Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Frequency-learning adversarial networks based on transfer learning for cross-scenarios signal modulation classification*

Qinyan MA†1, Jing XIAO[1], Zeqi SHAO[1], Duona ZHANG†‡2, Yufeng WANG‡3, Wenrui DING[3]

[1]*School of Electrical and Information Engineering, Beihang University, Beijing 100191, China*

[2]*School of Information Science and Technology, North China University of Technology, Beijing 100144, China*

[3]*Unmanned Systems Research Institute, Beihang University, Beijing 100191, China*

†E-mail: maqinyan17373036@buaa.edu.cn; zhangduona@buaa.edu.cn

Received Feb. 2, 2024; Revision accepted June 25, 2024; Crosschecked

Abstract: Automatic modulation classification (AMC) serves a challenging yet crucial role in wireless communication. Despite deep learning-based approaches being widely used in signal processing, they are challenged by signal distribution variations, especially in various channel conditions. In this paper, we introduce Frequency-learning adversarial networks (FLANs) based on transfer learning for cross-scenario signal classification. This method utilizes the stability in the frequency spectrum by introducing a frequency adaptation (FA) technique to incorporate target channel information into source domain signals. To address the unpredictable interference in the channel, a fitting channel adaptation (FCA) module is used to reduce the differences between the two domains caused by variations in the channel environment. Experimental results illustrate that FLANs outperforms state-of-the-art transfer approaches, demonstrating an improved top-1 classification accuracy by 5.2% in high signal-to-noise ratio (SNR) scenes on a cross-scenario real collected dataset (CSRC2023), which will be made publicly available soon.

Key words: Frequency spectrum; Generative adversarial network; Transfer learning; Automatic modulation classification; Wireless communication

https://doi.org/10.1631/FITEE.2400080 **CLC number:**

Unedited

1 Introduction

Automatic modulation classification (AMC), serving as an integral component of signal direction and demodulation in wireless communication, plays a pivotal role in a range of civil and military applications including spectrum monitoring, interference identification, and spectrum management. The importance and application of modulation classification have been extensively explored over several decades.

More recently, deep learning (DL) techniques have shown remarkable success in various complex tasks. Particularly, end-to-end models have been recognized for their exceptional performance (O'Shea et al., 2016), enabling a DL system to learn suitable features through a data-driven approach, without needing hand-crafted features. This advancement in DL has been effectively utilized in communications, particularly in the area of modulation classification. The use of convolutional neural networks (CNNs) in complex-valued radio signal classification has displayed competitive results when compared to traditional approaches based on expert features. With a substantial number of labeled signals, models can "study" the characteristics of the signals autonomously. Another effective tool is the recurrent neural network (RNN), designed specifically for processing time-sequenced signals and their interconnections, which, despite requiring large volumes of data, has proved effective in signal modulation recognition (Rajendran et al., 2018).

‡ Corresponding authors

* Project supported by the National Natural Science Foundation of China (No. U20B2042)

© Zhejiang University Press 2024

However, wireless radio signals can drastically vary across different channel conditions. The existence of a line of sight (LOS) between the transmitter and the receiver can significantly affect wireless communications, leading to deviations in the statistical characteristics of the received radio samples from the training data. DL algorithms necessitate large volumes of training data that adhere to the independent and identically distributed (i.i.d) condition with the test data (Pan and Yang, 2010). Unfortunately, obtaining sufficient training data for all potential radio scenarios is not practical, which can compromise the effectiveness of the model. The impact of scenario variation, particularly noticeable in the time domain, can also compromise model effectiveness. Conversely, the frequency domain offers more stability. For instance, when collecting signals indoors using universal software radio reripheral (USRP), the presence of walls and obstacles introduces multipath interference, which leads to the presence of strong high-frequency components in the received signal. However, significant variations in these components do not affect the identification of the modulation format used in the signal. Additionally, due to the numerous factors in real channel conditions, accurately estimating the variation in the characteristics of received signals from different channels is a formidable task.

To address this challenge, transfer learning (TL) is introduced to AMC. This relaxes the previous constraint, allowing the use of knowledge from existing channel conditions to solve issues in unknown scenarios. Fine-tuning, an extensively used method, shows promising performance (Zhou et al., 2022; Lin WS et al., 2023). It involves freezing the entire network trained in the source domain, and then updating the parameters of the last few fully connected layers with samples from the target domain. Similarly, weight sharing is used to initialize the parameter before inputting target signals. To solve more challenging problems, the generative adversarial network (GAN) is integrated into TL and applied to modulation classification. Domain adaptation, aiming to minimize the difference between source and target domains, is widely used to address insufficient data issues. The domain adversarial neural network (DANN) combines a deep transfer network with an adversarial network, consisting of a feature extractor, label classifier, and domain determiner (Ganin et al., 2016).

Inspired by adversarial networks and TL, we propose an adversarial transfer framework named Frequency-learning adversarial networks (FLANs). It leverages a frequency adaptation (FA) method and fitting channel adaptation (FCA) module to minimize the differences between two domains caused by channel environment variations (Fig. 1). Considering the stability of spectral characteristics during these changes, our FA module adjusts the high frequency spectrum of signals from two different domains to enforce their similarity, resulting in an approximately 10% improvement in accuracy. In addition, to address the uncertain factors in real channel conditions, the FCA module establishes a virtual channel and applies adversarial training to limit the distance between signal vectors after transmission, thereby further reducing differences. This approach has shown superior performance compared to existing adversarial networks, demonstrating an approximately 5% accuracy gap on the cross-scenario real collected dataset, CSRC2023.

The contributions of this work are summarized as follows:

1. We propose an adversarial training framework named FLANs to substantially enhance the performance of TL-based modulation classification of cross-scenario signals.

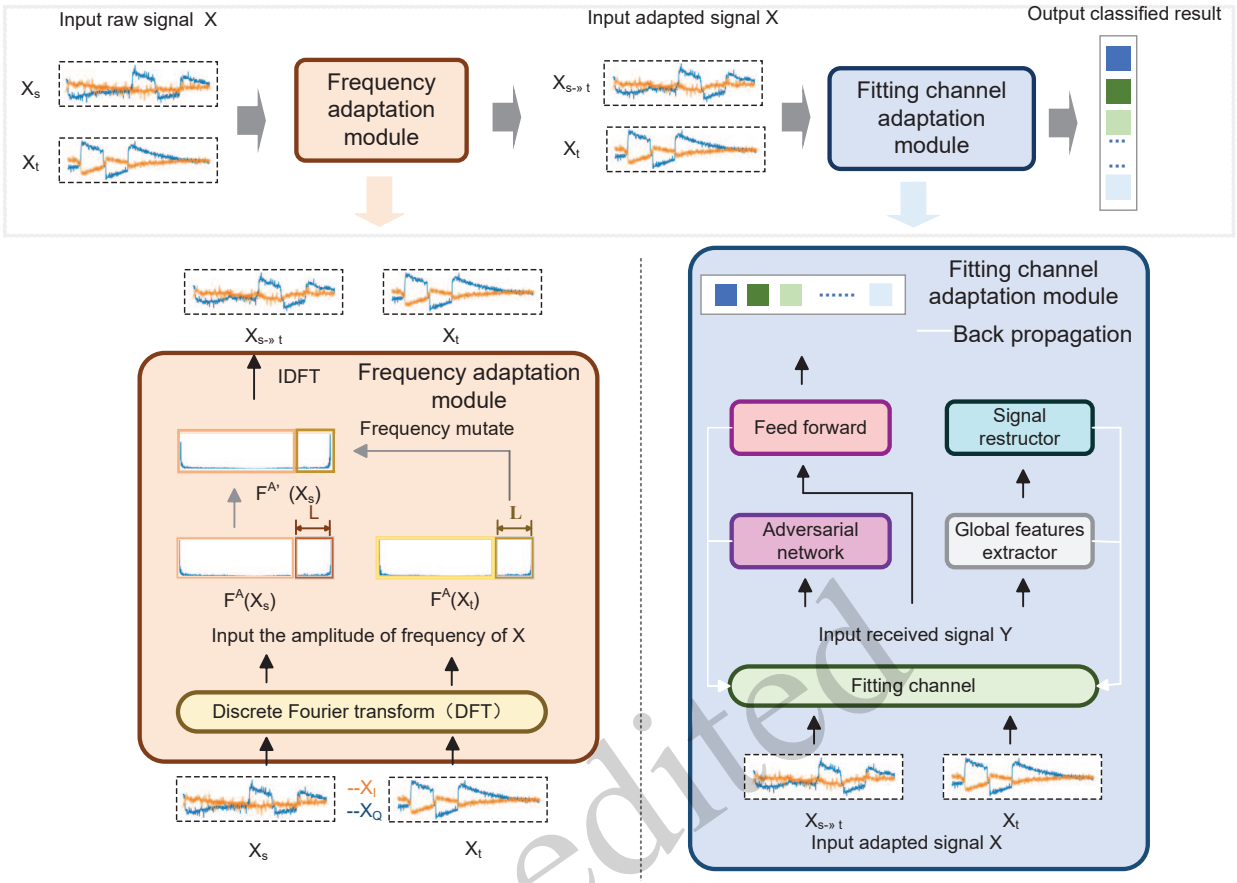
2. The proposed FLANs consist of two modules: the FA module and the FCA module. These modules are effectively used to minimize the differences between the two channels,

3. The proposed FLANs demonstrate superior performance compared to existing adversarial transfer methods, enhancing migration ability and extending application scenarios.

The remainder of this paper is organized as follows:

Section 2 presents relevant work pertaining to AMC and TL. Section 3 describes the structure of FLANs, including the FA and the FCA modules applied in AMC when channel conditions change from LOS to non-LOS (NLOS). In Section 4, the experimental results are presented and analyzed. Finally, the discussion and conclusion are provided in Section 5.

unedited



Unedited

Fig. 1 Pipeline of FLANs, which utilize the FA module and the FCA module to reduce the difference between two domains caused by channel environments. The FA module utilizes the stability of spectral characteristics during scenario changes. The FCA module applies adversarial training, aiming to address the uncertain factors and their variations in real channel conditions

2 Related works

2.1 AMC with DL

AMC is crucial in communication technology. With the continuous evolution of DL, the amalgamation of DL and AMC has grown into a significant field of research with various related studies. Unlike traditional feature-based AMC algorithms (Alarabi and Alkishriwo, 2021; G et al., 2021) and support vector machines (SVMs) (Lukito et al., 2021; Salama et al., 2022), which separate classes of data using an optimal hyperplane, DL is a data-driven approach that relies on an end-to-end model, eliminating the need for manual feature extraction.

DL methods can be broadly categorized into two main classes: RNNs and CNNs. RNNs excel at temporal feature extraction, while CNNs are adept at learning spatial features. O'Shea et al. (2016) studied the feasibility of CNNs on complex-valued radio signal classification, comparing end-to-end CNN models with traditional expert feature-based approaches. The former showed superior performance. Chen et al. (2022) proposed a novel DL framework named SigNet for wireless signal modulation classification, which adopts the signal-to-matrix (S2M) transformation method and integrates CNNs for classification purposes. Lin Y et al. (2021) studied the impact of carefully designed adversarial perturbations added to input signals on modulation recognition performance in practical non-cooperative communication scenarios. They evaluated the effectiveness of adversarial attack methods on signals and empirically assessed the reliability of CNNs. Their findings lay a crucial foundation for enhancing CNN resilience against adversarial attacks. Additionally, a data-driven sub-sampling approach (Ramjee et al., 2021) was presented for received

wireless signals, which significantly reduced classifier training time and improved classification accuracy. Other works utilized residual neural networks (ResNet) (He et al., 2016; O'Shea et al., 2018) and RNNs (Rajendran et al., 2018) for modulation classification, demonstrating effectiveness (Huang et al., 2020).

Despite the absence of prior knowledge, DL effectively captures complex patterns and relationships in signals using DNNs. However, it requires a substantial amount of labeled signal data, which is often unrealistic for most signals (Tan et al., 2022). When the distribution of training and testing data varies, performance deterioration occurs, leading to the introduction of TL (Perenda et al., 2021).

2.2 AMC with TL

TL eases the constraint that training data must follow the same distribution as test data, but its applications in AMC are scarce (Wang MY et al., 2021).

Signals as time series data present challenges due to weak semantic information, making manual annotation difficult. The initial application of TL to time series classification research (Wang ZG et al., 2017) is represented by a fully convolutional network known for its high accuracy and robustness. The technique includes two steps: pre-training on the source domain and fine-tuning on the target domain. This approach has been referred to as fine-tuning (Meng et al., 2018) and has been widely applied (Zhou et al., 2022; Lin WS et al., 2023). Yao et al. (2023) use a few-shot learning approach aimed at addressing the challenge of identifying specific emitters (e.g., communication devices or transmitters) using limited training data. Similarly, weight sharing (Pan and Yang, 2010) is used to address transfer problems with signals under different sample rates (Wang Q et al., 2019), initializing the classifier of the target source. This has also been adopted to address unstable model performance in low SNR (E_s/N_0) scenarios (Xu Y et al., 2019).

With the proposal of the GAN (Goodfellow et al., 2014), scholars have gradually integrated it into TL, applying it to modulation classification. Domain adaptation aims to minimize the difference between source and target domains and is universally applied to tackle insufficient data issues (Xu ZW et al., 2023). Adversarial discriminative domain adaptation (ADDA) and a general framework for adversarial unsupervised adaptation methods have been proposed (Tzeng et al., 2017). Further work incorporates adversarial training with knowledge transmission to decrease domain deviation (Bu et al., 2020). Taking into account the considerable shortage of labeled data in the target domains in cross-scenario cases, auxiliary classifier GANs (ACGANs) has been proposed as a generator to expand the dataset (Tang et al., 2018). TL based on AlexNet, called AlexNet-TL (Zhao et al., 2022), has been proposed with the pretraining and fine-tuning method, using 60 labeled data, showing an obvious improvement in accuracy (14%). Zhu et al. (2021) have proposed the deep subdomain adaptation network (DSAN), which, by introducing a subdomain feature alignment mechanism in DNNs, effectively reduces the distribution discrepancy between the source and target domains, thereby enhancing classification performance. Nam et al. (2021) propose a method Style-Agnostic Networks (SagNets) to reduce the domain gap by mitigating style bias. The authors introduce a framework that incorporates domain adaptation and style transfer techniques between a source domain and a target domain via an intermediary domain. This approach aims to balance stylistic differences across domains, thereby enhancing model generalization and performance.

The introduction of a deep transfer network into adversarial networks led to the proposal of DANN (Ganin et al., 2016), consisting of three parts: feature extractor, label classifier, and domain determiner. This network structure forms a fundamental framework for adversarial TL, prompting many research studies based on this framework (Long et al., 2015; Xiao et al., 2021). Inspired by DANN, we propose an adversarial TL architecture FLANs for TL-based AMC issues.

3 FLANs approach

3.1 Problem statement

This section uses the random model method to discuss the challenges that DL encounters when dealing with modulation classification issues. The wireless communication system consists of a transmitter, a channel, and a receiver. The received signal $x(n)$ can be given as follows:

$$r(n) = s(n) * h(n) + w(n), n = 1, 2, ..., N, \quad (1)$$

Unedited

where $s(n)$ symbolizes a transmission signal. After undergoing filtration, the transmission signal proceeds to a communication channel with an impulse function $h(n)$ and is accompanied by additive Gaussian white noise $w(n)$. The sampling rate of the receiver is denoted as N.

As the impulse function $h(n)$ varies with the channel environment, we use the random modeling method in practical channels to further demonstrate the effect of the channel environment on received signals. Rayleigh and Rician channels model signal statistics in intricate environments. The latter portrays a scenario with a LOS path in the channel, assuming the received signal as a combination of a principal path signal and randomly reflected path signals, thereby symbolizing multipath signals in a stochastic process.

The Rician channel model typically uses an amplitude response function with exponential decay and a sine function with a random phase offset to depict the primary path signal, as in Eq. (2):

$$\begin{aligned} h(t) &= \sqrt{\frac{K}{1+K}}\mathrm{e}^{-\alpha t}\mathrm{e}^{\mathrm{j}2\pi f_c t} \\ &+ \sqrt{\frac{1}{1+K}}\sum_{n=1}^{L} a_n \mathrm{e}^{-(\alpha+\mathrm{j}2\pi f_n)t}\mathrm{e}^{\mathrm{j}\phi_n}, \end{aligned} \tag{2}$$

where K signifies the average power ratio between the primary path signal and the randomly reflected path signals (i.e., the power ratio of the LOS path and the multipath power), α denotes the attenuation factor of the primary path signal, f_c symbolizes the frequency of the primary path signal, L indicates the number of randomly reflected paths, and a_n, f_n, and ϕ_n denote the amplitude, frequency, and phase of the corresponding n^{th} randomly reflected path, respectively. When K=0, the Rician channel model devolves into the Rayleigh channel model, which indicates the absence of a LOS path.

This paper addresses the AMC task under cross-scenario conditions, taking into account both Rician and Rayleigh channels. Based on these theories, we can view the presence or absence of LOS as the primary reason for distributional differences between the domains. We consider the former as the source domain and the latter as the target domain. Under these circumstances, the data-driven DL method, reliant on the consistency of data distribution between training and testing data, will likely exhibit poor classification performance. Thus, we introduce TL with the hope of transferring existing knowledge from source domain S to target domain T and enhancing the classification accuracy of target signals.

Given the variant channel conditions, it is impossible to establish a well-labeled dataset in the target domain. Therefore, this problem can be perceived as an unsupervised problem. Here, we exploit the stability of the frequency spectrum and the adversarial training method to minimize the difference between the two domains, ultimately aiming to boost the classification accuracy.

3.2 Baseline method

In this study, we use the DANN from literature (Ganin et al., 2016) as our baseline. DANN, characterized by its residual structure, is a classic neural model designed to tackle domain adaptation problems. It can effectively reduce differences between source and target domains, thereby enhancing the generalization capabilities of model in the target domain. After the introduction of DANN, similar domain adversarial training techniques have been adopted by many domain adaptation methods, and these have shown promising results in various tasks.

DANN comprises three components: the feature extractor G_f, the label predictor G_y, and the domain classifier G_d. Their functions are as follows:

The feature extractor G_f learns a function that maps an example into a new D-dimensional representation, parameterized by a matrix-vector pair $(W, b) \in R^{D\times m}\times R^D$, $X = R^m$, as defined by Eq. (3):

$$\begin{aligned} G_f(x; W, b) &= \mathrm{sigmoid}(Wx + b), \\ \mathrm{sigmiod}(a) &= \left[\frac{1}{1+\exp(-a_i)}\right]_{i=1}^{|a|}. \end{aligned} \tag{3}$$

The label predictor G_y learns a function parameterized by a pair $(V, c) \in R^{L\times D}\times R^L$, $G_f(x) = R^D$. Using the softmax function, each component of the vector $G_y(G_f(x))$ indicates the conditional probability that the neural network assigns x to the class in Y represented by that component:

$$\begin{aligned} G_y(G_f(x); V, c) &= \mathrm{softmax}(VG_f(x) + c), \\ \mathrm{softmax}(a) &= \left[\frac{\exp(-a_i)}{\sum_{j=1}^{|a|}\exp(-a_j)}\right]_{i=1}^{|a|}. \end{aligned} \tag{4}$$

Unedited

The domain classifier G_d learns a logistic regressor, parameterized by a vector-scalar pair $(u, z) \in R^D \times R$, which models the probability that a given input originates from either the source domain D_s or the target domain D_t:

$$G_d(G_f(x); u, z) = \text{sigmoid}(u^{\mathrm{T}} G_f(x) + z). \quad (5)$$

Both the label predictor G_y and the domain classifier G_d take the output of the feature classifier G_f as their input. This ensures that the features extracted from the signals are indistinguishable for G_d, while G_y is still able to classify the modulation format of the signals. The introduction of adversarial training aims to optimize the following equation:

$$\begin{aligned} E(W, V, b, c, u, z) = & \frac{1}{n} \sum_{i=1}^{n} L_y^i(W, b, V, c) \\ & - \lambda(\frac{1}{n} \sum_{i=1}^{n} L_d^i(W, b, u, z) \\ & + \frac{1}{n'} \sum_{i=n+1}^{N} L_d^i(W, b, u, z)). \end{aligned} \quad (6)$$

During training, G_f and G_y compete with G_d in an adversarial manner. Here, G_f learns to map signals into a new D-dimensional representation, enabling G_y to classify the modulation format of the signals while diminishing the capability of G_d to differentiate the domain of signals.

Despite its several advantages in domain adaptation learning, DANN also has its limitations:

1. High feature consistency requirement: The performance of DANN heavily depends on the consistency of features between the source and the target domains. When there are substantial differences in features between these domains, DANN may struggle to adapt effectively. For instance, Rician fading in Rician channels combines multipath fading with a strong LOS component, leading to less-severe signal strength fluctuation compared to Rayleigh fading, which is purely random and exhibits severe signal fading. These distinctions in signal distribution between the source and the target domains cause DANN to struggle in effectively handling cross-scenario problems.

2. Weak interpretability: Due to the complex nature of neural networks, DANN often yields highly abstract and nonlinear feature representations. This poses a challenge in establishing a connection between these features and interpretable signal characteristics, such as the signal frequency spectrum and constellation. Moreover, the theoretical aspects behind signal transmission, channel modeling, and signal processing can be hard to understand when applied within the neural network input, thereby making the decision-making process of model challenging to explain.

In this paper, we propose the FLANs, addressing these limitations of DANN, aiming to reduce the difference between training and test data. The FA module utilizes the stability of the frequency spectrum, while the FCA module introduces the concept of fitting channels and global feature extractor to strengthen the interpretability of the network.

3.3 FA module

This section primarily elucidates the methodology used for the FA module and the process of the FA module. During scenario variations, differences in diffusion are particularly noticeable in the time domain. Compared to that, the frequency domain exhibits stability to some extent. In modulation classification issues, for signals utilizing low-order phase shift keying (PSK) and most other modulation formats, the majority of high-frequency components are mainly composed of noise, which undergo intense fluctuations and do not significantly affect the recognition of the signal's modulation format, i.e., they do not contribute to the task. These interferences can be eliminated from the signal at the outset and do not need to be learned by the network.

Furthermore, this portion of noise also contains some channel information due to factors such as the transmission characteristics of the channel, multipath effects, and the absence of LOS paths.

We introduce the discrete fourier transform (DFT) at this point. The equation for a complex signal x_n is given as follows:

$$X_k = F(x) = \sum_{n=0}^{N-1} x_n \mathrm{e}^{-\mathrm{j}2\pi \frac{k}{N} n}, \mathrm{j}^2 = -1, \quad (7)$$

where F represents the DFT; accordingly, F^{-1} is the inverse process of DFT (IDFT). Additionally, F^A and F^P are the amplitude and phase components of the process F of a complex signal, respectively. Further, we designate a mask code $M_{L,n} = 1_{n \in [LN:]}$, the value of which is zero except for the margin region, where $L \in (0, 1)$.

Note that L represents a ratio instead of the

unedited

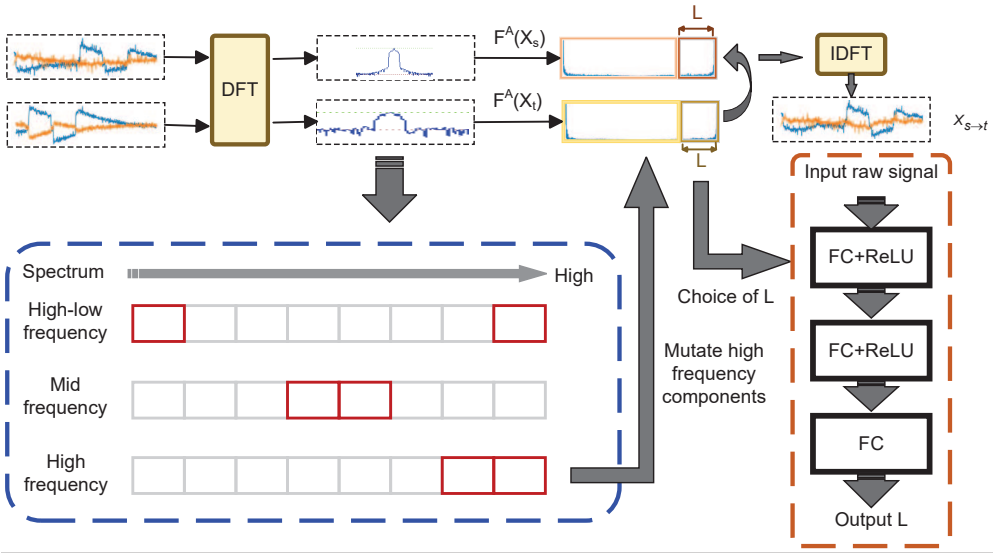


Fig. 2 Process of the FA module. By mutating high-frequency components, the FA module attaches source signals with target-domain channel characteristics without altering the modulation format. A regression network is introduced to obtain the optimal exchange length ratio L for every single source signal

number of sample points. For two signals from two domains $x_s \sim D_s$, $x_t \sim D_t$, the process of frequency spectrum mutation can be formalized as follows:

$$\begin{aligned} x_{s\to t} &= F^{-1}\left(\left[M_L \circ F^A(x_t) + \bar{M}_L \circ F^A(x_s), F^P(x_s)\right]\right) \\ \bar{M}_L &= 1 - M_L \end{aligned} \tag{8}$$

This implies that the high-frequency component of the source signal's amplitude, $F^A(x_s)$, is replaced with that of the target signal x_t. The modified amplitude component, along with its unaltered phase component, undergoes processing through the IDFT module and is mapped back to the signal $x_{s\to t}$, whose modulation format is the same as x_s but will resemble the appearance of a sample from D_t. The process is illustrated in Fig. 2, where the mask code M_L is shown with a length ratio L.

We also introduce a regression network to select an appropriate value for L in the FA module for each individual signal. As depicted in Fig. 2, if L =0, the signal $x_{s\to t}$ remains identical to the original signal x_s. Correspondingly, L =1 represents that the amplitude of x_s is completely replaced by x_t. In Fig. 2, the output of the IDFT module is attached with target-domain channel characteristics with L=0.01, which is applied as an initial value for the FLANs.

3.4 FCA module

In this study, due to the abundance of uncertain interferences in the channel, accurately estimating the distribution characteristics of the received signals is challenging. To counter this, we use a neural network to construct a virtual channel condition. Alongside this, we implement a global feature extractor to enhance the network's interpretability and minimize the variations in received signals simultaneously. Fig. 3 showcases the overview of the FCA module.

Received data in a wireless communication system can be described as shown in Eq. (1). We construct a network as a fitting channel, where $H(n)$ is its impulse response. The complete channel matrix is composed of a large-scale fading (LSF) part and a small-scale fading (SSF) part (Zheng et al., 2021). The LSF consists of path loss (PL), shadowing (SH), blockage loss (BL), and gas absorption loss (AL) and is presented as follows:

$$\mathbf{H} = [PL \cdot SH \cdot BL \cdot AL]^{1/2}\mathbf{H}_s, \tag{9}$$

where $\mathbf{H}_s$ is the SSF matrix and can be further rep-

unedited

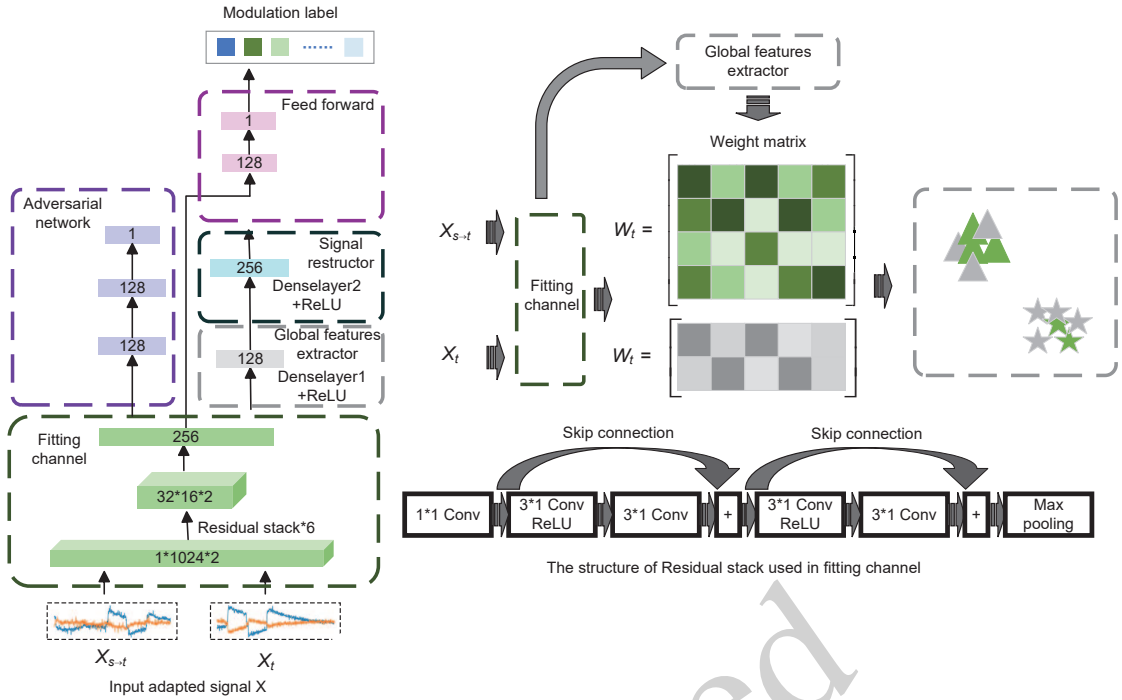


Fig. 3 Overview of the FCA module, which applies a neural network to build a virtual channel condition. By incorporating a global feature extractor, the interpretability of the network is enhanced. During the adversarial training process, the received signal is weighted with the extracted global feature to effectively minimize the differences between the two domains

Unedited

resented as follows:

$$\mathbf{H}_s = h_{qp}(t,\tau) = \sqrt{\frac{K_{\mathrm{RF}}(t)}{K_{\mathrm{RF}}(t)+1}} h_{qp}^{L}(t,\tau) + \sqrt{\frac{1}{K_{\mathrm{RF}}(t)+1}} h_{qp}^{N}(t,\tau) \tag{10}$$

where $K_{\mathrm{RF}}(t)$ is the Rician factor, $h_{qp}^{L}(t,\tau)$ is the LOS component, and $h_{qp}^{N}(t,\tau)$ is the NLOS components. These can be further represented as follows:

$$\begin{aligned} h_{qp}^{L}(t,\tau) = & \begin{bmatrix} F_{q,V_p}\left(\phi_{E,\mathrm{LOS}}^{R}(t),\phi_{A,\mathrm{LOS}}^{R}(t)\right) \\ F_{q,H_p}\left(\phi_{E,\mathrm{LOS}}^{R}(t),\phi_{A,\mathrm{LOS}}^{R}(t)\right) \end{bmatrix}^{\mathrm{T}} \\ & \cdot \begin{bmatrix} \mathrm{e}^{\mathrm{j}\theta_{\mathrm{LOS}}^{V_pV_p}} & 0 \\ 0 & -\mathrm{e}^{\mathrm{j}\theta_{\mathrm{LOS}}^{H_pH_p}} \end{bmatrix} \\ & \cdot \begin{bmatrix} F_{p,V_p}\left(\phi_{E,\mathrm{LOS}}^{T}(t),\phi_{A,\mathrm{LOS}}^{T}(t)\right) \\ F_{p,H_p}\left(\phi_{E,\mathrm{LOS}}^{T}(t),\phi_{A,\mathrm{LOS}}^{T}(t)\right) \end{bmatrix} \\ & \cdot \mathrm{e}^{\mathrm{j}2\pi f_c \tau_{qp}^{\mathrm{LOS}}(t)} \delta\left(\tau-\tau_{qp}^{\mathrm{LOS}}(t)\right) \end{aligned} \tag{11}$$

$$\begin{aligned} h_{qp}^{N}(t,\tau) = & \sum_{n=1}^{N_{qp}(t)} \sum_{m=1}^{M_n(t)} \begin{bmatrix} F_{q,V_p}\left(\phi_{E,m_n}^{R}(t),\phi_{A,m_n}^{R}(t)\right) \\ F_{q,H_p}\left(\phi_{E,m_n}^{R}(t),\phi_{A,m_n}^{R}(t)\right) \end{bmatrix}^{\mathrm{T}} \\ & \cdot \begin{bmatrix} \mathrm{e}^{\mathrm{j}\theta_{m_n}^{V_pV_p}} & \sqrt{\kappa_{m_n}^{-1}(t)}\mathrm{e}^{\mathrm{j}\theta_{m_n}^{V_pH_p}} \\ \sqrt{\kappa_{m_n}^{-1}(t)}\mathrm{e}^{\mathrm{j}\theta_{m_n}^{H_pV_p}} & \mathrm{e}^{\mathrm{j}\theta_{m_n}^{H_pH_p}} \end{bmatrix} \\ & \cdot \begin{bmatrix} F_{p,V_p}\left(\phi_{E,m_n}^{T}(t),\phi_{A,m_n}^{T}(t)\right) \\ F_{p,H_p}\left(\phi_{E,m_n}^{T}(t),\phi_{A,m_n}^{T}(t)\right) \end{bmatrix} \\ & \cdot \sqrt{P_{qp,m_n}(t)}\mathrm{e}^{\mathrm{j}2\pi f_c \tau_{qp,m_n}(t)} \delta\left(\tau-\tau_{qp,m_n}(t)\right) \end{aligned} \tag{12}$$

where $(\cdot)^{\mathrm{T}}$ denotes the transpose operation; $F_{p(q),V_p}^{T(R)}$ and $F_{p(q),H_p}^{T(R)}$ represent the vertical polarization and horizontal polarization at Tx (Rx) side, respectively; and κ_{m_n} denotes the cross-polarization ratio.

It is crucial to note that the process of a signal passing through the network is regarded as passage through a unique type of complex channel. This channel differs from conventional channels since it is not implemented through a physical medium. Instead, it is represented by a mathematical model with more complex and abstract properties.

The fitting channel serves a similar purpose to the feature extractor in the DANN architecture. It

can be given as follows:

$$y = H(x; W, b) = \mathrm{ReLU}(Wx + b), \quad \mathrm{ReLU}(x) = \max(0, x). \tag{13}$$

In this context, we consider the input of the fitting channel as the transmission signal and its output as the received signal.

The received signal is then sent to a global feature extractor F_e to obtain the overall distribution characteristics, such as the mean, variance, and periodic components, of the received signal:

$$F_e(y; M, n) = \mathrm{ReLU}(My + n). \tag{14}$$

To ensure that the extracted global feature is complete and contains comprehensive information of the received signal, we introduce the reconstructor F_r. This component avoids problems of dimensionality and overfitting by sharing a symmetric structure with the global feature extractor:

$$F_r(y; M', n') = \mathrm{ReLU}(M'y + n'). \tag{15}$$

Concerning adversarial TL, it is similar to the domain classifier G_d :

$$G_d(y; u, z) = \mathrm{ReLU}(u^{\mathrm{T}}y + z). \tag{16}$$

This classifier uses a measurement method to quantify the domain distance and applies an adversarial training method to minimize this distance d:

$$d = |p_s({x_s}', y_s; N_s), (p_t({x_t}', y_t; N_t))|_D, \tag{17}$$

where $|\cdot|_D$ signifies that the domain distance is reflected in the discriminator D's ability to distinguish domains, and P represents the probability distribution. We primarily assume that, through the fitting channel H, the received signals exhibit less domain differentiation compared to transmission signals.

Upon completion of adversarial training, we postulate that the received signal primarily contains modulation information, making it challenging to determine the signal's origin domain. Therefore, the received signal is directly input into a feed-forward network to determine the modulation format. This function is akin to the label predictor G_y:

$$G_y(y; V, c) = \mathrm{ReLU}(Vy + c). \tag{18}$$

With the addition of a global feature extractor, adversarial training aims to optimize the following equation:

$$E(W, V, b, c, u, z, M, n, M', n') = \frac{1}{n}\sum_{i=1}^{n} L_y^i(W, b, V, c) + \frac{1}{n}\sum_{i=1}^{n} L_e^i(W, b, M, n) + \frac{1}{n}\sum_{i=1}^{n} L_r^i(W, b, M', n') - \lambda\left(\frac{1}{n}\sum_{i=1}^{n} L_d^i(W, b, u, z) + \frac{1}{n'}\sum_{i=n+1}^{N} L_d^i(W, b, u, z)\right). \tag{19}$$

3.5 Loss function

Within each iteration, the parameters of the fitting channel are adjusted using the back-propagation algorithm. The loss function consists of four components, as follows:

$$L = L_c + L_D + \alpha L_R + \beta L_w. \tag{20}$$

where L_c represents the classification loss function, L_D the domain loss function, L_R the reconstruction loss function, and L_w the weight loss. Hyperparameters α and β are used to balance global feature discovery and alignment, respectively. Each loss function's specific features are detailed in the following text, with S representing the source domain and T the target domain.

The classification loss function L_c and the reconstruction loss function L_R are designed for the fitting channel network and re-constructor respectively. They maintain the ability to distinguish categories and performance of reconstruction, as follows:

$$L_c = L_{c_s} + L_{c_t} = L_{\mathrm{cross-entropy}}(y_s, {y_s}') + L_{\mathrm{entropy}}({y_t}'), \tag{21}$$

where L_{c_s} and L_{c_t} represent the losses for source and target domain label modulation classification, respectively. $L_{\mathrm{cross-entropy}}$ is the cross-entropy loss between the computed values:

$$L_{\mathrm{cross-entropy}}(Y, f(x)) = -\sum_{i=1}^{N} Y_i \log f(x_i), \tag{22}$$

where ${y_s}'$ represents the classification label of the source domain signal's modulation by the label discriminator, y_s represents the true label of the source domain data, and L_{entropy} denotes the computed entropy used as a loss function:

$$L_{\mathrm{entropy}}(Y) = -\sum_{i=1}^{N} Y_i \log Y_i. \tag{23}$$

unedited

Similarly, $y_t{}'$ denotes the classification label of the target domain signal's modulation by the label discriminator.

Reconstruction loss function L_R is given by:

$$L_R = L_{R_s} + L_{R_t} = L_{\text{mse}}(f_s, f_s{}') + L_{\text{mse}}(f_t, f_t{}'), \tag{24}$$

where L_{R_s} represents the loss function for the source domain's received signal f_s input to the global feature extractor and the reconstructed signal $f_s{}'$. L_{mse} denotes the mean squared error between f_s and $f_s{}'$:

$$L_{\text{mse}}(Y, f(x)) = \frac{1}{n}\sum_{i=1}^{N}(Y_i - f(x_i))^2. \tag{25}$$

The domain loss function L_D, commonly used in GAN, is as follows:

$$L_D = L_{D_s} + L_{D_t} = L_{\text{BCE}}(d_s, d_s{}') + L_{\text{BCE}}(d_t, d_t{}'), \tag{26}$$

where, L_{D_s} and L_{D_t} represent the losses when performing domain discrimination on the source and target domain signals, respectively. L_{BCE} represents the binary cross-entropy loss between d_s and $d_s{}'$:

$$L_{\text{BCE}}(Y, f(x)) = -(Y\log f(x) + (1-Y)\log(1-f(x)), \tag{27}$$

where $d_s{}'$ is the domain discriminator's classification of the source domain signal, and d_s is the domain label for the source domain signal. In this context, "0" signifies that the signal originates from the source domain, while "1" means that it comes from the target domain. Similarly, $d_t{}'$ represents the domain discriminator's classification of the target domain signal, and d_t is the domain label for that signal. The domain loss function aims to minimize the domain label discrimination error.

The weight loss function L_w is used to reduce the distinction between received signals from different domains, as shown below:

$$L_w = L_{\text{mse}}(\overline{w_s}, \overline{w_t}), \tag{28}$$

where, $\overline{w_s}$ denotes the average weight of each received signal for each type of global feature in the source domain. Its dimension matches the number of received signals:

$$\overline{w_s} = [\overline{w_{s_1}}, \overline{w_{s_2}}, ..., \overline{w_{s_N}}], \tag{29}$$

where $\overline{w_{s_1}}$ is the average of the column sums of the weight matrix W_s of the extracted global feature and received signals:

$$W_s = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,N} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N,1} & w_{N,2} & \cdots & w_{N,N} \end{bmatrix}, \tag{30}$$

where N is the number of received signals and global features. Each row represents the weights of the received signals concerning each global feature.

4 Experiments

This section first focuses on the selection of our dataset, CSRC2023, and implementation details of experiments. Then, it compares the proposed method with the baseline approach across a variety of SNRs. The performance index for modulation classification used in this study is accuracy. This section also showcases constellations of select samples from the transmitter and the receiver under different channel conditions.

4.1 Dataset and implementation details

4.1.1 Dataset

This study considers two distinct channel environments: the Raleigh multipath fading channel and the Rician multipath fading channel. The primary difference between these channels is the presence or absence of LOS. Accordingly, we positioned the transmitter and the receiver as shown in Fig. 4 to satisfy these channel conditions.

The collected dataset comprises in-phase and quadrature (IQ) samples, each of size 1024*2, indicating the inclusion of both real and imaginary components of the signal. The digital signal's center frequency is set at 195.1 MHz. We adopted a root-raised cosine shaping filter with a roll-off coefficient of 0.35 as the pulse-shaping filter. The sampling frequency and the number of sampling points for each symbol are 320 KHz and eight, respectively.

The dataset includes five common signals that need to be identified, namely, '2PSK', 'QPSK', '8PSK', '4QAM', and 'GFSK'. These signal symbols are modulated by random bit data generated by the NumPy-supported random library. We created a new dataset using the GNU radio, which provides numerous tools for creating robust datasets. The signals were modulated and transmitted using a USRP

unedited

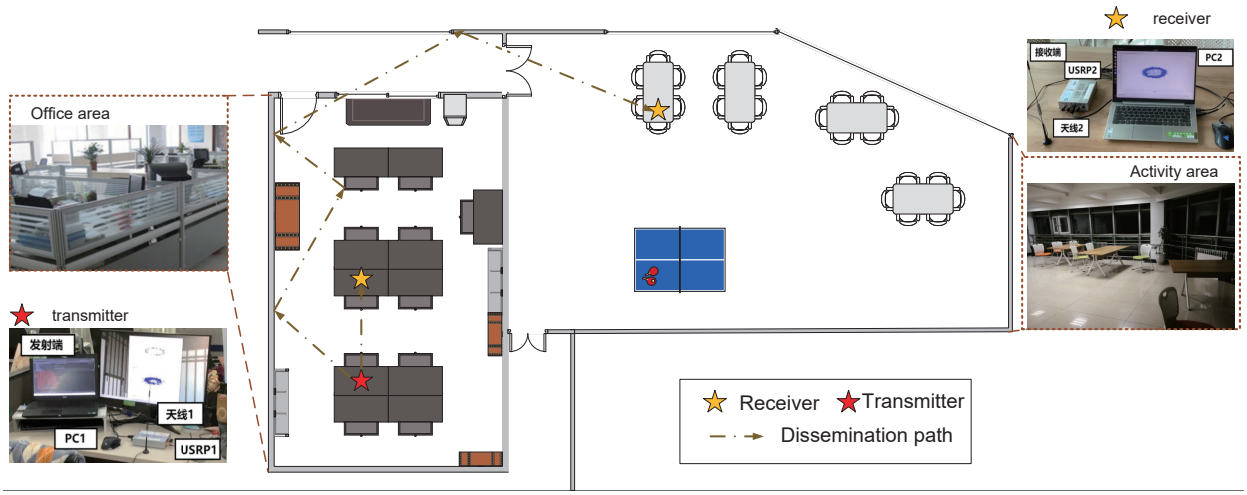


Fig. 4 Signal acquisition channel environment. The transmitter is placed in the office area, while the receiver is located in the same area and activity area to establish the LOS and the NLOS channels, respectively. In the planar graph, yellow lines represent potential dissemination paths of the signal

B210 software defined radio (SDR). The signals were then sent through a finite impulse response (FIR) filter, transmitted via USRP B210, and subsequently received by a second B210.

Moreover, we introduced additive Gaussian white noise to the signals at the receiving end. The SNR ranged from -20 dB to 30 dB in steps of 2 dB. For the training dataset, we had 80 samples for each SNR and each modulation method, resulting in a total of 80*26*5=10 400 samples. The testing dataset was similarly structured, yielding a total of 400*25*5=52 000 samples.

4.1.2 Implementation details

This paper proposes an adversarial network called the FLANs, consisting of two modules: the FA module and the FCA module. The latter one is composed of five components: the fitting channel, the global feature extractor, the signal reconstructor, the adversarial network, and the feed-forward network. The FLANs framework is depicted in Figs. 2 and 3. The design of the fitting channel, which comprises six residual stacks, draws inspiration from the ResNet model as proposed by G et al. (2021).

For this study, we created a well-annotated source dataset, D_s, comprising sufficient signal samples X_s and corresponding labels Y_s. We also established a target dataset D_t that had an equal number of signal samples X_t but lacked labels. Given the variations in channel conditions, it is often impossible to create well-labeled datasets for all conditions. Therefore, we treat this as an unsupervised problem. We define the source dataset as $D_s = (x_i, y_i, SNR_i)_{i=1}^{n_s}$, where x_i represents the raw signal, y_i signifies the modulation label, and SNR_i denotes the SNR of each signal. Likewise, the target dataset is represented as $D_t = (x_j, SNR_j)_{j=1}^{n_t}$, which lacks modulation labels.

Except for the nontransfer method, which uses TensorFlow, all other evaluated AMC methods are implemented using PyTorch. As the optimizer, we chose AdamW with a learning rate of 0.0001. This rate offers a balance between slow convergence at lower rates and inaccurate results at higher rates. The model was trained through 20 000 iterations, and testing was carried out every 200 iterations. Both processes utilized a batch size of 36. The model was trained and tested on a graphics processing unit (GPU) server equipped with a single Nvidia RTX 2070 card. We calculated the classification accuracies under all SNR conditions and used the top-1 accuracy as the criterion for comparing techniques and tuning the network.

4.2 Performance analysis of FLANs classification

4.2.1 Comparison with state-of-the-art methods

To emphasize the significance of our proposed method in modulation classification, we compare it with several representative prior works termed

unedited

Table 1 Comparison of the average and top-1 classification accuracies of FLANs with baseline methods

Method	ResNet_ST	ResNet_TT	DANN	DSAN	SagNets	FLANs w/o FA	FLANs
Top-1 accuracy	0.3731	0.8540	0.5415	0.6654	0.8645	0.6806	**0.9167**
Average accuracy	0.2556	**0.6450**	0.4029	0.4474	0.5038	0.6591	0.5676

DANN, DSAN, and SagNets. DANN also uses adversarial training and a residual structure. DSAN and SagNets are further studies based on DANN conducted in recent years. To ensure a fair comparison, we maintain identical training configurations and datasets for the baseline models and our proposed model FLANs.

Fig. 5 and Table. 1 present the comparison results. The results reveal that FLANs without (w/o)

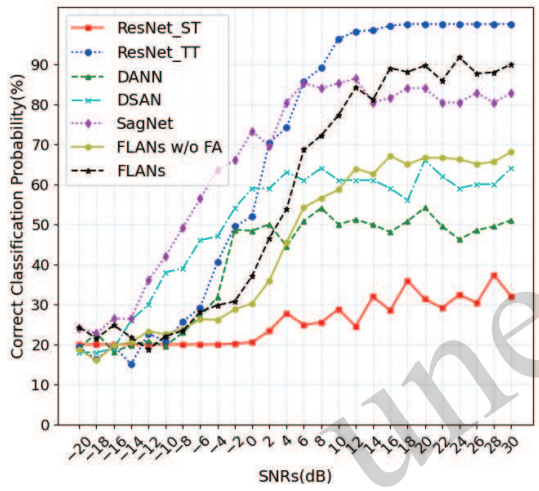


Fig. 5 Comparison results of FLANs with baseline methods

FA, by extracting the global features of signals processed through the fitting channel and balancing between the global feature and the received signal, significantly improves the modulation classification performance. It surpasses both baseline methods, achieving 18% over ResNet_ST (deep residual neural network trained on source dataset and tested on target dataset) and 4% over DANN across SNRs ranging from -20 dB to 30 dB. Under these conditions, the top-1 accuracy of FLANs w/o FA reaches 68%, outperforming DANN, which achieves 54%, and DSAN, which achieves 66%. SagNets perform excellently in low SNRs, but with the addition of our proposed FA module, the top-1 accuracy of FLANs jumps to 92%, about 23% higher than FLANs w/o FA and 6% higher than SagNets. ResNet_TT is trained on the target dataset (with labels) and is considered as an upper bound. The proposed FLANs significantly performs well in high SNR, and differs from the upper bound by only 6%.

Table 2 Mapping table for transfer scenarios

No.	Transfer scenario	Color	Labels
1	WLOS to NLOS	Blue	NLOS_model
2	WLOS to SLOS	Red	SLOS_model

4.2.2 Comparison with transfer channels

For further investigation of the model's generality in various transfer channel scenarios, a strong LOS (SLOS) channel (Strong Line of Sight) channel is involved in this section as another target domain. We selected a spacious indoor scene in the process of real signal collection with a direct line between the transmitter and the receiver with minimal or no obstruction, allowing for almost undisturbed signal transmission. The LOS channel data used in Section 4.2.1 were collected from the weak LOS (WLOS) channel in a crowded indoor scene, with more objects obstructing the LOS path. In this section, we compared the performance of FLANs and the baseline under two transfer scenarios: WLOS to NLOS; and WLOS to SLOS channels.

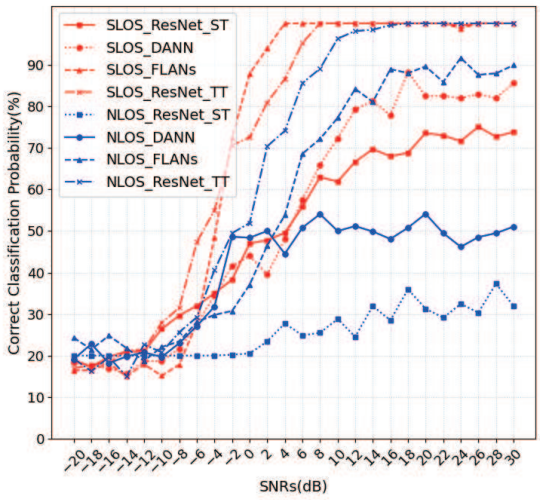


Fig. 6 Comparison of different transfer channels: WLOS channel to SLOS channel; and WLOS channel to NLOS channel

Unedited

Fig. 6 presents the comparison results wherein each transfer scenario includes four models: ResNet_ST, ResNet_TT, DANN and FLANs. The blue one represents transfer scenario 1, which is the transfer from the WLOS to the NLOS channel, the same as in Section 4.2.1. The red one represents transfer scenario 2, which is the transfer from the WLOS to the SLOS channel (Table. 2). Apparently, the differences between channels in transfer scenario 2 are less pronounced than those in transfer scenario 1, and the results also conform to expectations. In transfer scenario 2, due to minor differences, ResNet_ST without transfer achieved an accuracy of 70% while SNR was >14 dB, far surpassing its performance in scenario 1 (32%). However, with the introduction of the adversarial training method, DANN achieved 10% over ResNet_ST. It is notable that both the proposed FLANs and the upper bound ResNet_TT reached 100% accuracy when SNR was >8 dB. Furthermore, while -2 dB < SNRs < 8 dB, the proposed method exceeded the upper bound by up to 14%.

4.2.3 Classification performance by modulation mode

Fig. 9 illustrates the classification performance of FLANs for individual modulation modes. When SNR was <-2 dB, it only distinguished between BPSK and GFSK. While SNR < 10 dB, the classification accuracy decreased rapidly. The GFSK modulation mode, characterized by its distinct constellation structure, consistently showed promising results across the exhibited SNR range, maintaining accuracy >80%. As the SNR increased, other modulation modes also exhibited a steady increase in classification accuracy, achieving a classification accuracy >80% except for 8PSK at around 14 dB SNR.

We also delve into the details of misclassifications to determine which modulation format has high classification accuracy. Fig. 10 presents the confusion matrices for FLANs across all five classes with varying SNRs. These matrices indicate that the primary cause of classification error is the confusion between QPSK, 8PSK, and 4QAM, due to the similar modulation process.

Selected samples and their constellations from the transmitter and the receiver under different channel conditions are depicted in Figs. 7 and 8.

4.3 Performance analysis of FA module

In this section, we primarily investigate the two parameters of the FA module, and their influence on the cross-scenario issue performance. Following this, we incorporate the optimal initial value into the FA module and proceed to analyze its performance further.

Two factors play a significant role in influencing the outcome. (1) firstly, the position of the exchange in the frequency spectrum - broadly divided into high-, mid-, and low-frequency components. (2) secondly, the spectral exchanged length, represented as a ratio L.

When applied to our proposed FLANs, the effect of the FA module is shown in Fig. 11. When we include high-frequency components in the exchange, the classification top-1 accuracy of FLANs significantly improves from 67.18% to 91.38%, a surge of approximately 25 percentage points. Even when compared to the spectral exchange of combined high-low-frequency components, which results in a top-1 accuracy of 87.88%, high-frequency component exchange still yields an improvement of 3 points. Therefore, high-frequency components contain more channel information, and exchanging them benefits the network in extracting modulation characteristics and subsequently improving classification accuracy.

Moreover, the use of the FA module primarily enhances classification at low SNRs in contrast to conditions without mutation. This result suggests that the high-frequency signal components are more susceptible to noise interference and have a closer relationship with the channel, thereby holding more channel characteristic information. Consequently, by exchanging the high-frequency components of signals from different domains, we imbue source domain signals with target channel knowledge, reducing the differences between signals from both domains.

We further scrutinize the spectral exchanged length ratio L, adjusting its value to enhance the learning speed of FLANs and generalization capability. Fig. 12 and Table. 3 depict the results of experiments conducted with different L values.

We observe that as the exchanged ratio L decreases within $L \in [0.001, 0.1]$, the recognition accuracy tends to increase. However, at L =0.008, the top-1 accuracy slightly declines; at L =0.005, the average accuracy significantly reduces. Moreover,

unedited

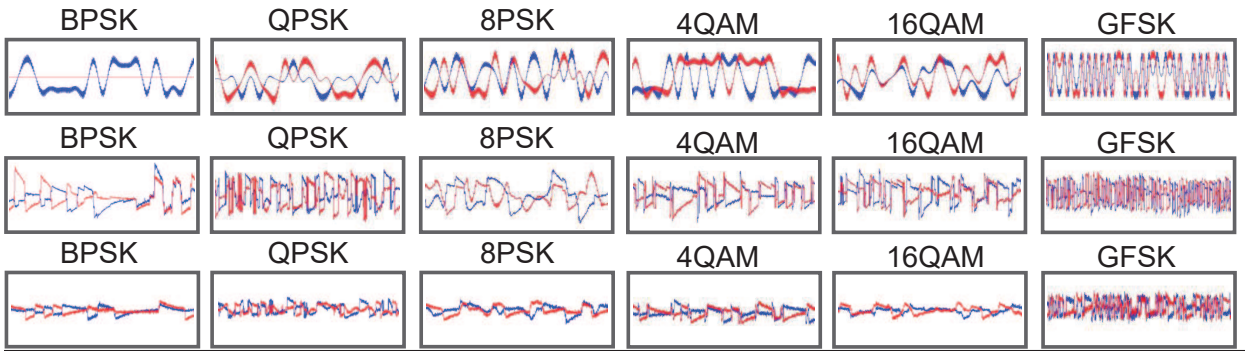
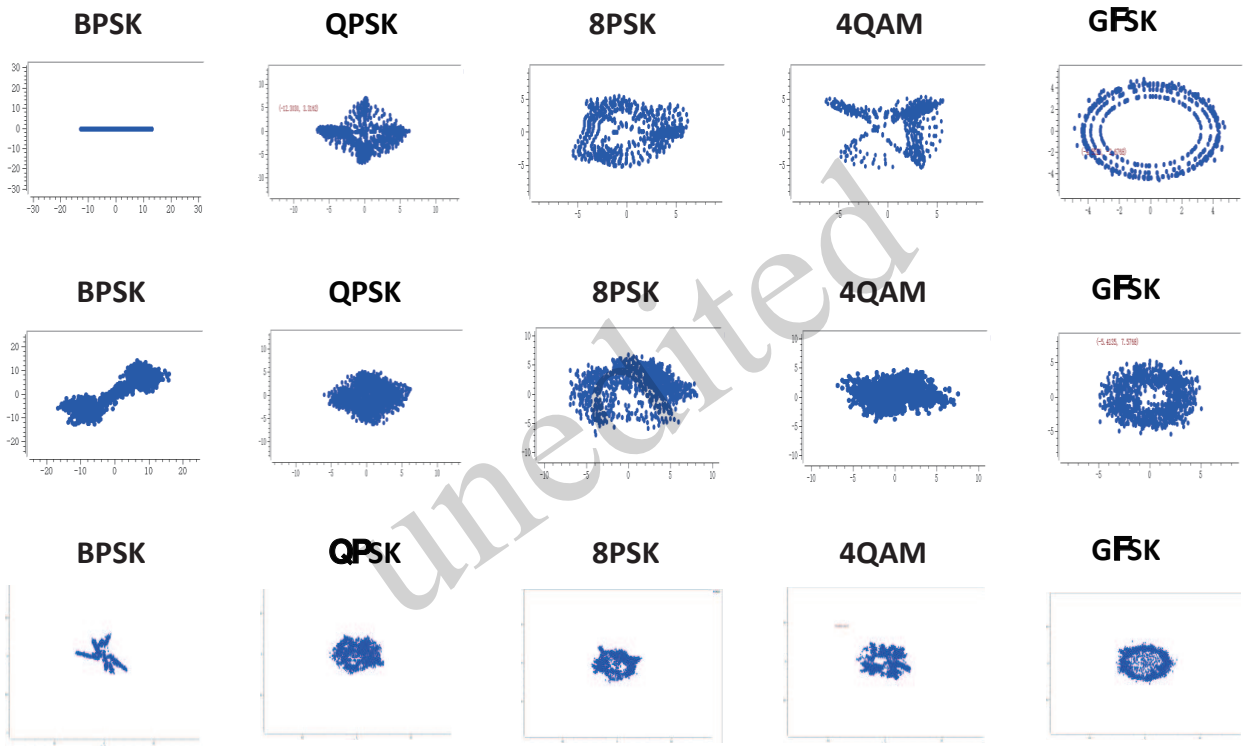


Fig. 7 Time domain plot of each modulation mode from transmitters, LOS channel, and NLOS channel



unedited

Fig. 8 Selected radio signal constellations of each modulation mode from transmitters, LOS channel, and NLOS channel

L=0.01 is the peak value of the group, which has the most obvious lifting effect. Therefore, we suggest L=0.01 as an appropriate initial value for the FA module. We also observe that the impact of length is predominantly apparent at high SNRs.

Lastly, we evaluate the efficacy of the FA module in DANN and FLANs (Fig. 13). The results reveal that application of the FA module leads to significant improvements in both networks when SNR >4 dB. Interestingly, the extent of improvement shows significant variations. DANN is improved by 8% while the SNR >8 dB. However, FLANs demonstrate a greater extent of improvement (approximately 25%) in a broader range of SNRs (> -2 dB), which indicates that the compatibility between the FA module and the FLANs is higher.

5 Conclusions

In this paper, we have proposed a novel TL framework, FLANs, for AMC. Our method focuses on signals that have traversed different channel environments. By using the FA and the FCA modules within the FLANs, we manage to reduce the dif-

Table 3 Average and top-1 classification accuracy versus the value of L*

L value	L=0.005	L=0.008	L=0.01	L=0.012	L=0.02	L=0.05	L=0.1
Average accuracy	0.6641	**0.6722**	0.6668	0.6488	0.6110	0.5237	0.5230
Top-1 accuracy	0.8358	0.8507	**0.9132**	0.8464	0.7883	0.619	0.6295

Here L is a ratio, eg.L=0.02 represents the exchanged number of sample point is 1024*2*0.02=20

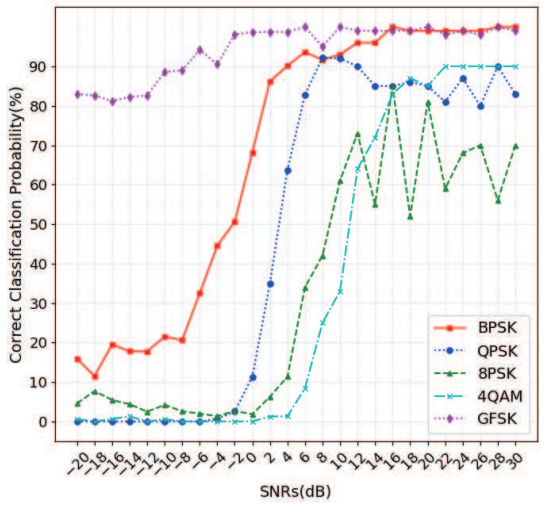


Fig. 9 Classification results of FLANs for individual modulation mode

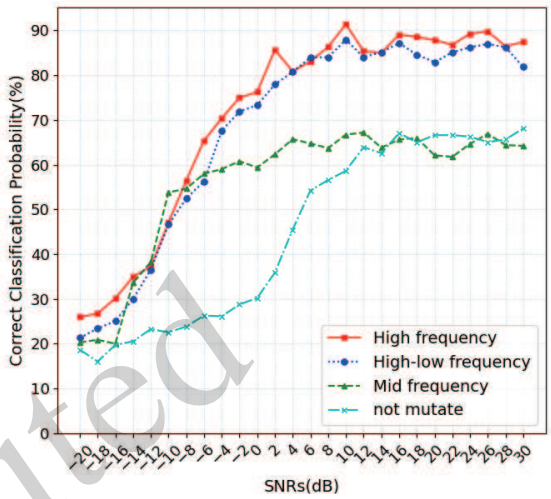


Fig. 11 Accuracy while position exchanged in the frequency spectrum

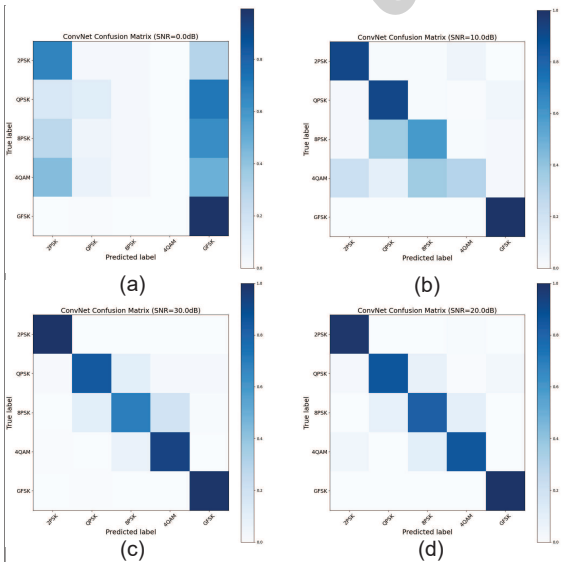


Fig. 10 Confusion matrix of FLANs with different SNRs. (a) 0 dB; (b) 10 dB; (c) 20 dB; (d) 30 dB

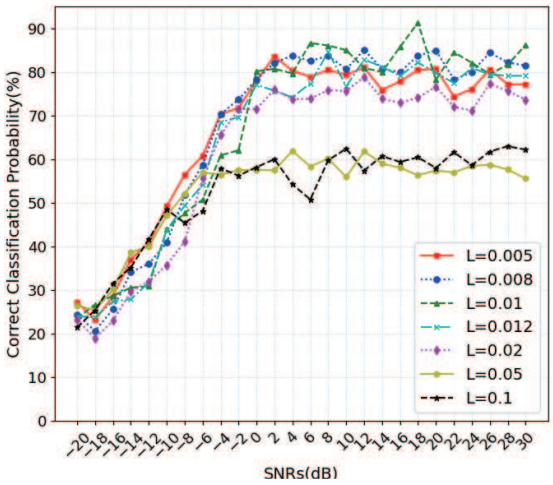


Fig. 12 Accuracy under different value of L with the variant of SNRs

unedited

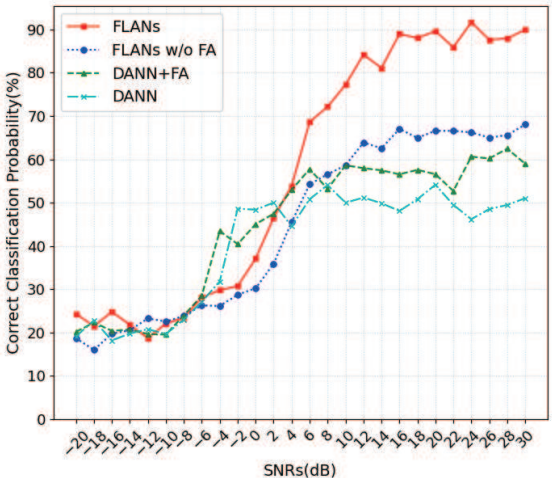


Fig. 13 Comparison results of DANN and FLANs without FA module. up: DANN, down: FLANs

ferences between signals from two distinct channels, thereby enhancing the classification performance of signals from the target domain. Our experimental analysis underscores the superior performance of our approach in addressing the cross-scenario issue. It surpasses existing adversarial training methods in terms of transfer capability and tolerance of dataset bias. Intriguingly, our method also increases the upper bound of classification performance in high SNRs, highlighting its significant application potential and research value. In our future work, we plan to expand the set of target modulations and intensify our focus on high-order AMC.

Contributors

Qinyan MA designed the research. Jing XIAO and Zeqi SHAO processed the data. Duona ZHANG and Yufeng WANG drafted the manuscript. Wenrui DING helped organize the manuscript. Qinyan MA and Duona ZHANG revised and finalized the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

References

Alarabi A, Alkishriwo OAS, 2021. Modulation classification based on statistical features and artificial neural network. Proc IEEE 1st Int Maghreb Meeting of the Conf on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA, p.748-751. https://doi.org/10.1109/MI-STA52233.2021.9464363

Bu K, He Y, Jing XJ, et al., 2020. Adversarial transfer learning for deep learning based automatic modulation classification. *IEEE Signal Process Lett*, 27:880-884. https://doi.org/10.1109/LSP.2020.2991875

Chen ZZ, Cui H, Xiang JY, et al., 2022. SigNet: a novel deep learning framework for radio signal classification. *IEEE Trans Cognit Commun Netw*, 8(2):529-541. https://doi.org/10.1109/TCCN.2021.3120997

G N, Vijayan V, Jose R, 2021. Performance analysis of modulation classification using machine learning. Proc 8th Int Conf on Smart Computing and Communications, p.70-74. https://doi.org/10.1109/ICSCC51209.2021.9528172

Ganin Y, Ustinova E, Ajakan H, et al., 2016. Domain-adversarial training of neural networks. *J Mach Learnng Res*, 17(1):2096-2030.

Goodfellow IJ, Pouget-Abadie J, Mirza M, et al., 2014. Generative adversarial nets. Proc 27th Int Conf on Neural Information Processing Systems, p.2672-2680.

He KM, Zhang XY, Ren SQ, et al., 2016. Deep residual learning for image recognition. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.770-778. https://doi.org/10.1109/CVPR.2016.90

Huang S, Dai R, Huang JJ, et al., 2020. Automatic modulation classification using gated recurrent residual network. *IEEE Internet Things J*, 7(8):7795-7807. https://doi.org/10.1109/JIOT.2020.2991052

Lin WS, Hou DB, Huang JS, et al., 2023. Transfer learning for automatic modulation recognition using a few modulated signal samples. *IEEE Trans Veh Technol*, 72(9):12391-12395. https://doi.org/10.1109/TVT.2023.3267270

Lin Y, Zhao HJ, Ma XF, et al., 2021. Adversarial attacks in modulation recognition with convolutional neural networks. *IEEE Trans Reliab*, 70(1):389-401. https://doi.org/10.1109/TR.2020.3032744

Long MS, Cao Y, Wang JM, et al., 2015. Learning transferable features with deep adaptation networks. Proc 32nd Int Conf on Int Conf on Machine Learning, p.97-105.

Lukito WD, Rashad FE, Hamid EY, 2021. Multi features-based baseband modulation classification using support vector machine. Proc Int Conf on Radar, Antenna, Microwave, Electronics, and Telecommunications, p.227-231. https://doi.org/10.1109/ICRAMET53537.2021.9650496

Meng F, Chen P, Wu LN, et al., 2018. Automatic modulation classification: a deep learning enabled approach. *IEEE Trans Veh Technol*, 67(11):10760-10772. https://doi.org/10.1109/TVT.2018.2868698

Nam H, Lee H, Park J, et al., 2021. Reducing domain gap by reducing style bias. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.8686-8695. https://doi.org/10.1109/CVPR46437.2021.00858

O'Shea TJ, Corgan J, Clancy TC, 2016. Convolutional radio modulation recognition networks. Proc 17th Int Conf on Engineering Applications of Neural Networks, p.213-226. https://doi.org/10.1007/978-3-319-44188-7_16

O'Shea TJ, Roy T, Clancy TC, 2018. Over-the-air deep learning based radio signal classification. *IEEE J Sel Top Signal Process*, 12(1):168-179. https://doi.org/10.1109/JSTSP.2018.2797022

Pan SJ, Yang Q, 2010. A survey on transfer learning. *IEEE Trans Knowl Data Eng*, 22(10):1345-1359. https://doi.org/10.1109/TKDE.2009.191

unedited

Perenda E, Rajendran S, Bovet G, et al., 2021. Learning the unknown: Improving modulation classification performance in unseen scenarios. Proc IEEE Conf on Computer Communications, p.1-10. https://doi.org/10.1109/INFOCOM42981.2021.9488835

Rajendran S, Meert W, Giustiniano D, et al., 2018. Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. *IEEE Trans Cognit Commun Netw*, 4(3):433-445. https://doi.org/10.1109/TCCN.2018.2835460

Ramjee S, Ju ST, Yang DY, et al., 2021. Ensemble wrapper subsampling for deep modulation classification. *IEEE Trans Cognit Commun Netw*, 7(4):1156-1170. https://doi.org/10.1109/TCCN.2021.3108809

Salama AA, Morsy ME, Darwish SH, et al., 2022. A novel SVM-based automatic modulation classifier. Proc Int Telecommunications Conf, p.1-3. https://doi.org/10.1109/ITC-Egypt55520.2022.9855683

Tan X, Xie ZD, Yuan XW, et al., 2022. Small sample signal modulation recognition based on higher-order cumulants and CatBoost. Proc 7$^{\text{th}}$ Int Conf on Communication, Image and Signal Processing, p.324-329. https://doi.org/10.1109/CCISP55629.2022.9974568

Tang B, Tu Y, Zhang ZY, et al., 2018. Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks. *IEEE Access*, 6:15713-15722. https://doi.org/10.1109/ACCESS.2018.2815741

Tzeng E, Hoffman J, Saenko K, et al., 2017. Adversarial discriminative domain adaptation. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.2962-2971. https://doi.org/10.1109/CVPR.2017.316

Wang MY, Lin Y, Tian Q, et al., 2021. Transfer learning promotes 6G wireless communications: recent advances and future challenges. *IEEE Trans Reliab*, 70(2):790-807. https://doi.org/10.1109/TR.2021.3062045

Wang Q, Du PF, Yang JY, et al., 2019. Transferred deep learning based waveform recognition for cognitive passive radar. *Signal Process*, 155:259-267. https://doi.org/10.1016/j.sigpro.2018.09.038

Wang ZG, Yan WZ, Oates T, 2017. Time series classification from scratch with deep neural networks: a strong baseline. Proc Int Joint Conf on Neural Networks, p.1578-1585. https://doi.org/10.1109/IJCNN.2017.7966039

Xiao WX, Ding ZM, Liu HF, 2021. Implicit semantic response alignment for partial domain adaptation. Proc 35$^{\text{th}}$ Int Conf on Neural Information Processing Systems, p.1059.

Xu Y, Li DZ, Wang ZY, et al., 2019. A deep learning method based on convolutional neural network for automatic modulation classification of wireless signals. Wireless Netw, 25(7):3735-3746. https://doi.org/10.1007/s11276-018-1667-6

Xu ZW, Han GJ, Liu L, et al., 2023. A lightweight specific emitter identification model for IIoT devices based on adaptive broad learning. *IEEE Trans Ind Inf*, 19(5):7066-7075. https://doi.org/10.1109/TII.2022.3206309

Yao ZS, Fu X, Guo LT, et al., 2023. Few-shot specific emitter identification using asymmetric masked auto-encoder. *IEEE Commun Lett*, 27(10):2657-2661. https://doi.org/10.1109/LCOMM.2023.3312669

Zhao ZQ, Khan FN, Li YB, et al., 2022. Application and comparison of active and transfer learning approaches for modulation format classification in visible light communication systems. *Opt Express*, 30(10):16351-16361. https://doi.org/10.1364/OE.456269

Zheng Y, Yu L, Yang RR, et al., 2021. A general 3D non-stationary massive MIMO GBSM for 6G communication systems. Proc IEEE Wireless Communications and Networking Conf, p.1-6. https://doi.org/10.1109/WCNC49053.2021.9417601

Zhou HJ, Wang X, Bai J, et al., 2022. Modulation signal recognition based on selective knowledge transfer. Proc IEEE Global Communications Conf, p.1875-1880. https://doi.org/10.1109/GLOBECOM48099.2022.10001238

Zhu YC, Zhuang FZ, Wang JD, et al., 2021. Deep subdomain adaptation network for image classification. *IEEE Trans Neural Netw Learn Syst*, 32(4):1713-1722. https://doi.org/10.1109/TNNLS.2020.2988928

Unedited