities of the CNN and the RNN such that the deep features containing both the locally invariant information of CSI and the contextual messages between different frequencies in CSI can be well extracted and further be exploited for user authentication.

As schematically illustrated in Fig. 4, our employed CRNN is consisted of multiple convolutional layers (together with pooling layers), several recurrent layers, and a few fully connected layers. The mechanisms of these layers have been discussed above. In the CRNN, the convolutional layers can capture middle-level features, which are useful for the dependencies modeling at the recurrent layers. At the same time, the contextual information learned by the recurrent layers can lead to better representations at the convolutional layers during backpropogation. The CRNN takes good advantage of both the discriminative representation capability of the CNN and the contextual information extraction power of the RNN and is therefore expected to outperform both of them in CSI-based authentication.

B. NETWORK TRAINING

So far, we have illustrated the architectures of the introduced DNNs. For ease of the subsequent description, define $f_d(H(t), \boldsymbol{w})$ as the network function of a DNN, where all the weights and biases are grouped together into a vector $\boldsymbol{w}$ and $H(t)$ is the network input. Since the activation function of the output layer is a logistic sigmoid, we have $0 \leq f_d(H(t), \boldsymbol{w}) \leq 1$. One can interpret $f_d(H(t), \boldsymbol{w})$ as the conditional probability $p(I_t = 1|H(t), \boldsymbol{w})$, with $p(I_t = 0|H(t))$ derived as $1 - f_d(H(t), \boldsymbol{w})$. The conditional distribution of the class label given the input channel is a Bernoulli distribution [26], i.e.,

$$p(I_t|H(t), \boldsymbol{w}) = f_d(H(t), \boldsymbol{w})^{I_t}[1 - f_d(H(t), \boldsymbol{w})]^{1-I_t}. \tag{4}$$

Accordingly, a DNN-enabled authenticator can be written as

$$f_a(H(t)) = \lceil f_d(H(t), \boldsymbol{w}) - 1/2 \rceil , \tag{5}$$

in which $\lceil x \rceil$ denotes the ceiling function that maps x to the least integer greater than or equal to x. Now, the structure of our authenticator $f_a(\cdot)$ has been specified.

Given a training set of channels $\{H(t)\}_{t=1}^{T}$, together with a corresponding set of labels $\{I_t\}_{t=1}^{T}$, where T denotes the number of training samples, we train the network to minimize the error function, which is defined as a *cross-entropy* error function of the form

$$E(\boldsymbol{w}) = -\sum_{t=1}^{T}\{I_t \ln f_d(H(t), \boldsymbol{w})\} - \sum_{t=1}^{T}\{(1 - I_t)\ln[1 - f_d(H(t), \boldsymbol{w})]\}. \tag{6}$$

To train the network, we exploit the stochastic gradient descent (SGD) method. The gradients in the convolutional layers and the recurrent layers are calculated by the backpropagation algorithm and the backpropagation through time (BPTT) algorithm, respectively. To reduce the error fluctuation, our implementation utilizes a mini-batch strategy, that is, the gradients are calculated based on mini-batches. $\boldsymbol{w}$ will be iteratively updated until the training and validation loss converges. After the training process, the CSI-based authenticator $f_a(\cdot)$ can be fully derived. The DNNs-based authentication process is summarized in Algorithm 1.

Algorithm 1 DNN-Based Authentication Algorithm

```
Initialize: $f_a(\cdot)$ with $\boldsymbol{w}$
Record historical channel estimations $\{H(t)\}_{t=1}^{T}$ and the corresponding labels $\{I_t\}_{t=1}^{T}$
repeat
    Update $\boldsymbol{w}$ to minimize $E(\boldsymbol{w})$
until convergence achieved
Obtain the trained network $\hat{f}_a(\cdot)$
Receive a signal claimed to be Alice at time $T + 1$
Estimate the channel to derive $H(T + 1)$
Calculate $\hat{I}_{T+1}$ as $\hat{I}_{T+1} = \hat{f}_a(H_{T+1})$
if $\hat{I}_{T+1} = 0$ then
    Accept the current transmitter
Else
    Send spoofing alarm
End if
```

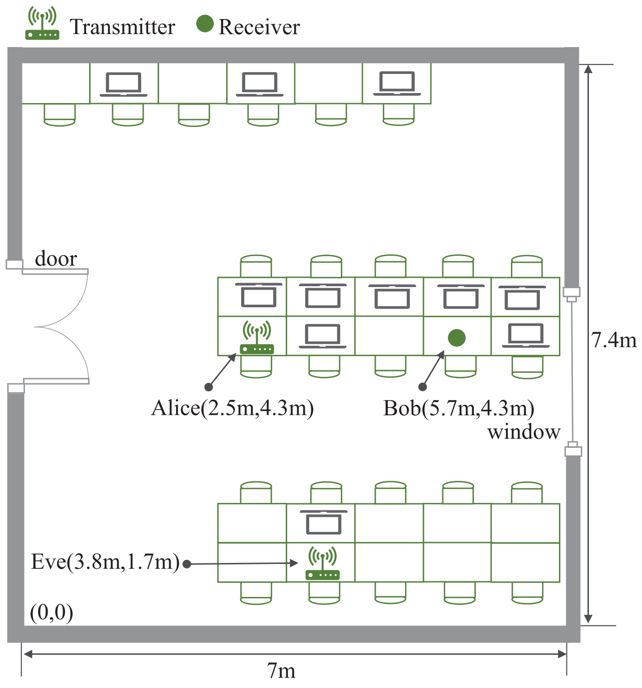


FIGURE 5. Network topology of the experiments.

IV. EXPERIMENTS

In our study, experiments on USRPs are conducted in an office so as to obtain real-world channel estimates. The real-world dataset is used to validate the performance of the proposed methods and other benchmark designs.

A. TESTBED SETUP

As illustrated in Fig. 5, the testbeds are deployed in a $7.4 \times 7 \times 5\text{m}^3$ office room, wherein there are two transmitters, i.e., Alice and Eve, and one receiver, namely, Bob. Specifically, we utilize a USRP-2955, a USRP-2954R, and a USRP-2944R to work as Bob, Alice, and Eve, respectively. Every USRP is connected to a computer through a Peripheral Component Interconnect Express (PCIE) bus.

TABLE 1. Summary of the network configurations.

CRNN	CNN	RNN	skip-layer CNN
conv1x3-32 maxpooling recur-128 FC-128-1	conv1x3-32 maxpooling FC-2048-1	recur-128 recur-256 FC-256-64 FC-64-1	conv1x3-32 conv1x3-32 max-pooling conv1x3-32 max-pooling FC-8192-1

TABLE 2. Performance for Different Methods on Real-World Data.

Method	Accuracy	False Alarm	Miss Detection
NP test	54.8%	16%	74.5%
DTW	59.1%	30.3%	51.5%
CNN	97.0%	3.0%	3.0%
RNN	98.6%	1.3%	1.5%
skip-layer CNN	99.1%	0.7%	1.1%
CRNN	99.7%	0.4%	0.3%

In such way, Laboratory Virtual Instrumentation Engineering Workbench [27] can be employed on the computer to generate the digital baseband transmission signal. With the digital baseband transmission signal received via the PCIE bus, the USRP-2954R and the USRP-2944R can perform digital-to-analog conversion and frequency upconversion on it and then send the converted signal via an omnidirectional antenna. After receiving the radio frequency signal, the USRP-2955 will downconvert it to baseband and do analog-to-digital conversion on the received baseband signal. Last, the channel is estimated by the least squares method at $N = 128$ tones. The experimental settings are as follows: Each USRP employs a single antenna. The transmit power is equal to 0.6mW for both the USRP-2954R and the USRP-2944R. 1000 channel observations are obtained at bob, 500 of which correspond to the legitimate user, while the rest of which is relevant to the illegitimate channel.

B. EXPERIMENTAL SETUP

We employ a workstation with two 1.7-GHz Intel(R) Xeon(R) E5-2603 v4 CPUs to perform experiments, in which the neural networks are implemented based on the TensorFlow framework [28]. We carry out the 20-fold Stratified cross validation to test the performance of the proposed machine learning schemes. After a careful grid search, our employed CNN has one convolutional layer, one pooling layer, and one fully connected layers. The RNN utilizes two recurrent layers and two fully connected layers. Besides, the CRNN has one convolutional layer, one pooling layer, one recurrent layer and one fully connected layers. The configurations of these networks are summarized in Table 1, in which “conv1×3-n_1” denotes a convolutional layer with a receptive field size of 1×3 and n_1 filters, “maxpooling” is a maxpooling layer, “recur-n_2” represents a recurrent layer whose feature dimension is n_2, and “FC-n_3-n_4” denotes a fully connected layer with n_3 input units and n_4 output units. The number of network parameters can be derived according to the settings in Table 1. To avoid overfitting, the recurrent layers in the RNN and the CRNN are applied with dropout rate [29] of 0.8 and 0.6, respectively. Z-score normalization is performed on the input data before it is feeded into the network. At the beginning of the training process, the network weights are randomly chosen, and the learning rate is set to be 10^{-3}. Then, we run the SGD algorithm for 100 epoches to update the network parameters, where each epoch utilizes all the training data in the mini-batch manner with a batch size of 256 and the learning rate halves every 20 epoches.

C. BENCHMARK DESIGNS

To make a comparison, we also consider the usage of a heuristic NP test, and the dynamic time warping (DTW) technique [25] for comparison. The NP test we consider is given by

$$L \triangleq ||H - \bar{H}_A|| \gtrless \gamma, \tag{7}$$

wherein $||X||$ is the Frobenius norm of the matrix X, H denotes the to-be-authenticated channel observation, $\bar{H}_A$ is the mean of the historical Alice-to-Bob channels, and γ represents a specially chosen threshold. The estimated identity will be Eve if $L > \gamma$, otherwise the transmitter will be authenticated as Alice. It needs to be mentioned that the authentication accuracy varies with γ and the authentication performance given in Table 2 is obtained with the optimal γ.

D. EXPERIMENTAL RESULTS

Table 2 presents the authentication accuracy, the false alarm rate and miss detection rate for different methods Note that for each method, the accuracy presented is obtained according to the authentication results on the whole test set. As shown in the table, there are huge performance gaps between the NP test and the machine learning-based methods. This is because one may need precise channel variation information to design a NP test-based algorithm that can deliver decent performance, while the machine learning-based algorithms can analyze the invariant channel structure intelligently from historical CSI. Also, Table 2 shows that DTW is no match for the proposed neural networks in dealing with the signal transformations existed in the channel observations. The significant performance gains achieved by the proposed schemes are attributed to the fact that our adopted DNNs not only have remarkable modeling abilities but also apply to the CSI-based authentication problem. As seen, the RNN yields lower false alarm rate and miss detection rate than the CNN. In addition, the CRNN-enabled method can achieve the lowest false alarm rate and miss detection rate among the propose DNNs-enabled algorithms. This is expected since the CRNN possesses both the representation ability of the CNN and the sequence modeling power of the RNN.

According to the suggestions of the reviewers, we try to further enhance the authentication performance by incorporating late deep networks [30], [31]. Fortunately, we find that the authentication performance of the CNN can be improved by utilizing the skip-layer architecture, the result of which is also presented in Table 2. The network configuration and

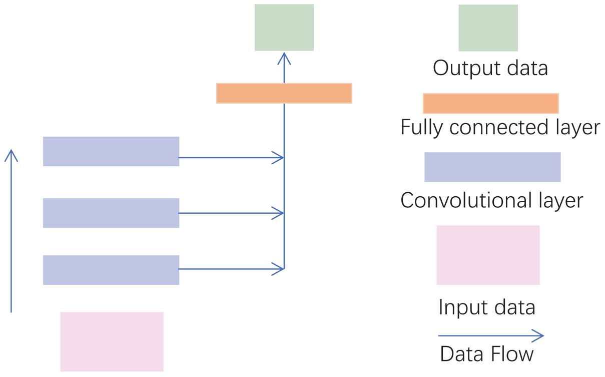


FIGURE 6. The architecture of the skip-layer CNN.

architecture of the skip-layer CNN are presented in Table 1 and Fig 6, respectively.

V. CONCLUSION

This work studied CSI-based authentication algorithms in a time-variant communication environment. Without knowledge of the underlying channel variation pattern, the DNNs were introduced to build authenticators that connect CSI to its authenticated identity. Regarding the innovation of this work, we first pointed out that the channel observations between two nodes at different time slots are transformed versions of each other, based upon which the CNN is proposed to extract the authentication features that are insensitive to transformations such as scaling and shifting. Next, we noticed the spectral dependencies as a natural characteristic of the CSI and accordingly, we propose to employ the RNN to exploit the dependencies in the CSI so as to achieve reliable authentication bases. Last but not least, we propose a combinatorial network, i.e., CRNN, which can harvest both of the above merits. According to the experimental results, the proposed DNNs-enabled authenticators can have significant performance gains over the benchmark schemes, while the combination of the CNN and the RNN can further enhance the authentication accuracy.

REFERENCES

[1] N. Yang, L. Wang, G. Geraci, M. Elkashlan, J. Yuan, and M. Di Renzo, "Safeguarding 5G wireless communication networks using physical layer security," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 20–27, Apr. 2015.

[2] 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.

[3] 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.

[4] P. Hao, X. Wang, and A. Behnad, "Relay authentication by exploiting I/Q imbalance in amplify-and-forward system," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 613–618.

[5] A. C. Polak, C. Dolatshahi, and D. L. Goeckel, "Identifying wireless users via transmitter imperfections," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 7, pp. 1469–1479, Aug. 2011.

[6] W. Hou, X. Wang, J.-Y. Chouinard, and A. Refaey, "Physical layer authentication for mobile systems with time-varying carrier frequency offsets," *IEEE Trans. Commun.*, vol. 62, no. 5, pp. 1658–1667, May 2014.

[7] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Using the physical layer for wireless authentication in time-variant channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2571–2579, Jul. 2008.

[8] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Channel-based spoofing detection in frequency-selective Rayleigh channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5948–5956, Dec. 2009.

[9] J. Liu and X. Wang, "Physical layer authentication enhancement using two-dimensional channel quantization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4171–4182, Jun. 2016.

[10] Y. Chen, J. Yang, W. Trappe, and R. P. Martin, "Detecting and localizing identity-based attacks in wireless and sensor networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2418–2434, Jun. 2010.

[11] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1106–1114.

[13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[14] W. Wang and J. Shen, "Deep visual attention prediction," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2368–2378, May 2018.

[15] X. Dong, J. Shen, W. Wang, Y. Liu, L. Shao, and F. Porikli, "Hyperparameter optimization for tracking with continuous deep Q-learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 518–527.

[16] X. Dong and J. Shen, "Triplet loss in siamese network for object tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 459–474.

[17] X. Dong, J. Shen, D. Wu, K. Guo, X. Jin, and F. Porikli, "Quadruplet network with one-shot learning for fast visual object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3516–3527, Jul. 2019.

[18] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.

[19] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017.

[20] H. Wu and S. Prasad, "Convolutional recurrent neural networks for hyperspectral data classification," *Remote Sens.*, vol. 9, no. 3, p. 298, 2017.

[21] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, "PHY-layer spoofing detection with reinforcement learning in wireless networks," *IEEE Trans. Vehic. Technol.*, vol. 65, no. 12, pp. 10037–10047, Dec. 2016.

[22] L. Xiao, X. Wan, and Z. Han, "PHY-layer authentication with multiple landmarks with reduced overhead," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1676–1687, Mar. 2017.

[23] Z. Zuo, B. Shuai, G. Wang, X. Liu, X. Wang, B. Wang, and Y. Chen, "Convolutional recurrent neural networks: Learning spatial dependencies for image representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 18–26.

[24] S. Mathur, R. Miller, A. Varshavsky, W. Trappe, and N. Mandayam, "Proximate: Proximity-based secure pairing using ambient wireless signals," in *Proc. 9th Int. Conf. Mobile Syst. Appl. Services*, Jun./Jul. 2011, pp. 211–224.

[25] J. Wang and D. Katabi, "Dude, where's my card?: RFID positioning that works with multipath and non-line of sight," in *Proc. ACM SIGCOMM*, Aug. 2013, pp. 51–62.

[26] C. M. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Heidelberg, Germany: Springer, 2006.

[27] (2007). *Developing Remote Front Panel LabVIEW Applications*. [Online]. Available: https://www.ni.com/white-paper/3277/en/

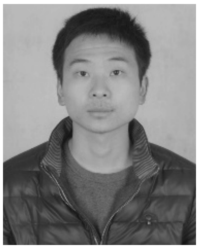
[28] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2016, *arXiv:1603.04467*. [Online]. Available: https://arxiv.org/abs/1603.04467

[29] 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. 1929–1958, 2014.

[30] W. Wang, J. Shen, and L. Shao, "Video salient object detection via fully convolutional networks," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 38–49, Jan. 2018.

[31] W. Wang, J. Shen, and H. Ling, "A deep network solution for attention and aesthetics aware photo cropping," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1531–1544, Jul. 2019. doi: 10.1109/TPAMI.2018.2840724.

QIAN WANG received the B.E. degree in communication engineering from the University of Electronic Science and Technology of China (UESTC), in 2015, where she is currently pursuing the Ph.D. degree with the National Key Laboratory of Science and Technology on Communications. Her research and study interests include network-connected UAV communications and physical layer security.



HANG LI received the B.E. and M.E. degrees in communication engineering from the University of Electronic Science and Technology of China (UESTC), in 2016 and 2019, respectively. His research and study interests include network-connected UAV communications and machine learning.

DOU ZHAO received the B.E. degree in communication engineering from Southwest Jiaotong University. She is currently pursuing the M.E. degree with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China. Her research and study interest includes physical layer authentication.

ZHI CHEN (M'08–SM'16) received the B.E., M.E., and Ph.D. degrees in electrical engineering from the University of Electronic Science and Technology of China (UESTC), in 1997, 2000, and 2006, respectively. In April 2006, he joined the National Key Laboratory of Science and Technology on Communications, UESTC, where he has been a Professor, since 2013. He was a Visiting Scholar with the University of California, Riverside, from 2010 to 2011. His current research interests include 5G mobile communications, tactile internet, and terahertz communication. Dr. Chen has served as a Reviewer for various international journals and conferences, including the IEEE Transactions on Vehicular Technology and the IEEE Transactions on Signal Processing.

SHUANG YE received the B.E. degree in communication engineering from the Huazhong University of Science and Technology. He is currently pursuing the M.E. degree with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China. His research and study interest includes physical layer authentication.

JIANSHENG CAI received the B.E. degree in communication engineering from Southwest Jiaotong University. He is currently pursuing the M.E. degree with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China. His research and study interest includes physical layer authentication.