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-scenario signal modulation classification*

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

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

[2]*School of Artificial Intelligence and Computer Science, North China University of Technology, Beijing 100144, China*

[3]*Institute of Unmanned System, Beihang University, Beijing 100191, China*

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

Received Feb. 2, 2024; Revision accepted June 25, 2024; Crosschecked Mar. 27, 2025

Abstract: Automatic modulation classification (AMC) serves a challenging yet crucial role in wireless communications. 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 an adversarial transfer framework named frequency-learning adversarial networks (FLANs) based on transfer learning for cross-scenario signal classification. This method uses 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 difference between the source and target 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 about 5.2 percentage points in high signal-to-noise ratio (SNR) scenes on a cross-scenario real collected dataset CSRC2023.

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

https://doi.org/10.1631/FITEE.2400080 **CLC number:** TN911.72

1 Introduction

Automatic modulation classification (AMC), serving as an integral component of signal direction and demodulation in wireless communications, 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 used 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

‡ Corresponding authors

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

ORCID: Qinyan MA, https://orcid.org/0009-0007-0266-9727; Duona ZHANG, https://orcid.org/0000-0002-5567-0816; Yufeng WANG, https://orcid.org/0000-0001-8713-3153

processing time-sequenced signals and their interconnections; despite requiring large volumes of data, RNN has proved effective in signal modulation recognition (Rajendran et al., 2018).

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 higher stability. For instance, when collecting signals indoors using universal software radio peripheral (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 the 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 the 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, a label classifier, and a domain determiner (Ganin et al., 2015).

Inspired by adversarial networks and TL, we propose an adversarial transfer framework named frequency-learning adversarial networks (FLANs). It leverages a frequency adaptation (FA) module and a fitting channel adaptation (FCA) module to minimize the difference 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 inter-domain 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 5.2 percentage points (PPs) 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 framework consists of two modules: the FA module and the FCA module. These modules are effectively used to minimize the difference between the two channels.

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

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

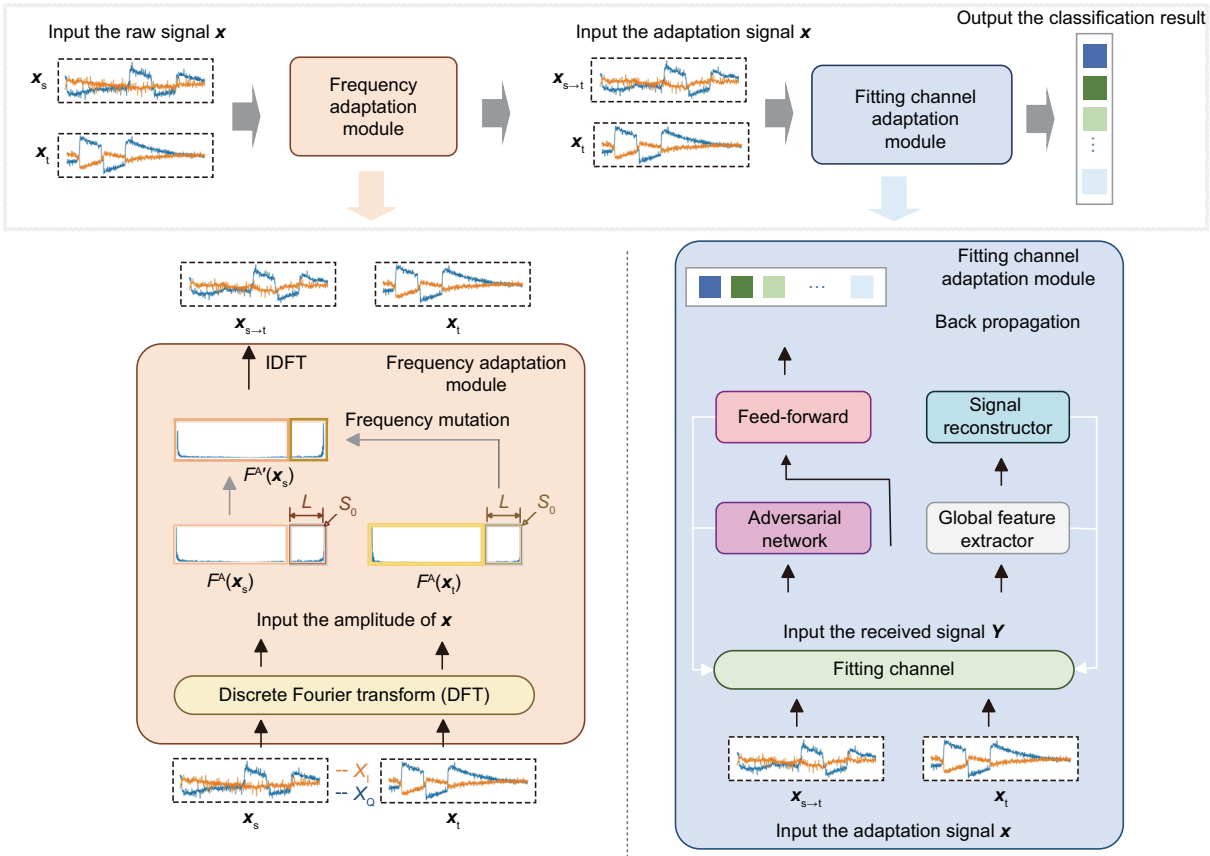

Fig. 1 Pipeline of FLANs framework, which uses the FA module and the FCA module to reduce the differences between two domains caused by channel environments. The FA module uses the stability of spectral characteristics during scenario changes. The FCA module applies adversarial training, aiming to address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

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 was presented for received wireless signals (Ramjee et al., 2021), which significantly reduced classifier training time and improved classification accuracy. Other works used residual neural networks (ResNets) (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 deep neural networks (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 test 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) used 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) was used to address transfer problems with signals under different sample rates (Wang Q et al., 2019), initializing the classifier of the target domain. This has also been adopted to address unstable model performance in low signal-to-noise ratio (SNR) scenarios (Xu Y et al., 2019).

With the proposal of 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 incorporated 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 domain in cross-scenario cases, auxiliary classifier GANs (ACGANs) have 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 pre-training and fine-tuning method, using 60 labeled samples, showing an obvious improvement in accuracy (14%). Zhu et al. (2021) proposed a 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) proposed a method named style-agnostic networks (SagNets) to reduce the domain gap by mitigating the style bias. The authors introduced a framework that incorporates domain adaptation and style transfer techniques between the source domain and the target domain via an intermediary domain. This approach aims to balance stylistic differences across domains, thereby enhancing model generalization capability and performance.

The introduction of a deep transfer network into adversarial networks leads to the proposal of DANN (Ganin et al., 2015), which consists 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 Proposed approach FLANs

3.1 Problem statement

In this subsection, the random modeling method is used 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 $\boldsymbol{x}(n)$ can be given as follows:

$$\boldsymbol{x}(n)=\boldsymbol{s}(n)\boldsymbol{h}(n)+\boldsymbol{w}(n),\ n=1,2,...,N, \tag{1}$$

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

As the impulse function $\boldsymbol{h}(n)$ varies with the channel environment, we use the random modeling method in practical channels to further demonstrate the effect of the channel environment on the 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 primary path signal and the 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):

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

where K signifies the average power ratio of the primary path signal to the randomly reflected path signals (i.e., the power ratio of the LOS path to the multipath), α denotes the attenuation factor of the primary path signal, f_{c} symbolizes the frequency of the primary path signal, L_{p} indicates the number of randomly reflected paths, t denotes the time, 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 test data, will likely exhibit poor classification performance. Thus, we introduce TL with the hope of transferring existing knowledge from the source domain to the target domain 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 Domain adversarial neural network

DANN (Ganin et al., 2015), 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 models 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: feature extractor $\boldsymbol{G}_{\mathrm{f}}$, label predictor $\boldsymbol{G}_{\mathrm{l}}$, and domain classifier $\boldsymbol{G}_{\mathrm{d}}$. Their functions are as follows:

The feature extractor $\boldsymbol{G}_{\mathrm{f}}$ learns a function that maps an example into a new D-dimensional representation, parameterized by a matrix–vector pair $(\boldsymbol{W},\boldsymbol{b}) \in \mathbb{R}^{D\times m}\times\mathbb{R}^{D}$ and $\boldsymbol{x}\in\mathbb{R}^{m}$, as defined by Eq. (3):

$$\begin{cases} \boldsymbol{G}_{\mathrm{f}}(\boldsymbol{x};\boldsymbol{W},\boldsymbol{b}) = \mathrm{sigmoid}(\boldsymbol{W}\boldsymbol{x}+\boldsymbol{b}), \\ \mathrm{sigmoid}(\boldsymbol{a}) = \left[\dfrac{1}{1+\exp(-a_i)}\right]_{i=1}^{|\boldsymbol{a}|}. \end{cases} \tag{3}$$

The label predictor $\boldsymbol{G}_{\mathrm{l}}$ learns a function parameterized by a pair $(\boldsymbol{V},\boldsymbol{c}) \in \mathbb{R}^{L\times D}\times\mathbb{R}^{L}$ and $\boldsymbol{G}_{\mathrm{f}}(\boldsymbol{x})\in\mathbb{R}^{D}$. Using the softmax function, each component of the vector $\boldsymbol{G}_{\mathrm{l}}(\boldsymbol{G}_{\mathrm{f}}(\boldsymbol{x}))$ indicates the conditional probability that the neural network assigns $\boldsymbol{x}$ to the class represented by that component:

$$\begin{cases} \boldsymbol{G}_{\mathrm{l}}(\boldsymbol{G}_{\mathrm{f}}(\boldsymbol{x});\boldsymbol{V},\boldsymbol{c}) = \mathrm{softmax}(\boldsymbol{V}\boldsymbol{G}_{\mathrm{f}}(\boldsymbol{x})+\boldsymbol{c}), \\ \mathrm{softmax}(\boldsymbol{a}) = \left[\dfrac{\exp(-a_i)}{\sum\limits_{j=1}^{|\boldsymbol{a}|}\exp(-a_j)}\right]_{i=1}^{|\boldsymbol{a}|}. \end{cases} \tag{4}$$

The domain classifier $\boldsymbol{G}_{\mathrm{d}}$ learns a logistic regressor, parameterized by a vector–scalar pair $(\boldsymbol{u},z)\in\mathbb{R}^{D}\times\mathbb{R}$, which models the probability that a given input originates from either the source domain D_{s} or the target domain D_{t}:

$$\boldsymbol{G}_{\mathrm{d}}(\boldsymbol{G}_{\mathrm{f}}(\boldsymbol{x});\boldsymbol{u},z) = \mathrm{sigmoid}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{G}_{\mathrm{f}}(\boldsymbol{x})+z). \tag{5}$$

Both the label predictor $\boldsymbol{G}_{\mathrm{l}}$ and the domain classifier $\boldsymbol{G}_{\mathrm{d}}$ take the output of the feature classifier $\boldsymbol{G}_{\mathrm{f}}$ as their input. This ensures that the features extracted from the signals are indistinguishable for $\boldsymbol{G}_{\mathrm{d}}$, while $\boldsymbol{G}_{\mathrm{l}}$ is still able to classify the modulation format of the signals. The introduction of adversarial training aims to optimize the following equation:

$$\begin{aligned}&E(\boldsymbol{W},\boldsymbol{V},\boldsymbol{b},\boldsymbol{c},\boldsymbol{u},z)\\&=\frac{1}{n}\sum_{i=1}^{n}L_{\mathrm{l}}^{i}(\boldsymbol{W},\boldsymbol{b},\boldsymbol{V},\boldsymbol{c})-\lambda\left(\frac{1}{n}\sum_{i=1}^{n}L_{\mathrm{d}}^{i}(\boldsymbol{W},\boldsymbol{b},\boldsymbol{u},z)\right.\\&\left.+\frac{1}{n'}\sum_{i=n+1}^{N}L_{\mathrm{d}}^{i}(\boldsymbol{W},\boldsymbol{b},\boldsymbol{u},z)\right),\end{aligned}\tag{6}$$

where n is the number of classes and $n'=N-(n+1)$.

During training, $\boldsymbol{G}_{\mathrm{f}}$ and $\boldsymbol{G}_{\mathrm{l}}$ compete with $\boldsymbol{G}_{\mathrm{d}}$ in an adversarial manner. Here, $\boldsymbol{G}_{\mathrm{f}}$ learns to map signals into a new D-dimensional representation, enabling $\boldsymbol{G}_{\mathrm{l}}$ to classify the modulation format of the signals while diminishing the capability of $\boldsymbol{G}_{\mathrm{d}}$ to differentiate the domain of signals.

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

1. High feature consistency requirement. The performance of DANN heavily depends on the consistency of features between the source and 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 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 challenging to explain.

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

3.3 FA module

This subsection 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 using low-order phase shift keying (PSK) and most other modulation formats, majority of high-frequency components are composed mainly of noise components, 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 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 $\boldsymbol{x}_n$ is given as follows:

$$\boldsymbol{X}_k=F(\boldsymbol{x})=\sum_{n=0}^{N-1}\boldsymbol{x}_n\mathrm{e}^{-\mathrm{j}2\pi\frac{k}{N}n},\tag{7}$$

where F represents the DFT; accordingly, F^{-1} is the inverse DFT (IDFT). Additionally, F^{A} and F^{P} are the amplitude and phase components of the process F of a complex signal, respectively. The value of k ranges from 0 to $N-1$, representing different frequency components. Further, we designate a mask code $M_L=1_{n\in[L*N:]}$, 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 number of sample points. For two signals from two domains $\boldsymbol{x}_{\mathrm{s}}\sim D_{\mathrm{s}}$, $\boldsymbol{x}_{\mathrm{t}}\sim D_{\mathrm{t}}$, the process of frequency spectrum mutation can be formalized as follows:

$$\boldsymbol{x}_{\mathrm{s}\to\mathrm{t}}=F^{-1}([M_L\circ F^{\mathrm{A}}(\boldsymbol{x}_{\mathrm{t}})+\bar{M}_L\circ F^{\mathrm{A}}(\boldsymbol{x}_{\mathrm{s}}),F^{\mathrm{P}}(\boldsymbol{x}_{\mathrm{s}})]),\tag{8}$$

where $\bar{M}_L = 1 - M_L$.

This implies that the high-frequency component of the source signal's amplitude, $F^{\mathrm{A}}(\boldsymbol{x}_{\mathrm{s}})$, is replaced with that of the target signal $\boldsymbol{x}_{\mathrm{t}}$. The modified amplitude component, along with its unaltered phase component, undergoes processing through the IDFT module and is mapped back to the signal $\boldsymbol{x}_{\mathrm{s}\to\mathrm{t}}$, whose modulation format is the same as that of $\boldsymbol{x}_{\mathrm{s}}$ 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 $\boldsymbol{x}_{\mathrm{s}\to\mathrm{t}}$ remains identical to the original signal $\boldsymbol{x}_{\mathrm{s}}$. Correspondingly, L =1 represents that the amplitude of $\boldsymbol{x}_{\mathrm{s}}$ is completely replaced by that of $\boldsymbol{x}_{\mathrm{t}}$. In Fig. 2, the output of the IDFT module is attached with target-domain channel characteristics with L=0.01 which is applied as the 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 the received signals simultaneously. Fig. 3 showcases an 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 $\boldsymbol{H}$ 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). LSF consists of path loss (PL), shadowing loss (SL), blockage loss (BL), and gas absorption loss (AL), and is presented as follows (Wang XC et al., 2022):

$$\boldsymbol{H} = [\mathrm{PL}\cdot\mathrm{SL}\cdot\mathrm{BL}\cdot\mathrm{AL}]^{1/2}\boldsymbol{H}_{\mathrm{s}}, \tag{9}$$

where $\boldsymbol{H}_{\mathrm{s}}$ is the SSF matrix and its element can be further represented as follows:

$$\begin{aligned} &h_{q,p}(t,\tau)\\ =&\sqrt{\frac{K_{\mathrm{RF}}(t)}{K_{\mathrm{RF}}(t)+1}}h^{\mathrm{L}}_{q,p}(t,\tau)+\sqrt{\frac{1}{K_{\mathrm{RF}}(t)+1}}h^{\mathrm{N}}_{q,p}(t,\tau), \end{aligned} \tag{10}$$

where $h_{q,p}(t,\tau)$ represents the channel response from the transmitting antenna q to the receiving antenna p, $K_{\mathrm{RF}}(t)$ is the Rician factor, $h^{\mathrm{L}}_{q,p}(t,\tau)$ is the LOS component, and $h^{\mathrm{N}}_{q,p}(t,\tau)$ is the non-line-of-sight (NLOS) component. These can be further represented as follows:

$$\begin{aligned} h^{\mathrm{L}}_{q,p}(t,\tau) = &\begin{bmatrix} F_{q,\mathrm{V}}\left(\phi^{\mathrm{R}}_{E,\mathrm{LOS}}(t),\phi^{\mathrm{R}}_{A,\mathrm{LOS}}(t)\right)\\ F_{q,\mathrm{H}}\left(\phi^{\mathrm{R}}_{E,\mathrm{LOS}}(t),\phi^{\mathrm{R}}_{A,\mathrm{LOS}}(t)\right)\end{bmatrix}^{\mathrm{T}}\\ &\cdot\begin{bmatrix} \mathrm{e}^{\mathrm{j}\theta^{\mathrm{V,V}}_{\mathrm{LOS}}} & 0\\ 0 & -\mathrm{e}^{\mathrm{j}\theta^{\mathrm{H,H}}_{\mathrm{LOS}}}\end{bmatrix}\\ &\cdot\begin{bmatrix} F_{p,\mathrm{V}}\left(\phi^{\mathrm{T}}_{E,\mathrm{LOS}}(t),\phi^{\mathrm{T}}_{A,\mathrm{LOS}}(t)\right)\\ F_{p,\mathrm{H}}\left(\phi^{\mathrm{T}}_{E,\mathrm{LOS}}(t),\phi^{\mathrm{T}}_{A,\mathrm{LOS}}(t)\right)\end{bmatrix}\\ &\cdot\mathrm{e}^{\mathrm{j}2\pi f_c\tau^{\mathrm{LOS}}_{q,p}(t)}\delta\left(\tau-\tau^{\mathrm{LOS}}_{q,p}(t)\right), \end{aligned} \tag{11}$$

$$\begin{aligned} &h^{\mathrm{N}}_{q,p}(t,\tau)\\ =&\sum_{n=1}^{N_{q,p}(t)}\sum_{m=1}^{M_n(t)}\begin{bmatrix} F_{q,\mathrm{V}}\left(\phi^{\mathrm{R}}_{E,m_n}(t),\phi^{\mathrm{R}}_{A,m_n}(t)\right)\\ F_{q,\mathrm{H}}\left(\phi^{\mathrm{R}}_{E,m_n}(t),\phi^{\mathrm{R}}_{A,m_n}(t)\right)\end{bmatrix}^{\mathrm{T}}\\ &\cdot\begin{bmatrix} \mathrm{e}^{\mathrm{j}\theta^{\mathrm{V,V}}_{m_n}} & \sqrt{\kappa^{-1}_{m_n}(t)}\mathrm{e}^{\mathrm{j}\theta^{\mathrm{V,H}}_{m_n}}\\ \sqrt{\kappa^{-1}_{m_n}(t)}\mathrm{e}^{\mathrm{j}\theta^{\mathrm{H,V}}_{m_n}} & \mathrm{e}^{\mathrm{j}\theta^{\mathrm{H,H}}_{m_n}}\end{bmatrix}\\ &\cdot\begin{bmatrix} F_{p,\mathrm{V}}\left(\phi^{\mathrm{T}}_{E,m_n}(t),\phi^{\mathrm{T}}_{A,m_n}(t)\right)\\ F_{p,\mathrm{H}}\left(\phi^{\mathrm{T}}_{E,m_n}(t),\phi^{\mathrm{T}}_{A,m_n}(t)\right)\end{bmatrix}\\ &\cdot\sqrt{P_{q,p,m_n}(t)}\mathrm{e}^{\mathrm{j}2\pi f_c\tau_{q,p,m_n}(t)}\delta\left(\tau-\tau_{q,p,m_n}(t)\right), \end{aligned} \tag{12}$$

where $(\cdot)^{\mathrm{T}}$ denotes the transpose operation; $F^{\mathrm{T(R)}}_{p(q),\mathrm{V}}$ and $F^{\mathrm{T(R)}}_{p(q),\mathrm{H}}$ represent the vertical polarization and horizontal polarization at the Tx (Rx) side, respectively; κ_{m_n} denotes the cross-polarization ratio on the subpath m within the multipath route n.

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

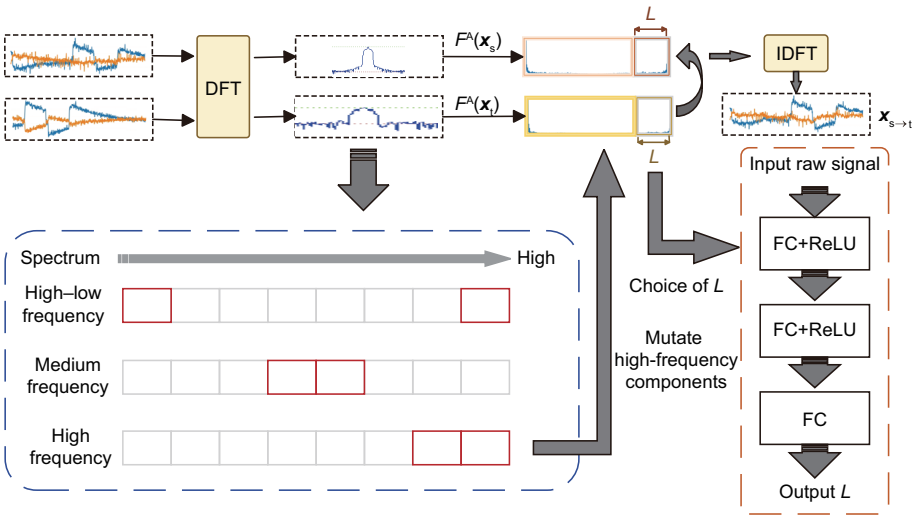

Fig. 2 Process of the FA module. By mutating high-frequency components, the FA module attaches source signals with target-domain channel characteristics without altering the modulation format. A regression network is introduced to obtain the optimal exchange length ratio L for every single source signal

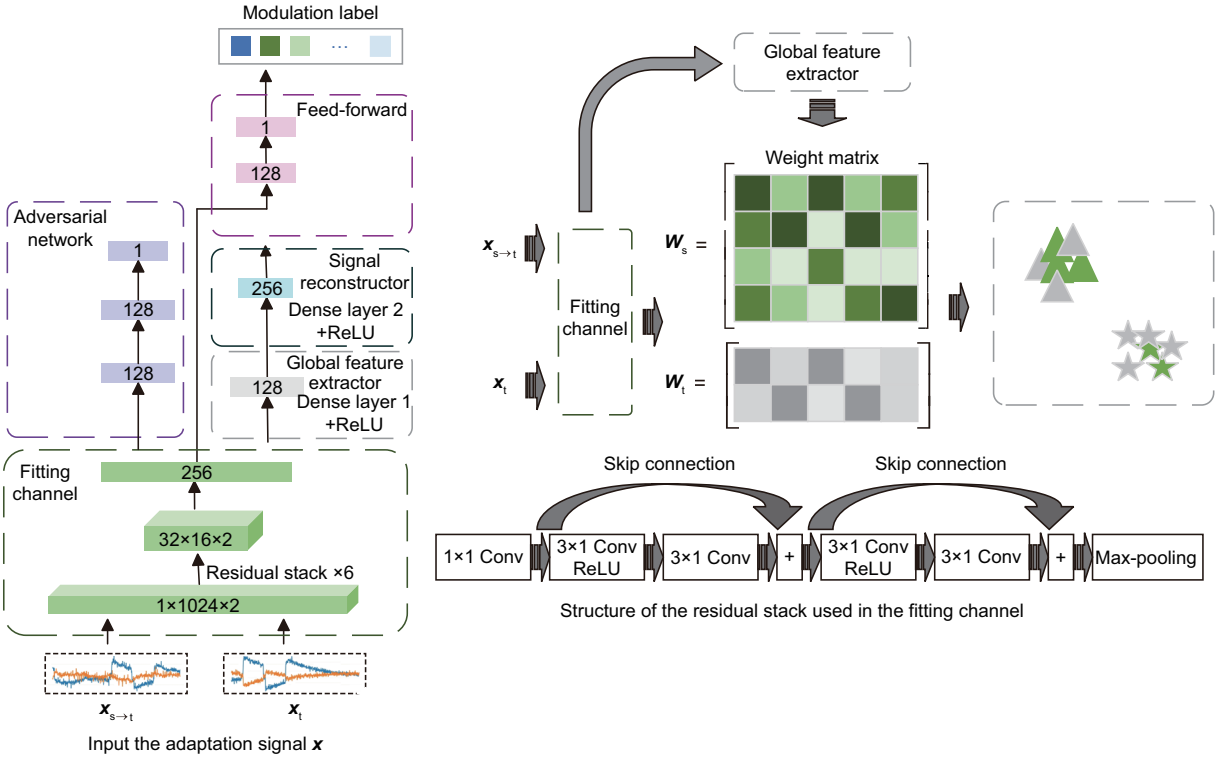

Fig. 3 Overview of the FCA module, which applies a neural network to build a virtual channel condition. By incorporating a global feature extractor, the interpretability of the network is enhanced. During the adversarial training process, the received signal is weighted with the extracted global feature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

can be given as follows:

$$\begin{cases} \boldsymbol{y} = H(\boldsymbol{x}; \boldsymbol{W}, \boldsymbol{b}) = \mathrm{ReLU}(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}), \\ \mathrm{ReLU}(\boldsymbol{x}) = \max(\boldsymbol{0}, \boldsymbol{x}). \end{cases} \tag{13}$$

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:

$$F_{\mathrm{e}}(\boldsymbol{y};\boldsymbol{M},\boldsymbol{n})=\mathrm{ReLU}(\boldsymbol{M}\boldsymbol{y}+\boldsymbol{n}), \tag{14}$$

where $\boldsymbol{M}$ and $\boldsymbol{n}$ represent the weight matrix and bias vector of F_{e}, respectively. 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:

$$F_{\mathrm{r}}(\boldsymbol{y};\boldsymbol{M}',\boldsymbol{n}')=\mathrm{ReLU}(\boldsymbol{M}'\boldsymbol{y}+\boldsymbol{n}'), \tag{15}$$

where $\boldsymbol{M}'$ and $\boldsymbol{n}'$ represent the weight matrix and bias vector of F_{r}, respectively. Concerning adversarial TL, it is similar to the domain classifier $\boldsymbol{G}_{\mathrm{d}}$:

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

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

$$d=|p_{\mathrm{s}}(\boldsymbol{x}'_{\mathrm{s}},\boldsymbol{y}_{\mathrm{s}};n_{\mathrm{s}}),p_{\mathrm{t}}(\boldsymbol{x}'_{\mathrm{t}},\boldsymbol{y}_{\mathrm{t}};n_{\mathrm{t}})|_D, \tag{17}$$

where $|\cdot|_D$ signifies that the domain distance is reflected in the discriminator D's ability to distinguish domains, n_{s} and n_{t} are the numbers of signals in the source and target domains respectively, and p represents the probability distribution. We primarily assume that, through the fitting channel, 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 original domain. Therefore, the received signal is directly inputted into a feedforward network to determine the modulation format. This function is akin to the label predictor $\boldsymbol{G}_{\mathrm{l}}$:

$$\boldsymbol{G}_{\mathrm{l}}(\boldsymbol{y};\boldsymbol{V},\boldsymbol{c})=\mathrm{ReLU}(\boldsymbol{V}\boldsymbol{y}+\boldsymbol{c}). \tag{18}$$

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

$$\begin{aligned}&E(\boldsymbol{W},\boldsymbol{V},\boldsymbol{b},\boldsymbol{c},\boldsymbol{u},z,\boldsymbol{M},\boldsymbol{n},\boldsymbol{M}',\boldsymbol{n}')\\&=\frac{1}{n}\sum_{i=1}^{n}L_{\mathrm{l}}^{i}(\boldsymbol{W},\boldsymbol{b},\boldsymbol{V},\boldsymbol{c})+\frac{1}{n}\sum_{i=1}^{n}L_{\mathrm{e}}^{i}(\boldsymbol{W},\boldsymbol{b},\boldsymbol{M},\boldsymbol{n})\\&\quad+\frac{1}{n}\sum_{i=1}^{n}L_{\mathrm{r}}^{i}(\boldsymbol{W},\boldsymbol{b},\boldsymbol{M}',\boldsymbol{n}')\\&\quad-\lambda\left(\frac{1}{n}\sum_{i=1}^{n}L_{\mathrm{d}}^{i}(\boldsymbol{W},\boldsymbol{b},\boldsymbol{u},z)\right.\\&\quad\left.+\frac{1}{n'}\sum_{i=n+1}^{N}L_{\mathrm{d}}^{i}(\boldsymbol{W},\boldsymbol{b},\boldsymbol{u},z)\right).\end{aligned} \tag{19}$$

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:

$$L=L_{\mathrm{c}}+L_{\mathrm{d}}+\alpha L_{\mathrm{r}}+\beta L_{\mathrm{w}}, \tag{20}$$

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 the global feature discovery and alignment, respectively. Each loss function's specific features are detailed in the following, 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 reconstructor, respectively. They maintain the ability to distinguish categories and performance of reconstruction, as follows:

$$L_{\mathrm{c}}=L_{\mathrm{c_s}}+L_{\mathrm{c_t}}=L_{\text{cross-entropy}}(\boldsymbol{y}_{\mathrm{s}},\boldsymbol{y}'_{\mathrm{s}})+L_{\text{entropy}}(\boldsymbol{y}'_{\mathrm{t}}), \tag{21}$$

where $L_{\mathrm{c_s}}$ and $L_{\mathrm{c_t}}$ represent the losses for source and target-domain label modulation classification, respectively. $\boldsymbol{y}'_{\mathrm{s}}$ represents the classification label of the source-domain signal's modulation by the label discriminator, and $\boldsymbol{y}_{\mathrm{s}}$ represents the true label of the source-domain data. Similarly, $\boldsymbol{y}'_{\mathrm{t}}$ denotes the classification label of the target-domain signal's modulation by the label discriminator. $L_{\text{cross-entropy}}$ is the cross-entropy loss between the computed values:

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

L_{entropy} denotes the computed entropy used as a loss function:

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

Reconstruction loss function L_{r} is given by

$$L_{\text{r}} = L_{\text{r}_\text{s}} + L_{\text{r}_\text{t}} = L_{\text{MSE}}(f_{\text{s}}, f'_{\text{s}}) + L_{\text{MSE}}(f_{\text{t}}, f'_{\text{t}}), \tag{24}$$

where L_{r_s} represents the loss function for the source domain's received signal f_{s} inputted to the global feature extractor and the reconstructed signal f'_{s}. L_{r_t} is the counterpart of L_{r_s} in the target domain. L_{MSE} denotes the mean-squared error (MSE) between f_{s} and f'_{s}:

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

The domain loss function L_{d}, commonly used in GAN, is as follows:

$$L_{\text{d}} = L_{\text{d}_\text{s}} + L_{\text{d}_\text{t}} = L_{\text{BCE}}(d_{\text{s}}, d'_{\text{s}}) + L_{\text{BCE}}(d_{\text{t}}, d'_{\text{t}}), \tag{26}$$

where L_{d_s} and L_{d_t} represent the losses when performing domain discrimination on the source- and target-domain signals, respectively. $L_{\text{BCE}}(d_{\text{s}}, d'_{\text{s}})$ represents the binary cross-entropy (BCE) loss between d_{s} and d'_{s}:

$$\begin{aligned} &L_{\text{BCE}}(\boldsymbol{Y}, f(\boldsymbol{x})) \\ &= -\frac{1}{M}\sum_{i=1}^{M} (Y_i \log f(x_i) + (1 - Y_i)\log(1 - f(x_i))). \end{aligned} \tag{27}$$

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 the received signals from different domains, as shown below:

$$L_{\text{w}} = L_{\text{MSE}}(\overline{\boldsymbol{w}}_{\text{s}}, \overline{\boldsymbol{w}}_{\text{t}}), \tag{28}$$

where $\overline{\boldsymbol{w}}_{\text{s}}$ 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:

$$\overline{\boldsymbol{w}}_{\text{s}} = [\overline{w}_{\text{s}_1}, \overline{w}_{\text{s}_2}, ..., \overline{w}_{\text{s}_N}], \tag{29}$$

where $\overline{w}_{\text{s}_i}$ is the average of the column sums of the weight matrix $\boldsymbol{W}_{\text{s}}$ of the extracted global feature and the received signals:

$$\boldsymbol{W}_{\text{s}} = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,N} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,N} \\ \vdots & \vdots & & \vdots \\ w_{N,1} & w_{N,2} & \cdots & w_{N,N} \end{bmatrix}, \tag{30}$$

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

$\overline{\boldsymbol{w}}_{\text{t}}$ is the counterpart of $\overline{\boldsymbol{w}}_{\text{s}}$ in the target domain.

4 Experiments

This section first focuses on the selection of our dataset CSRC2023 and implementation details of experiments. Then, the proposed method is compared with the baseline approaches across a variety of SNRs. The performance index for modulation classification used in this study is accuracy. This section also showcases the constellations of the selected 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 position 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 adopt 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 8, respectively.

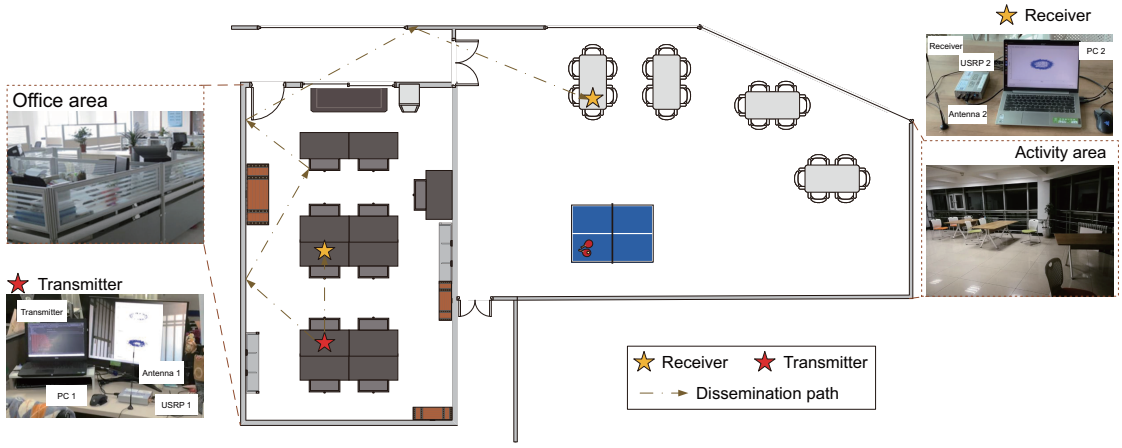

Fig. 4 Signal acquisition channel environment. The transmitter is placed in the office area, while the receiver is located in the same area and activity area to establish the LOS and NLOS channels, respectively. In the planar graph, yellow lines represent potential dissemination paths of the signal.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

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 creat a new dataset using the GNU radio, which provides numerous tools for creating robust datasets. The signals are modulated and transmitted using USRP B210 software defined radio (SDR). The signals are then sent through a finite impulse response (FIR) filter, transmitted via USRP B210, and subsequently received by a second B210.

Moreover, we introduce additive Gaussian white noise to the signals at the receiving end. The SNR ranges from -20 dB to 30 dB in a step of 2 dB. For the training dataset, we have 80 samples for each SNR and each modulation method, resulting in a total of $80 \times 26 \times 5 = 10\,400$ samples. The test dataset is similarly structured, yielding a total of $400 \times 26 \times 5 = 52\,000$ samples.

4.1.2 Implementation details

This paper proposes an adversarial network called FLANs, consisting of two modules: the FA module and the FCA module. The latter one is composed of five components: fitting channel, global feature extractor, signal reconstructor, adversarial network, and 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 create a well-annotated source dataset D_{s}, comprising sufficient signal samples $\boldsymbol{X}_{\mathrm{s}}$ and the corresponding labels $\boldsymbol{Y}_{\mathrm{s}}$. We also establish a target dataset D_{t} that has an equal number of signal samples $\boldsymbol{X}_{\mathrm{t}}$ but lacks 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 $D_{\mathrm{s}} = \{\boldsymbol{x}_i, \boldsymbol{y}_i, \mathrm{SNR}_i\}_{i=1}^{n_{\mathrm{s}}}$, where $\boldsymbol{x}_i$ represents the raw signal, $\boldsymbol{y}_i$ signifies the modulation label, and SNR_i denotes the SNR of each signal. Likewise, the target dataset is represented as $D_{\mathrm{t}} = \{\boldsymbol{x}_j, \mathrm{SNR}_j\}_{j=1}^{n_{\mathrm{t}}}$, which lacks modulation labels.

Except for the nontransfer method, which uses TensorFlow, all other evaluated AMC methods are implemented using PyTorch. For the optimizer, we choose 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 is trained through 20 000 iterations, and test is carried out every 200 iterations. Both processes use a batch size of 36. The model is trained and tested on a graphics processing unit (GPU) server equipped with a single NVIDIA RTX 2070 card. We calculate the classification accuracies under all SNR conditions and use the top-1 accuracy as the criterion for comparing techniques and tuning the network.

4.2 Performance analysis of FLANs

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 DANN, DSAN, and SagNets. DANN also uses adversarial training and a residual structure. DSAN and SagNets are further studies based on the 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) FA, by extracting the global features of signals processed through the fitting channel and balancing the global feature and the received signal, significantly improves the modulation classification performance. It surpasses the baseline methods ResNet_ST and DANN in the classification accuracy, achieving 30.75 PPs over ResNet_ST (deep residual neural network trained on the source dataset and tested on the target dataset) and 13.90 PPs over DANN. Under these conditions, the top-1 accuracy of FLANs w/o FA reaches 68.06%, outperforming DANN which achieves 54.15%, and DSAN which achieves 66.54%. SagNets method performs excellently in low SNRs, but with the addition of our proposed FA module, the top-1 accuracy of FLANs jumps to 91.67%, about 23.61 PPs higher than that of FLANs w/o FA and 5.22 PPs higher than that of SagNets. ResNet_TT is trained on the target dataset (with labels). The proposed FLANs performs significantly well in high SNRs, and its performance differs from that of ResNet_TT by only 8.33 PPs.

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

Method	Top-1 accuracy (%)	Average accuracy (%)
ResNet_ST	37.31	25.56
ResNet_TT	100.00	66.24
DANN	54.15	40.29
DSAN	66.54	50.38
SagNets	86.45	65.90
FLANs w/o FA	68.06	44.74
FLANs	91.67	56.76

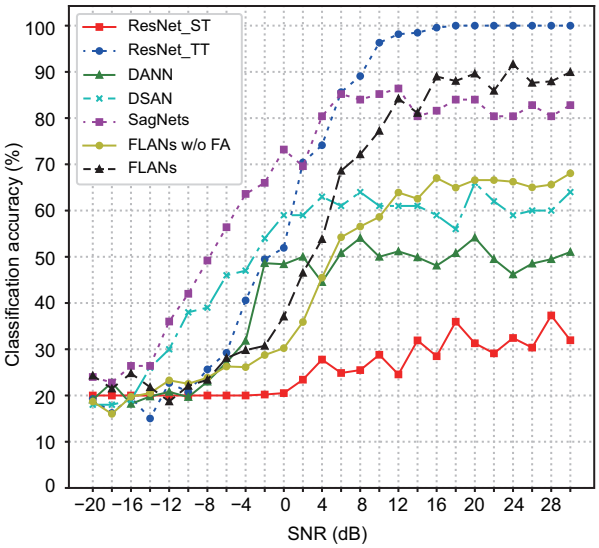

Fig. 5 Comparison results of FLANs and baseline methods

4.2.2 Comparison with transfer channels

For further investigation of the model's generalization capability in various transfer channel scenarios, a strong LOS (SLOS) channel is involved in this subsection as another target domain. We select 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 are collected from the weak LOS (WLOS) channel in a crowded indoor scene, with more objects obstructing the LOS path. In this subsection, we compare the performance of FLANs and the baselines under two transfer scenarios: WLOS to NLOS and WLOS to SLOS.

Fig. 6 presents the comparison results wherein each transfer scenario includes four models: ResNet_ST, ResNet_TT, DANN, and FLANs. The blue ones represent transfer scenario 1, which is the transfer from WLOS to NLOS. The red ones represent transfer scenario 2, which is the transfer from WLOS to SLOS (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 achieves an accuracy of 70% when SNR is >14 dB, far surpassing its performance in scenario 1 (32%). However, with the introduction

of the adversarial training method, DANN achieved about 10 PPs over ResNet_ST. It is notable that both the proposed FLANs and ResNet_TT reached 100% accuracy when SNR is >8 dB. Furthermore, while −2 dB < SNR < 8 dB, the accuracy of the proposed method exceeds that of the ResNet_TT by up to 14 PPs.

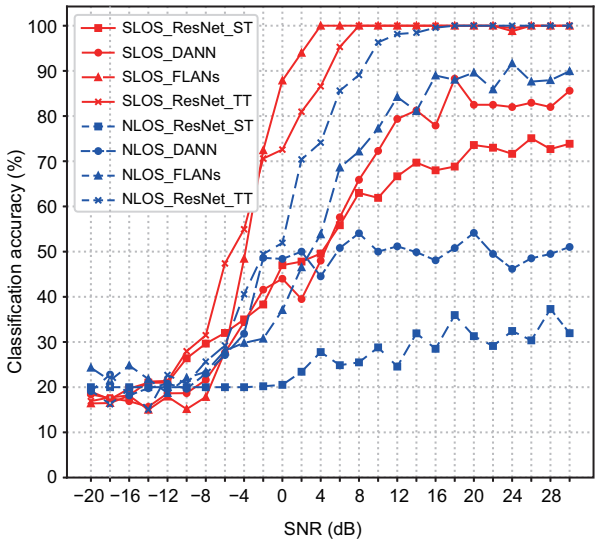

Fig. 6 Comparison of different transfer channels: WLOS to SLOS and WLOS to NLOS. References to color refer to the online version of this figure

Table 2 Mapping table for transfer scenarios

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

4.2.3 Classification performance by modulation mode

Fig. 7 illustrates the classification performance of FLANs for individual modulation modes. When SNR was <−2 dB, it only distinguished between BPSK and GFSK. 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

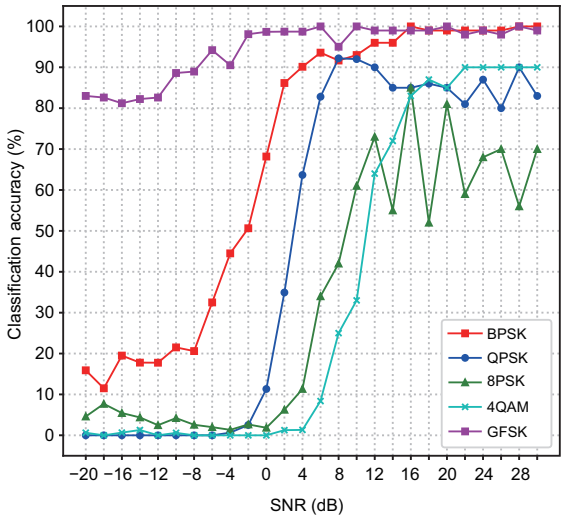

Fig. 7 Time domain plot of each modulation mode from transmitters, LOS channel, and NLOS channel

classification accuracy. Fig. 8 presents the confusion matrices for FLANs across all five classes with varying SNR. These matrices indicate that the primary cause of classification error is the confusion between QPSK, 8PSK, and 4QAM, due to the similar modulation process.

The selected samples and their constellations from the transmitter and the receiver under different channel conditions are depicted in Figs. 9 and 10.

4.3 Performance analysis of FA module

In this subsection, we primarily investigate 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) the position of the exchange in the frequency spectrum—broadly divided into high-, medium-, and low-frequency components; (2) the spectral exchanged length, represented as a ratio L.

The effect of the FA module is shown in Fig. 11. When high-frequency components are included in the exchange, the classification top-1 accuracy of FLANs significantly improves from 67.18% to 91.38%, a surge of 24.2 PPs. Even when compared to the spectral exchange of combined high- and low-frequency components, which results in a top-1 accuracy of 87.88%, the high-frequency-only component exchange still yields an improvement of 3.5 PPs.

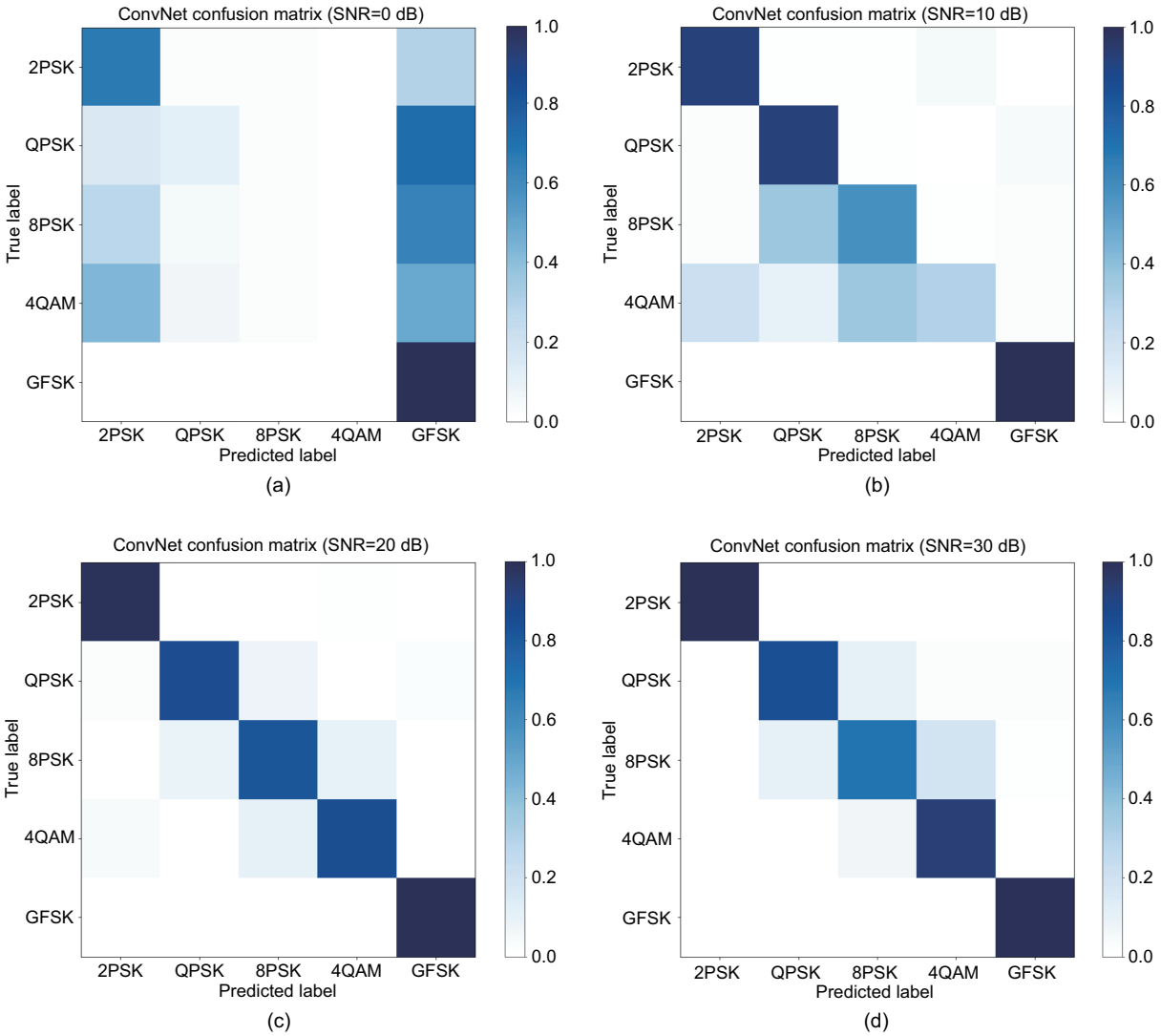

Fig. 8 Confusion matrix of FLANs with different SNRs: (a) 0 dB; (b) 10 dB; (c) 20 dB; (d) 30 dB

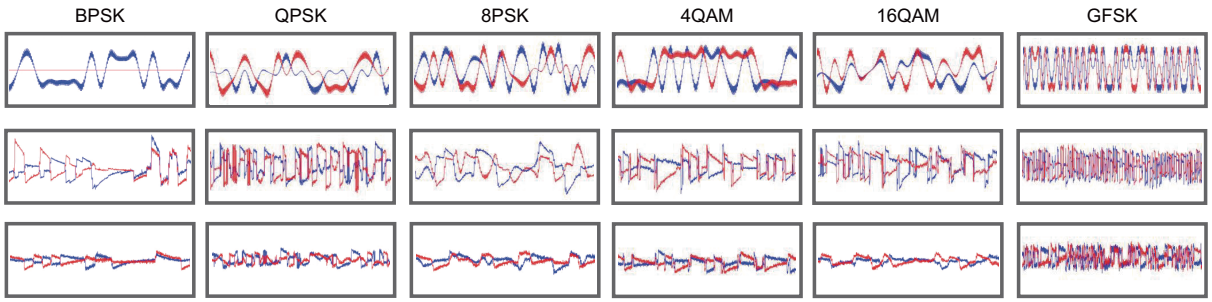

Fig. 9 Classification results of FLANs for individual modulation model

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 accuracy at low SNRs in contrast to conditions without mutation. This result suggests that the high-frequency signal components

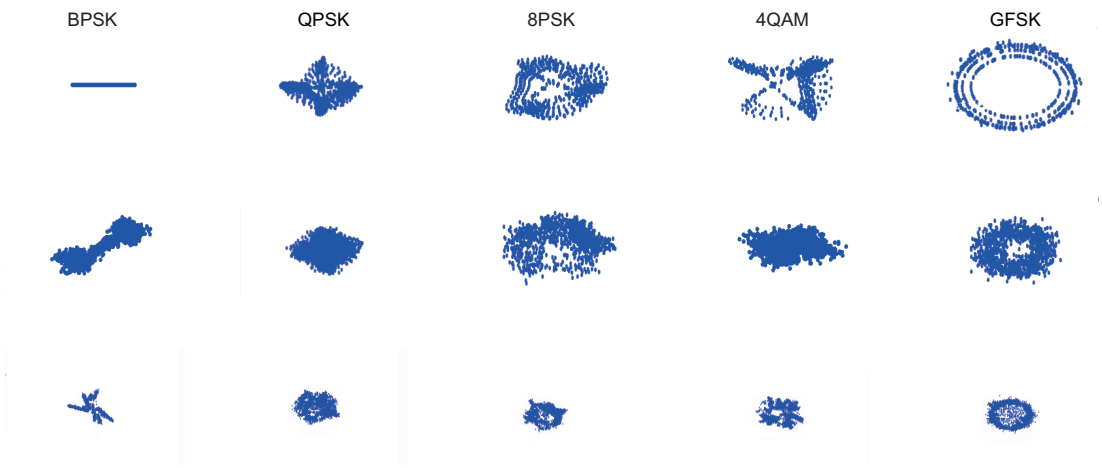

Fig. 10 Selected radio signal constellations of each modulation mode from transmitters, LOS channel, and NLOS channel

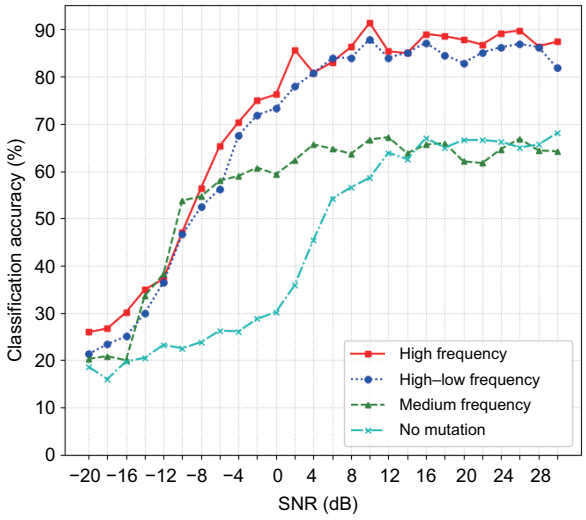

Fig. 11 Accuracy with position exchanging in the frequency spectrum

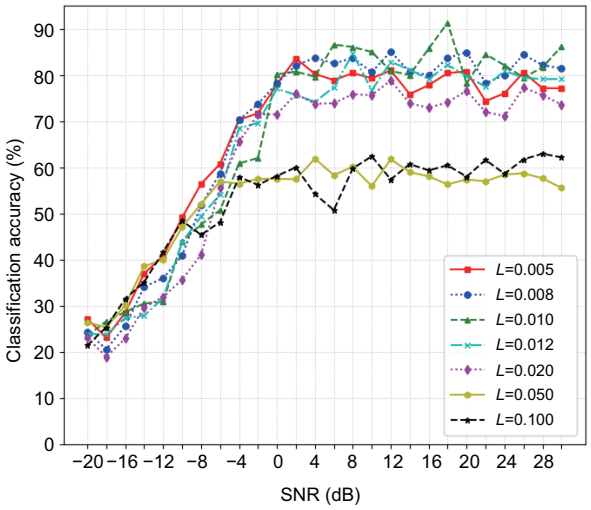

Fig. 12 Accuracy under different values of L with the variant of SNR

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 $L \in [0.005, 0.100]$, the recognition accuracy tends to increase at first. However, at $L = 0.010$, the top-1 accuracy declines; at $L = 0.008$, the average accuracy decreases. The most obvious lifting effect is obtained at $L = 0.010$. Therefore, we suggest $L = 0.010$ as the appropriate initial value for the FA module. We also observe that the impact of length is predominantly apparent at high SNRs.

Last, we evaluate the efficacy of the FA module in DANN and FLANs (Fig. 13). The results reveal that the application of the FA module leads to significant improvements in FLANs when SNR>4 dB. Interestingly, the extent of improvement shows significant variations. The accuracy of DANN is improved

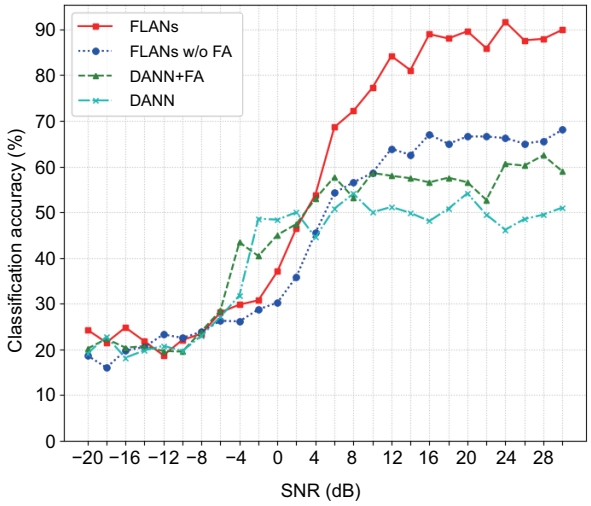

Fig. 13 Comparison results of DANN and FLANs with or without the FA module

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

L	Average accuracy (%)	Top-1 accuracy (%)
0.005	66.41	83.58
0.008	**67.22**	85.07
0.010	66.68	**91.32**
0.012	64.88	84.64
0.020	61.10	78.83
0.050	52.37	61.90
0.100	52.30	62.95

The best results are in bold

by 8 PPs when 8 dB<SNR<20 dB. However, The accuracy of FLANs demonstrates a greater extent of improvement (approximately 25 PPs) in a broader range of SNR, which indicates that the compatibility between the FA module and FLANs is higher.

5 Conclusions

In this paper, we propose a novel TL framework, FLANs, for AMC. Our method focuses on signals that have traversed different channel environments. By using the FA and FCA modules within FLANs, we manage to reduce the differences between signals from two distinct channels, thereby enhancing the performance of classifying 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 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. Qinyan MA, Duona ZHANG, and Yufeng WANG drafted the paper. Wenrui DING helped organize the paper. Qinyan MA and Duona ZHANG revised and finalized the paper.

Conflict of interest

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

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

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 Cogn 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., 2015. Domain-adversarial training of neural networks. https://arxiv.org/abs/1505.07818

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 Int 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 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

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 Cogn 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 Cogn 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 7th 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 XC, Lv Z, Gao X, et al., 2022. Pervasive wireless channel modeling theory and applications to 6G GBSMs for all frequency bands and all scenarios. *IEEE Trans Veh Technol*, 71(9):9159-9173. https://doi.org/10.1109/TVT.2022.3179695

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 35th 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. *Wirel 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 Inform*, 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 Expr*, 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 Neur Netw Learn Syst*, 32(4):1713-1722. https://doi.org/10.1109/TNNLS.2020.2988928