

# Multi-representation domain attentive contrastive learning based unsupervised automatic modulation recognition

Received: 2 December 2023

Accepted: 6 June 2025

Published online: 01 July 2025

 Check for updatesYu Li<sup>1</sup>, Xiaoran Shi<sup>1</sup>✉, Haoyue Tan<sup>1</sup>, Zhenxi Zhang<sup>1</sup>, Xinyao Yang<sup>1</sup> & Feng Zhou<sup>1,2</sup>

The rapid advancement of B5G/6G and wireless technologies, combined with rising end-user numbers, has intensified radio spectrum congestion. Automatic modulation recognition, crucial for spectrum sensing in cognitive radio, traditionally relies on supervised methods requiring extensive labeled data. However, acquiring reliable labels is challenging. Here, we propose an unsupervised framework, Multi-Representation Domain Attentive Contrastive Learning, which extracts high-quality signal features from unlabeled data via cross-domain contrastive learning. Inter-domain and intra-domain contrastive mechanisms enhance mutual modulation feature extraction across domains while preserving source domain self-information. The domain attention module dynamically selects representation domains at the feature level, improving adaptability. The experiments through public datasets show that the proposed method outperforms existing modulation recognition methods and can be extended to accommodate various representation domains. This study bridges the gap between unsupervised and supervised learning for radio signals, advancing Internet of Things and cognitive radio development.

The 5G wireless communications have become the social infrastructure for Internet of Things (IoT) and mobile internet applications<sup>1,2</sup>. The powerful feature extraction capability offered by the stacked layers of artificial neurons has sparked a significant expansion of research in reforming or even revolutionizing the design of communication systems in B5G and 6G IoT paradigms<sup>3,4</sup>. Among diverse efforts made by both academia and industry, automatic modulation recognition (AMR) using deep learning (DL) methods have attracted much attention<sup>5–8</sup>. AMR is of great significance as a key intermediate link in multiple natural wireless communication fields such as cognitive radio<sup>9,10</sup>, wireless sensor networks<sup>11</sup>, ensuring the physical layer security<sup>12,13</sup> and dynamic IoT environments<sup>14,15</sup>.

The purpose of AMR, as a key step in signal reception and demodulation, is to identify the modulation scheme employed in wireless signals and discern their types and characteristics<sup>16</sup>. Data-driven DL-AMR approaches typically employ supervised learning,

where well-designed deep neural networks are trained by a large number of labeled samples. From the perspective of network structure, CNN-type networks are particularly effective at extracting spatial correlation features from signals<sup>17–19</sup>. In addition, wireless signals also exhibit temporal correlation features<sup>20</sup>, which can be effectively learned using RNN<sup>20,21</sup>. Pure CNN or RNN models focus solely on either the spatial or temporal dimension of wireless signals. However, an increasing number of researchers have started investigating hybrid models that combine both CNN, RNN, and other architectures for AMR<sup>22–24</sup>. From the perspective of the input form of data, the aforementioned approach directly feeds time-domain IQ sequences into the network for feature extraction and recognition. Scholars have also considered leveraging features from different representational domains to enhance the performance of AMR<sup>25–28</sup>. Both manual and deep learning-based feature extraction methods have demonstrated that modulation information in wireless signals is distributed across

<sup>1</sup>Key Laboratory of Electronic Information Countermeasure and Simulation Technology, School of Electronic Engineering, Xidian University, Xi'an, China.

<sup>2</sup>School of Aerospace Science and Technology, Xidian University, Xi'an, China. ✉ e-mail: [xrshi@xidian.edu.cn](mailto:xrshi@xidian.edu.cn)

multiple representation domains. This process parallels the way humans gather information through multiple sensory modalities, including vision, hearing, touch, and taste. Such integration from different views enhances cognitive abilities. Methods that leverage information from multi-representation domains often achieve superior recognition performance. However, the choice of representational domains under different modulation schemes has currently become a critical issue for researchers in AMR.

In addition, supervising AMR method training requires a large number of high-quality samples. The process of annotating a large volume of labeled samples necessitates substantial investments in terms of time and financial resources. Moreover, in civilian scenarios, considering user privacy and security presents inherent difficulties in acquiring massive and accurate modulation labels<sup>16,29</sup>. In military scenarios, the reconnaissance side can intercept only a minimal number of labeled signals from the non-cooperative target<sup>30,31</sup>. The performance of the supervised AMR model cannot be fully guaranteed in the presence of insufficient labels<sup>32</sup>.

The mechanism of unsupervised learning offers a novel approach to tackle the aforementioned problem. Over the years, as large amounts of unlabeled wireless signals have accumulated in the radio stations of civilians and the military, we can collect more and more wireless signals. By constructing appropriate proxy tasks, high-quality representation information is extracted from unlabeled samples. These unsupervised representations can be applied to recognition and other downstream tasks either by fine-tuning with a few labeled samples<sup>33</sup> or through metric-based querying<sup>34</sup>. This prompts us to study the benefits of utilizing unlabeled wireless signals.

The key to unsupervised automatic modulation recognition (UAMR) critically depends on the approach to signal representation learning. Recent research<sup>16,32,35</sup> has provided compelling evidence of the superior performance achieved through the utilization of contrastive learning in unsupervised signal representation learning. The choice of signal augmentation is critical to the success of contrastive learning. Recently, innovative efforts have been directed towards semi-supervised learning<sup>35</sup>. Yihong Dong et al.<sup>36</sup> developed a semi-supervised signal recognition convolutional neural network (SSRCNN) that employs multiple loss functions for unlabeled samples. Dongxin Liu et al.<sup>32</sup> pioneered the use of self-supervised contrastive learning for the representational learning of modulation signals, creating positive and negative sample pairs by manipulating unlabeled IQ signals for self-supervised training. After pre-training, the encoder parameters are frozen, and labeled samples are used to refine the classifier. Weisi Kong et al.<sup>16</sup> explored the use of a self-supervised Transformer encoder, advancing the discussion on the application of pre-trained models in semi-supervised AMR to achieve superior recognition accuracy. However, the data augmentation strategies employed in existing works for constructing positive and negative sample pairs predominantly originate from computer vision or audio signal domains, rather than being specifically tailored for radio signals. Moreover, these augmentation methods remain confined to a single dimension (time-domain I-Q), neglecting the complex cross-dimensional correlations inherent in the high-dimensional modulation characteristic space, as shown in Fig. 1a. This oversight restricts the potential for deeper understanding and improvement in AMR representation learning.

In this work, we propose an unsupervised framework for UAMR called multi-representation attentive domain contrastive learning (MAC). The framework integrates multi-domain signal representation with contrastive learning. Unlike previous methods that were limited to contrastive learning within a single dimension (time-domain I-Q), MAC is capable of correlating signal modulation characteristics across multiple high-dimensional spaces, as shown in Fig. 1b. Similar to observing an object from multiple perspectives, MAC provides a more comprehensive understanding of the subject by integrating diverse

viewpoints. The domain attention (DA) module is proposed to emphasize features from multiple representation domains based on their contextual relevance. The selection of appropriate transformation domains for different modulation types is finished at the feature level, which is more robust than relying on expert knowledge during the signal preprocessing level. MAC maximizes the mutual information between representations of the signal in different domains, constructing dual-domain feature dictionaries for wireless signals. Furthermore, we propose the “I-Q single-centering” optimization strategy to extend MAC to an arbitrary number of representation domains. By leveraging proxy tasks, MAC has achieved efficient signal multi-domain representation learning in unlabeled scenarios, and promoted the construction of large communication semantic models and advances in cognitive radio.

## Results

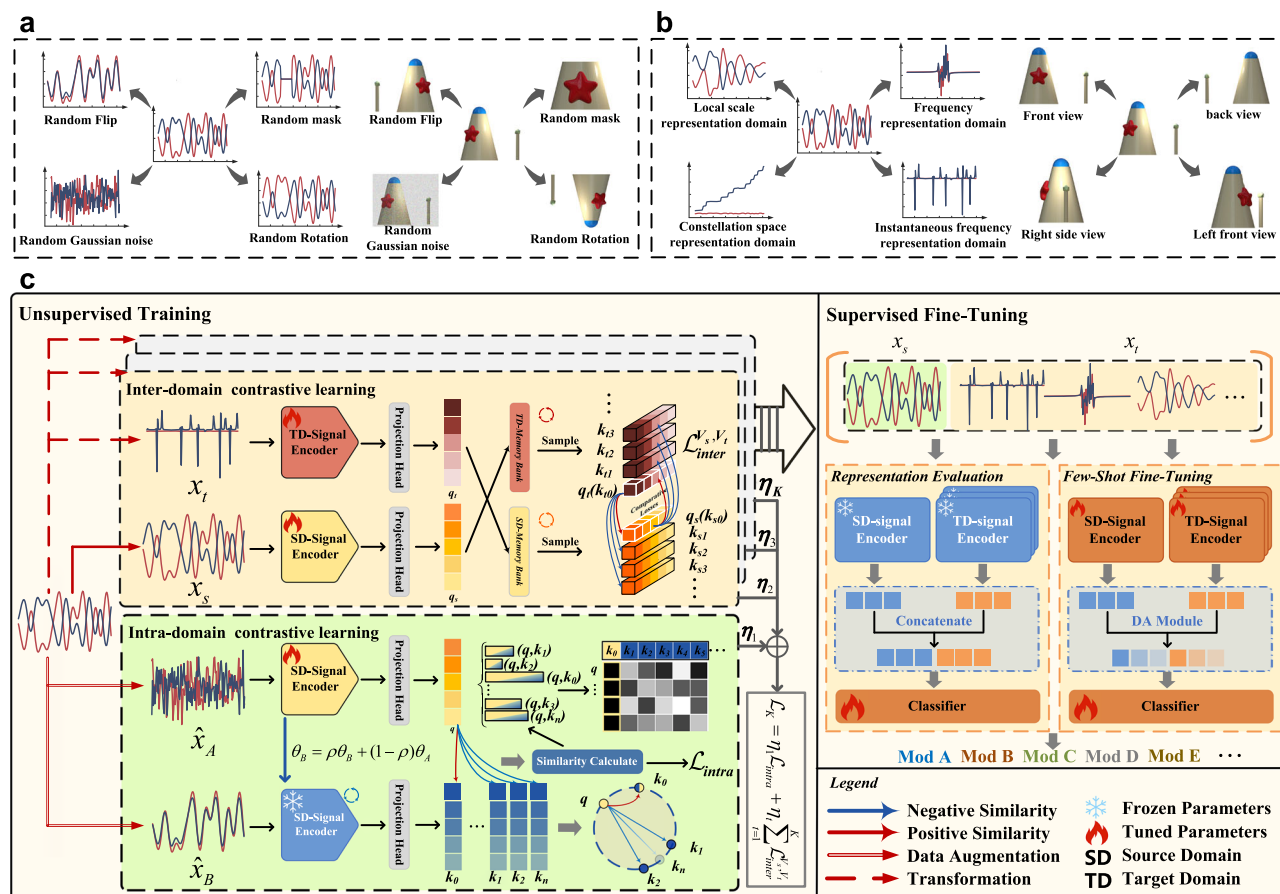
### The proposed MAC-based UAMR framework

We propose a MAC-based UAMR framework, as illustrated in Fig. 1c. Firstly, unsupervised learning of MAC can be divided into two parts: inter-domain contrastive learning and intra-domain contrastive learning. In the inter-domain contrastive learning part, multi-representation domain signals are taken to construct positive and negative sample pairs, which focus on the similarities and differences between different representation domains. Then, we establish an intra-domain contrastive learning within the representation domain by leveraging the augmented samples, which maintain robust features within the source domain (SD) during the process of maximizing inter-domain mutual information. Additionally, we extend MAC to handle any number of representation domains by “I-Q single centralization”. The DA module is proposed to leverage attention mechanisms to shift the selection of signal preprocessing forms from the signal level to the feature level. Finally, in the supervised fine-tuning stage, we validated the AMR performance through representation evaluation (linear evaluation and prototype evaluation) and few-shot fine-tuning. Three publicly available datasets, including RML2016.10 A, RML2016.10B, RML2018.01 A, were used to evaluate the proposed MAC. (refer to Supplementary Note 2 for details.)

### Performance of representation learning

We froze the feature embeddings obtained after unsupervised training and employed two evaluation paradigms, linear evaluation and prototype evaluation, to demonstrate the superior representation learning performance of the MAC unsupervised framework. We conducted a comparative analysis with several previous unsupervised learning frameworks, including SAE<sup>37</sup>, MoCo<sup>33</sup>, SimSiam<sup>38</sup>, MoCoV2<sup>39</sup>, and CPC<sup>40</sup>. For fairness, we denote the MAC-MT4 without the DA module as MC-MT4, and use it for comparison with the unsupervised framework, employing the same classifier. Significantly, we maintained consistency in hyperparameters such as  $k_m\rho$  across MC-MT4, MoCo, and MoCoV2. Additionally, in all comparative experiments except for SAE, we utilized the same backbone as the feature extractor to ensure fair performance comparison among unsupervised frameworks with similar feature extraction capabilities.

Figure 2a, b illustrate the accuracy curves of various methods at all SNRs in two evaluation paradigms. Linear evaluation and prototype evaluation yield similar results, with MC-MT4 achieving the best recognition performance. This strongly demonstrates the superiority of multi-domain representation learning. Among the comparison methods, CPC achieves reasonably good results in two evaluation paradigms, which can be attributed to the contrastive learning task of encoding predictions, which captures the temporal dependencies of the signal. However, the information obtained from a single domain remains inherently limited. SimSiam performs better under the prototype evaluation paradigm than under linear evaluation. This is due to SimSiam discarding negative sample pairs and directly maximizing the



**Fig. 1 | Design of MAC-based UAMR framework. a** Signal observation comparative dimensions based on IQ single dimension. **b** MAC considers multi-domain representation perspectives. **c** The workflow of MAC. Firstly, the unsupervised training module performs inter-domain and intra-domain representation learning on unlabeled signals, bringing positive samples closer and negative samples farther

apart in the feature space, respectively. Secondly, the AMR tasks are performed by supervised fine-tuning. Representation evaluation (linear evaluation and prototype evaluation) and few-shot fine-tuning are used to assess the quality of representation learning and semi-supervised performance, respectively.

similarity between two augmented views of the same sample, which aligns with the instance-based verification of the prototype evaluation. The proposed MC-MT4 outperforms the previous UAMR models SAE, indicating that the contrastive unsupervised learning approach exhibits superior capabilities in extracting meaningful representations compared to the simpler generative unsupervised sparse auto-encoders. The comprehensive utilization of multi-domain features in MC-MT4 contributes to its improved accuracy, indicating that crucial modulation information in wireless signals resides in multiple representation domains within a high-dimensional space. By effectively leveraging this multi-domain information, the proposed model can capture the discriminative features underlying the signals and achieve enhanced classification accuracy.

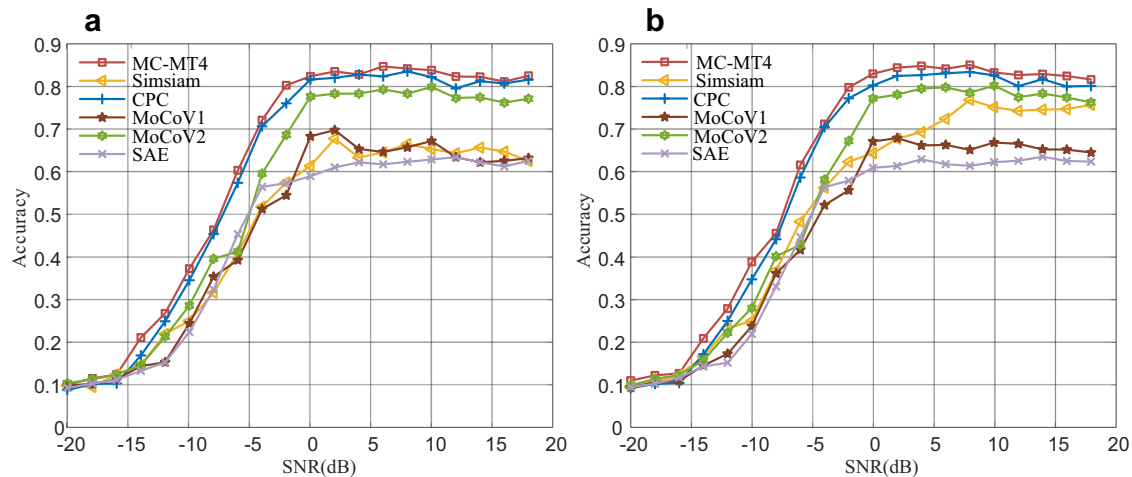
Furthermore, we visualize and analyze the effectiveness of the proposed MAC framework utilizing t-Distributed Stochastic Neighbor Embedding (t-SNE)<sup>41</sup> to map the high-dimensional feature to a two-dimensional space for visualization. As shown in Fig. 3a–h, we consider the t-SNE results for both linear evaluation and prototype evaluation (three sub-centroids) at 0 dB (see Supplementary Fig. S1) and 8 dB. As the number of training epochs increases, the intra-class distances for both paradigms are gradually reduced, and the inter-class distances are expanded, indicating that MAC is able to get high-quality signal representations using unlabeled samples. We found that the difference in classification hierarchy between linear evaluation (Fig. 3a–d) and prototype evaluation (Fig. 3e–h) can lead to changes in feature discrimination. In addition, for some modulation types, an excessive

number of sub-centroids may not contribute to classification. (Detailed discussion see Supplementary Note 3 and Note 7.)

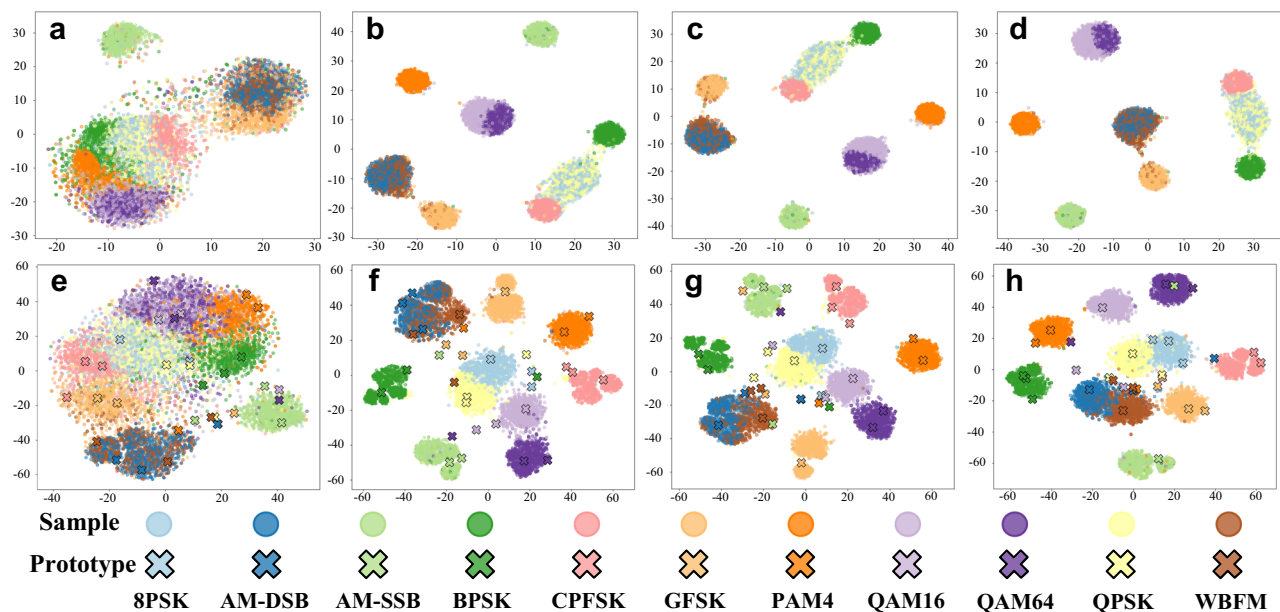
### Performance of few-shot fine-tuning recognition

To intuitively demonstrate the utilization of unlabeled samples in representation learning, we conducted few-shot fine-tuning experiments using  $N=2, 5, 10, 20, 50, 100$  labeled samples. To ensure fairness, in all datasets, the number of labeled samples for each modulation type at each SNR remains the same. To eliminate the impact of individual sample differences on the results, all fine-tuning and real-time inference results are averaged over ten Monte Carlo experiments.

Firstly, we conducted a thorough analysis of the fine-tuning performance of MAC using different numbers of labeled samples. The results are shown in Fig. 4, where the curves in different colors represent different values, and the shading around the curves represents the fluctuation in accuracy due to random sample selection. It is evident that as the number of labeled samples increases, the recognition accuracy of MAC continuously improves. And, the impact of individual sample differences on MAC fine-tuning weakens across all three datasets. When we use 100 labeled samples to fine-tune, meaning 10% labeled ratio in the RML2016.10A and 1.67% labeled ratio in the RML2016.10B, MAC achieves an average recognition accuracy of over 78.93% and 74.15% for signals above 0 dB, respectively. For the RML2018.01A dataset, using the 2.44% labeled samples, the average recognition accuracy for signals above 12 dB was over 79.24%.



**Fig. 2 | Unsupervised representation learning evaluation for MAC. a, b** Comparison Accuracy(%) of existing unsupervised frameworks and the proposed MC-MT4 under linear evaluation and prototype evaluation paradigms on RML2016.10 A dataset. Source data are provided as a Source Data file.

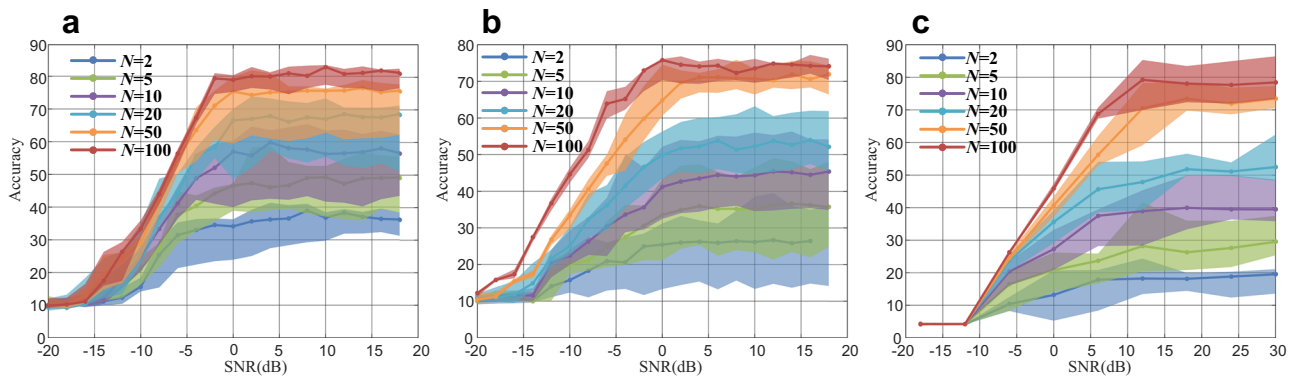


**Fig. 3 | Feature visualization results of the MAC-MT4 model when SNR = 8 dB at different training epochs of RML2016.10 A dataset. a–d** Linear evaluation when epoch = 1, 20, 40, 80. **e–h** Prototype evaluation when epoch = 1, 20, 40, 80.

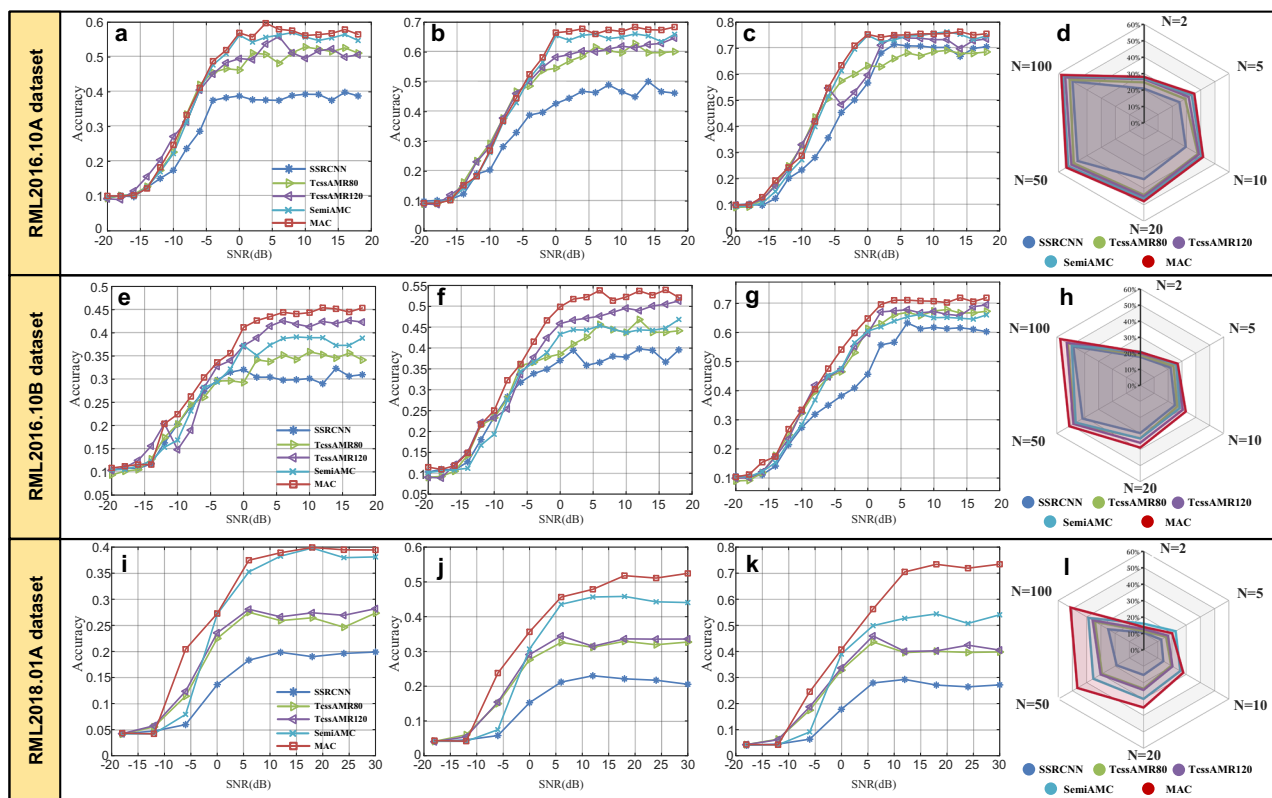
Additionally, we compared our method with the current popular semi-supervised AMR methods, including SSRCNN<sup>36</sup>, SemiAMC<sup>32</sup>, and TcssAMR<sup>16</sup>. The comparison results on RML2016.10 A, RML2016.10B and RML2018.01 A datasets are shown in Fig. 5. It is evident that MAC achieved the best semi-supervised recognition performance across three datasets in 90% of cases (Fig. 5d, h, i). We observe that the performance gain of MAC compared with the semi-supervised methods mainly comes from the signals with high SNR, while MAC performs similarly with the semi-supervised methods when the SNRs are low. When  $N > 10$  (Fig. 5a–c, e–g, i–k and Supplementary Fig. S2), MAC achieves significant performance improvements at high SNRs across all three datasets. Particularly on the RML2018.01 A dataset, when  $N = 50$  (Fig. 5k), using only 1.22% of the labeled samples from the dataset, MAC can achieve an over 70% recognition accuracy for signals above 12 dB. Compared to SemiAMC and TcssAMR, recognition accuracy has increased by 17.8% and 30.5% respectively, illustrating a improvement over existing semi-supervised methods.

The backbone of TcssAMR is based on Transformer. We increased the fine-tuning epochs for TcssAMR. However, even after 120 epochs of training, the accuracy remained lower than MAC. On the RML2018.01 A dataset, TcssAMR might require more labeled samples. When  $N = 2$  and  $N = 5$  (Supplementary Fig. S2), SemiAMC achieved higher accuracy than MAC on the RML2018.01 A dataset. As the number of labeled samples increased, the accuracy improvement of SemiAMC increased slowly. Note that, when  $N = 20, 50, 100$  (Fig. 5j, k and (Supplementary Fig. S2), using 0.49%, 1.22%, and 2.44% labeled signals from the RML2018.01 A dataset for fine-tuning, the recognition accuracy of MAC is higher than SemiAMC by 5.98%, 19.03%, and 19.68%, respectively, at a SNR of 18 dB. These phenomenon may be due to the multi-domain representation extraction pre-trained through unsupervised learning requires a certain number of labeled samples for activation. MAC significantly improved the recognition accuracy of semi-supervised AMR through the multi-domain contrastive learning and the DA mechanism.





**Fig. 4 | Few-shot fine-tuning recognition accuracy(%) of semi-supervised experiments under different SNRs on three datasets. a–c represent RML2016.10 A, RML2016.10B and RML2018.01 A dataset. Source data are provided as a Source Data file.**



**Fig. 5 | Comparison experiment with semi-supervised AMR methods. a–d** Recognition accuracy(%) on RML2016.10A dataset for  $N=10, 20, 50$  and the overview radar map respectively. **e–h** Recognition accuracy(%) on RML2016.10B dataset

for  $N=10, 20, 50$  and the overview radar map respectively. **i–l** Recognition accuracy(%) on RML2018.01A dataset for  $N=10, 20, 50$  and the overview radar map respectively. Source data are provided as a Source Data file.

The comparison of recognition accuracy between MAC and popular supervised AMR models, including I-Q CNN<sup>42</sup>, ResNet<sup>43</sup>, CLDNN<sup>44</sup>, CLDNN2<sup>43</sup>, LSTM<sup>20</sup>, MCLDNN<sup>23</sup>, PET-CGDNN<sup>45</sup> in few-shot scenarios are shown in the (Supplementary Table S1). It is evident that the performance of supervised methods significantly decreases with fewer labeled samples, while MAC, pre-trained with unlabeled samples, significantly improves the AMR performance under  $N \leq 100$ . When  $N=2$  and  $N=5$  (Supplementary Table S1), MAC improves recognition accuracy by 7.22% and 9.92%, respectively, compared to the SOTA supervised AMR methods. It is worth noting that PET-CGDNN achieves the best performance under fully supervised conditions (when  $N=ALL$ , thanks to the proposed frequency estimation module. However, as the number of labeled samples decreases (when  $N=2, 5, 10, 20, 50, 100$ ), PET-CGDNN found it difficult to

accurately estimate the frequency offset, resulting in a significant decrease in the recognition accuracy. Compared to SP-MAC, MAC significantly improves recognition accuracy across all values of  $N$ . When  $N=100$  (Supplementary Table S1), using 10% of the dataset labeled samples, MAC (58.12%) nearly achieves the same results as SP-MAC (58.35%) using all dataset labeled samples. This fully demonstrates the effective utilization and exploitation of unlabeled samples through the proposed multi-representation domain unsupervised pre-training. Additionally, even when using all labeled samples, MAC still improves accuracy by 2.99% compared to SP-MAC (Supplementary Table S1). This demonstrates that unsupervised pre-training can provide robust pre-trained weights for downstream AMR tasks.

The inability of an AMR method to accurately classify modulations arises when it fails to identify the fundamental AMR characteristics

within its training set. In order to thoroughly validate the network's generalization ability, our experiment is designed to ensure a clear demarcation between training and testing phases by employing distinct datasets. In each method, the corresponding model pre-trained in RML2016.10 A will be tested in RML2016.10B to verify the generalization abilities of the model.

Considering the results in (Supplementary Table S2), most of the supervised AMR methods based on I-Q signals have shown poor performance in generalization accuracy. It can be seen that an AMR method can no longer classify modulations well if this method cannot find the essential AMR characteristics in its training set. The ResNet architecture, incorporating multiple layers of residual structure connections, effectively mitigates overfitting on the training dataset, thereby preserving an accuracy of 86.98% on the generalization dataset (Supplementary Table S2). MCLDNN exploits the complementary information from I-Q multi-channel, I channel, and Q channel data to extract robust signal features, resulting in an accuracy of 89.16%. The proposed MAC-MT4 has achieved the best performance and its accuracies are over 91% when the SNR = 8 dB (Supplementary Table S2). This demonstrates the strong AMR generalization ability of MAC-MT4. The possible reasons are as follows. Firstly, during unsupervised training, MAC-MT4 does not establish a direct mapping between signal samples and class labels but discretizes the feature vectors of the samples in the sample space. This enables MAC-MT4 to avoid overfitting samples in a single dataset. Secondly, the utilization of multi-domain information assists MAC-MT4 in identifying the distinctive features of signals in different representation domains. Most importantly, the proposed MAC model selects relevant domain information through domain attention mechanisms to emphasize key representations. For PSK and QAM signals in particular, the constellation space representation domain is a 2D statistical distribution diagram result of symbol values of a signal on the I-Q plane, and its shape, number, and array of constellations are generally unchanged, even when the transmission symbols of the same modulation are ordered differently in different datasets. These explicit features to represent modulations are not difficult to find by MAC.

The confusion matrices in Fig. 6a, b depict the performance of MAC-MT4 on the generalization task at SNR = 0 dB and 8 dB. Overall, the diagonal elements of the confusion matrices are clear, indicating that MAC-MT4 maintains robust feature extraction capabilities and achieves satisfactory classification performance when trained and tested on two different datasets, thereby ensuring accurate and reliable identification of modulation schemes in diverse and dynamic IoT transmission environments.

To assess the effectiveness of MAC in achieving favorable generalization performance, we evaluate the quality of features extracted by both supervised and unsupervised model encoders. The distribution of feature vectors in the feature space serves as a reliable indicator of the signal representation quality. In this regard, we introduce the similarity ratio  $\mathcal{R}$ , which provides a comprehensive assessment of both intra-class and inter-class distributions, and can be expressed as

$$\mathcal{R} = \frac{1}{L} \sum_{l=1}^L \frac{P_{l}^{intra-class}}{P_{l}^{inter-class}} \quad (1)$$

A higher similarity ratio between intra-class and inter-class indicates that the features within the same class are more tightly clustered, which is beneficial for subsequent classification. Figure 6c, d illustrates the distribution of the proposed unsupervised method and a supervised model with the same backbone on the generalization task. In Fig. 6c, when the training and testing data belong to the same dataset, the density distribution of the unsupervised model is similar to the supervised model. However, in Fig. 6d, when tested with signals that are not in the same dataset as the ones used for training, the similarity ratio of the supervised model peaks around  $\mathcal{R} = 1$ . This suggests that

some signals have similar intra-class similarity  $P_{intra-class}$  and inter-class similarity  $P_{inter-class}$ , posing significant challenges for subsequent classification. In contrast, the proposed unsupervised model exhibits a higher similarity ratio, indicating that MAC-MT4 can learn more robust signal representations and has pleasant generalization ability.

### Ablation study

We discuss the impact of hyperparameters as well as key modules of MAC through ablation studies. The symbol for the variant is shown in Supplementary Note 4. Prior to this, we demonstrate the effects of the momentum update coefficient and feature dictionary stack size on the effectiveness of signal representation learning. We compared the recognition accuracy and training time of the proposed MAC-MT4 and representative MAC-D1 for 0 dB signals on two datasets.

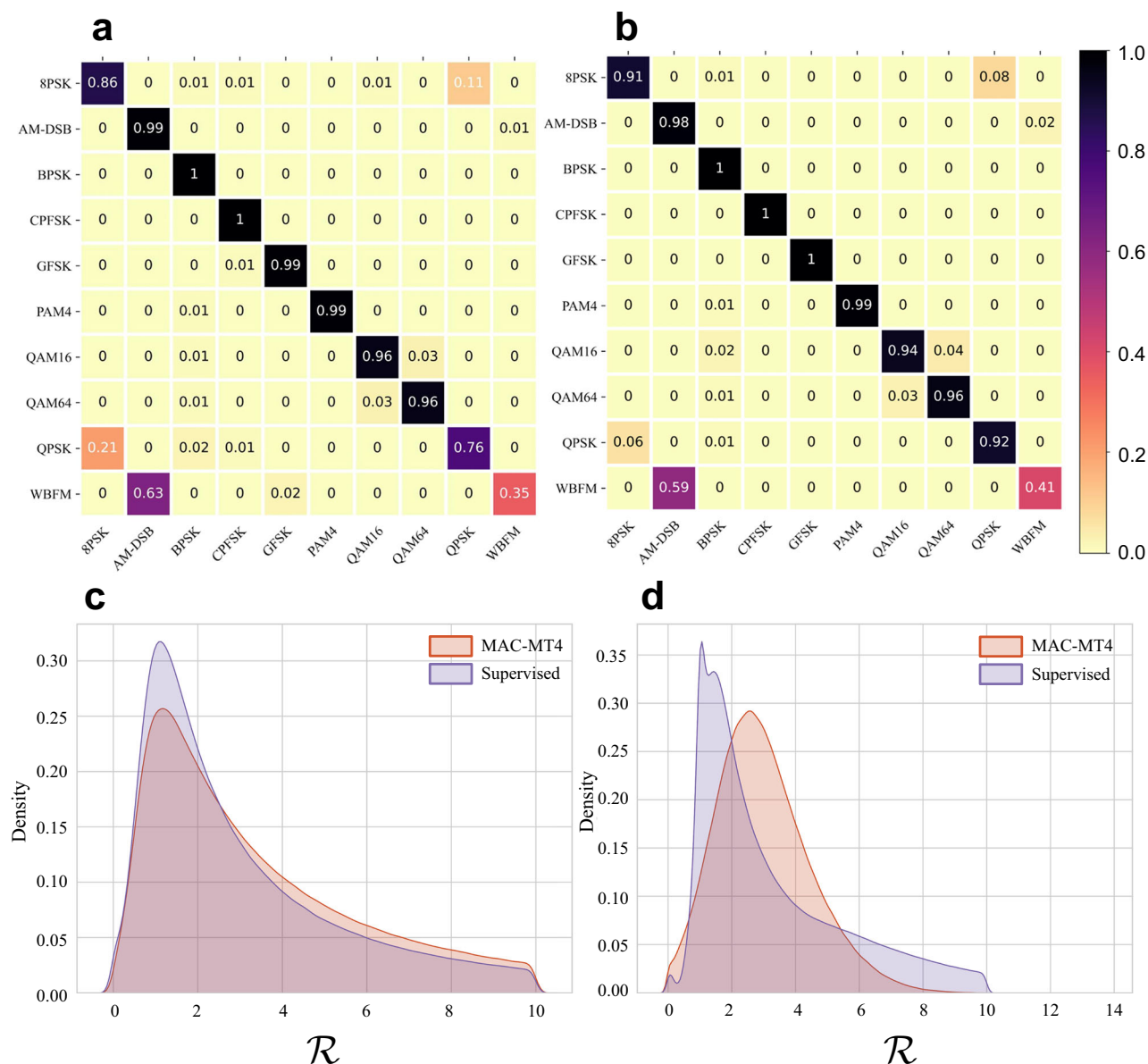
Supplementary Table S3 illustrates the impact of momentum update coefficients on the consistency of features within the range of 0 to 0.999. When  $\rho = 0.9$ , both MAC-MT4 and MAC-D1 achieve optimal performance on two datasets, indicating that a moderately slow update of the key encoder is beneficial. This can be attributed to the fact that rapid updates of encoder parameters (i.e., too small coefficients) lead to the loss of consistency between consecutive iterations of features within the dictionary over time. Conversely, slow updates (i.e., too large coefficients) result in significant distribution differences among features in different dictionaries; at the extreme of no momentum ( $\rho = 0$ ), the training loss oscillates and fails to converge. These results support our motivation to build a consistent dictionary.

Notably, the proposed method outperforms RML2016.10B on the RML2016.10 A dataset when using a smaller momentum update coefficient ( $\rho < 0.7$ ). This can be attributed to the smaller sample size of the RML2016.10 A dataset, where individual identification-based proxy tasks are relatively easy to accomplish in unsupervised training, thereby requiring less consistency in sample features. Furthermore, the momentum update strategy does not introduce any additional trainable parameters.

Figures 7, 8 shows the results of the ablation study on MAC. Firstly, Fig. 7 illustrate the influence of  $k_m$  on the effectiveness of signal representation learning. In general, the accuracy of identification benefits from a larger  $k_m$ , akin to the concept of a memory bank. The inclusion of a greater number of negative sample features in the field of comparison facilitates the learning of signal representation. However, excessive negative samples not only prolong training time but also increase the difficulty of contrastive learning.

The results indicate that the improvement in accuracy becomes negligible compared to the additional computational resources when the number of negative samples exceeds 16,384. The proposed method achieves the optimal trade-off between accuracy and training time on the RML2016.10 A and RML2016.10B datasets when  $k_m = 8192$  and 16,384, respectively. Building upon these findings, we will conduct subsequent experiments based on the aforementioned hyperparameters.

The different representation domains serve as different perspectives of the same modulation type, where the sample from SD and TD can alternate as contrastive query vectors. Figure 8a shows the contrastive loss curves between different representation domains according to Fig. 9b on the validation set in training MAC-MT4. The four groups of contrastive loss curves consistently decrease and converge as the unsupervised training epoch goes on. This indicates that MAC can simultaneously complete the task of positive sample screening in four representation domains.  $L_{s \sim t}^{V_{s'} V_t}$  converges to nearly the minimum value. This is attributed to the locally scaled representation domains  $\{WT\}$  are obtained from the SD through wavelet transformation, resulting in high similarity in overall waveform trends that facilitate effective contrastive learning by the network. Benefiting from the "I-Q single-centering" strategy,  $L_{s \sim t}$  with SD samples as query vectors, reaches a better solution than  $L_{t \sim s}$ . Figure 8b demonstrates



**Fig. 6 | Confusion matrix and similarity ratio density map of generalization experiment.** The confusion matrix of the proposed MAC-MT4 at SNR = 0 dB (a), 8 dB (b). Distribution of similarity ratio  $\mathcal{R}$  density of supervised frameworks and the

proposed MAC-MT4. c RML2016.10 A trained; RML2016.10 A tested. (d) RML2016.10 A trained; RML2016.10B tested. Source data are provided as a Source Data file.

the convergence of the loss function for MAC-MT4, MC-MT4, and TAC during the linear evaluation stage. TAC-MT4, which lacks the intra-domain contrastive learning, only converges to a local minimum. Due to the presence of DA, MAC-MT4 outperforms MC-MT4 and achieves the comparatively best solution.

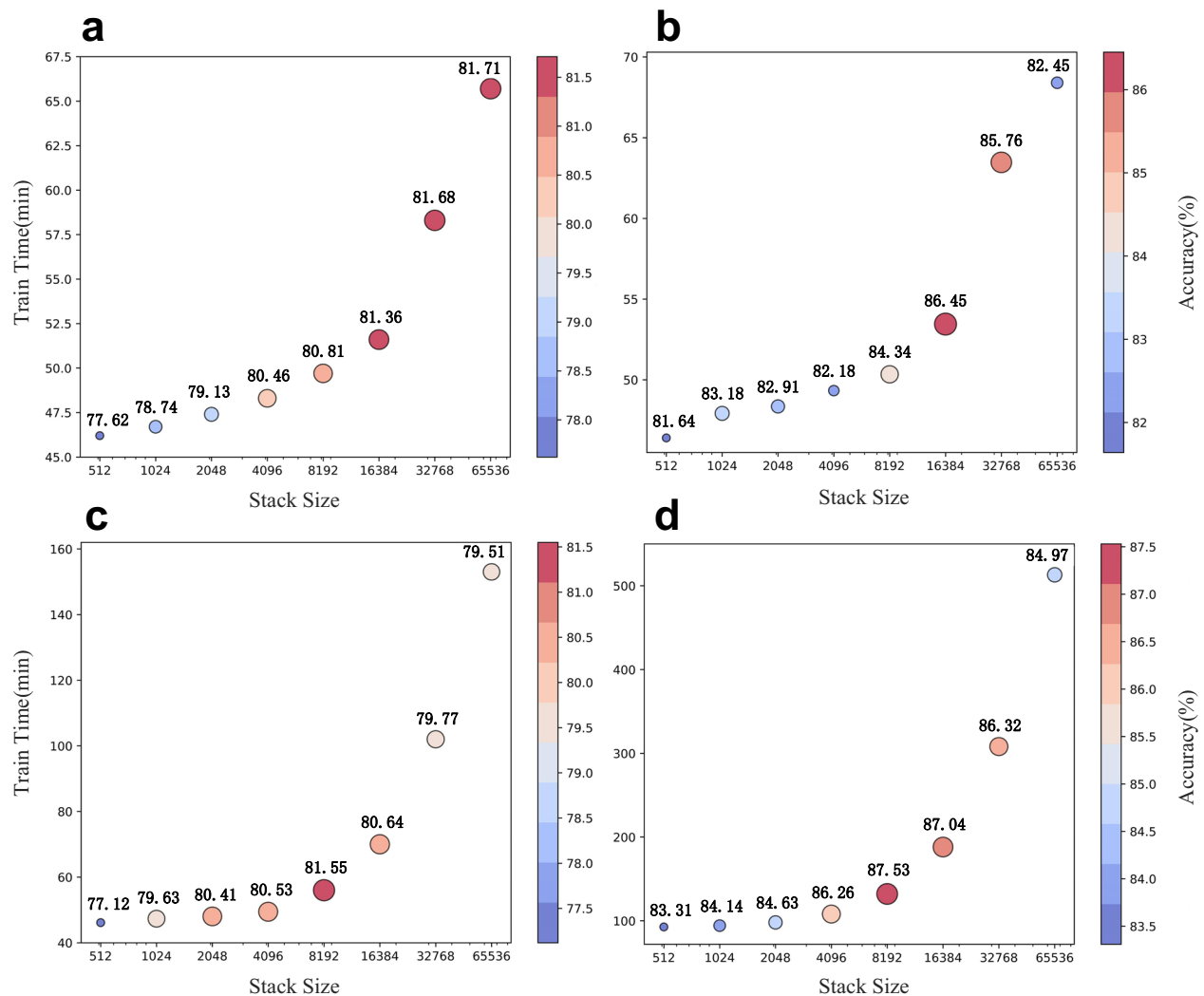
Subsequently, a detailed analysis will be undertaken to assess the impact of each representation domain's ablation and the specific contributions made by SD representation learning and DA to MAC.

We conducted ablation experiments on intra-domain and inter-domain contrastive learning, as well as the domain attention mechanism, comparing the classification performance of MAC-MT4, MAC-MT4, SRC, TAC, and MAC-DX(MAC-D1-MAC-D4) on two datasets.

The linear evaluation results on the RML2016.10 A and RML2016.10B datasets, as shown in Supplementary Table S4 and Supplementary Table S5, respectively, indicate that overall, the recognition accuracy is higher on the RML2016.10B dataset due to its larger sample size. The performances for both datasets demonstrate

that MAC-MT4 is the most effective among all SNRs. Consistent with the analysis in Fig. 8b the method of combining TDs and SD under the MAC framework (MAC-DX) significantly outperforms SRC in terms of testing accuracy. Compared to SRC, the accuracy of MAC-MT4 improved by 12.94% on the RML2016.10 A dataset and 12.92% on the RML2016.10B dataset (Supplementary Table S4 and Supplementary Table S5). This confirms the effectiveness of the MAC unsupervised framework, where inter-domain contrastive learning can comprehensively utilize information from both SD and TDs.

The results of MAC-DX, which use only a specific representation domain, are lower than those of MAC-MT4. Notably, among the contributions of different specific representation domains, MAC-D2 shows lower recognition accuracy, which can be attributed to the challenges posed by the rapid fluctuations of instantaneous frequency, making feature learning by the encoder more difficult. Additionally, MAC-D4 exhibits good performance at low SNRs, benefiting from the high- and low-frequency decomposition effects of wavelet transforms. Overall,



**Fig. 7 | The ablation study of stack size of MAC framework. a–d** Performance comparison of the proposed method with different domain stack sizes on the RML2016.10 A and RML2016.10B datasets. **a** MAC-D1 on RML2016.10 A, **b** MAC-MT4

on RML2016.10 A, **c** MAC-D1 on RML2016.10B, **d** MAC-MT4 on RML2016.10B. Source data are provided as a Source Data file.

these results highlight the advantages of integrating multiple representation domains to effectively explore and leverage information from different domains.

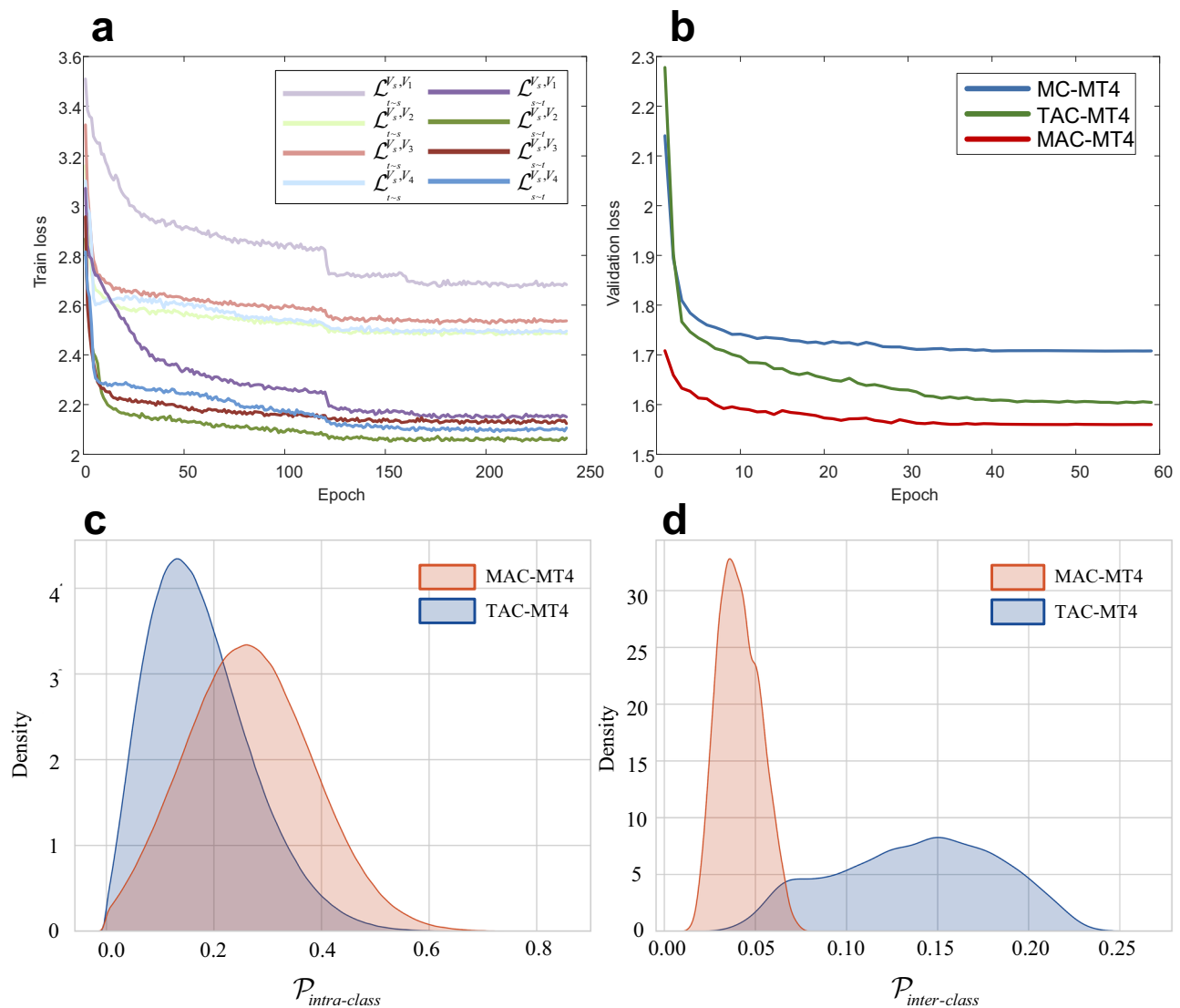
The confusion matrices for the proposed MAC-MT4 on the RML2016.10 A and RML2016.10B datasets are shown in Supplementary Fig. S3. The proposed MAC-MT4 demonstrates satisfactory discriminability for the challenging recognition problem of QAM16 and QAM64, which has been a major challenge for most existing AMR methods<sup>23,24</sup>. Additionally, through the DA module, the MAC framework effectively reduces confusion between 8PSK, AM-DSB, and QPSK signals. These modulation types exhibit significant distinguishability in constellation space and instantaneous amplitude representation, which can be appropriately captured by MAC. However, differentiating between AM-DSB and WBFM poses some difficulties. Due to the strong spectral similarity between these two types of signals<sup>46</sup>, WBFM also has periods of audio silence<sup>21</sup>.

For a more straightforward visualization of the impact of SD representation learning on feature vectors, we compute the density distribution of intra-class similarity  $P_{\text{intra-class}}$  and inter-class similarity  $P_{\text{inter-class}}$  for modulation signal classes. The specific definition for intra-class and inter-class similarity measurement is shown in the Supplementary Note 5.

Pleasant contrastive learning results should maximize the intra-class sample similarity while minimizing the inter-class similarity. We define  $N_s = N$  as the number of inter-class samples used for inter-class similarity calculations. As depicted in Fig. 8c,d, with the addition of intra-domain contrastive learning, intra-class sample similarity is increased, while inter-class sample similarity is decreased. Consequently, SD representation learning is crucial for preserving the IQ characteristics and distinguishing amplified samples within the feature space.

Undifferentiated handling of information from various representation domains, as the classification network treats all inputs equally, makes it difficult to select appropriate representations. MC-MT4, lacking attention guidance, achieves lower performance compared to MAC-MT4 (Supplementary Tables S4,S5). MC-MT4 has limited capability to filter out irrelevant information and emphasize relevant information among signals from multi-representation domains. Furthermore, we delve into a detailed display of the effectiveness of the DA module in selecting signal representation domains at the feature level in the Supplementary Fig. S4. Comparing the visual attention weights to the modulation signal types, MAC correctly chooses the transformation domains for signal preprocessing at the feature level and emphasizes the feature vectors of the appropriate





**Fig. 8 | The ablation study of core module of MAC framework.** **a** Training loss curves of unsupervised training on the RML2016.10 A dataset.  $V_1, V_2, V_3, V_4$  representing the combination of transformation domains  $\{A\phi\}, \{Af\}, \{F\}, \{WT\}$  separately.

**b** Linear evaluation validation loss on ablation of different modules. Distribution of intra-class **(c)** and inter-class **(d)** similarity density of MAC-MT4 and TAC-MT4. Source data are provided as a Source Data file.

representation domains for different modulation types in the final classification module. (Refer to Supplementary Note 6 for details.)

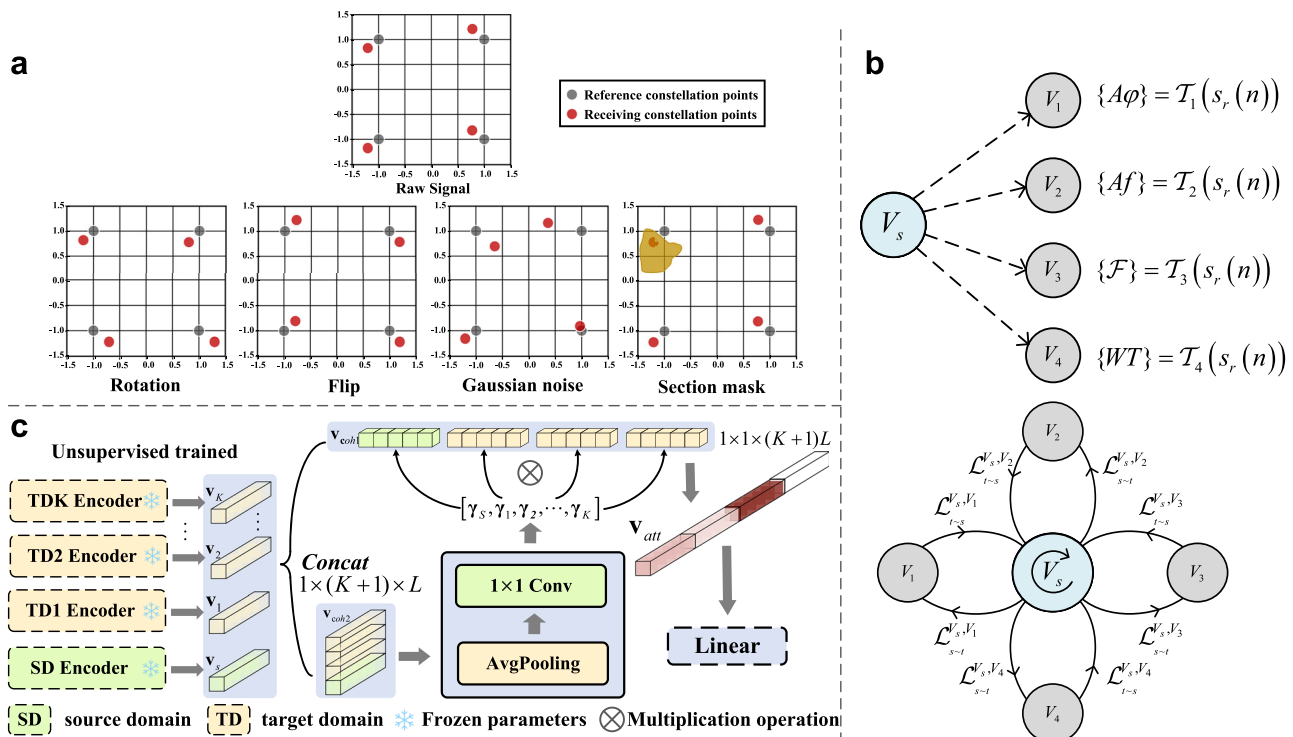
In summary, the extensive incorporation of multi-representation domains in MAC-MT4 leads to superior classification performance. Intra-domain contrastive learning assists MAC in preserving robust SD features, and the DA module within the MAC framework plays a pivotal role. Effective domain attention weights aid in selecting the most suitable representation domain forms at the feature level, laying a strong foundation for final recognition.

## Discussion

Previous experiments on publicly available datasets have shown that multi-representation attentive domain contrastive learning (MAC) has the ability to leverage a substantial quantity of unlabeled signal samples. Creating positive and negative sample pairs using signal multi-representation domains and data augmentation, signal representation learning is performed unsupervised. The experimental results of linear evaluation and prototype evaluation in Fig. 2a, b indicate that MAC achieves the mining of information from multiple characterization domains and obtains discriminative modulation features.

Subsequently, the results of few-shot fine-tuning highlight the impressive unsupervised learning capability and interpretability of the MAC framework. The results under different sample label sizes in Figs. 4 and 5 indicate that MAC can reduce the number of labeled samples required. The comparison results in Supplementary Table S1, S2 with the supervised model show that MAC reduces the disparity with supervised models in terms of classification results and exhibits robust generalization performance.

The results of the ablation study indicate the contributions of different representation domains and components in MAC. We found that the characteristic differences of modulated signals in the representation domain are related to their performance contributions, as shown in Supplementary Fig. S5. For instance, in distinguishing QPSK and 8PSK signals, as well as QAM16 and QAM64 signals, the constellation space representation domain and the instantaneous frequency representation domain, which consider amplitude information, achieved better clustering results. In general, the visual results of using specific representation domains align closely with the inherent modulation features of the signals. In addition, the domain attention (DA) module allocates suitable



**Fig. 9 | Modules in MAC Framework.** **a** Different data augmentation methods for modulation signals used for SD representation learning. **b** “I-Q single centralization” multi-representation domain loss calculation strategy. **c** DA module for attention inference.

attention to multi-representation domains, facilitating the selection of the optimal signal representation domain. The contribution of specific representation domains to different modulation types is in good alignment with the attention scores obtained by the DA module, shown in Supplementary Fig. S4, providing strong evidence of the interpretability of MAC.

Our approach enables effective spectrum utilization without relying on labeled training samples, particularly in dynamic and heterogeneous environments. Most importantly, unsupervised signal representation learning is not confined to a specific feature task. Our objective is to leverage MAC for transferring across multiple downstream tasks, such as modulation recognition, estimation of key signal parameters, SNR estimation, and communication signal behavior recognition. By leveraging unlabeled data, our method contributes to the development of more efficient, adaptive, and intelligent wireless communication systems. These advancements have the potential to facilitate a wide range of innovative applications and services in the era of IoT devices and next-generation wireless networks. In addition, optimizing the representation learning process by considering multiple similarity measurement methods and mining quantitative indicators that can evaluate the level of unsupervised representation learning of signals could be further investigated.

## Methods

### Signal model and problem formulation

Consider the baseband wireless signal model after down-conversion at the receiver. At any discrete time instant, the relationship between the transmitted signal and the received signal can be represented as

$$s_r(n) = s_t^l(n) * h(n) e^{j(2\pi n \Delta f + \varphi_0)} + w(n) \quad (2)$$

where  $s_r(n)$  is the received baseband signal,  $s_t^l(n)$  is the modulated signal generated from one of  $L$  modulation schemes  $\{s^1(n), s^2(n), \dots, s^L(n)\}$ ,  $h(n)$  is the pulse response of the wireless transmission channel,  $*$  denotes the convolution operation,  $w(n)$  is the

noise,  $\Delta f$  and  $\varphi_0$  represents the additional carrier frequency offset and phase jitter during the transmission process.  $n = 0, 1, \dots, N_L - 1$ ,  $N_L$  represents the total length of the signal.

The AMR methods based on multi-input use different representation domains and achieve pleasant performance when the input form and characteristics of the signal can be selected properly. The modulation recognition problem based on signals from multi-domain can be formulated as

$$\hat{l} = \arg \max_{l \in \{1, \dots, L\}} P(M_l | \sum_{k=1}^K T_k(s_r(n))) \quad (3)$$

where  $T_k(\cdot)$  denotes the  $k$ -th preprocessing transformation operation for I-Q sequence of baseband signal. The four signal representation domains we consider are introduced in Supplementary Method 1.

### Unsupervised learning of MAC

Considering a collection of representation domain datasets, denoted as  $V_D = \{V_1, V_2, \dots, V_K\}$ . For each domain, we define  $x_t^i$  as the  $i$ -th signal sample in the  $t$ -th domain dataset.  $V_t, V_s$  from  $V_D$  represent the two representation domain datasets for wireless signal.  $V_s$  represents the I-Q SD.  $V_t$  represents the outcome of a target signal processing transformation, referred to as TD. The dataset comprises sample pairs  $\{x_s^i, x_t^i\}_{i=1}^N$  with a total of  $N$  pairs, forming a contrastive domain of signal samples between domains  $V_s$  and  $V_t$ .

Positive sample pairs of inter-domain contrastive learning are those originating from the joint distribution  $\alpha = \{x_s^i, x_t^i\}$ , while negative sample pairs come from the marginal product  $\beta = \{x_s^i, x_t^i\} | i \neq t$ . (Refer to Supplementary Method 2 for details.) For the proxy task of signal individual identification between the SD and TDs, the similarity score for a pair of positive sample can be expressed as

$$S_{[s,t]}^i = \exp\left(\frac{g_s(f_s(x_s^i))g_t(f_t(x_t^i))}{\|g_s(f_s(x_s^i))\| \cdot \|g_t(f_t(x_t^i))\| \cdot \mu}\right) \quad (4)$$

where the hyperparameter  $\mu$  acts as a temperature coefficient that scales the range of similarity scores. A higher value of  $\mu$  shifts the emphasis towards negative sample pairs with smaller similarity differences. Our aspiration is to train a discriminator to identify a single positive sample from a batched contrastive set  $U_{s \sim t} = \{\alpha, \beta_1, \beta_2, \dots, \beta_\zeta\}$  which includes  $\zeta$  negative samples. The “SD-TD” contrastive loss function can be defined as

$$L_{s \sim t} = -E_{U_{s \sim t}} \left[ \log \frac{S_{[s,t]}^i}{S_{[s,t]}^i + \sum_{j=1}^{\zeta} S_{[s,t]}^{i,j}} \right] \quad (5)$$

However, when viewed from a high-dimensional feature space, each representation domain of the signal is merely a form of representation. Symmetrically, we not only consider the negative sample similarity  $\sum_{j=1}^{\zeta} S_{[s,t]}^{i,j}$  when traversing the TD  $V_t$  using the SD  $V_s$  as the query set, but also take into account the negative sample discriminative score  $\sum_{j=1}^{\zeta} S_{[t,s]}^{i,j}$  obtained by swapping the query relationship between the SD and TDs.

$$L_{inter}^{V_s, V_t} = L_{s \sim t} + L_{t \sim s} \quad (6)$$

where  $L_{inter}^{V_s, V_t}$  represents the inter-domain contrastive loss between the two representation domains.

Considering the potential impacts encountered in real-world channel transmission, we selected four data augmentation methods to construct positive and negative sample pairs. Figure 9a illustrates the impact of various data augmentation techniques on the signal in the constellation diagram. Similar to the construction of inter-domain contrastive similarity (refer to Supplementary Method 3), we define the similarity between a pair of positive samples within the same domain as

$$S_{[A,B]}^i = \exp \left( \frac{g_A(f_A(\hat{x}_A^i)) g_B(f_B(\hat{x}_B^i))}{\|g_A(f_A(\hat{x}_A^i))\| \cdot \|g_B(f_B(\hat{x}_B^i))\| \cdot \mu} \right) \quad (7)$$

The intra-domain contrastive loss can be represented as

$$L_{intra} = -E_{U_{AB}} \left[ \log \frac{S_{[A,B]}^i}{S_{[A,B]}^i + \sum_{j=1}^{\zeta} S_{[A,B]}^{i,j}} \right] \quad (8)$$

The correlation between different representation domains of wireless signals decreases with nonlinear signal processing steps. Therefore, we propose the “I-Q single-centering” strategy to facilitate MAC to effectively handle any number of representation domains as illustrated in Fig. 9b. (Refer to Supplementary Method 4 for details.) The multi-domain joint contrastive loss function under the “I-Q single-centering” strategy can be expressed as

$$L_K = \eta_1 L_{intra} + \sum_{t=2}^K \eta_t L_{inter}^{V_s, V_t} \quad (9)$$

where  $\eta_1, \eta_2, \dots, \eta_K$  are the weighting coefficients for the loss functions between the SD and each TD, and  $\sum_{t=1}^K \eta_t = 1$ . “I-Q single-centering”, which effectively amplifies the distinctiveness of samples in the feature space while striking a balance between computational efficiency and operational effectiveness.

### Domain attention module

DA module was proposed to distinguish which feature vectors obtained from different representation domains provide the most helpful information for AMR, as illustrated in Fig. 9c. DA performs two distinct fusion operations to obtain consolidated features

$$\mathbf{v}_{coh1} \in R^{1 \times (K+1)L}, \mathbf{v}_{coh2} \in R^{(K+1) \times L}.$$

$$\begin{aligned} \mathbf{v}_{coh1} &= J(\mathbf{v}_s, \mathbf{v}_1, \dots, \mathbf{v}_K), \text{ s.t. } \mathbf{v}_{coh1}(1, n + (k+1) \times L) = \mathbf{v}_k(n) \\ \mathbf{v}_{coh2} &= J(\mathbf{v}_s, \mathbf{v}_1, \dots, \mathbf{v}_K), \text{ s.t. } \mathbf{v}_{coh2}(k+1, n) = \mathbf{v}_k(n) \end{aligned} \quad (10)$$

where  $J$  represents the concatenate operation,  $\mathbf{v} : \{\mathbf{v}_s, \mathbf{v}_1, \dots, \mathbf{v}_K\}$  represents the set of feature vectors derived from different representation domains. Subsequently, DA applies global average pooling, squeezing and convolution to map it to the representation domain weight scores  $\boldsymbol{\gamma} : [\gamma_s, \gamma_1, \dots, \gamma_K]$ . The entire process can be represented as

$$\mathbf{v}_{att} = \mathbf{v}_{coh1} \otimes \sigma(\text{cov}^{1 \times 1}(\text{AvgPool}(\mathbf{v}_{coh2}))) \quad (11)$$

where  $\sigma$  represents the sigmoid activation function.  $\mathbf{v}_{att} \in R^{1 \times (K+1)L}$  is capable of disregarding redundant information across multiple representation domains, ensuring that the final modulation classification employs the most suitable representation domain features. DA achieves the selection of signal representation domains at the feature level, thereby significantly reducing reliance on prior information and expert experience.

### Experimental setup

In experiments, the proportions of the unlabeled training set, test set, and validation set are divided into a 6:3:1 ratio in each dataset. All unlabeled signals in the training set are used for unsupervised pre-training. For the labeled training set, during the representation evaluation phase, the number of labeled samples is equal to that of unlabeled samples. In the few-shot fine-tuning phase,  $N$  labeled signals are randomly selected from the unlabeled dataset. Note that our method’s effectiveness analysis and ablation validation were conducted on RML2016.10A and RML2016.10B. Few-shot fine-tuning experiments were validated on RML2016.10A, RML2016.10B and RML2018.01A datasets. The comparison with supervised AMR methods experiments were validated on the RML2016.10A and RML2016.10B datasets.

**Unsupervised representation learning:** In the unsupervised training phase, we conduct a total of 240 training epochs with a batch size of 256. The initial learning rate is 0.03, and after the first 120 epochs, the learning rate is reduced by a factor of 0.1 every 40 epochs. For calculating the contrastive loss, the temperature coefficient  $\mu$  is set to 0.07.

**Supervised fine-tuning:** There exist two paradigms within this stage. One is called representation evaluation, where the pre-trained encoder is kept frozen. Linear evaluation<sup>33</sup> and prototype evaluation<sup>47</sup> are implemented by training linear classifiers or prototype classifiers, which can evaluate the representation learning ability of MAC. Another one is few-shot fine-tuning, where both the encoder and classifier are fine-tuned by few-shot labeled signals, which can assess the semi-supervised AMR performance. In the supervised fine-tuning stage, we conduct a total of 80 training epochs with a batch size of 128. The initial learning rate is set to 0.01, and after the first 40 epochs, the learning rate is reduced by a factor of 0.2 every 10 epochs.

### Data availability

All experimental datasets used in this paper were in DeepSIG’s publicly available datasets RML2016.10A, RML2016.10B, RML2018.01A and generated using GNU Radio, which are available in <https://www.deepsig.ai/datasets/>. Source data are provided with this paper.

### Code availability

Codes used in this work are available from the corresponding authors on request. A permanent version is released on Zenodo: <https://doi.org/10.5281/zenodo.15599189>.

## References

- Choi, H. W. et al. Smart textile lighting/display system with multi-functional fibre devices for large scale smart home and IoT applications. *Nat. Commun.* **13**, 814 (2024).
- Giribaldi, G., Colombo, L., Simeoni, P. & Rinaldi, M. Compact and wideband nanoacoustic pass-band filters for future 5G and 6G cellular radios. *Nat. Commun.* **15**, 304 (2024).
- Ding, R. et al. Data and knowledge dual-driven automatic modulation classification for 6G wireless communications. *IEEE Trans. Wirel. Commun.* **23**, 4228–4242 (2024).
- Zhao, H. et al. Underwater wireless communication via TENG-generated Maxwell's displacement current. *Nat. Commun.* **13**, 3325 (2022).
- Qi, P., Zhou, X., Zheng, S. & Li, Z. Automatic modulation classification based on deep residual networks with multimodal information. *IEEE Trans. Cognit. Commun. Netw.* **7**, 21–33 (2021).
- Yang, W., Ren, K. & Du Modulation recognition method of mixed signals based on cyclic spectrum projection. *Sci. Rep.* **13**, 21459 (2023).
- Zhang, D., Lu, Y., Li, Y., Ding, W. & Zhang, B. High-order convolutional attention networks for automatic modulation classification in communication. *IEEE Trans. Wirel. Commun.* **22**, 4600–4610 (2023).
- Chang, S., Huang, S., Zhang, R., Feng, Z. & Liu, L. Multitask-learning-based deep neural network for automatic modulation classification. *IEEE Internet Things J.* **9**, 2192–2206 (2022).
- Liang, Y.-C., Chen, K.-C., Li, G. Y. & Mahonen, P. Cognitive radio networking and communications: an overview. *IEEE Trans. Veh. Technol.* **60**, 3386–3407 (2011).
- Qiu, K., Zheng, S., Zhang, L., Lou, C. & Yang, X. DeepSIG: A hybrid heterogeneous deep learning framework for radio signal classification. *IEEE Trans. Wirel. Commun.* **23**, 775–788 (2024).
- Hou, S. et al. Multi-domain-fusion deep learning for automatic modulation recognition in spatial cognitive radio. *Sci. Rep.* **13**, 10736 (2023).
- Lin, Y., Tu, Y., Dou, Z., Chen, L. & Mao, S. Contour stella image and deep learning for signal recognition in the physical layer. *IEEE Trans. Cognit. Commun. Netw.* **7**, 34–46 (2021).
- He, J. et al. Channel agnostic radio frequency fingerprint identification using spectral quotient constellation errors. *IEEE Trans. Wirel. Commun.* **23**, 158–170 (2024).
- Hou, C. et al. Multisignal modulation classification using sliding window detection and complex convolutional network in frequency domain. *IEEE Internet Things J.* **9**, 19438–19449 (2022).
- Wang, C. et al. Universal attack against automatic modulation classification DNNs under frequency and data constraints. *IEEE Internet Things J.* **10**, 12938–12950 (2023).
- Kong, W., Jiao, X., Xu, Y., Zhang, B. & Yang, Q. A transformer-based contrastive semi-supervised learning framework for automatic modulation recognition. *IEEE Trans. Cognit. Commun. Netw.* **9**, 950–962 (2023).
- O'Shea, T. J., Corgan, J. & Clancy, T. C. Convolutional radio modulation recognition networks. In *Proc. Int. Conf. Eng. Appl. Neural Netw.* 213–226 (IEEE, 2016).
- Wang, Y., Gui, G., Ohtsuki, T. & Adachi, F. Multi-task learning for generalized automatic modulation classification under non-gaussian noise with varying SNR conditions. *IEEE Trans. Wirel. Commun.* **20**, 3587–3596 (2021).
- Tekbiyik, K. et al. Robust and fast automatic modulation classification with CNN under multipath fading channels. In *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, 1–6 (IEEE, 2020).
- Hong, D., Zhang, Z. & Xu, X. Automatic modulation classification using recurrent neural networks. In *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, 695–700 (IEEE, 2017).
- Rajendran, S. & Meert, W. Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. *IEEE Trans. Cognit. Commun. Netw.* **4**, 433–445 (2018).
- Huang, S. et al. Automatic modulation classification using contrastive fully convolutional network. *IEEE Trans. Wirel. Commun.* **8**, 1044–1047 (2019).
- Xu, J., Luo, C., Parr, G. & Luo, Y. A spatiotemporal multi-channel learning framework for automatic modulation recognition. *IEEE Wirel. Commun. Lett.* **9**, 1629–1632 (2020).
- Ke, Z. & Vikalo, H. Real-time radio technology and modulation classification via an LSTM auto-encoder. *IEEE Trans. Wirel. Commun.* **21**, 370–382 (2022).
- Wu, S. Communication modulation recognition algorithm based on STFT mechanism in combination with unsupervised feature-learning network. *Peer-Peer Netw. Appl.* **12**, 1615–1623 (2019).
- Mao, Y. et al. Attentive siamese networks for automatic modulation classification based on multitiming constellation diagrams. *IEEE Trans. Cognit. Commun. Netw.* **34**, 5988–6002 (2023).
- Daldal, N., Comert, Z. & Polat, K. Automatic determination of digital modulation types with different noises using convolutional neural network based on time-frequency information. *Appl. Soft Comput.* **86**, 105834 (2020).
- Wang, T., Hou, Y., Zhang, H. & Guo, Z. Deep learning based modulation recognition with multi-cue fusion. *IEEE Wirel. Commun. Lett.* **10**, 1757–1760 (2021).
- Wang, Y. et al. Transfer learning for semi-supervised automatic modulation classification in ZF-MIMO systems. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **10**, 231–239 (2020).
- Tu, Y., Lin, Y., Wang, J. & Kim, J.-U. Semi-supervised learning with generative adversarial networks on digital signal modulation classification. *Comput. Mater. Contin.* **55**, 243–254 (2019).
- Li, Y., Shi X., Yang, X. & Zhou, F. Unsupervised modulation recognition method based on multi-domain representation contrastive learning. In *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, 1–6 (IEEE, 2023).
- Liu, D., Wang, P., Wang, T. & Abdelzaher, T. Self-contrastive learning based semi-supervised radio modulation classification. In *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, 777–782 (IEEE, 2021).
- He, K. et al. Momentum contrast for unsupervised visual representation learning. In *Proc. IEEE Conf. Comput. Vis. Pattern. Recognit. (CVPR)*, 9276–9735 (IEEE, 2020).
- Wang, W., Liang, J. & Liu, D. Learning equivariant segmentation with instance-unique querying. In *Proc. Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, 12826–12840 (Springer, 2022).
- Xiao, C., Yang, S., Feng, Z. & Jiao, L. MCLHN: Towards automatic modulation classification via masked contrastive learning with hard negatives. *IEEE Trans. Wirel. Commun.* **23**, 14304–14319 (2024).
- Dong, Y., Jiang, X., Cheng, L. & Shi, Q. SSRCNN: A semi-supervised learning framework for signal recognition. *IEEE Trans. Cognit. Commun. Netw.* **7**, 780–789 (2021).
- Ali, A. & Yangyu, F. k-Sparse autoencoder-based automatic modulation classification with low complexity. *IEEE Commun. Lett.* **21**, 2162–2165 (2017).
- Chen, X. & He, K. Exploring simple siamese representation learning. In *Proc. IEEE Conf. Comput. Vis. Pattern. Recognit. (CVPR)*, 15750–15758 (IEEE, 2021).
- Chen X., Fan H., Girshick R., and He K. Improved baselines with momentum contrastive learning. Preprint at <https://arxiv.org/abs/2003.04297> (2020).
- Oord, A.V.D., Li, Y. & Vinyals, O. Representation learning with contrastive predictive coding. Preprint at <https://arxiv.org/abs/1807.03748> (2019).
- Maaten, L. V. D. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).



42. O'Shea, T. J. & West, N. Radio machine learning dataset generation with GNU radio. In *Proc. GNU Radio Conf.* (IEEE, 2016).
43. Liu, X. D., Yang, D. & Gamal, A. E. Deep neural network architectures for modulation classification. In *Proc. 51st Asilomar Conf. Signals Syst. Comput.*, 915–919 (IEEE, 2017).
44. West, N. E. & O'Shea, T. Deep architectures for modulation recognition. In *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, 1–6 (IEEE, 2017).
45. Zhang, F., Luo, C., Xu, J. & Luo, Y. An efficient deep learning model for automatic modulation recognition based on parameter estimation and transformation. *IEEE Commun. Lett.* **25**, 3287–3290 (2021).
46. Sathyanarayanan, V., Gerstoft, P. & Gamal, A. E. RML22: realistic dataset generation for wireless modulation classification. *IEEE Trans. Wirel. Commun.* **22**, 7663–7675 (2023).
47. Wang, W., Han, C., Zhou, T. & Liu, D. Visual recognition with deep nearest centroids. In *Proc. Int. Conf. Learn. Represent. (ICLR)*, 1–30 (MIT Press, 2023).

## Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 61971332, Grant 62001350, Grant 61801347, Grant 61801344, and Grant 61631019; in part by the Fundamental Research Funds for the Central Universities under Grant XJS220211, Grant XJS210210.

## Author contributions

X.S. and F.Z. suggested the designs, planned and supervised the work. Y.L. in consultation with H.T. conceptualized ideas, conducted MAC model construction, created algorithm design and visualizations, analyzed experimental verification. Z.Z. and X.Y. performed the signal pre-processing and data analysis. All authors discussed and reviewed the results, edited, and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-60921-z>.

**Correspondence** and requests for materials should be addressed to Xiaoran Shi.

**Peer review information** *Nature Communications* thanks Dongfang Liu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025