electronics

MDPI

Article

Controllable Wireless Spoofing Attack Based on Conditional BEGAN and Auxiliary Channel Sensing

Mingjun Ma, Yan Zhang, Tianyu Zhao, Wancheng Zhang * and Zunwen He

School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
* Correspondence: zhangwancheng@bit.edu.cn

Abstract: This paper investigates how to build a controllable wireless spoofing attack launch framework that is driven by fundamental channel modeling and practical wireless datasets. First, we propose a wireless spoofing attack scheme against the defense mechanism with adversarial deep learning. To obtain channel characteristics and facilitate offline training of the attack model, auxiliary channel sensing is proposed with fundamental channel modeling. Based on these, a conditional boundary equilibrium generative adversarial network (CBEGAN) is designed with adversarial autoencoder (AAE), which takes true labels of signals and channel characteristics as conditions and enables the generation of controllable spoofing signals to fool the protected legitimate classifier. We verify the performance of the proposed spoofing attack scheme with CBEGAN and channel sensing by using wireless datasets, which contain signal data of multiple emitters and modulation types. Results show that the proposed scheme outperforms random attack, replay attack, and the recent attack scheme based on generative adversarial network (GAN) when a single legitimate emitter sends a fixed modulation type. It is also shown that the average attack success probability of the proposed CBEGAN attack model can reach more than 80% while mimicking multiple emitters and modulation types. The performance of the proposed scheme on different channel conditions including signal-to-noise ratio (SNR) and *K*-factor of the Rician fading channel is evaluated.

Keywords: adversarial machine learning; spoofing attack; general adversarial network (GAN); channel sensing; wireless security

Citation: Ma, M.; Zhang, Y.; Zhao, T.; Zhang, W.; He, Z. Controllable Wireless Spoofing Attack Based on Conditional BEGAN and Auxiliary Channel Sensing. *Electronics* **2023**, *12*, 84. https://doi.org/10.3390/electronics12010084

Academic Editor: Hirokazu Kobayashi

Received: 16 November 2022
Revised: 7 December 2022
Accepted: 19 December 2022
Published: 26 December 2022

1. Introduction

Due to the open and shared nature of propagation channels, wireless communication systems are vulnerable to being attacked by adversaries. Adversaries can launch spoofing attacks by simulating transmissions from legitimate users at the physical layer. These attacks can be used for various adversarial purposes, such as emulating the primary user in a wireless communication system and fooling the authentication system to attack a protected wireless network. Adversaries tried to send random signals from other transmitters as a naive spoofing attack. In addition, the replay attack was considered a common spoofing attack in previous studies. By recording a legitimate user's transmission and replaying the signal later, the replay attack can achieve reasonable results in traditional spoofing tasks.

As deep learning represents a very successful machine learning paradigm in the last decade, it has been widely applied in the fields of computer vision and natural language processing [1]. Deep learning can automatically build high–dimensional data models by processing raw data without handcrafted feature extraction. Therefore, deep learning is integrated into wireless communication systems, and automated approaches are provided. For example, ref. [2–4] used deep learning for spectrum sensing and radio signal classification, which includes two main techniques: automatic modulation classification (AMC) and specific emitter identification (SEI). The protection mechanism for wireless communication systems is built based on these two techniques for signal authentication. In [5–9], a deep learning–based approach, such as deep neural networks (DNN) or convolutional neural

networks (CNN), was used as a classifier to build signal authentication systems. With the successful applications of deep learning, the protection level of wireless communication systems is constantly improved, which makes replay attacks perform poorly. In [10,11], the limitations of replay attacks were analyzed and a scheme to detect replay attacks was proposed.

Consequently, the adversaries also need to improve their own level of intelligence to cope with the improvement of the defense capability of wireless communication systems. Such security issues have been studied in the field of adversarial machine learning [12–15]. There are three broad attack categories based on adversarial machine learning: exploratory attacks, poisoning attacks, and evasion attacks. For the first category, during the launching of exploratory attacks, adversaries can learn the transmit behavior of legitimate users to interfere with the data transmission process by training deep neural networks which learn the underlying transmission behavior [16,17]. In addition, adversaries can interfere with transmissions during the data collection process with poisoning attacks. In [18,19], poisoning attacks influence the input data during the training and testing of legitimate machine learning classifiers. Poisoning attacks are also studied for cooperative spectrum sensing, which provides incorrect sensing results to the machine learning classifier to launch spectrum sensing data falsification (SSDF) attacks [20,21]. Both exploratory attacks and poisoning attacks need to implement attacks during the training process of the legitimate classifier, which is difficult to realize in practice.

In addition, evasion attacks can fool classifiers to make incorrect classification decisions by adding adversarial perturbation to real samples [22–24]. Within this kind of method, spoofing attacks can be launched on the well-trained classifier without influencing the training process. In [25–30], the authors generated adversarial jamming inputs to fool machine learning and deep learning–based wireless signal classifiers by adding perturbations and inserting Trojans. Nevertheless, these methods assume that perturbations are directly added to real wireless signals, and thus they cannot be readily applied to evasion attacks.

In [31], researchers introduced generative adversarial network (GAN) into the use of spoofing attacks by generating synthetic signals as evasion attacks to fool classifiers. It has been verified that the GAN-based spoofing attacks outperformed previous works such as random attacks and replay attacks in radio signal classification tasks [31]. The attack scheme in [31] can only mimic a single emitter with a single modulation type in the attack launching process. However, there are multiple legitimate transmitters in actual wireless communication application scenarios. In addition, signals from legitimate transmitters may be sent with different modulation types. That makes the above attack scheme have difficulty meeting the needs of complex wireless communication applications. Furthermore, in [31], the attack scheme compensates for the effect of the transmission channel to generate more realistic spoofing signals by transferring training data and feedback from the GAN model during both the training and attack process. This online compensating scheme causes high overhead. It also makes adversaries easy to detect during the attack process because of the larger communication footprint.

To tackle the above problems, we propose a controllable wireless spoofing attack launching scheme based on conditional adversarial deep learning. By feeding the channel sensing data and signal labels as conditions into the designed boundary equilibrium generative adversarial network (BEGAN) based on adversarial autoencoder (AAE), the attack model can control the output category to launch controllable spoofing attacks. Compared to the recent comparative work, the proposed scheme provides innovative thinking for wireless signal spoofing. First, the proposed scheme can mimic multiple emitters with different modulation types in the attack launching process, whereas the existing attack scheme [31] can only mimic a single emitter with a single modulation type. Secondly, the proposed scheme can carry out the model training and channel compensation offline with auxiliary channel sensing. The existing attack scheme in [31] achieves the above purpose by transferring training data and feedback from the GAN model online,

which leads to a larger communication footprint and is easily detected. Moreover, the introduction of AAE also solves the problem that the performance of the attack model is affected due to the limited signal samples. The contributions of this work are summarized as follows.

- We propose a novel, controllable wireless spoofing attack scheme based on conditional BEGAN (CBEGAN) and auxiliary channel sensing. The CBEGAN is designed to take the channel characteristics and signal labels as conditions and to generate spoofing signals by mimicking multiple emitters with different modulation types.
- We design an AAE–based CBEGAN network to improve the quality of synthetic signals, which supports training with few samples. Auxiliary channel sensing is introduced to compensate for the effect of the transmission channel. With auxiliary channel sensing, the attack model can be trained offline, which is more covert and prevents interception by legitimate pairs due to long-term online training and transmission.
- Simulations are carried out to evaluate the performance of the proposed scheme. With different channel conditions including signal-to-noise ratio (SNR) and K-factor of the Rician fading channel, it is shown that the success probability of the proposed attack scheme outperforms random attacks, replay attacks, and the comparative scheme in [31]. It is also illustrated that the proposed scheme can realize controllable wireless spoofing attacks, i.e., it can mimic different legitimate emitters and modulation types while the comparative method can only simulate a single emitter with a fixed modulation type. Moreover, the performance of the proposed scheme under different channel conditions including SNR and K-factor of the Rician fading channel is evaluated.

The remainder of this paper is organized as follows. Section 2 introduces the system and channel model. Section 3 presents the proposed network and scheme for controllable spoofing attack launching based on CBEGAN with auxiliary channel sensing. Simulation setups, results, and discussions are provided in Section 4. Conclusions are drawn in Section 5.

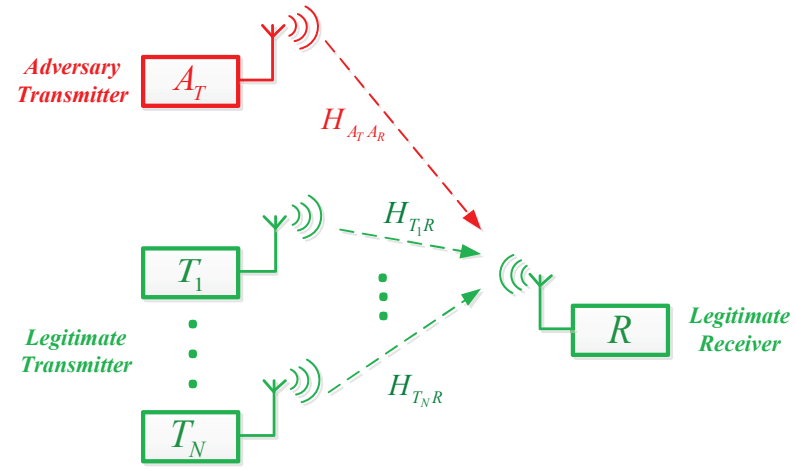
2. System and Channel Model

As illustrated in Figure 1, we consider a typical wireless communication system that consists of N legitimate transmitters and a legitimate receiver. Each legitimate transmitter can send signals with different modulation types. The signals from legitimate transmitters or illegitimate ones are distinguished at the legitimate receiver with a deep learning classifier trained by signal data collected in advance. In addition to distinguishing whether a transmission is legitimate or not, the receiver can also classify which legitimate transmitter the signal comes from and the modulation type of the transmitted signal.

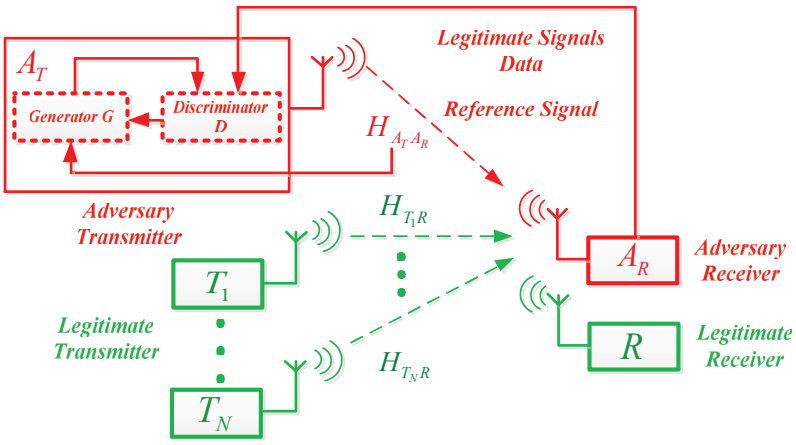
To realize the spoofing attack on the legitimate receiver, we introduce an adversary transmitter–receiver pair. The adversary transmitter attempts a spoofing attack on the legitimate receiver and mimics any legitimate transmitter with different modulation types. The adversary receiver is used to collect legitimate signals for training the attack model in the adversary transmitter. Therefore, the purpose of the spoofing attack is that the adversary can fool the legitimate receiver. The spoofing signal sent by the adversary can be classified as a legitimate signal at the legitimate receiver. The controllable spoofing attack launched by the adversary can also mimic each legitimate transmitter with different modulation types.

The attack and training processes of the proposed spoofing attack based on the system model are shown in Figure 1a,b. As shown in Figure 1a, legitimate transmitters T_i ($i = 1, \ldots, N$) send signals to the legitimate receiver R. The transmitted signals can conform to multiple modulation types. The adversary transmitter A_T attempts a spoofing attack on R and mimics any legitimate transmitter with different modulation types. To achieve the attack process above, the adversary receiver A_R is placed close to R which is the same as that in [28]. Therefore, signals from different legitimate transmitters, which are similar to those from T_i ($i = 1, \ldots, N$) to R, can be also collected at A_R. To achieve a controllable spoofing attack, a training process is needed. In the training process (see Figure 1b), A_T

sends a few reference signals to A_R for sensing the current adversary channel characteristics, which are able to compensate for the adversary transmission. A complete GAN network including the generator and discriminator is deployed in A_T for spoofing signal generation. With the channel characteristics from A_T to A_R implementing auxiliary GAN training, the discriminator compares the signals from T_i to A_R and those from A_T to A_R, and then gives the feedback to the generator. The generator improves synthetic signals to make them closer to legitimate signals. Once the GAN in A_T is trained, only the generator is used to generate spoofing signals in the attack process. A_T can mimic the signals sent by any legitimate transmitter to launch controllable spoofing attacks.



(a)



(b)

Figure 1. Network topology during the (**a**) attack and (**b**) training processes of spoofing attacks.

In the considered wireless scenario, the legitimate signal at R is

$$s_{T_i R}^{M_j} = x_{T_i}^{M_j} \cdot H_{T_i R} + \eta_i, \quad i \in [1, \ldots, N], \quad j \in [1, \ldots, L], \tag{1}$$

where $x_{T_i}^{M_j}$ is the signal sent by legitimate transmitter T_i with modulation type M_j, $H_{T_i R}$ is the channel impulse response from T_i to R, N is the number of legitimate transmitters, and L is the number of modulation types. Moreover, an additive white Gaussian noise (AWGN)

η_i, is considered. Similarly, the synthetic signal sent by the adversary transmitter A_T at receiver R with channel characteristics from A_T to A_R can be given as

$$s_{A_T A_R} = x_{A_T} \cdot H_{A_T A_R} + \eta_0, \tag{2}$$

where x_{A_T} is the synthetic signal sent by the adversary transmitter A_T and $H_{A_T A_R}$ is the channel impulse response from A_T to A_R. We aim to achieve to make $s_{A_T A_R}$ as similar as possible to $s_{T_i R}^{M_j}$.

The channel between any transmitter–receiver pair is depicted by the frequency-selective fading. It should be noted that the proposed spoofing attack scheme can be used in other channels as well because the auxiliary channel sensing can acquire the knowledge of the channel characteristics. For the fading channel model, the first discrete path experiences Rician fading and the other discrete paths follow independent Rayleigh fading. The Rician fading model takes into account the presence of the line of sight (LOS) path and the non-line of sight (NLOS) multipath components. The Rician K-factor is defined as the ratio between the power of the LOS path signal and the sum of the powers of the remaining NLOS multipath components. In this case, channel impulse response H can be expressed according to [32] as

$$H = \sqrt{\frac{1}{K+1}} \cdot H_{NLOS} + \sqrt{\frac{K}{K+1}} \cdot H_{LOS}, \tag{3}$$

where H_{NLOS} is the channel impulse response of the remaining NLOS multipath components, H_{LOS} is the channel impulse response of the LOS path, and K is the Rician K-factor. The Rician fading model turns into the Rayleigh fading model when K equals zero.

Within the proposed spoofing attack scheme, the adversary transmitter does not need to assume any prior knowledge about the legitimate transmission channel and the legitimate signal parameters such as modulation rate. By generating legitimate signals for each type ($s_{T_i R}^{M_j}$), the adversary can mimic different legitimate transmitters with different modulation types. The purpose of controllable spoofing attacks is achieved.

3. Controllable Spoofing Attack Launching

In this section, we introduce how to launch the controllable spoofing attack, and then present the structure of the AAE–based CBEGAN network. Due to the introduction of conditions including the auxiliary channel sensing data and signal labels, the designed AAE–based CBEGAN can generate synthetic signals from different transmitters with different modulation types even if the legitimate transmission channel condition is variant. By feeding the channel characteristics into the network, the network can be trained in an offline learning manner.

3.1. CBEGAN and Auxiliary Channel Sensing

The spoofing attack launching scheme and the functions of the component modules are shown in Figure 2. The proposed scheme consists of CBEGAN and auxiliary channel sensing. CBEGAN is used for the generation of synthetic signals and the launching of spoofing attacks. Auxiliary channel sensing is deployed to acquire the channel characteristics to assist the training of the attack model. BEGAN is applied as a boosting model for generative adversarial networks in the body of the attack model for training and generation of spoofing signals. Unlike the optimization objective of conventional generative adversarial networks, BEGAN [33] makes the auto–encoding error distribution of the generated samples match that of the real samples as much as possible. This is more conducive to the generation of time sequence data such as modulated communication signal data.

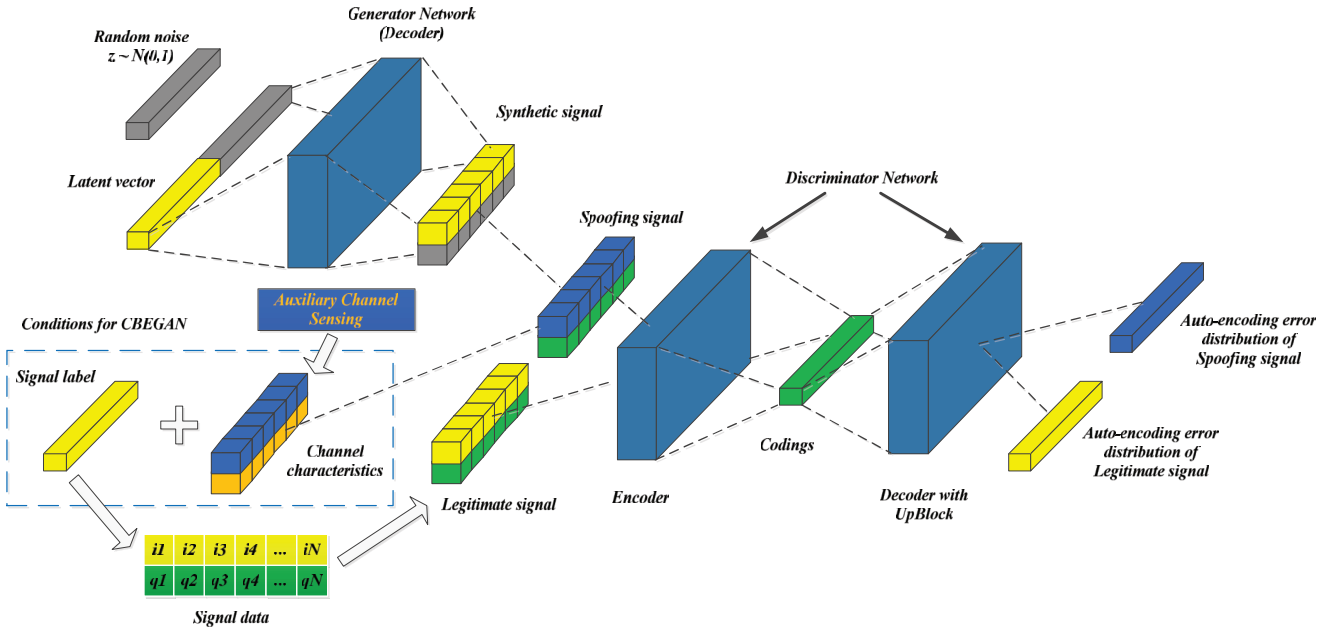


Figure 2. The structure of the proposed scheme based on CBEGAN and auxiliary channel sensing.

In addition, auxiliary channel sensing is used to obtain the channel characteristics during the training and launching of spoofing attacks. The introduction of auxiliary channel sensing allows training to be performed only in the training stage while reducing the exposure risks in the attack launching stage. Furthermore, channel characteristics obtained by auxiliary channel sensing and the information about the legitimate transmission (the true labels of signals transmitted) are fed into CBEGAN as conditions. This completes the architecture of the CBEGAN attack model, which enables controllable spoofing attack launching.

The detailed process of launching a controllable spoofing attack based on CBEGAN and auxiliary channel sensing is presented as follows. The controllable spoofing attack procedure is divided into two stages: the model training stage and the attack launching stage.

In the training stage, the adversary receiver A_R collects the signal samples from transmitter T and the adversary transmitter A_T, and marks their transmissions with labels so that the adversary receiver A_R can obtain the true label of the transmission. Channel sensing characteristics are also fed into the discriminator D. The discriminator D then starts training based on the collected data samples, with the objective of maximizing the auto–encoding error, i.e.,

$$\max_D L\left(s_{T_iR}; l_{T_iR}; \theta_D\right) - L(G(z; l_{T_iR}; \theta_G) \cdot H_{A_TA_R}; \theta_D), \tag{4}$$

where z is the random noise input of the generator G, $\mathbf{s}_{T_iR}$ is the legitimate signal of category i, l_{T_iR} is the signal label of s_{T_iR}, and $H_{A_TA_R}$ is the channel sensing data of the adversary transmission channel. $G(z)$ is the generator output, which is also the synthetic signal sent by the adversary transmitter. θ_G and θ_D are the current model weights. L is the auto–encoding error.

The generator G collects the feedback of the discriminator D and updates the generator weights to improve the quality of the synthetic data. The objective is to minimize the auto−encoding error of the sample generated by the generator G, which can be expressed as

$$\min_G L\left(G\left(z; l_{T_iR}; \theta_G\right) \cdot H_{A_TA_R}; \theta_G\right). \tag{5}$$

The above training process continues with updated generator G and discriminator D. The weights are continuously updated with each round of training as well until CBEGAN reaches equilibrium. Besides, it should be noted that when the generator G is trained, the gradient of G vanishes rapidly, which makes the training of CBEGAN very difficult. To relax the equilibrium boundary and solve the vanishing gradient, the hyperparameter γ is introduced as

$$\gamma = \frac{E\left[L\left(G\left(z; l_{T_iR}; \theta_G\right) \cdot H_{A_TA_R}; \theta_G\right)\right]}{E\left[L\left(s_{T_iR}; l_{T_iR}; \theta_D\right)\right]}, \gamma \in [0, 1]. \tag{6}$$

The whole process forms a complete generative adversarial network, which performs a minimax game as

$$\begin{cases} \max\limits_{D} L\left(s_{T_iR}; l_{T_iR}; \theta_D\right) - k_t \cdot L(G(z; l_{T_iR}; \theta_G) \cdot H_{A_TA_R}; \theta_D), \\ \min\limits_{G} L\left(G\left(z; l_{T_iR}; \theta_G\right) \cdot H_{A_TA_R}; \theta_G\right), \\ k_{t+1} = k_t + \lambda_t(\gamma \cdot L\left(s_{T_iR}; l_{T_iR}; \theta_D\right) - L\left(G\left(z; l_{T_iR}; \theta_G\right) \cdot H_{A_TA_R}; \theta_G\right). \end{cases} \tag{7}$$

λ_t is a learning rate parameter of k_{t+1}, and k_t controls the gradient drop degree of D where t denotes the training step. The parameter k_t is continuously updated with the above objective function optimized.

Once the model converged, the adversary transmitter A_T can use the generator to generate synthetic signals in the attack stage. After passing through the transmission channel, the synthetic signal is received by the legitimate receiver which is similar to the signal sent from the legitimate transmitter. The controllable spoofing attack is successfully launched. The entire controllable spoofing attack launching procedure is organized into an Algorithm 1 shown below.

Algorithm 1: Controllable Spoofing Attack Launching Procedure.

Input:
Random noise input of the generator G: z_G
Sample of the legitimate signal i: s_{T_iR}
Label of the legitimate signal i: l_{T_iR}
Channel sensing data of the adversary transmission channel: $H_{A_TA_R}$
Output:
Synthetic signal of generator G: $G(z_G; l_{T_iR}; \theta_G) \cdot H_{A_TA_R}$
Attack Model Training Stage:
Feed the input data into the CBEGAN network and set the training hyperparameters of model;
for $t = 0$ *to steps - 1* **do**
  Initialize the network parameters θ_G and θ_D;
  Generate synthetic signal based on channel sensing data and signal label;
  Updates the discriminator D parameters θ_D with maximizing the auto-encoding error:
  $L\left(s_{T_iR}; l_{T_iR}; \theta_D\right) - L(G(z_G; l_{T_iR}; \theta_G) \cdot H_{A_TA_R}; \theta_D)$;
  Updates the generator G parameters θ_G by minimizing the autoencoding error:
  $L\left(G\left(z_G; l_{T_iR}; \theta_G\right) \cdot H_{A_TA_R}; \theta_G\right)$;
  Calculate the hyperparameter γ and update the parameters k_t by γ and λ_t;
end
Save the generator parameters θ_G;
Attack Launching Stage:
Load the network parameters θ_G into the generator of CBEGAN;
Identify the transmission state of the channel;
Launch a controllable spoofing attack based on channel sensing data and the signal labels.

3.2. AAE-Based CBEGAN Model

In the above wireless application scenario, the invisibility and security of the spoofing attack launching need to be fully considered. The adversary needs to launch wireless spoofing attacks based on the collected legitimate signals. If the data collection process for attack model training lasts too long, it makes adversaries easy to detect because of the larger communication footprint. Thus, only limited training samples can be collected and utilized, which leads to the need to solve the problem of the training of the attack model with a limited number of samples. AAE [34] is a promising deep neural network architecture for computer vision-related tasks, which enables learning with few samples. In this paper, we introduce the AAE to our CBEGAN to solve the problem of not having access to a large number of training samples.

The AAE–based CBEGAN architecture designed for spoofing signal generation task is introduced as follows. Unlike the image generation task, within which the training samples are two-dimensional matrices, the samples of the spoofing signal generation task are one-dimensional vectors. In addition, the signal samples are signals sampled in the baseband. For each signal sample, I and Q are taken as two channels. A one-dimensional vector with two channels is formed as the final training sample. Similarly, according to the characteristics of the signal samples, we consider one-dimensional convolution kernels as the main operational units of the generative adversarial network.

Specifically, for the encoder in autoencoder, the network structure of multiple one-dimensional convolution layers followed by fully connected layers is selected. In order to accelerate the convergence of the model, a batch-normalization operation is added after each fully connected layer. The detailed structure of the encoder is shown in Figure 3. The generator and the decoder take the same network structure, i.e., a network structure of two fully connected layers added several UpBlock modules and convolutional output layers. The overall architecture of the generator/decoder is given in Figure 4. To enhance the fineness of the generated signals, we add the output of the fully connected layer to the output of each UpBlock module by channel splicing. Figure 4 shows the detailed structure of each UpBlock module. There are two inputs and two outputs in the UpBlock module. One input goes through several convolutional output layers before being concatenated with another input.

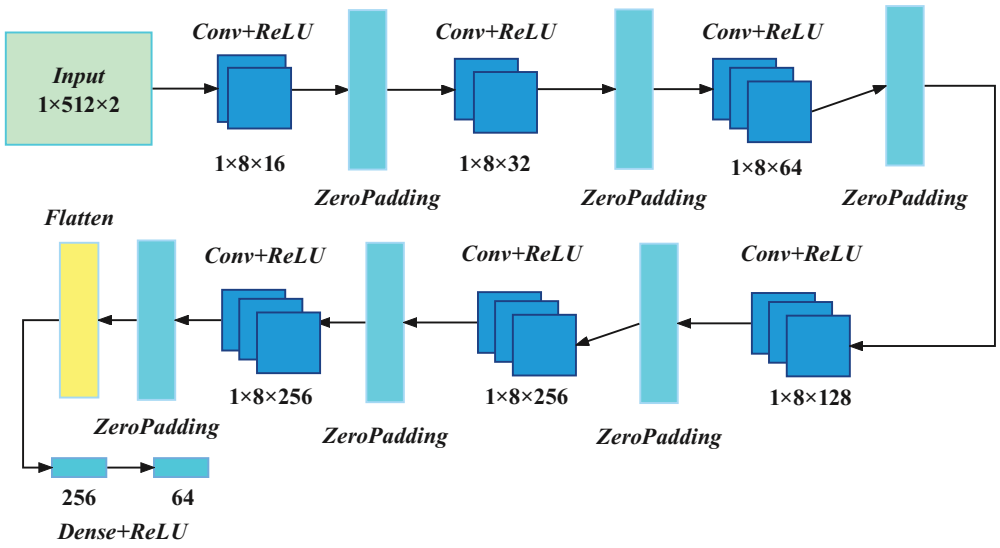


Figure 3. The architecture of the discriminator (encoder).

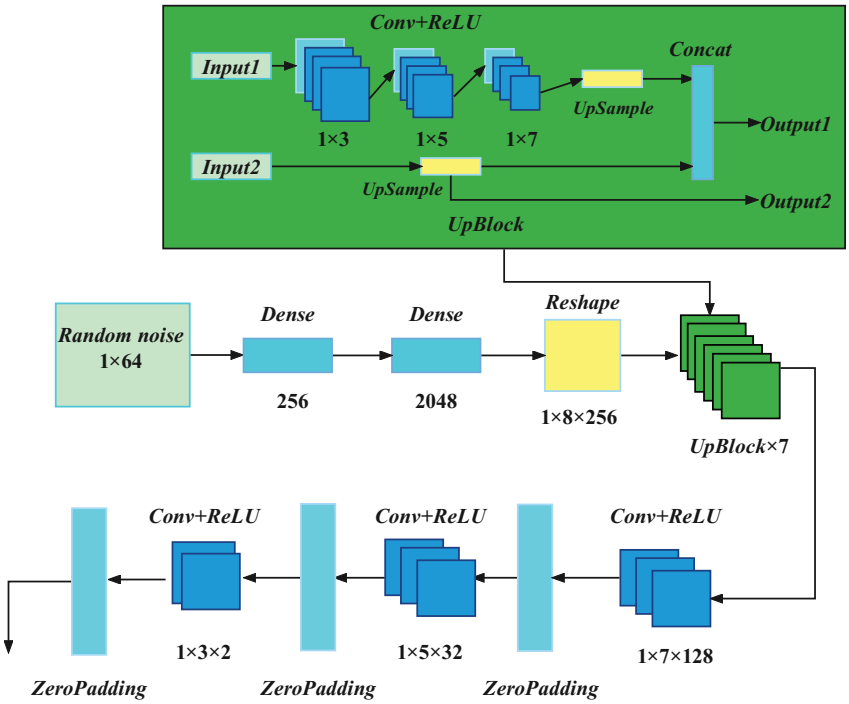


Figure 4. The architecture of the generator (decoder).

4. Numerical Results and Analysis

In this section, simulations are carried out to verify the performance of the proposed scheme. Its ability to mimic multiple emitters or multiple modulation types is also demonstrated. The performance of the proposed method with auxiliary channel sensing under different channel conditions is discussed.

4.1. Dataset and Network Setups

According to the system model in Figure 1, 8 universal software radio peripherals (USRPs) of X310 produced by Ettus with a single antenna are taken as transmitters in the wireless scenario. Five different USRPs are taken as legitimate transmitters (T_1 to T_5) and one for the legitimate receiver (R). The other two USRPs are used as the adversary transmitter–receiver pair (A_T and A_R). Each transmitter and receiver is composed of GNU radio and USRP. The USRP radio frequency parameters are configured with GNU radio. During the measurements, all USRPs work at 900 MHz central frequency with 40 MHz bandwidth. The gain value is set to 89 dB. All transmitters can send the signal data with the same preamble, modulation rate of 120 kBaud, and receiving sampling rate of 2.4 MSps. The received signals are upsampled by a factor of 10, and $2N$ real samples are collected to form one $1 \times N$ complex–valued input vector, which can be decomposed into two $1 \times N$ real–valued vectors. In the experiments, N is set to 512. There are five different modulation types transmitted by USRPs including BPSK, 2ASK, QPSK, 16QAM, and GMSK.

For each emitter, radio signals with different modulation types are sent under fading channels with AWGN. Specifically, frequency–selective fading [32] is considered, and the number of discrete paths is set to three. The first discrete path experiences Rician fading. The Rician K-factor is set to a range from 5 to 25 dB. The other two discrete paths follow independent Rayleigh fading. The SNR of radio signals is also set to a range from 5 to 25 dB. For each emitter and each modulation type, 1000 samples are generated under different Rician K-factors and SNRs. These data are used for the training of legitimate classifiers as well as attack models in subsequent simulations.

The AAE–based CBEGAN is designed based on the structure and parameters in the previous section. The training hyperparameters of the AAE–based CBEGAN include

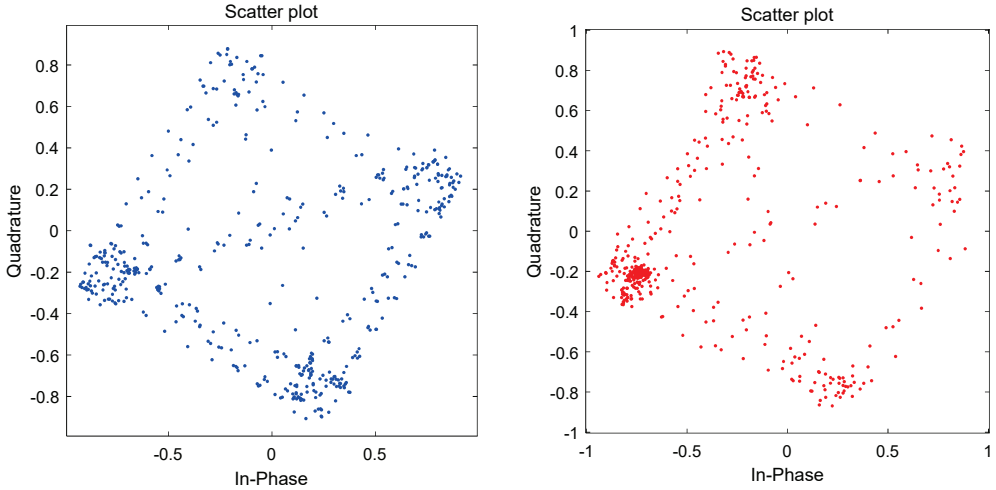
batchsize, the learning rate of the generator and discriminator, γ, k_0, and λ_t. The training hyperparameters are optimized through Grid Search to achieve the highest success probabilities of spoofing attacks. The final training hyperparameters of AAE–based CBEGAN are illustrated in Table 1. The same settings of training hyperparameters are used in all subsequent simulations.

Table 1. The training hyperparameters of CBEGAN.

Hyperparameter	Value
Batchsize	8
Learning rate of G	0.0001
Learning rate of D	0.0001
γ	0.9
k_0	0
λ_t	0.001

4.2. Results on Spoofing Signal Generating

A comparison between the constellation diagrams of the legitimate signal and the spoofing signal generated by the proposed scheme collected at R is shown in Figure 5. The constellation diagram of the spoofing signal conforms to the modulation characteristics of the QPSK signal in Figure 5. The rotation of the spoofing signal constellation is similar to that of the legitimate signal. It can be seen that the spoofing signal is similar to the legitimate QPSK signal.



(a) (b)

Figure 5. A comparison between the constellation diagrams of (**a**) legitimate and (**b**) spoofing QPSK signals.

Figure 6 shows the constellation diagrams of BPSK signals from five different legitimate emitters and the corresponding spoofing signals. The rotation of the constellation reflects the phase shift resulting from the influence of the hardware device. It is a unique property inherent to each emitter. These emitter properties can be used for multiemitter classification tasks. In Figure 6, each emitter has its own property and the proposed scheme mimics each emitter successfully.

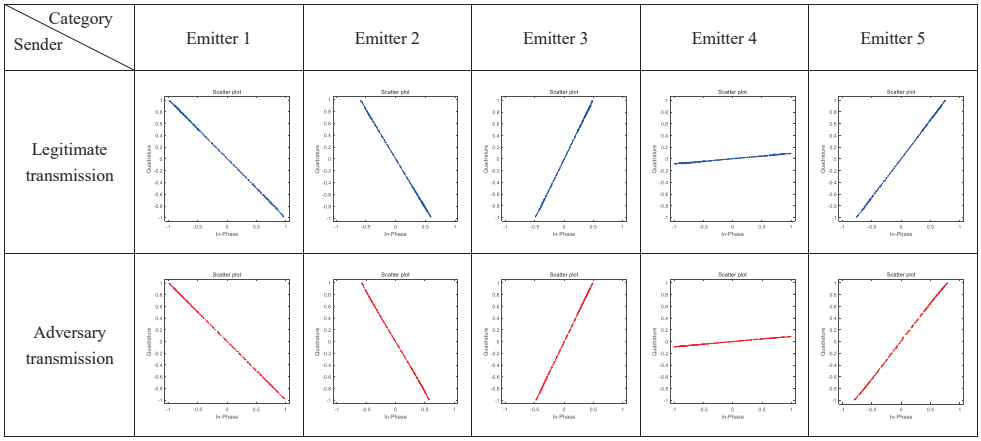


Figure 6. The constellation diagrams of legitimate and spoofing BPSK signals with different legitimate emitters.

4.3. Performance on Spoofing Attack against Related Works

This simulation aims to compare the performance of the proposed spoofing attack scheme with the related works. Random attack and replay attack [10] are considered baselines in the simulation. In addition, a recent wireless signal spoofing method [31] is included as a comparison in the simulation. This method in [31] is based on GAN and its good performance on spoofing deep learning–based classifier has been verified. Because previous works can only distinguish whether the signal is legitimate or not, only two transmitters are selected here to send modulated signals as legitimate and illegitimate transmitters. As for the binary classification, the legitimate classifier is designed as a DNN network, as was [31]. The structure of the DNN network is shown in Figure 7, where the model parameters are set to three fully connected layers with 50 neurons. ReLU is used as the activation function of each fully connected layer and Softmax is used as that of the output layer with two neurons. For classifying legitimate and illegitimate signals, 1000 signal samples are used for testing. The recognition accuracy rate is 96.9%, the false detection rate is 2.6%, and the false alarm rate is 4.6%, indicating that the classifier can achieve effective signal classification. In addition, the parameter settings of the GAN network refer to the settings in [31]. Denote sp as the success probability of spoofing attacks, which is calculated by $sp = \frac{n_T}{n}$. For a given test data with n attacks, denote n_T as the number of attacks classified as legitimate.

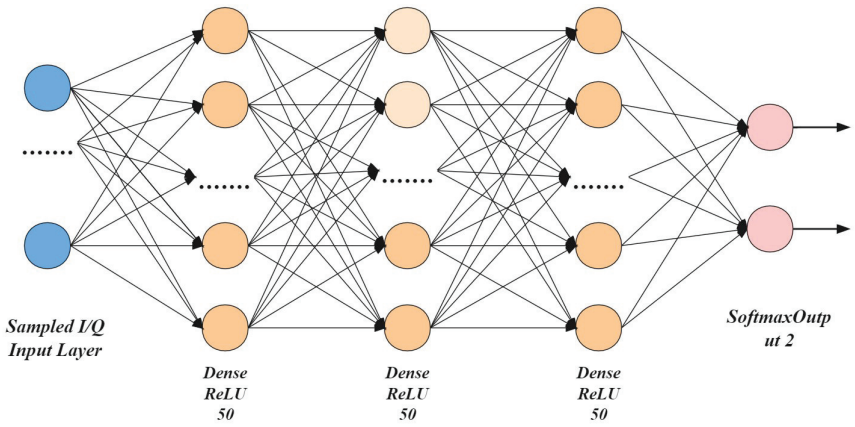


Figure 7. The architecture of the DNN network for classifying legitimate/illegitimate signals.

Each of the above–mentioned spoofing attacks is launched against the pretrained legitimate DNN classifier. In order to reproduce the results of related works, the purpose is achieved by adjusting the transmission channel conditions including SNR and Rician

K-factor. The results show that the replay attack is better than transmitting random signals, i.e., the success probability of the replay attack is 36.2%, which is much larger than the success probability 7.71% if A_T transmits random signals. Moreover, the comparative attack scheme can achieve a high success probability with 76.2%. With the same channel conditions, the success probability of the proposed spoofing attack scheme increased to 85.7%. The improvement in the success probability is due to the introduction of AAE–based CBEGAN to further improve the quality of signal generation. To further verify the advantages of the proposed scheme under different channel conditions, each type of spoofing attack is tested 1000 times under different channel conditions to ensure the reliability of the results. In Figure 8, the proposed scheme achieved the highest success probability of attack when SNR and Rician *K*-factor come to 20 dB. As the channel conditions get worse, the success probabilities of other attack schemes drop sharply, especially wireless signal spoofing based on GAN, which has fallen to less than 50%. In contrast, the proposed scheme can still achieve a success probability of nearly 60% with SNR and Rician *K*-factor of 10 dB. It is illustrated that the proposed scheme outperforms other related works with a single emitter and modulation type under different channel conditions.

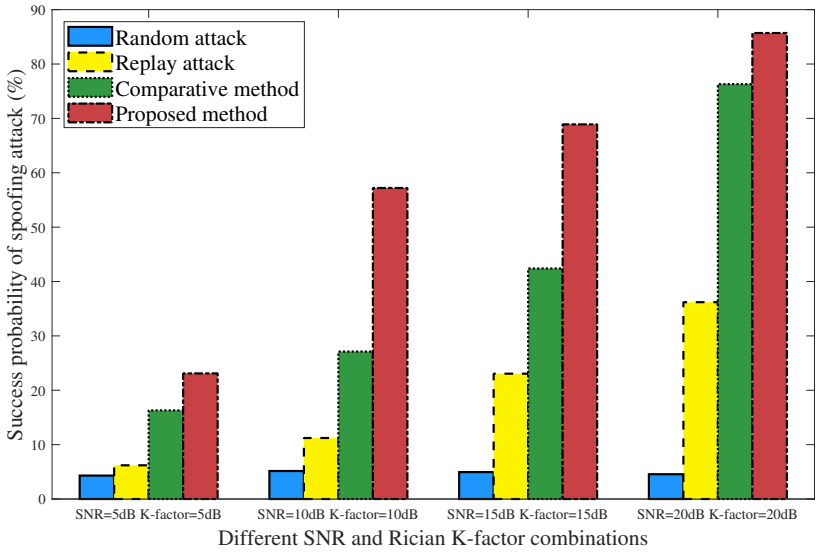


Figure 8. The success probability of spoofing attack versus attack schemes for different channel conditions (different SNRs and Rician *K*-factors).

4.4. Performance on Controllable Spoofing Attack

This simulation aims to test the effectiveness of the controllable spoofing attack scheme in multiple emitters and multiple emitters' scenarios. As there are five legitimate transmitters, a legitimate deep learning multiclassifier is constructed by Tensorflow for the five-class classification. The legitimate classifier is the six–layer CNN which has been widely used in the field of radio signal classification [3]. The main structure is the linear connection of the convolutional layer and the fully connected layer. ReLU and SoftMax are used as the activation function of the output layer. The detailed structure of CNN used here is shown in Figure 9. For both multiemitter and multimodulation classification tasks, 5000 signal samples are used for the training of the CNN models. The training dataset uniformly contains each modulation type as well as the emitter. The recognition accuracy rate of the multiemitter classification is 95.5%, whereas that of the multimodulation classification is 94.1%. For each type of signal, the success probability of spoofing attacks can be interpreted as the same as above. Because related works focused on a single modulation type or a single emitter, no comparison simulation is set up here.

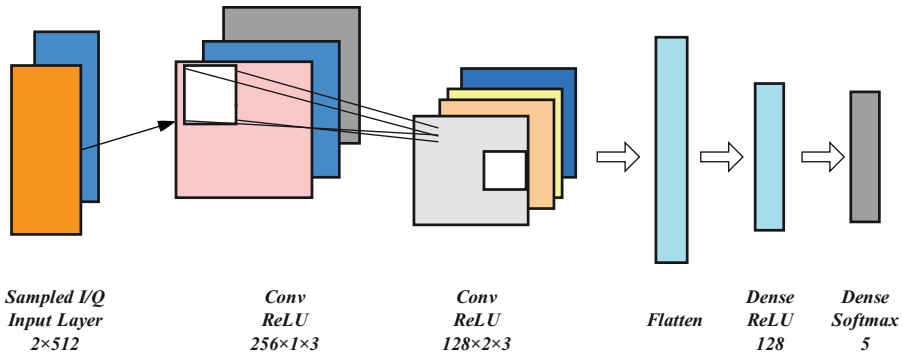


Figure 9. The architecture of the CNN network for classifying multiple modulation types and multiple emitters.

In the simulation of dealing with multimodulation and multiemitter scenarios, the performance of the controllable spoofing attack is verified by adequate numerical results.

Figure 10a,b show the performance of the proposed scheme under different SNRs and different Rician *K*-factors in scenarios with multiple modulation types. In the analysis simulation of the effect of the SNR, we set Rician *K*-factor to 20 dB. It is shown that the success probabilities of the proposed scheme with all modulation types increase with the increase of the SNR. In addition, when the SNR reaches 15 dB, the success probabilities of spoofing attack are over 80% for all modulations which means that a relatively good spoofing attack effect has been achieved. Moreover, the success probabilities of 2ASK and BPSK signals can both exceed 90% and the former even reaches 92% when SNR is 25 dB. In order to fully verify the effect of Rician *K*-factor on the success probability of spoofing attacks in multiclassification scenarios, the SNR is set to 20dB to reduce the effect of noise. It can be seen from Figure 10b that in the lower Rician *K*-factor case (5dB), the success probabilities of all modulation types drop sharply to about 30%. On the other hand, the success probability of the proposed scheme exceeds 80% for all modulation types when Rician *K*-factor reaches 20 dB.

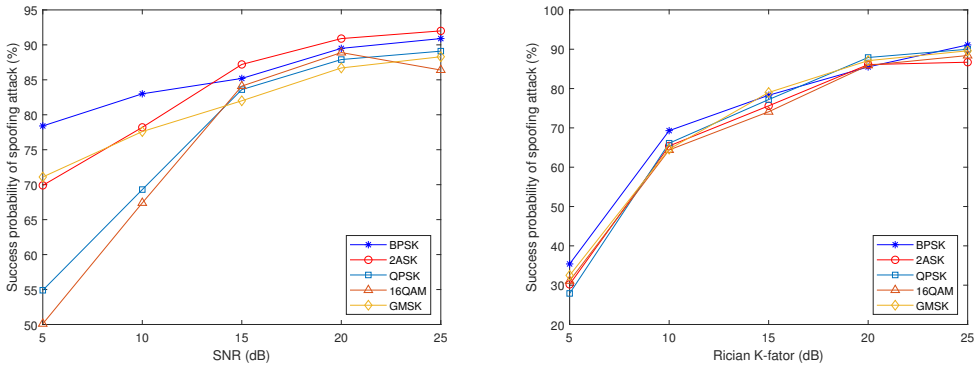
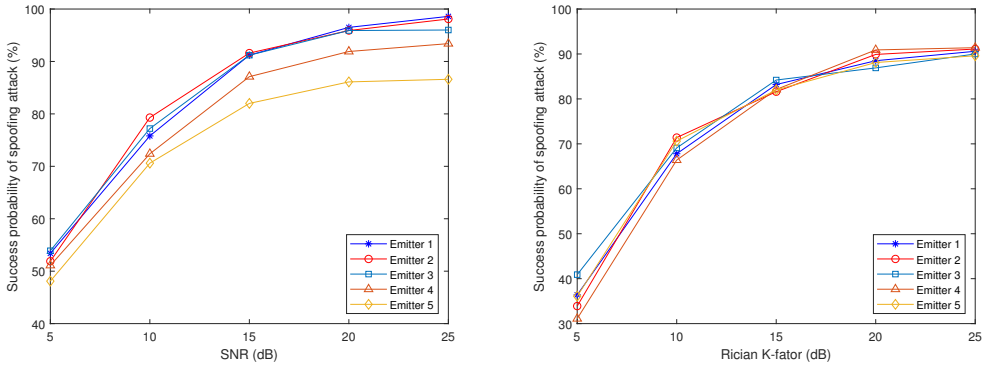


Figure 10. The success probability of spoofing attack versus (**a**) SNR and (**b**) Rician *K*-factor for multiple modulation types.

Figure 11a,b shows the performance of the proposed scheme under different SNRs and different Rician *K*-factors in scenarios with multiple emitters. In comparison, multiemitter spoofing attacks achieve exceptional performance with SNR of 25 dB. Most multiemitter spoofing attacks realize a high success probability of over 95%. On the contrary, with lower SNR (10 dB), the success probabilities of most emitters are within 70–80% which is still acceptable. Moreover, it can be observed that the success probabilities of the proposed

scheme with all emitters increase with the increase of Rician *K*-factor following a similar trend. In addition, when the Rician *K*-factor ratio reaches 15 dB, the success probabilities of all emitters are over 80%. It is illustrated that our proposed scheme can be practicable under different channel conditions.



(a) (b)

Figure 11. The success probability of spoofing attack versus (**a**) SNR and (**b**) Rician *K*-factor for multiple emitters.

5. Conclusions

In this paper, we proposed a controllable wireless spoofing attack scheme against the deep learning-based classifier with AAE–based CBEGAN and auxiliary channel sensing. By training a CBEGAN with the joint conditions including channel sensing characteristics and embedding signal labels, the spoofing attack could be launched controllable. AAE was introduced in the architecture of CBEGAN to train the networks with few samples. Simulations were conducted by using USRPs to capture the wireless signal and propagation data for a multiemitter wireless scenario. Numerical results verified the advantages of the proposed scheme over the random attack, replay attack, and the comparative attack scheme based on GAN in the scenario where a single emitter sent signals with only one modulation type. It was also shown that the proposed scheme could mimic different emitters with different modulation types. The performance of the proposed scheme under different channel conditions was evaluated as well.

Author Contributions: Conceptualization, M.M. and Y.Z.; methodology, M.M.; software, M.M. and T.Z.; validation, M.M. and T.Z.; formal analysis, M.M. and Y.Z.; investigation, M.M.; resources, M.M.; data curation, M.M. and T.Z.; writing—original draft preparation, M.M.; writing—review and editing, Y.Z.; visualization, M.M.; supervision, Y.Z. and W.Z.; project administration, W.Z.; funding acquisition, Z.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Key Research and Development Program of China (No.2019YFE0196400) and the National Natural Science Foundation of China (No.62271051 and No.61871035).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
2. Lee, W.; Kim, M.; Cho, D. Deep Cooperative Sensing: Cooperative Spectrum Sensing Based on Convolutional Neural Networks. *IEEE Trans. Veh. Technol.* **2019**, *68*, 3005–3009. [CrossRef]
3. O'Shea, T.J.; Corgan, J.; Clancy, T.C. Convolutional Radio Modulation Recognition Networks. In Proceedings of the 2016 International Conference on Engineering Applications of Neural Networks, Aberdeen, UK, 2–5 September 2016.
4. O'Shea, T.J.; Roy, T.; Clancy, T.C. Over-the-Air Deep Learning Based Radio Signal Classification. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 168–179. [CrossRef]

5. Li, R.; Li, L.; Yang, S.; Li, S. Robust Automated VHF Modulation Recognition Based on Deep Convolutional Neural Networks. *IEEE Commun. Lett.* **2018**, *22*, 946–949. [CrossRef]
6. Lin, C.; Huang, J.; Huang, S.; Yao, Y.; Guo, X. Features Fusion based Automatic Modulation Classification Using Convolutional Neural Network. In Proceedings of the IEEE INFOCOM 2020—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Piscataway, NJ, USA, 2020; pp. 1099–1104.
7. Ding, L.; Wang, S.; Wang, F. Specific Emitter Identification via Convolutional Neural Networks. *IEEE Commun. Lett.* **2018**, *22*, 2591–2594. [CrossRef]
8. Wong, L.J.; Headley, W.C.; Andrews, S. Clustering Learned CNN Features from Raw I/Q Data for Emitter Identification. In Proceedings of the MILCOM 2018—2018 IEEE Military Communications Conference (MILCOM), Los Angeles, CA, USA, 29–31 October 2018.
9. Tian, F.; Wang, L.; Xia, M. Signals Recognition by CNN Based on Attention Mechanism. *Electronics* **2022**, *11*, 2100. [CrossRef]
10. Hoehn, A.; Zhang, P. Detection of replay attacks in cyber-physical systems. In Proceedings of the 2016 American Control Conference (ACC), Boston, MA, USA, 6–8 July 2016; pp. 290–295.
11. Kinnunen, T.; Sahidullah, M.; Delgado, H.; Todisco, M.; Evans, N.; Yamagishi, J.; Lee, K.A. The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection. In Proceedings of the Interspeech, Stockholm, Sweden, 20–24 August 2017.
12. Shi, Y.; Sagduyu, Y.E.; Davaslioglu, K.; Levy, R. Vulnerability detection and analysis in adversarial deep learning. In *Guide to Vulnerability Analysis for Computer Networks and Systems—An Artificial Intelligence Approach*; Springer: Cham, Switzerland, 2018.
13. Shi, Y.; Sagduyu, Y.E.; Erpek, T.; Davaslioglu, K.; Lu, Z.; Li, J.H. Adversarial Deep Learning for Cognitive Radio Security: Jamming Attack and Defense Strategies. In Proceedings of the 2018 IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.
14. Kong, Y.; Li, X.; Hao, G.; Liu, C. Face Anti-Spoofing Method Based on Residual Network with Channel Attention Mechanism. *Electronics* **2022**, *11*, 3056. [CrossRef]
15. Inayat, U.; Zia, M.F.; Mahmood, S.; Khalid, H.M.; Benbouzid, M. Learning-Based Methods for Cyber Attacks Detection in IoT Systems: A Survey on Methods, Analysis, and Future Prospects. *Electronics* **2022**, *11*, 1502. [CrossRef]
16. Erpek, T.; Sagduyu, Y.E.; Shi, Y. Deep learning for launching and mitigating wireless jamming attacks. *IEEE Trans. Cogn. Commun. Netw.* **2019**, *5*, 2–14. [CrossRef]
17. Liang, Y.; Cai, Z.; Han, J.; Yu, Q.; Li, Y. Deep learning based inference of private information using embedded sensors in smart devices. *IEEE Netw.* **2018**, *32*, 8–14. [CrossRef]
18. Shi, Y.; Erpek, T.; Sagduyu, Y.E.; Li, J.H. Spectrum Data Poisoning with Adversarial Deep Learning. In Proceedings of the MILCOM 2018—2018 IEEE Military Communications Conference (MILCOM), Los Angeles, CA, USA, 29–31 October 2018; pp. 407–412.
19. Sagduyu, Y.E.; Shi, Y.; Erpek, T. Adversarial Deep Learning for Over-the-Air Spectrum Poisoning Attacks. *IEEE Trans. Mob. Comput.* **2021**, *20*, 306–319. [CrossRef]
20. Penna, F.; Sun, Y.; Dolecek, L.; Cabric, D. Detecting and Counteracting Statistical Attacks in Cooperative Spectrum Sensing. *IEEE Trans. Signal Process.* **2012**, *60*, 1806–1822. [CrossRef]
21. Sagduyu, Y.E. Securing Cognitive Radio Networks with Dynamic Trust against Spectrum Sensing Data Falsification. In Proceedings of the 2014 IEEE Military Communications Conference, Baltimore, MD, USA, 6–8 October 2014; pp. 235–241.
22. Flowers, B.; Buehrer, R.M.; Headley, W.C. Evaluating Adversarial Evasion Attacks in the Context of Wireless Communications. *IEEE Trans. Inf. For. Secur.* **2020**, *15*, 1102–1113. [CrossRef]
23. Kokalj-Filipovic, S.; Miller, R.; Vanhoy, G. Adversarial Examples in RF Deep Learning: Detection and Physical Robustness. In Proceedings of the 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Ottawa, ON, Canada, 11–14 November 2019; pp. 1–5.
24. Jilani, S.A.; Koner, C.; Nandi, S. Security in Wireless Sensor Networks: Attacks and Evasion. In Proceedings of the 2020 National Conference on Emerging Trends on Sustainable Technology and Engineering Applications (NCETSTEA), Durgapur, India, 7–8 February 2020; pp. 1–5
25. Sadeghi, M.; Larsson, E.G. Adversarial Attacks on Deep-Learning Based Radio Signal Classification. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 213–216. [CrossRef]
26. Kim, B.; Sagduyu, Y.E.; Davaslioglu, K.; Erpek, T.; Ulukus, S. Over-the-air adversarial attacks on deep learning based modulation classifier over wireless channels. In Proceedings of the Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 18–20 March 2020.
27. Kim, B.; Sagduyu, Y.E.; Davaslioglu, K.; Erpek, T.; Ulukus, S. Channel-Aware Adversarial Attacks Against Deep Learning-Based Wireless Signal Classifiers. *arXiv* **2020**, arXiv:2005.05321.
28. Kim, B.; Sagduyu, Y.E.; Davaslioglu, K.; Erpek, T.; Ulukus, S. How to Make 5G Communications 'Invisible': Adversarial Machine Learning for Wireless Privacy. *arXiv* **2020**, arXiv:2005.07675.
29. Flowers, B.; Buehrer, R.M.; Headley, W.C. Communications Aware Adversarial Residual Networks for over the Air Evasion Attacks. In Proceedings of the MILCOM 2019—2019 IEEE Military Communications Conference (MILCOM), Norfolk, VA, USA, 12–14 November 2019; pp. 133–140

30. Davaslioglu, K.; Sagduyu, Y.E. Trojan Attacks on Wireless Signal Classification with Adversarial Machine Learning. In Proceedings of the 2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Newark, NJ, USA, 11–14 November 2019; pp. 1–6
31. Shi, Y.; Davaslioglu, K.; Sagduyu, Y.E. Generative Adversarial Network in the Air: Deep Adversarial Learning for Wireless Signal Spoofing. *IEEE Trans. Cogn. Commun. Netw.* **2021**, *7*, 294–303. [CrossRef]
32. ITU. *Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced, Mobile, Radio Determination*; ITU-R M.2135-1.M Series; ITU: Geneva, Switzerland, 2009.
33. Berthelot, D.; Schumm, T.; Metz, L. BEGAN: Boundary Equilibrium Generative Adversarial Networks. *arXiv* **2017**, arXiv:1703.10717.
34. Makhzani, A.; Shlens, J.; Jaitly, N.; Goodfellow, I.; Frey, B. Adversarial autoencoders. *arXiv* **2015**, arXiv:1511.05644.