Universal and Complementary Representation Learning for Automatic Modulation Recognition

Bohan Liu,[1] Ruixing Ge,[1] Yuxuan Zhu,[1] Bolin Zhang,[2] and Yanfei Bao[1]

[1]*Institute of Systems Engineering, Academy of Military Science of the People's Liberation Army, Beijing 100083, China*

[2]*National Key Laboratory of Science and Technology on Communication, University of Electronic Science and Technology of China, 611731, Chengdu, China*

Email: bohanliu99@sina.com

Automatic Modulation Recognition (AMR) is a fundamental research topic in the field of signal processing and wireless communication, which has widespread applications in cognitive radio, non-collaborative communication, etc. However, current AMR methods are mostly based on uni-modal inputs, which suffer from incomplete information and local optimization. In this paper, we focus on the modality utilization in AMR. The proxy experiments show that different modalities achieve a similar recognition effect in most scenarios, while the personalities of different inputs are complementary to each other for particular modulations. Therefore, we mine the universal and complementary characteristics of the modality data in the domain-agnostic and domain-specific aspects, yielding the **U**niversal and **C**omplementary subspaces accordingly (dubbed as **UC**Net). To facilitate the subspace construction, we propose universal and complementary losses accordingly, where the former minimizes the heterogeneous feature gap by an adversarial constraint and the latter consists of an orthogonal constraint between universal and complementary features. The extensive experiments on the RadioML2016.10A dataset demonstrate the effectiveness of UCNet, which has achieved the highest recognition accuracy of 93.2% at 10 dB, and the average accuracy is 92.6% at high SNR greater than zero.

Introduction: With the increasing demand for wireless spectrum bandwidth, improving the utilization of wireless spectrum is an inevitable requirement. In order to reduce illegal occupation, Automatic Modulation Recognition (AMR) is widely studied in signal confirmation [1], spectrum sensing [2], and signal monitoring of spectrum management [3] in the non-collaborative environment. Currently, deep learning models that send signals directly into the network for end-to-end learning have achieved superior performance for AMR (DL-AMR) [4]. However, most DL-AMR methods take one single modality as input, such as In-phase/Quadrature (I/Q) [5], Amplitude/Phase series (A/P) [6], welch/square/fourth power spectrum [7]. These different modalities contained discriminative information from different domains.

To integrate the advantages of different modalities, many researchers focus on multi-modality fusion methods for AMR. Qi et al. propose a Waveform Spectrum Multi-modality Fusion (WSMF) method, which relies on a deep Residual Network (ResNet) and a concise concatenation layer [8]. An optimized Product-based Neural Networks (PNN) model [9] cross-combines the features extracted from I/Q, A/P, and spectrum. However, the above methods simply carry out the cross-connect or direct concatenation of features instead of further capturing the underlying information.

To explore the modality-wise differences, proxy experiments on AMR are conducted with one single modality input (I/Q, A/P, and spectrum). We have found the following two properties: 1) **University**: Features extracted from different modalities have universal performance for most modulations. For example, the I/Q-based method and A/P-based method share similar performance on most modulations (WBFM, BPSK, CPFSK, AM-DSB), which can be well identified with over 98% accuracy at 18db either with I/Q data or A/P data input as shown in Fig. 1a and Fig. 1b. 2) **Complementarity**: For certain modulations, specific discriminative information can be extracted only from the particular modality, and these complementary features of different modalities are supplementary to each other. The features of I/Q data are significantly more distinguishable than A/P data for QAM64 and GFSK, while A/P is able to compensate for the lack of I/Q's representational ability in AM-SSB. At the same time, spectrum has achieved outstanding performance in PAM4 in Fig. 1c, which evidences particular identification characteristics that are not retrievable in other domains.

Based on the preceding analysis, we focus on the modality properties and design two subspaces for multi-modality inputs, which model the university and complementarity of the different modalities. Specifically, universal subspace further aggregates similar common features, while complementary subspace separates out the discriminative characteristics from each domain. To construct subspaces, we propose a universal loss and complementary loss. Concretely, the universal loss minimizes the heterogeneous feature gap through an adversarial constraint, and complementary loss is composed of orthogonal constraints between universal features and complementary features.

The Proposed Method: In this section, we first elaborate on the modality data preparation. Then we introduce the universal and complementary subspace construction. Finally, we present the loss functions for the model optimization.

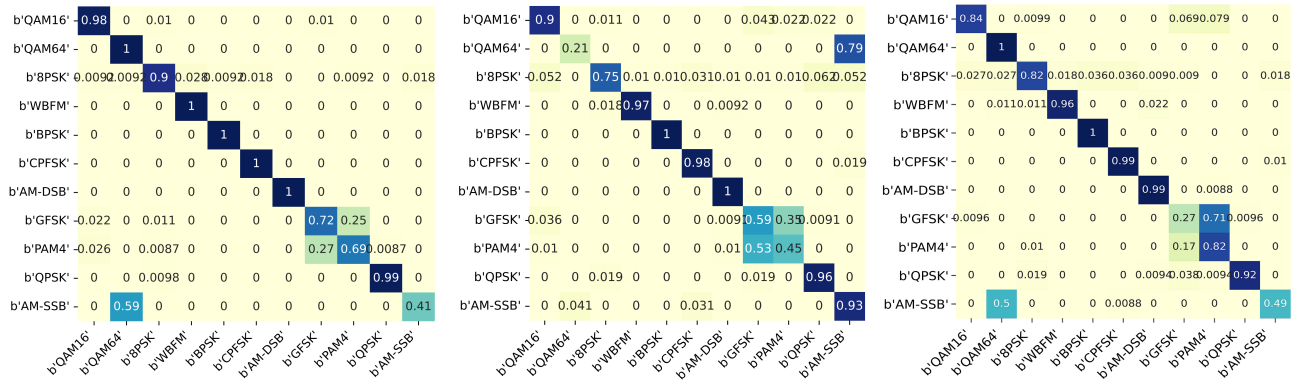
Modality embedding: Inspired by [9], the original signal symbol is transferred into three modalities, i.e., In-phase/Quadrature (IQ), Amplitude/Phase(AP), and Spectrum (SP). IQ, AP, and SP represent information on signal frequency, waveform, and spectrum analysis. For the original signal S, it is pre-processed into three modalities inputs, denoted as $V_{IQ} \in \mathcal{R}^{l_{seq} \times d_{IQ}}$, $V_{AP} \in \mathcal{R}^{l_{seq} \times d_{AP}}$, $V_{SP} \in \mathcal{R}^{l_{seq} \times d_{SP}}$, where l_{seq} indicates the length of signal sequence as 128 and d is the dimension of different modalities (two for IQ and AP, three for SP).

In order to obtain the processable features, the respective modality data are firstly projected into embedded vectors v_m ($m \in (IQ, AP, SP)$) as shown in Fig.2. Following [5, 6, 10], CLDNN, LSTM, and CNN are used to conduct feature embedding for IQ, AP, and SP modalities, respectively. The respective embedded vectors were normalized into (batch-size x 128) and then input into encoding layers.

University & Complementary Encoding: After embedding, the embedded vectors are encoded into universal and complementary subspaces to inspect the two properties of the data. Specifically, two separate liner layers are utilized to conduct the feature encoding:

$$r_m^u = M_u(v_m; \omega^u) \tag{1}$$

$$r_m^c = M_c(v_m; \omega_m^c), \tag{2}$$



(a) CLDNN [5] model with IQ input (b) LSTM [6] model with AP input (c) CNN model with spectrum input

Fig 1 *The recognition performance of different modes to 11 modulations*

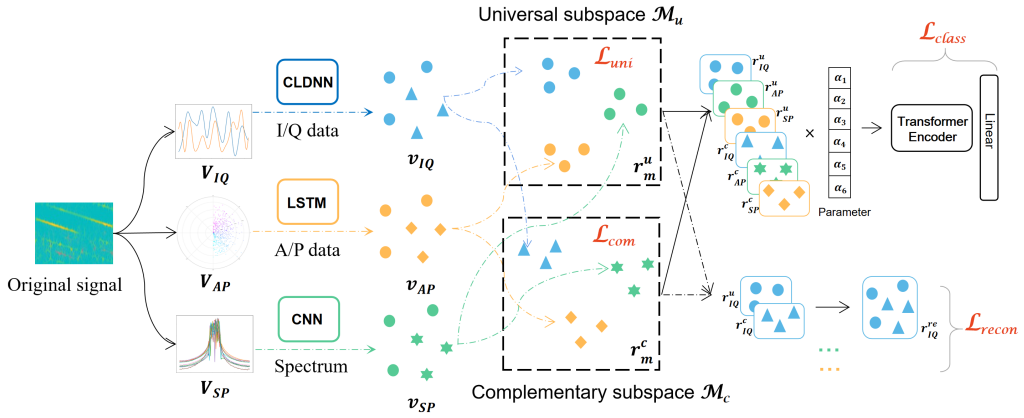


Fig 2 *Architecture of the proposed UCNet. The input data is first preprocessed into trimodal information V_m and then embedded to embedded vectors v_m. The universal and complementary subspace encoding separates universal features r^u_m and complementary features r^c_m from the embedded vectors v_m, which are trained by the combination of four losses. Different colors represent different modalities, where the same shape in different colors represents similar universal features between different modalities, and different shapes in different colors represent modality-specific complementary features.*

where r^u is the universal feature, and r^c represents the complementary feature respectively. These obtained features combined with learnable parameters α_i to form the final multi-modal feature.

Model Optimization: We apply four losses during the model training. The construction of two subspaces is achieved by minimizing universal loss and complementary loss. After the multi-modal fusion, the categories are predicted using classification losses, while reconstruction losses are designed to avoid losing valid information. The overall loss function is denoted as follows:

$$L = L_{class} + \alpha L_{uni} + \beta L_{com} + \eta L_{recon} \tag{3}$$

Universal loss: The purpose of L_{uni} is to maximize the similarity of features obtained from different modal features after a universal encoding layer consisting of a Gradient Reversal Layer (GRL) [11] and an adversarial classifier. Firstly, the multi-modal embedded vectors are put into the GRL. Secondly, the adversarial classifier consisting of linear layers divides embedded vectors into three classes.

For the GRL, the forward propagation is similar to the usual, but the loss will be multiplied by $(-\lambda)$ in backward propagation to achieve the reverse update and enable adversarial learning. Additionally, λ is a dynamically varying parameter, which is expressed as:

$$\lambda = \frac{2}{1 + exp(-\gamma * p)} - 1, \tag{4}$$

where p represents the relative value of the iterative process, the ratio of current iterations to the total number of iterations, and γ is generally set as a constant of 10 [11]. As training proceeds, the universal encoding layer progressively generates similar features that can confuse the classifier with the original modality, which is trained by a cross-entropy function as follows:

$$L_{uni} = -y \log \widehat{y^u} - (1 - y) \log(1 - \widehat{y^u}), \tag{5}$$

where y is the distribution of real labels, $\widehat{y^u}$ is the distribution of predicted results.

Complementary loss: With the orthogonal constraint in the L_{com}, universal features from different modalities are separated from complementary features, and different complementary features are clearly distinguished. For each embedded vector, r^u_m and r^c_m of each modality are used as rows of a matrix to form R^u_m and R^c_m. For the purpose of differentiating features, orthogonal constraints are used for L_{com} to train the extraction of complementary features:

$$\begin{aligned} L_{com} = & \sum_{m \in (IQ,AP,SP)} \left\| [f(R^u_m)^\top f(R^c_m)] \right\|^2_F \\ & + \frac{1}{3} \sum_{m \in (IQ,AP,SP)} \left\| f(R^c_{m_1})^\top f(R^c_{m_2}) \right\|^2_F, \end{aligned} \tag{6}$$

where $f(\cdot) = R^\top R \odot (1 - I)$, $\|\|^2_F$ denotes squared Frobenius norm, and I is the identity matrix. The parameters of each feature extractor are not shared, ensuring diverse domain information of the input is captured.

Classification loss: The L_{class} uses a cross-entropy function to assess the classification accuracy.

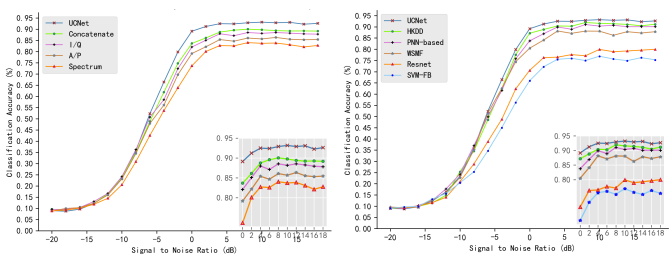
Reconstruction loss: L_{recon} ensures that the universal and complementary encoder learns valid features to avoid trivial solutions. The reconstructed vectors r^{re}_m are generated by respective encoders of three linear layers, using Mean Squared Error (MSE) constraint between the reconstructed vectors r^{re}_m and the embedded vectors v_m, which is denoted as:

$$L_{recon} = \frac{1}{3} \sum_{m \in (IQ,AP,SP)} \frac{\left\| r^{re}_m - v_m \right\|^2}{N} \tag{7}$$

Experiment Results and Discussion: In this section, experiments are conducted to compare with single modality input and other fusion methods. Besides, the effectiveness of universal and complementary features is verified. Subsequently, ablation experiments are carried out to ensure the optimal performance of the model structure design.

Datasets and implement details: The experiments in this paper are conducted on the RadioML2016.10A benchmark dataset [10], which consists of 11 modulations including BPSK, QPSK, 8PSK, 16QAM, 64QAM, BFSK, CPFSK, PAM4, WB-FM, AM-SSB, and AM-DSB. The dataset is extracted in the format of 128 samples per step and shifted by 64 samples, so each single data size is 2×128. The 220,000 data in the dataset are divided into the training set, testing set, and validation set in the ratio of 6: 2: 2. The classification accuracy is chosen as the evaluation metric in subsequent experiments, which represents the proportion of correctly recognized samples to the total.

Uni-modal and multi-modal inputs: In order to intuitively demonstrate the superiority of the multi-modal methods, we conduct comparison experiments with uni-modal and multi-modal inputs, respectively. As shown in Fig. 3a, the classification accuracies of uni-modal models are much lower than that of multi-modal fusion or direct concatenation. The highest accuracy of UCNet is 93.2% at 10 dB, and the average value is 92.6% at high SNR greater than zero.

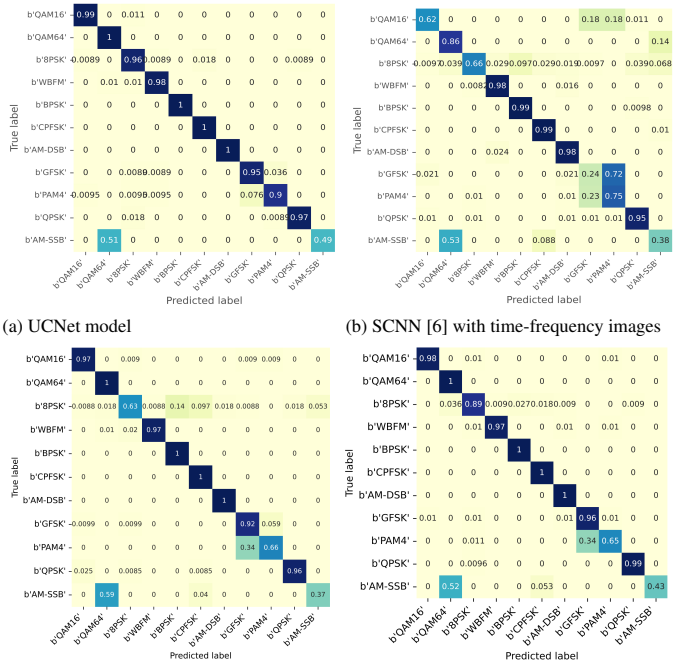


(a) Comparison with uni-modal inputs. (b) Comparison with previous methods.

Fig 3 *Performance in different modalities and models.*

Comparisions with previous methods: As shown in Fig. 3b, it is obvious that our UCNet structure is significantly superior to traditional feature-

based approaches such as SVM-FB [12] and Resnet [13] or previous fusion methods like HKDD [14], WSMF, and the PNN-based model. The identification accuracies of most modulations on UCNet exceeds 95% except for AM-SSB and PAM4 at 18 dB. It may be attributed to the fact that UCNet not only extracts similar discriminative features but also fuses the superiority of different modalities, which obtains a plentiful and comprehensive view.



(a) UCNet model (b) SCNN [6] with time-frequency images

(c) Concatenation with IQ, AP, and SP (d) PNN-based fusion with IQ, AP, and SP

Fig 4 *The recognition performance in different models in 18dB*

Selection of different modalities as inputs: Firstly, we select four commonly used modalities as options for input formats, i.e., IQ, AP, SP, and the time-frequency diagram after the Short-Time Fourier Transform (STFT). As shown in Table 1, the recognition performance of IQ, AP, SP, and STFT is found to be from high to low. As for STFT in Fig. 4b, the recognition on QAM16, 8BSK, GFSK, PAM4, and AM-SSB is lower than 0.7, which does not seem to provide enough modulation information to be selected as inputs.

Secondly, we arrange and combine these modalities to further explore their complementarity to each other. As shown in Table 1, the combination of IQ and AP effectively aids to form a more holistic view as we predicted, especially with the addition of SP. From the confusion matrix in Fig. 1, it can be seen that the information of IQ, AP, and SP are complementary, particularly for QAM64, AM-SSB, and PAM4. Therefore, IQ, AP, and SP are eventually selected as inputs to construct a complete view of the multi-modal data.

Table 1. The ablation results of different input combinations (average accuracy at full SNR).

Input	Accuracy	Input	Accuracy
IQ	0.5892	IQ+AP	0.6092
AP	0.5643	IQ+AP+STFT	0.6104
SP	0.5425	IQ+A/P+SP	0.6263
STFT	0.5132	IQ+AP+SP+STFT	0.6189

Ablations of universal and complementary features: Previous research[15] only feeds complementary features into the classifier, while universal features are not used for classification due to the inclusion of noise highly correlated with the universal representation. However, as seen in Table 2, the universal and complementary features both show superior performance when fed into the classifier respectively. Both of them have achieved at least 1.5% higher accuracy than direct concatenation without encoding.

Ablations of single subspace: In order to explore the differences in capture ability between universal and complementary subspace, two subspaces are used separately for comparison. As shown in Table 2, universal and complementary subspaces respectively provide partial views for fusion, and the accuracy decreases with a single subspace but is still higher than concatenation without encoding. Complementary subspace is 2% higher in accuracy than universal subspace when encoding alone, due to the complementarity that can synthesize richer information.

Table 2. The ablations results of a single feature or subspace.

Input	Accuracy	Input	Accuracy
Universal feature	0.6132	Universal subspace	0.6082
Complementary feature	0.6145	Complementary subspace	0.604
Concatenation	0.5994	UCNet	0.6263

Table 3. The ablation results of feature length.

Feature length	64	128	256
Accuracy	0.6132	0.6263	0.6112

Ablations of feature length: The original signal data is in the format of 128 samples and moves 64 samples per step, so 64, 128, and 256 are taken as the length of the embedded vectors to include the sampled samples and contextual information as much as possible. As shown in Table 3, it is found that the 128 sequence length contains the full information of each sampled sample, while too short or too long length causes information folding or introduces contextual information noise.

References

1. Jiang, K., et al.: A novel digital modulation recognition algorithm based on deep convolutional neural network. Applied Sciences 10(3), 1166 (2020)
2. Bhatti, F.A., et al.: Shared spectrum monitoring using deep learning. IEEE Transactions on Cognitive Communications and Networking 7(4), 1171–1185 (2021)
3. Jagannath, A., Jagannath, J.: Multi-task learning approach for modulation and wireless signal classification for 5g and beyond: Edge deployment via model compression. Physical Communication 54, 101793 (2022)
4. Ke, Z., Vikalo, H.: Real-time radio technology and modulation classification via an lstm auto-encoder. IEEE Transactions on Wireless Communications 21(1), 370–382 (2021)
5. Liu, X., Yang, D., El Gamal, A.: Deep neural network architectures for modulation classification. In: 2017 51st Asilomar Conference on Signals, Systems, and Computers, pp. 915–919. IEEE (2017)
6. Rajendran, S., et al.: Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. IEEE Transactions on Cognitive Communications and Networking 4(3), 433–445 (2018)
7. Zeng, Y., et al.: Spectrum analysis and convolutional neural network for automatic modulation recognition. IEEE Wireless Communications Letters 8(3), 929–932 (2019)
8. Qi, P., et al.: Automatic modulation classification based on deep residual networks with multimodal information. IEEE Transactions on Cognitive Communications and Networking 7(1), 21–33 (2020)
9. Zhang, X., et al.: Modulation recognition of communication signals based on multimodal feature fusion. Sensors 22(17), 6539 (2022)
10. O'Shea, T.J., Corgan, J., Clancy, T.C.: Convolutional radio modulation recognition networks. In: Engineering Applications of Neural Networks: 17th International Conference, EANN 2016, Aberdeen, UK, September 2-5, 2016, Proceedings 17, pp. 213–226. Springer (2016)
11. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International conference on machine learning, pp. 1180–1189. PMLR (2015)
12. RB, C., KARA, A., TORA, H.: Hierarchical classification of analog and digital modulation schemes using higher-order statistics and support vector machines (2022)
13. Shi, F., et al.: Combining neural networks for modulation recognition. Digital Signal Processing 120, 103264 (2022)
14. Zheng, S., et al.: Towards next-generation signal intelligence: A hybrid knowledge and data-driven deep learning framework for radio signal classification. IEEE Transactions on Cognitive Communications and Networking (2023)
15. Bousmalis, K., et al.: Domain separation networks. Advances in neural information processing systems 29 (2016)