A Semantic-Consistent Few-Shot Modulation Recognition Framework for IoT Applications

Jie Su, Peng Sun, *Student Member, IEEE*, Yuting Jiang, Zhenyu Wen, *Senior Member, IEEE*, Fangda Guo, Yiming Wu, Zhen Hong, *Member, IEEE*, Haoran Duan, *Member, IEEE*, Yawen Huang, Rajiv Ranjan, and Yefeng Zheng, *Member, IEEE*

***Abstract*— The rapid growth of the Internet of Things (IoT) has led to the widespread adoption of the IoT networks in numerous digital applications. To counter physical threats in these systems, automatic modulation classification (AMC) has emerged as an effective approach for identifying the modulation format of signals in noisy environments. However, identifying those threats can be particularly challenging due to the scarcity of labeled data, which is a common issue in various IoT applications, such as anomaly detection for unmanned aerial vehicles (UAVs) and intrusion detection in the IoT networks. Few-shot learning (FSL) offers a promising solution by enabling models to grasp the concepts of new classes using only a limited number of labeled samples. However, prevalent FSL techniques are primarily tailored for tasks in the computer vision domain and are not suitable for the wireless signal domain. Instead of designing a new FSL model, this work suggests a novel approach that enhances wireless signals to be more efficiently processed by the existing state-of-the-art (SOTA) FSL models. We present the semantic-consistent signal pretransformation (ScSP), a parameterized transformation architecture that ensures signals with identical semantics exhibit similar representations. ScSP is designed to integrate seamlessly with various SOTA FSL models for signal modulation recognition and supports commonly used deep learning backbones. Our evaluation indicates that ScSP boosts the performance of numerous SOTA FSL models, while preserving flexibility.**

***Index Terms*— Deep learning, few-shot learning (FSL), Internet of Things (IoT), signal processing.**

Manuscript received 12 June 2023; revised 23 April 2024; accepted 5 August 2024. This work was supported in part by the National Nature Science Foundation of China under Grant 62302454, Grant 62072408, and Grant 62302485; in part by Zhejiang Provincial Science Fund for Distinguished Young Scholars under Grant LR24F020004; in part by China Postdoctoral Science Foundation under Grant 2023M743403 and Grant 2022M713206; in part by Zhejiang Provincial Natural Science Foundation of Major Program (Youth Original Project) under Grant LDQ24F020001; in part by the Key Research and Development Program of Zhejiang under Grant 2024C03288; and in part by the Chinese Academy of Sciences (CAS) Special Research Assistant Program. *(Corresponding authors: Zhenyu Wen; Fangda Guo.)*

Jie Su, Zhenyu Wen, and Yiming Wu are with the Institute of Cyberspace Security and the College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China, and also with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, Anhui 230026, China (e-mail: jiesu@zjut.edu.cn; wenluke427@gmail.com; wyiming@zjut.edu.cn).

Peng Sun and Yefeng Zheng are with the School of Engineering, Westlake University, Hangzhou 310030, China (e-mail: sp12138sp@gmail.com; yefengzheng@tencent.com).

Yuting Jiang and Zhen Hong are with the Institute of Cyberspace Security and the College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China (e-mail: 221123030164@zjut.edu.cn; zhong1983@zjut.edu.cn).

Fangda Guo is with the CAS Key Laboratory of AI Safety, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China (e-mail: guofangda@ict.ac.cn).

Haoran Duan is with the Department of Computer Science, Durham University, DH1 3LE Durham, U.K. (e-mail: haoran.duan@ieee.org).

Yawen Huang is with the Tencent Jarvis Lab, Shenzhen 518057, China (e-mail: yawenhuang@tencent.com).

Rajiv Ranjan is with the School of Computing Science, Newcastle University, NE1 7RU Newcastle upon Tyne, U.K. (e-mail: raj.ranjan@ncl.ac.uk).

Digital Object Identifier 10.1109/TNNLS.2024.3441597

I. Introduction

THE burgeoning field of the Internet of Things (IoT) has garnered immense interest due to its promise of extending Internet connectivity to everyday physical objects [1], [2], [3], [4], [5], [6]. This technology has led to the deployment of a plethora of interconnected devices, permeating both our personal lives and industrial processes. Due to the open nature inherent in the IoT devices, these interconnected devices frequently function in environments lacking trust, thereby leaving them vulnerable to numerous active, malicious attacks. Automatic modulation classification (AMC) [7] is the conventional approach to identifying physical-layer threats, such as anomalous unmanned aerial vehicle (UAV) jamming [8] and pilot jamming [9], by determining the modulation type and calculating the related modulation parameters for noise-affected signals within a complex radio environment [10].

Recently, deep learning-based (DL-based) AMC [11], [12], [13] approaches have achieved decent performance on various applications by learning representations from large-scale labeled datasets. Unfortunately, the situation of few samples or insufficient samples is common in many IoT applications, including but not limited to intrusion detection, anomaly detection, and fault diagnosis. For example, in the city surveillance applications, a model needs to a quick adaptation to detect new or unknown drone models in sensitive areas with limited data. However, the few-shot cases present unique challenge; as illustrated in Fig. 1, the anomaly detection model trained with insufficient signals may generate an inferior decision boundary (in red), overfitting to the existing observed abnormal samples. However, the distribution of real abnormal samples (in gray) is often larger than the observed abnormal samples, making the learned detection model less effective in detecting new abnormal signals from similar UAVs. Alternatively, developing a mechanism that allows the model to learn a superior decision boundary when insufficient samples are provided can significantly improve the robustness and generalization of the AMC system.

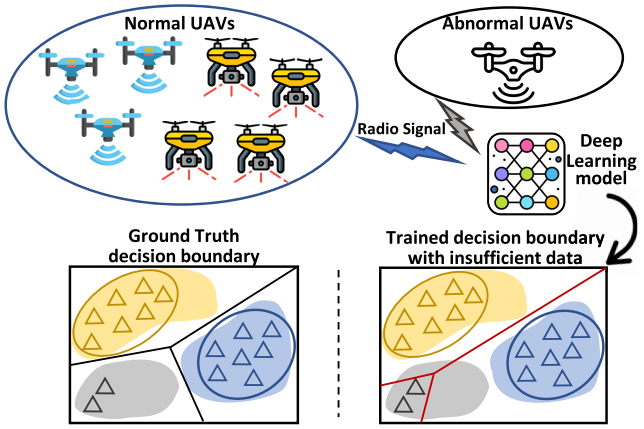


Fig. 1. Example of an inferior decision boundary in an abnormal UAV detection application when insufficient abnormal radio signals are provided. The inferior decision boundary (in red color) often leads to poor generalization ability for other abnormal UAVs that are presented in training samples.

In order to address the challenge of label scarcity by increasing the quantity and diversity of samples, various augmentation techniques and meta/metric-based approaches have been proposed, such as AFHN [14], MAML [15], and ProtoNet [16]. These meta/metric-based approaches typically aim to learn an efficient base model from a large dataset and adapt it to few-shot classes using distance measurement or gradient optimization. However, the availability of a substantial base class in the radio signal processing domain is often limited, and the direct application of these approaches may lead to the problem of "inadequate generalization." Specifically, the semantic information or signal patterns are not effectively learned by those methods with limited base samples, rendering the few-shot adaptation less generalizable. As depicted in Fig. 2, the conventional few-shot learning (FSL) framework would suffer from indistinguishable latent problems. The class distribution often exhibits considerable intraclass variation, and the phenomenon of interclass pattern overlapping obscures the decision boundary, thereby compromising the final classification performance [see Fig. 2(a), (c), and (e)].

Furthermore, the conventional augmentation-based approach primarily relies on spatial transformations or adversarial generation, ensuring semantic/pattern invariance visually. However, in the domain of radio signal processing, signal patterns are often invisible and noninterpretable. The direct adaptation of spatial augmentation approaches could destroy the inherent pattern, thus causing performance degradation. In addition, in addition to the previously mentioned challenges, inherent signal properties further complicate conventional FSL. These include the following.

1) Modulated signals often experience noise interference during transmission in open environments, complicating the representation learning process.
2) Radio signal data exhibits distinctive characteristics, such as periodicity and symmetry, which may pose challenges for deep learning models with limited sample sizes.

To address the aforementioned issues, we introduce a parameterized radio signal transformation framework, semantic-consistent signal pretransformation (ScSP). The primary concept behind ScSP involves the extraction of high-density constants, for instance, semantic information/signal pattern, while eliminating nonconstant elements, such as additional noises. This leads to improved intraclass concentration [19] of constants, enhancing the performance of downstream FSL methods. To achieve that, we design a framework to encourage the model to learn meaningful representations that capture invariant (i.e., semantic-consistent) features across augmented versions of the same instance while discriminating against representations from different instances. Specifically, we deeply investigate the signal pattern expression form (e.g., the constellation of the IQ signals) and propose the info-preserved augmentation module to generate diverse augmented signals without modifying its original semantic information (i.e., modulation type). Then, to minimize the noise interference, we present an adaptive noise filtering module that transforms the parameter of the conventional Gaussian noise filter into a learnable layer, enabling it to adaptively capture various patterns of noise. Furthermore, we present an amplitude–phase feature enhancement module to improve fine-grained/semantic feature extraction with amplitude–phase transformation. Finally, the InfoNCE [20] loss is applied to maximize the mutual information of the augmented versions of the same instance so to encourage the consistent pattern extraction. Based on our current understanding, ScSP stands as the pioneering framework that addresses few-shot automated modulation recognition by enhancing the pattern expressiveness of signals via expert knowledge-guided pretransformation. Comprehensive experiments and analyses have been conducted, resulting in the following primary insights.

1) The current SOTA FSL approaches achieve noteworthy performance improvements on the task of radio signal modulation FSL, by incorporating the suggested ScSP framework.
2) We show that the combination of the info-preserved augmentation and the adaptive noise filtering is more suitable under high-noise conditions, while the entire ScSP framework works better under low-noise conditions.
3) We propose a measurement method to comprehensively study the impact of the noise on radio signal representation after amplitude–phase transformation. The semantic information represented by the transformed signals positively correlates with the signal noise level.

II. Related Work

A. DL-Based Automatic Modulation Recognition

Automatic modulation recognition refers to the modulation category identification of the received radio signals. This technology is widely used in spectrum management [21], [22], interference identification [12], and electronic reconnaissance systems [11]. There have been many attempts to perform modulation recognition with DL-based methods. Hong et al. [23] proposed a two-layer GRU to capture the information from context level (i.e., time dimension) and feature level. Later, West and O'Shea [24] proposed CLDNN, which combines

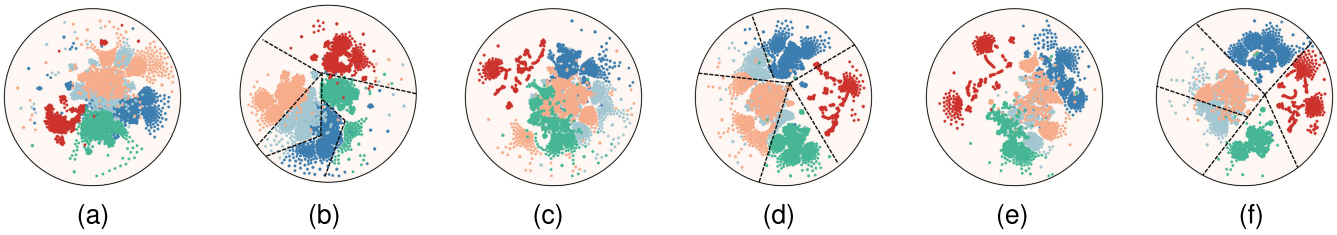


Fig. 2. Latent space visualization of the radio signal feature embedding space for test samples in the Signal-128 dataset, as reported [16], [17], [18]. The "/ScSP" represents the combination with the proposed ScSP framework. (a) RelatNet. (b) RelatNet/ScSP. (c) ProtoNet. (d) ProtoNet/ScSP. (e) MatchNet. (f) MatchNet/ScSP.

the convolutional neural network (CNN) and long short-term memory (LSTM) to enhance the feature-level information extraction. To further improve the feature extraction ability, Zhang et al. [10] utilized a deep residual network model to conduct the classification task, while the training process is time-consuming. Nonetheless, neither of these DL-based methods considers signal preprocessing (e.g., noise filtering) according to its particular properties. Previous work [25] presents that special properties, such as noise, often lead to a significant performance drop on DL-based methods. This issue becomes more critical when sufficient signal samples are unavailable (i.e., few-shot situations). Moreover, the DL-based frameworks with insufficient samples often suffer from the overfitting problem, which leads to poor generalization ability and unsatisfied results.

B. Few-Shot Learning

Typically, traditional FSL techniques are categorized into either inductive or transductive inference, depending on the inference setting. Inductive inference approaches classify individual unlabeled samples, whereas transductive inference ones classify multiple query samples simultaneously. The inductive inference strategies can be further subdivided into the following.

1) *Metric-Oriented Methods (e.g., Matching Network [18], Prototypical Networks [16], and ReNet [26]):* These techniques aspire to establish a series of projection functions (embedding functions) and metrics that quantify the similarity among samples.
2) *Meta-Oriented Methods (e.g., MAML [15] and ProtoMAML [27]):* These techniques leverage a model-agnostic meta-learner to develop an efficient base model across multiple training tasks, which can be adapted to a new task with a limited number of training samples via a few gradient steps, resulting in a model with decent generalization ability.
3) *Augmentation-Oriented Methods [14], [28]:* These methods aim at creating diverse sample generation strategies for unfamiliar classes to foster representation learning. Recently, transductive inference methods (TIMs [29] and LaplacianShot [30]) have emerged as an appealing approach to tackling few-shot tasks, which have better performance than inductive inference.

Although the abovementioned state-of-the-art (SOTA) FSL methods have achieved decent results on various vision-based tasks [31], [32], adapting these methods to process radio signal data remains challenges. A few works [10], [33] have recently been proposed to perform few-shot recognition on modulated signals. However, those methods mainly focus on extracting fine-grained information from signals with specific network structures. Zhou et al. [34] proposed a GAN-based signal sample generation method to solve the first challenge. However, this generation method only maintains the integrity and consistency of the generated signals, that is, to generate integral signals with a similar pattern. When sufficient samples are unavailable (i.e., low quantity), the generated samples tend to be identical while losing the diversity. In this article, we propose an efficient and flexible signal data transformation framework that allows the SOTA FSL algorithms can be easily applied to solve the radio signal modulation recognition FSL problem.

III. Problem Formulation

A. Background Knowledge

In wireless communication systems, modulation aims to add information to a set of signals by varying one or more properties of periodic electromagnetic waves (carriers), which can be transmitted [35]. A transmitted time modulation signal $r(t)$ can be illustrated as follows:

$$r(t) = \mathcal{S}(t) * h(t) \exp\left[j2\pi \Delta f t + \psi_0\right] + \text{noise}(t) \quad (1)$$

where $*$ represents the convolution operation, $\mathcal{S}(t)$ denotes the modulated signal, $h(t)$ represents the impulse response of the wireless channel, Δf indicates the carrier frequency offset, ψ_0 signifies the initial phase, and noise(t) refers to the environmental noise.

To facilitate signal information extraction and signal recovery, in-phase signals and the quadrature-phase signal are used to jointly characterize the relevant modulation information, i.e., I–Q data [36]. So, we define the received discrete complex signal as $x_{IQ} = \{x_I, x_Q\}$, which is sampled from $r(t)$

$$\begin{aligned} \left\{x_I, x_Q\right\} &= \text{sample}\left\{r_I(t), r_Q(t)\right\} \\ \left\{r_I(t), r_Q(t)\right\} &= \{\text{Re}\{r(t)\}, \text{Im}\{r(t)\}\} \end{aligned} \quad (2)$$

where x_I denotes the in-phase signal, x_Q represents the quadrature-phase signal, Re is the real part, and Im indicates the imaginary part. For the AMC task, the modulated signal segment x (i.e., sampled from x_{IQ}) and its corresponding label y are used for the training procedure (feature extraction).

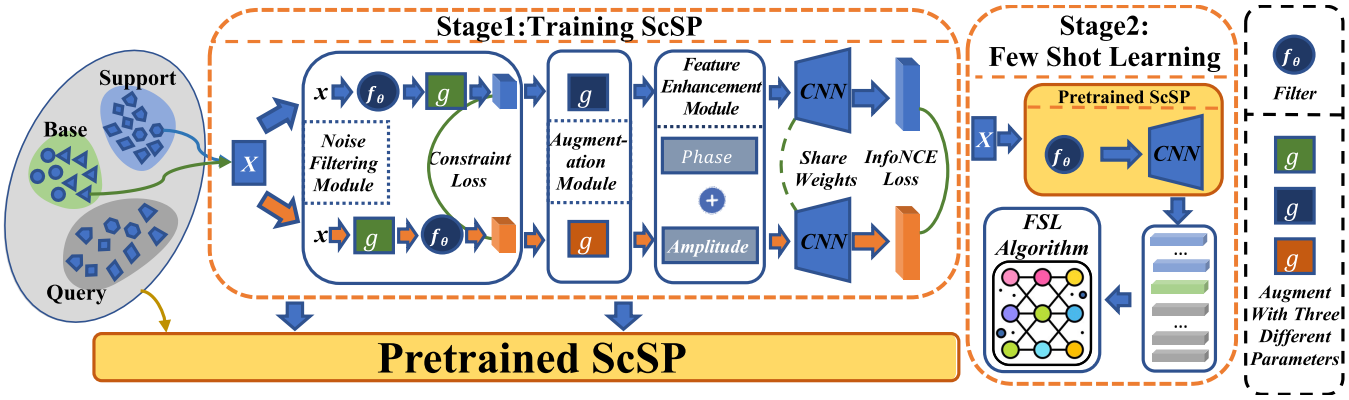


Fig. 3. Structure of the proposed ScSP framework. The ScSP framework consists of two stages of the training paradigm. *Stage 1:* The base and support set samples are fed into the noise filtering module for noise removal. Then, the augmentation module receives the noise-filtered samples for augmentation (i.e., creating two variations of one sample). Finally, the shared weights encoder is optimized by InfoNCE loss to extract pattern-consistent representations from the augmented samples. *Stage 2:* The pretrained ScSP framework takes the input samples and outputs the domain-specific signal representation for downstream FSL algorithm training.

B. Automatic Modulation FSL

The automatic modulation FSL (AMFSL) operates in two phases: 1) on the training stage, a model is trained on a set of *base* classes, and a new set of novel classes is defined as *support* set for novel classes learning and 2) then, a *query* set with identical classes as the support set is presented for novel classes prediction on the evaluation stage. We define the base training set as $D_b = \{x^b, y^b\}$, where x^b is the quadrature modulated signal segment and the corresponding label $y^b \subset \mathbb{R}^{C_B}$ belongs to a total of C_B base classes. The support set is denoted as $D_s = \{x^s, y^s\}$, $y^s \subset \mathbb{R}^{C_K}$, where C_K is the novel classes, and each class includes N samples. The remaining Q samples in C_K novel classes form the query set $D_q = \{x^q, y^q\}$, $y^q \subset \mathbb{R}^{C_K}$ [18]. It should be noticed that the novel classes on support set and query set are disjoint with base training set classes (i.e., $C_K \cap C_B = \emptyset$). As a result, the objective of the FSL settings could be formulated as follows:

$$\min(\epsilon_{\text{error}}) = \mathbb{E}_{(x^q, y^q)\sim D_q}\left[f'\left(x^q\right) \neq y^q\right] \tag{3}$$

where ϵ_{error} denotes the target error on *query* set and the f' represents the trained FSL using *base* and *support* set.

IV. Modulated Signal Pretransformation

To extract constant semantic content from the signals, our objective is to optimize the mutual information between two signal segments, x_1 and x_2, sharing similar semantics. This can be expressed as follows:

$$I(x_1, x_2) = \sum_{x_1, x_2} p(x_1, x_2) \log\left(\frac{p(x_1, x_2)}{p(x_1) \cdot p(x_2)}\right). \tag{4}$$

In order to achieve this, the ScSP framework is designed to minimize the InfoNCE loss [refer to (18)]. Fig. 3 illustrates the primary elements of the ScSP framework as follows: 1) the *info-preserved augmentation* module augments the input signals, maintaining the modulation type constant, to assist in the minimization of the InfoNCE loss; 2) the *adaptive noise filtering* module diminishes the impact of nonsemantic information (i.e., inherent Gaussian noise); and 3) an *amplitude–phase feature enhancement* module transforms the signal into amplitude and phase for semantic information enhancement. The last two components are served as the enhancing module for the InfoNCE loss, which aims to maximize the mutual information from two augmented signals to get constant semantic information.

Chen et al. [37] articulate that the effectiveness of InfoNCE in the extraction of semantic information can be significantly enhanced through data augmentation. The enhancements are realized via two primary avenues: 1) data augmentation serves to expand the volume of training data and 2) it also elevates the count of semantically similar data entries, thereby optimizing the mutual information. However, compared with image-like augmentations, signals' intrinsic properties (e.g., modulation type) are often invisible, making designing proper signal augmentation methods challenging. Inspired by the constellation diagram [38] in the signal processing community, where the representation of the digital modulation scheme is obtained, we draw the axial projection (i.e., similar to constellation diagram) of different signals to investigate the special properties. Fig. 4 illustrates the axial projection of signals in three different modulation types. We find it shares similar properties to the constellation diagram where the semantic information of signals is obtained. Therefore, we designed four types of axial projection-invariant augmentation to generate information preserved data samples.

A. Info-Preserved Augmentation

1) Flipping: Our left–right flipping is an asymmetrical adjustment of the signal timing relationship, while the frequency (i.e., modulation information) remains unchanged. Moreover, the left–right flipping will not affect the axial projection to preserve semantic information. For a given quadrature modulated signal x in original length L with t-index time step value $x(t)$ and a flipping augmentation operation $g_{\text{flip}}(x)$, the flipping operation can be formulated as follows:

$$g_{\text{flip}}(x) : x(t) \to x(L - t) \quad \forall t \in [0, L] \tag{5}$$

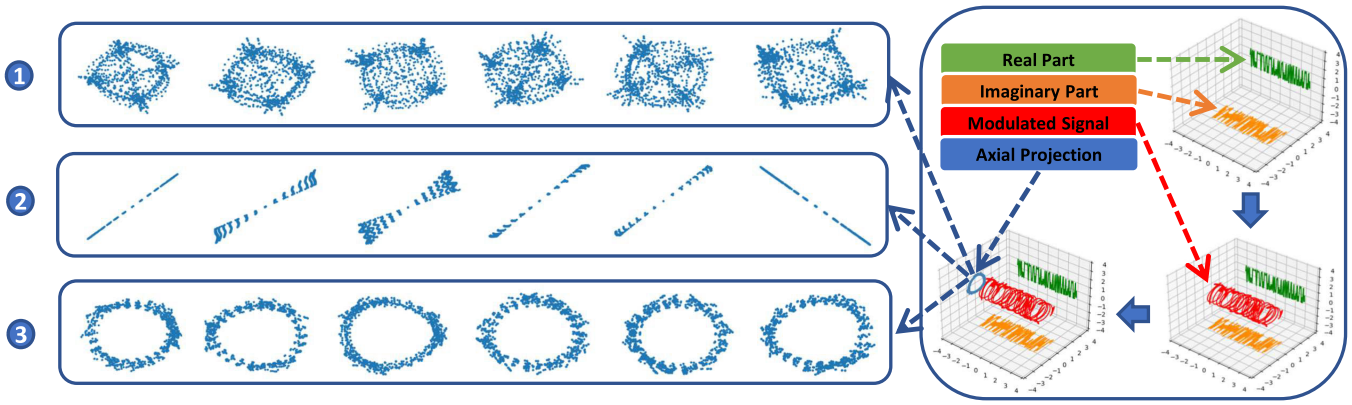


Fig. 4. In the right part of the figure, we show the process of obtaining a 2-D modulated signal from the 1-D real and imaginary parts of the signal, which leads to the axial projection. In the left part of the figure, we show three types of modulation, each with six axial projections.

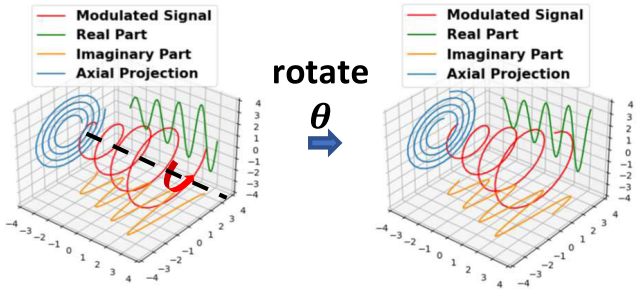


Fig. 5. Visualization of signal rotation operation.

2) Interception: The modulation is applied to the whole signal segment, so that the intercepted signals still share the same semantic information as the original signal. Therefore, the intercepted signal can be treated as a weaker representation of the original signal, since the axial projection of intercepted signal is nearly identical (sparse) to the original signals. For a segment of a quadrature modulated signal, denoted as x, with an initial length L, the process of interception can be represented as follows:

$$g_{\text{inter}}\left(x, a, L'\right) : x \rightarrow x_{a,L'} = \left[x(a), x(a+1), \ldots, x\left(a+L'\right)\right] \tag{6}$$

where $x_{a,L'}$ denotes the intercepted signal and L' signifies the set interception length. The beginning point of interception, a, should be chosen within the range $[0, L - L']$.

3) Rotation: To facilitate the illustration of rotation, a quadrature modulated signal/complex signal [39] segment x can be represented as the in-phase and quadrature-phase parts of signals, respectively. The relationship between these signals is given by

$$x = x_I + j \cdot x_Q \tag{7}$$

where j is a imaginary number, that is, $j = (-1)^{1/2}$.

Fig. 4 presents that the semantic information of the modulated signals is rotation-invariant, such that character is used in the signal generation. Hence, we utilize the Euler formula [40] for rotating the initial signals (refer to Fig. 5). For a signal modulated in quadrature, we can represent the rotation operation as follows:

$$\begin{aligned} g_{\text{rotate}}(x, \theta) &= x \cdot e^{\theta \pi j} \\ &= (x_R + j \cdot x_I)(\cos(\theta\pi) + j \cdot \sin(\theta\pi)) \end{aligned} \tag{8}$$

where $g_{\text{rotate}}(x, \theta)$ signifies the rotation augmentation operation and $\theta \sim U(0, 2\pi)$ illustrates the rotation angle originating from the uniform distribution.

4) Conjugate Transformation: Inspired by the generation of constellation diagram, we develop conjugate transformation as follows:

$$g_{\text{conjugate}}(x) = x_I - j \cdot x_Q \quad \text{where } j = \sqrt{-1}. \tag{9}$$

We only change the in-phase part x_I of signal, and its semantic information is preserved.

B. Adaptive Noise Filtering

This section introduces a parameterized Gaussian-based noise filter that can filter noise for various radio signals. Previous study [13] presents that adding a Gaussian noise filtering module will vastly reduce the negative effect of noise and improve the final recognition result. The Gaussian filter [41] can suppress high-frequency information to a certain degree, thereby generating smoother signals with less high-frequency noise. However, the conventional Gaussian noise filtering approach is ineffective in addressing the diverse changes in noise intensity and signal types when using fixed parameters. Hence, we transform the parameter of the conventional Gaussian noise filter into a learnable layer, enabling it to adaptively capture various patterns of noise. The Gaussian filtering operation $F(x)$ can be formulated as follows:

$$f(x) = x * G \quad \text{s.t. } G(n) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{n^2}{2\sigma^2}} \tag{10}$$

and $G(n)$ represents the n-indexed variable of the Gaussian filter kernel. The parameter σ can be fine-tuned through learning to accommodate varying signal types. Furthermore, the ideal filter $f_\theta^*(x)$ should display equivariance [37] to the augmentations $g(x)$, which implies that the order of augmentation and filtering can be interchanged without altering the outcome

$$g\left(f_\theta^*(x)\right) = f_\theta^*(g(x)). \tag{11}$$

Consequently, adopting this presumption, we employ the parameterized filter f_θ to mimic the optimal filter by reducing the constraint loss

$$\mathcal{L}_G = \|g(f_\theta(x)) - f_\theta(g(x))\|_2^2 \tag{12}$$

C. Amplitude–Phase Feature Enhancement

Rajendran et al. [42] observed that converting modulated signals to amplitude–phase format would benefit the representation learning of DL models. Furthermore, for modulation types, such as amplitude-shift keying (ASK) and phase-shift keying (PSK), the modulated signal $\mathcal{S}(t)$ uses the changes in amplitude and phase to carry the information of the binary bit stream $s_b(t)$, that is,

$$\begin{aligned} &\text{ASK}: \mathcal{S}(t) = A_c s_b(t) \exp\left[j2\pi f_c t\right] \\ &\text{PSK}: \mathcal{S}(t) = A_c \exp\left[j2\pi f_c t + j2\pi\left(\Delta_p s_b(t) + \psi_0\right)\right] \end{aligned} \tag{13}$$

where A_c denotes the carriers amplitude, f_c represents the carrier wave, and Δ_p is the phase modulation factor. From (13), we can observe that the instantaneous amplitude $A(t)$ and instantaneous phase $\psi(t)$ of the carriers are linearly dependent with $s_b(t)$, which can be represented by

$$\begin{aligned} A(t) &= A_c s_b(t) \propto s_b(t) \\ \psi(t) &= \Delta_p s_b(t) + \psi_0 \propto \Delta_p s_b(t) \end{aligned} \tag{14}$$

Equations (13) and (14) demonstrate that the variation of $A(t)$ and $\psi(t)$ carries the main information of the modulated signal. In this case, we present the amplitude–phase feature enhancement module to transform signals into amplitude and phase, allowing the downstream deep learning model to learn their features better. The discrete amplitude A_r and discrete phase ψ_r modal of the signal at the receiver side can be extracted by the following operations:

$$\begin{aligned} A_r &= \sqrt{x_I^2 + x_Q^2} \\ \psi_r &= \arctan\left(x_Q/(x_I + \epsilon)\right) \end{aligned} \tag{15}$$

where ϵ denotes the microconstants. More specifically, Fig. 6 illustrates the procedure of the transformation on the axial projection of signals. The point on the axial projection is represented by abscissa, and the ordinate is converted to radial length and angle.

Finally, we leverage the concatenation operation to extract information from discrete amplitude and discrete phase simultaneously, which can be represented as follows:

$$E(x) = \text{concat}\{A_r, \psi_r\} \tag{16}$$

where $E(x)$ represents the enhancement operation. concat denotes the concatenation operation.

The transformed signal has a simpler representation than it is presented as abscissa and ordinate. With the simpler representation, the information of the axial projection is easier captured by the feature extractor.

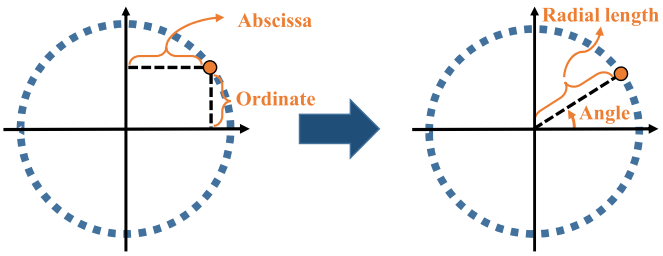


Fig. 6. Transformation of signals into amplitude and phase (right).

D. InfoNCE Loss-Based ScSP Pretraining

The pretraining of the ScSP is formulated as a unsupervised learning procedure. Mini-batch signal segments are first sampled from the base and support set to perform info-preserved augmentation. Then, the augmented signal segments will be processed by adaptive noise filtering and amplitude–phase feature enhancement modules to remove the carried noise and enhance the signal representation. Finally, a shared weight parametric model (e.g., CNN) is employed to transform the signals into advanced representations for training. It is noteworthy that all components in the ScSP framework undergo concurrent training, optimized by the constraint loss $\mathcal{L}_G$ and InfoNCE loss $\mathcal{L}_C$. This can be formulated as follows:

$$\mathcal{L} = \mathcal{L}_C + \mathcal{L}_G. \tag{17}$$

Specifically, the InfoNCE loss could be formulated as follows:

$$\begin{aligned} L_C &= -E_{x_1^i, x_2^i \sim p(x_1, x_2)}\left\{\log \frac{\text{h}\left(x_1^i, x_2^i\right)}{\sum_{j \neq i}^{N} h\left(x_1^i, x_2^j\right)}\right\} \\ h(x_1, x_2) &= \exp\left\{\text{sim}\left[g_\theta(x_1), g_\theta(x_2)\right]/\alpha\right\} \end{aligned} \tag{18}$$

We define the ScSP framework as g_θ, and the temperature parameter is represented by α. We measure the similarity between u and v using the cosine similarity, which is defined as $\text{sim}(u, v) = u^{\text{T}}v/\|u\|\|v\|$, serving as the dot product of l_2 normalized u and v. Augmented samples drawn from an identical signal, x_1^i and x_2^i, are deemed as positives, while augmented samples from different signals, denoted as x_1^i and x_2^j, are considered negatives. After the pretraining, the trained ScSP framework will extract the representation of all signal segments from the base, support, and query set for further few-shot learning usage.

V. Experiment

A. Datasets

Our proposed ScSP framework is assessed across three benchmark datasets: Signal-128 [7], Signal-512, and Signal-1024 [46]. The numbers in the dataset names correspond to the lengths of the respective signals (e.g., Signal-128 represents the signals with 128 lengths).

Signal-128 represents a comprehensive public radio dataset, comprising eight digital variations (BPSK, QPSK, 8PSK, 16QAM, 64QAM, BFSK, CPFSK, and PAM4) and three analog ones (WB-FM, AM-SSB, and AM-DSB). Each modulation type is associated with 20 distinct signal-to-noise

TABLE I

FIVE-WAY FEW-SHOT CLASSIFICATION ACCURACY ON THREE BENCHMARK DATASETS. "ORIGIN" SIGNIFIES THE RESULT OF FSL WITHOUT APPLYING THE SCSP FRAMEWORK, WHILE "+SCSP" IMPLIES THE OUTCOME OF FSL INCORPORATING THE SCSP FRAMEWORK. "+" OR "−" SYMBOL INDICATES A RESPECTIVE INCREASE OR DECREASE IN PERFORMANCE. PEAK PERFORMANCE IS HIGHLIGHTED USING AN UNDERLINE ("_")

Method	*Signal-128*				*Signal-512*				*Signal-1024*			
	1-shot		5-shot		1-shot		5-shot		1-shot		5-shot	
	Origin	+ScSP	Origin	+ScSP	Origin	+ScSP	Origin	+ScSP	Origin	+ScSP	Origin	+ScSP
CNN Backbone												
MatchNet(NeurIPS,2016) [18]	59.99%	+13.14%	72.77%	+9.33%	40.30%	+0.24%	41.11%	+3.29%	80.57%	+7.87%	82.15%	+8.81%
MAML(ICML,2017) [15]	30.39%	+27.56%	58.37%	+5.55%	23.35%	+0.15%	24.73%	+12.19%	58.37%	+15.97%	78.14%	+1.63%
ProtoNet(NeurIPS,2017) [16]	59.55%	+12.31%	68.83%	+13.61%	40.74%	-0.13%	40.98%	+2.97%	80.76%	+4.58%	81.65%	+9.03%
RelatNet(CVPR,2018) [17]	48.23%	+20.81%	67.05%	+14.18%	36.25%	+1.61%	39.78%	+2.31%	61.91%	+18.65%	80.07%	+8.12%
Tim-GD(NeurIPS,2020) [29]	62.33%	+13.18%	74.14%	+8.11%	40.29%	+0.06%	42.72%	+3.42%	80.94%	+5.24%	81.27%	+10.71%
Tim-ADM(NeurIPS,2020) [29]	67.46%	+10.21%	75.14%	+7.63%	40.02%	+0.10%	42.73%	+3.66%	80.91%	+0.96%	81.06%	+11.39%
ReNet(ICCV,2021) [26]	52.88%	+2.84%	70.04%	+7.77%	35.93%	+0.28%	40.54%	+0.78%	75.46%	+4.44%	82.00%	+10.59%
CM(TNNLS,2022) [43]	41.58%	+22.01%	47.29%	+7.43%	23.93%	+1.67%	38.63%	+8.84%	53.28%	+26.16%	34.68%	+19.19%
Meta-Proto(TNNLS,2022) [44]	34.27%	+7.81%	36.06%	+8.5%	41.78%	+14.32%	42.32%	+14.29%	71.79%	+6.09%	77.11%	+0.97%
DFR(TNNLS,2024) [45]	64.41%	+6.72%	68.27%	+7.65%	73.55%	+1.35%	75.71%	+6.90%	63.53%	+4.70%	70.62%	+1.45%
ResNet Backbone												
MatchNet(NeurIPS,2016) [18]	61.23%	+9.46%	74.94%	+8.71%	40.49%	+0.30%	40.91%	+2.46%	79.94%	+2.35%	82.76%	+7.54%
MAML(ICML,2017) [15]	35.77%	+20.88%	59.27%	+2.46%	23.01%	+3.43%	28.80%	+11.99%	59.43%	+15.43%	72.64%	+3.75%
ProtoNet(NeurIPS,2017) [16]	57.01%	+15.17%	70.65%	+12.34%	39.98%	+0.12%	41.16%	+3.86%	80.29%	+0.25%	81.40%	+8.33%
RelatNet(CVPR,2018) [17]	45.87%	+21.52%	65.36%	+15.75%	36.03%	+0.44%	36.42%	+4.32%	49.21%	+31.65%	79.98%	+7.67%
Tim-GD(NeurIPS,2020) [29]	60.15%	+13.54%	73.95%	+9.70%	41.38%	+0.60%	42.22%	+3.57%	80.47%	+0.45%	81.60%	+9.55%
Tim-ADM(NeurIPS,2020) [29]	62.17%	+14.08%	75.97%	+7.82%	41.49%	+0.40%	41.89%	+4.56%	80.28%	+0.54%	81.57%	+9.94%
ReNet(ICCV,2021) [26]	66.51%	+7.72%	72.74%	+3.72%	38.22%	+0.30%	40.80%	+1.16%	78.40%	+3.07%	82.09%	+8.12%
CM(TNNLS,2022) [43]	32.07%	+29.27%	58.96%	+12.33%	24.29%	+0.65%	38.11%	+8.14%	58.22%	+21.22%	74.63%	+8.52%
Meta-Proto(TNNLS,2022) [44]	33.79%	+7.81%	38.02%	+4.02%	45.27%	+7.65%	48.49%	+9.49%	70.21%	+8.19%	75.10%	+3.64%
DFR(TNNLS,2024) [45]	72.29%	+11.31%	76.01%	+12.50%	67.69%	+2.91%	73.63%	+5.76%	63.05%	+2.66%	64.38%	+5.02%

ratios (SNRs), each accompanied by 1000 samples. SNR, expressed as signal/noise, is a key measure of signal quality, where a high SNR value signifies less noise interference in the signal.

Signal-512 is a private dataset that factors in several complex aspects of communication systems, such as carrier phase, pulse shaping, frequency offsets, and noise. The dataset includes 12 different modulation types, namely, BPSK, QPSK, 8PSK, OQPSK, 2FSK, 4FSK, 8FSK, 16QKAM, 32QAM, 64QAM, 4PAM, and 8PAM. The SNR of each modulation type is uniformly distributed within a range from −20 to 30 dB. Each data sample comprises 64 symbols, oversampled at a rate of 8, leading to a total of 512 sampling points for each sample.

Signal-1024 is a publicly available radio signal dataset comprising 24 distinct types of both digital and analog modulations. These modulation methods include OOK, 4ASK, 8ASK, BPSK, QPSK, 8PSK, 16PSK, 32PSK, 16APSK, 32APSK, 64APSK, 128APSK, 16QAM, 32QAM, 64QAM, 128QAM, 256QAM, AM-SSB-WC, AM-SSB-SC, AM-DSB-WC, AM-DSB-SC, FM, GMSK, and OQPSK. The data in Signal-1024 was obtained from environments characterized by high SNR and low fading, which pose complex signal classification challenges. The dataset is structured, such that each modulation method is represented across 26 unique SNRs, with 4096 samples provided for each SNR.

B. Experimental Setup

1) Evaluation Models: We carried out the implementation of the proposed ScSP framework using PyTorch [47] and performed the training process on a Tesla V100. To evaluate the performance of our ScSP framework, we employed six existing FSL models for comparison. These included MAML, MatchNet, ProtoNet, RelatNet, TIM-GD, TIM-ADM, and ReNet. Previous studies [48], [49] present that decent performance of radio signal recognition can be achieved without a complex network structure. Moreover, the complex network structure (e.g., ResNet-18) with ample parameters may suffer from the overfitting problem when samples are insufficient, which affects the recognition performance. Therefore, we followed these studies and replaced the backbone of the existing FSL models with the modified CNN and ResNet.

2) Experimental Setting: Most radio signal datasets have very limited classes; for example, Signal-128 only has eight digital modulation classes and three analog modulations. The later may be impossible to conduct FSL task, due to the lack of base classes. Therefore, for all following experiments, we select two classes as the base training set, and the remaining classes form the support and query set. This setting is more challenge than that with sufficient base classes. Following previous studies [14], our experiments are carried out in two different setups: 5-way-5-shot and 5-way-1-shot. In these settings, we deal with five new classes, where each has only 5 and 1 instances, correspondingly. Each dataset is explored under four distinct SNRs: −4, 0, 10, and 18 dB.

3) Training Setting: We employ the Adam optimizer [50] with a learning rate set at 0.001 to train our framework. We limit the maximum training epoch to 50. Depending on the setup, the batch size for input varies—we use 70 for the 5-way-5-shot configuration and 42 for the 5-way-1-shot arrangement.

C. Experimental Results

Table I presents the efficacy of our ScSP framework in enhancing the performance of the existing SOTA FSL methods in both 5-way-1-shot and 5-way-5-shot modulation recognition

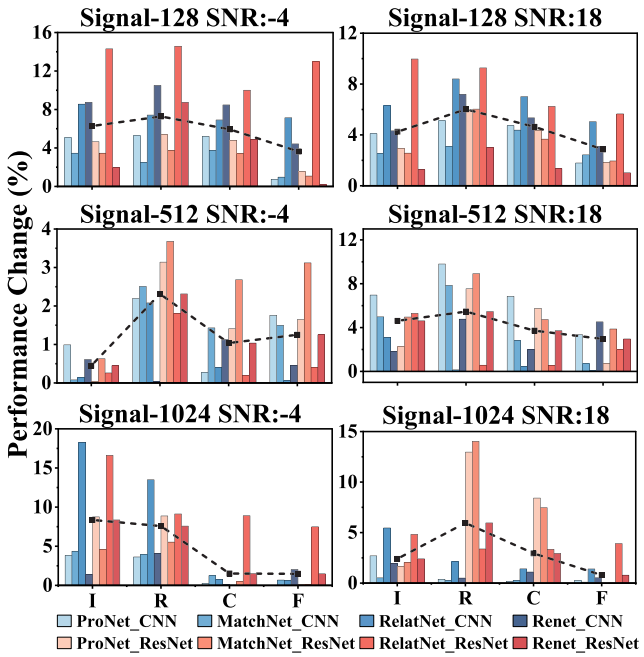


Fig. 7. Ablation study on single info-preserved augmentation operation under 5-way-5-shot setting. The dotted line indicates the average performance change, and "I," "R," "C," and "F" represents the "interception," "rotation," "conjugation," and "flipping" operations, respectively.

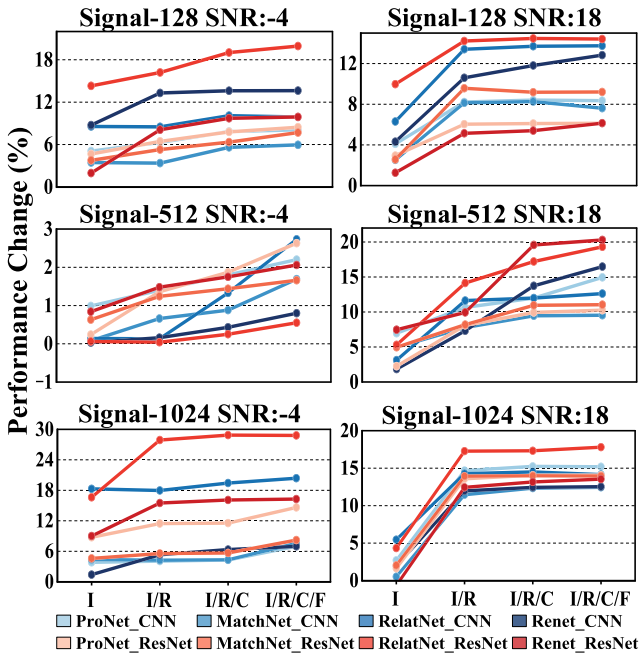


Fig. 8. Ablation study on stacking single info-preserved augmentation operation under 5-way-5-shot setting. The experiments are conducted on three datasets with −4- and 18-dB SNRs condition. The"/" indicates the combination (e.g., I/R means the combination of "interception" and "rotation").

tasks with varying backbone structures. Notably, our MSP framework, in combination with RelatNet, yields the highest performance improvement in six tasks, averaging a growth of 18.33%.

In addition, we note a marginal performance enhancement in the 5-way-1-shot task utilizing the Signal-512 dataset. A slight performance decline of approximately 0.13% is observed for ProtoNet with a CNN backbone. A potential explanation for this phenomenon could be the significant intraclass gaps within the Signal-512 dataset, making it challenging to learn the statistical properties of different categories from a solitary sample. This hypothesis is further substantiated, as the performance gain on the Signal-512 dataset improves when the number of samples is increased (i.e., 5-way-5-shot).

D. Ablation Study on Augmentation Operation

In this section, we carefully study the impact of each operation in augmentation for downstream FSL models. Our evaluation is performed on the aforementioned three datasets with −4- and 18-dB SNR, respectively.

1) Metrics: The *performance change* is presented as a percentage that is calculated by the FSL+ScSP recognition accuracy minus FSL recognition accuracy.

Fig. 7 shows each operation of ScSP can improve the performance gain for any downstream FSL models. In addition, we observed that rotation operation outperforms other operations, which brings the top improvement in five out of six groups. One possible assumption is that the semantic information of most signal modulation categories is rotation-invariant (i.e., mentioned in Section IV-A). Thus, the rotation operation can well preserve signals' semantic information

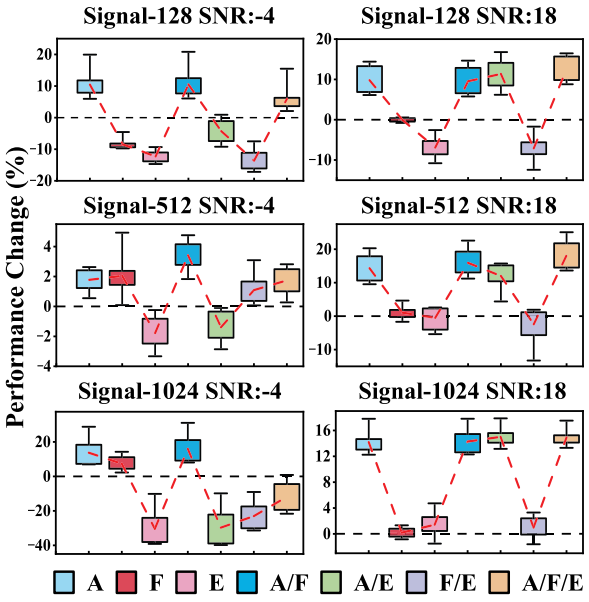


Fig. 9. Ablation study on 5-way-5-shot task for three datasets. "A," "F," and "E" represents the augmentation, noise filtering, and feature enhancement module, respectively. The dotted line indicates the average performance change. The "/" indicates the combination (e.g., A/F means the combination of augmentation and noise filtering).

while improving its diversity by randomly selecting the rotation angles. As a result, this operation outperforms other augmentation methods.

Furthermore, we observe that the RelatNet obtains more significant performance gains than other methods. This may be caused by the sensitivity of the complex network to the diversity of training samples. Also, the sample spaces

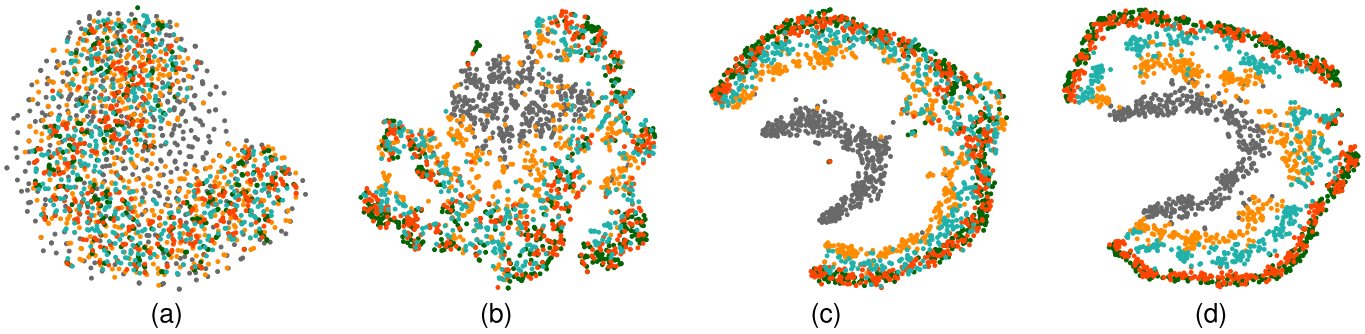


Fig. 10. Latent space visualization of the query set classes based on the feature extractor trained on base set classes. (a) Raw signal representations. (b) Pretrained ResNet (supervised) representation. (c) Contrastive pretrained (unsupervised) representation. (d) Pretrained ScSP (unsupervised) representation.

generated by our proposed augmentation operations do not overlap. Therefore, we believe that the different augmentation operations can be stacked to generate more samples. To verify the effectiveness of augmentation operation stacking in improving model performance, we test the performance gain by stacking augmentation operation sequentially.

From Fig. 8, we can observe that the performance gain of the different models is increased with the stacking of augmentation operations, which presents the importance of sample diversity to representation learning. Moreover, we can observe that the performance gain in -4 dB is relatively lower than that in 18 dB on the Signal-512 dataset. We conjecture that the noise factor in the Signal-512 dataset affects augmentation operations' abilities.

2) Takeaway: The aforementioned results present that both single and stacking augmentation operations can boost the performance of few-shot recognition under different noise conditions. Furthermore, the stacking of four augmentation operations can generate more diverse samples.

E. Ablation Study on ScSP Framework

In order to evaluate the individual and combined efficacy of components within the ScSP framework, we conduct experiments that isolate each component and its combinations. Fig. 9 shows that the following hold.

1) *Single Component Often (But Not Always) Contributes Positively to the Task:* For example, under the high SNR, where the input signals contain less noise, the improvement of the single noise filtering module is marginal. Furthermore, the feature enhancement might cause negative effects, especially under a low SNR (i.e., input signals contain more noise). One possible conjecture is that the FSL is overfitting to the enhanced base set when the sample is insufficient.
2) *Combination of "A, F" and "A, F, E" Works Better Under High-Noise and Low-Noise Conditions, Respectively:* Specifically, the enhancement module contributes positively to the FSL task when signals contain less noise, while it brings negative impacts in high-noise situations. Our conjecture is that the noise might affect the carried information in signals (e.g., in amplitude or phase), which makes the feature enchantment module amplify the negative effects of the noise. We will describe this effect in the subsequent SNR investigation section.

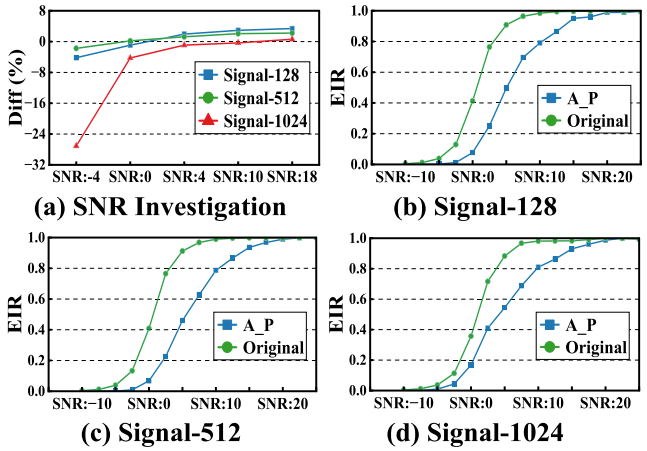


Fig. 11. SNR investigation experiments. (a) SNR investigation. (b)–(d) Effective information measurement on three datasets. "Diff" indicates the difference between the accuracy of (A/F/E) and (A/F) combination, while the "EIR" denotes the EIR. The "A_P" represents the amplitude–phase.

F. SNR Investigation

Following the conjecture from the previous section, that is, the feature enchantment module amplifies the negative effects of the noise, we conduct experiments for the combination of (A/F/E) and (A/E) under four different SNR scenarios (i.e., -4, 0, 10, 18 dB). From Fig. 11(a), we observed that the accuracy difference between A/F/E and A/E is gradually decreased, accompanied by the decrease of noises. This phenomenon presents that when the signal contains less noise, the feature enhancement module starts to bring a positive effect. Since the feature enhancement module provides a transformation to convert signals to a better representation processed by deep learning models, we assume that the semantic information represented by the transformed signals has a positive correlation with SNR.

Therefore, we define the effective information rate (EIR) for a noisy signal x as the normalized mutual information [51] between a noisy signal x and a corresponding pure signal $\hat{x}$ to measure the correlation between them (i.e., $\text{NMI} = I(x;\hat{x})/\max\{H(x), H(\hat{x})\}$), where I denotes the mutual information and H represents the information entropy.

Next, we use CLUB [52] to estimate I and H, and the measurement results are shown in Fig. 11(b)–(d).

Fig. 11(b)–(d) shows that EIR has a positive correlation with SNR. The transformed signals (i.e., amplitude–phase) are affected more by the noise. For example, Fig. 11(b) illustrates that, under SNR:5 condition, the EIRs are 0.93 and 0.52, corresponding to original and amplitude–phase. The large difference presents that the noise affects the carrying information in amplitude–phase, making representation learning more challenging. We believe this is the reason why the feature enhancement module does not work effectively when the SNR becomes low.

G. Latent Visualization

To further verify the effectiveness of our ScSP framework—(whether a better intraclass concentration is provided), we applied t-SNE to generate visualizations for latent features on the Signal-128 dataset. Fig. 10 illustrates the t-SNE plot for the raw signal data [Fig. 10(a)], ResNet pretrained feature [Fig. 10(b)], contrastive pretrained feature [Fig. 10(c)], and ScSP pretrained [Fig. 10(d)]. We can witness that the clusters of latent features under ScSP pretraining [Fig. 10(d)] are more distinct than the ResNet pretraining [Fig. 10(b)]. Furthermore, we observe that the ScSP pretraining latent features present more organized results (with intraclass and interclass concentration) than the normal contrastive pretrained latent features [Fig. 10(c)], which demonstrates the effectiveness of our ScSP framework.

VI. Conclusion

We introduce a novel radio signal preprocessing framework, ScSP, designed to supplement various SOTA FSL models for the task of modulation recognition. The ScSP framework employs info-preserved augmentations, generating a variety of signal segments and eliminating associated noises. A feature enhancement module within the framework aids in simplifying the signal representations, thereby fostering efficient representation learning. Comprehensive experimental outcomes validate the efficiency of the proposed ScSP framework.

References

[1] L. D. Xu, W. He, and S. Li, "Internet of Things in industries: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014.

[2] X. Zhang et al., "Distributed compressive sensing augmented wideband spectrum sharing for cognitive IoT," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 3234–3245, Aug. 2018.

[3] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapè, "MIMETIC: Mobile encrypted traffic classification using multimodal deep learning," *Comput. Netw.*, vol. 165, Dec. 2019, Art. no. 106944.

[4] S. He, K. Shi, C. Liu, B. Guo, J. Chen, and Z. Shi, "Collaborative sensing in Internet of Things: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1435–1474, 3rd Quart., 2022.

[5] Z. Yang, L. He, H. Yu, C. Zhao, P. Cheng, and J. Chen, "Detecting PLC intrusions using control invariants," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9934–9947, Jun. 2022.

[6] H. Pu, L. He, P. Cheng, M. Sun, and J. Chen, "Security of industrial robots: Vulnerabilities, attacks, and mitigations," *IEEE Netw.*, vol. 37, no. 1, pp. 111–117, Jan. 2023.

[7] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, Aberdeen, U.K. Cham, Switzerland: Springer, 2016, pp. 213–226.

[8] T. Li, Z. Hong, Q. Cai, L. Yu, Z. Wen, and R. Yang, "BISSIAM: Bispectrum Siamese network based contrastive learning for UAV anomaly detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12109–12124, Dec. 2021.

[9] C. Shahriar, S. Sodagari, and T. C. Clancy, "Performance of pilot jamming on MIMO channels with imperfect synchronization," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 898–902.

[10] Z. Zhang, Y. Li, Q. Zhai, Y. Li, and M. Gao, "Few-shot learning for fine-grained signal modulation recognition based on foreground segmentation," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2281–2292, Mar. 2022.

[11] Z. Zhang, C. Wang, C. Gan, S. Sun, and M. Wang, "Automatic modulation classification using convolutional neural network with features fusion of SPWVD and BJD," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 5, no. 3, pp. 469–478, Sep. 2019.

[12] M. Schmidt, D. Block, and U. Meier, "Wireless interference identification with convolutional neural networks," in *Proc. IEEE 15th Int. Conf. Ind. Informat. (INDIN)*, Jul. 2017, pp. 180–185.

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

[14] K. Li, Y. Zhang, K. Li, and Y. Fu, "Adversarial feature hallucination networks for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13470–13479.

[15] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, Aug. 2017, pp. 1126–1135.

[16] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 4077–4087.

[17] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1199–1208.

[18] O. Vinyals et al., "Matching networks for one shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 3630–3638.

[19] J. Mitrovic, B. McWilliams, and M. Rey, "Less can be more in contrastive learning," in *Proc. NeurIPS Workshops*, 2020, pp. 70–75.

[20] M. Gutmann and A. Hyvärinen, "Noise-contrastive estimation: A new estimation principle for unnormalized statistical models," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 297–304.

[21] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Commun. Mag.*, vol. 46, no. 4, pp. 40–48, Apr. 2008.

[22] Y. Zhao, Z. Hong, Y. Luo, G. Wang, and L. Pu, "Prediction-based spectrum management in cognitive radio networks," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3303–3314, Dec. 2018.

[23] D. Hong, Z. Zhang, and X. Xu, "Automatic modulation classification using recurrent neural networks," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2017, pp. 695–700.

[24] N. E. West and T. O'Shea, "Deep architectures for modulation recognition," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Mar. 2017, pp. 1–6.

[25] A. Ali and F. Yangyu, "Automatic modulation classification using deep learning based on sparse autoencoders with nonnegativity constraints," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1626–1630, Nov. 2017.

[26] D. Kang, H. Kwon, J. Min, and M. Cho, "Relational embedding for few-shot classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 8822–8833.

[27] E. Triantafillou et al., "Meta-dataset: A dataset of datasets for learning to learn from few examples," 2019, *arXiv:1903.03096*.

[28] E. Schwartz et al., "Delta-encoder: An effective sample synthesis method for few-shot object recognition," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 2850–2860.

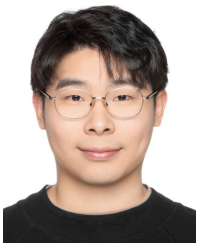
[29] M. Boudiaf, I. Ziko, J. Rony, J. Dolz, P. Piantanida, and I. B. Ayed, "Information maximization for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 2445–2457.

[30] I. Ziko, J. Dolz, E. Granger, and I. B. Ayed, "Laplacian regularized few-shot learning," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 11660–11670.

[31] B. Kang, Z. Liu, X. Wang, F. Yu, J. Feng, and T. Darrell, "Few-shot object detection via feature reweighting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 8420–8429.

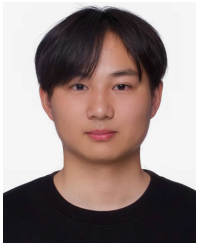
[32] Z. Peng, Z. Li, J. Zhang, Y. Li, G. Qi, and J. Tang, "Few-shot image recognition with knowledge transfer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 441–449.

[33] Z. Zhang, Y. Li, and M. Gao, "Few-shot learning of signal modulation recognition based on attention relation network," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 1372–1376.
[34] H. Zhou et al., "Few-shot electromagnetic signal classification: A data union augmentation method," *Chin. J. Aeronaut.*, vol. 35, no. 9, pp. 49–57, Sep. 2022.
[35] L. Peng, J. Zhang, M. Liu, and A. Hu, "Deep learning based RF fingerprint identification using differential constellation trace figure," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1091–1095, Jan. 2020.
[36] B. Flowers, R. M. Buehrer, and W. C. Headley, "Evaluating adversarial evasion attacks in the context of wireless communications," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1102–1113, 2020.
[37] T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. 37th Int. Conf. Mach. Learn.*, vol. 119, 2020, pp. 1597–1607.
[38] C. Zhendong, J. Weining, X. Changbo, and L. Min, "Modulation recognition based on constellation diagram for M-QAM signals," in *Proc. IEEE 11th Int. Conf. Electron. Meas. Instrum.*, vol. 1, Aug. 2013, pp. 70–74.
[39] O. Monerris-Belda et al., "High power RF discharge detection technique based on the in-phase and quadrature signals," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 12, pp. 5429–5438, Dec. 2021.
[40] J.-F. Dufourd, "Polyhedra genus theorem and Euler formula: A hypermap-formalized intuitionistic proof," *Theor. Comput. Sci.*, vol. 403, nos. 2–3, pp. 133–159, Aug. 2008.
[41] J. Issac, M. Wüthrich, C. G. Cifuentes, J. Bohg, S. Trimpe, and S. Schaal, "Depth-based object tracking using a robust Gaussian filter," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 608–615.
[42] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.
[43] P. Tian, W. Li, and Y. Gao, "Consistent meta-regularization for better meta-knowledge in few-shot learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7277–7288, Dec. 2022.
[44] R.-Q. Wang, X.-Y. Zhang, and C.-L. Liu, "Meta-prototypical learning for domain-agnostic few-shot recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6990–6996, Nov. 2022.
[45] H. Cheng, Y. Wang, H. Li, A. C. Kot, and B. Wen, "Disentangled feature representation for few-shot image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 8, pp. 10422–10435, Aug. 2023.
[46] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.
[47] A. Paszke, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 8024–8035.
[48] L. J. Wong and S. McPherson, "Explainable neural network-based modulation classification via concept bottleneck models," in *Proc. IEEE 11th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2021, pp. 191–196.
[49] X. Lu, M. Tao, X. Fu, G. Gui, T. Ohtsuki, and H. Sari, "Lightweight network design based on ResNet structure for modulation recognition," in *Proc. IEEE 94th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2021, pp. 1–5.
[50] Z. Zhang, "Improved Adam optimizer for deep neural networks," in *Proc. IEEE/ACM 26th Int. Symp. Quality Service (IWQoS)*, Jun. 2018, pp. 1–2.
[51] T. Kvålseth, "On normalized mutual information: Measure derivations and properties," *Entropy*, vol. 19, no. 11, p. 631, Nov. 2017.
[52] P. Cheng, W. Hao, S. Dai, J. Liu, Z. Gan, and L. Carin, "CLUB: A contrastive log-ratio upper bound of mutual information," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 1779–1788.



Jie Su received the B.S. degree in computer science and technology from China Jiliang University, Hangzhou, China, in 2017, the M.S. degree (Hons.) in data analytics from the University of Southampton, Southampton, U.K., in 2018, and the Ph.D. degree from Newcastle University, Newcastle upon Tyne, U.K., in 2023.

He is currently an Assistant Professor with the Institute of Cyberspace Security and the College of Information Engineering, Zhejiang University of Technology, Hangzhou. His research interests include deep learning, signal processing, and the IoT security.



Peng Sun (Student Member, IEEE) received the B.Eng. degree in electronic information engineering from Zhejiang University of Technology, Hangzhou, China, in 2023. He is currently pursuing the joint Ph.D. degree in computer science with Zhejiang University, Hangzhou, and Westlake University, Hangzhou.

His current research interests include optimization and generalization theory in machine learning and representation learning theory.

Yuting Jiang is currently pursuing the M.Sc. degree with the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China.

Her current research interests focus on deep learning applications, signal processing, and generation.

Zhenyu Wen (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from Newcastle University, Newcastle upon Tyne, U.K., in 2011 and 2016, respectively.

He is currently a Professor with the Institute of Cyberspace Security and the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. His current research interests include the IoT, crowd sources, AI system, and cloud computing.

Dr. Wen was awarded the IEEE TCSC Award for Excellence in Scalable Computing (Early Career Researchers) in 2020, for his contributions to the area of scalable data management for the Internet of Things.

Fangda Guo received the double M.S. degrees in software engineering and computer engineering from Northeastern University, Shenyang, China, and the University of Pavia, Pavia, Italy, in 2015 and 2016, respectively, and the Ph.D. degree in computer science from Northeastern University in 2021.

He was a joint Ph.D. Student with the University of Edinburgh, Edinburgh, U.K. He is currently an Assistant Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. His research interests include graph databases, query optimization, social media analysis, and computer vision.

Yiming Wu received the Ph.D. degree in cyberspace security from Zhejiang University, Hangzhou, China, in 2022.

She is currently an Assistant Professor at the Institute of Cyberspace Security and the College of Information Engineering, Zhejiang University of Technology, Hangzhou. Her research interests include data-driven security, software and system security, and AI security.

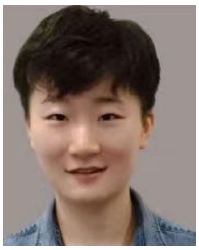
Zhen Hong (Member, IEEE) received the B.S. degree from Zhejiang University of Technology (ZJUT), Hangzhou, China, and the University of Tasmania, Hobart, TAS, Australia, in 2006, and the Ph.D. degree from ZJUT in January 2012.

He is currently a Full Professor with the Institute of Cyberspace Security and the College of Information Engineering, ZJUT. Before joining ZJUT, he was an Associate Professor with the Faculty of Mechanical Engineering and Automation, Zhejiang Sci-Tech University, Hangzhou. He has visited the Sensorweb Lab, Department of Computer Science, Georgia State University, Atlanta, GA, USA, in 2011. He also has been at CAP Research Group, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, as a Research Scholar, from 2016 to 2018. His research interests include the Internet of Things, wireless sensor networks, cyberspace security, and data analytics.

Dr. Hong is a Senior Member of CCF and CAA. He received the First Zhejiang Provincial Young Scientists Title in 2013 and Zhejiang Provincial New Century 151 Talent Project in 2014. He serves on the Youth Committee of Chinese Association of Automation and Blockchain Committee and CCF YOCSEF.

Haoran Duan (Member, IEEE) received the M.S. degree (Hons.) in data science from Newcastle University, Newcastle upon Tyne, U.K., in 2019, and the Ph.D. degree from Durham University, Durham, U.K., in 2024.

He was a Research Student at the OpenLab, Newcastle University. He is currently a Post-Doctoral Research Associate with the Networked and Ubiquitous Systems Engineering Group, School of Computing, Newcastle University, working on deep learning applications. His current research interests focus on the applications and theories of deep learning.



Yawen Huang received the M.Sc. and Ph.D. degrees from the Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield, U.K., in 2015 and 2018, respectively.

She is currently a Senior Scientist with the Tencent Jarvis Laboratory, Shenzhen, China. Her research interests include computer vision, machine learning, medical imaging, deep learning, and practical AI for computer-aided diagnosis.

Rajiv Ranjan received the Ph.D. degree from the Department of Computer Science and Software Engineering, The University of Melbourne, Parkville, VIC, Australia, in 2009.

He is currently a Full Professor of computing science at Newcastle University, Newcastle upon Tyne, U.K. Before moving to Newcastle University, he was a Julius Fellow from 2013 to 2015, a Senior Research Scientist, and a Project Leader of the Digital Productivity and Services Flagship of Commonwealth Scientific and Industrial Research Organization (CSIRO C Australian Government's Premier Research Agency). Prior to that, he was a Senior Research Associate (Lecturer Level B) with the School of Computer Science and Engineering, University of New South Wales (UNSW), Sydney, NSW, Australia.

Yefeng Zheng (Member, IEEE) received the B.E. and M.E. degrees from Tsinghua University, Beijing, China, in 1998 and 2001, respectively, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2005.

After graduation, he joined Siemens Corporate Research, Princeton, NJ, USA. He is currently the Director and the Distinguished Scientist with the Tencent Jarvis Laboratory, Shenzhen, China, leading the company's initiative on medical AI. His research interests include medical image analysis, graph data mining, and deep learning.

Dr. Zheng is a fellow of American Institute for Medical and Biological Engineering (AIMBE).